1 Notes on Matrix Derivation

Thursday, January 14, 2021 5:39 PM

Matrix Algebra for OLS Estimator

Big Picture

- Matrix algebra can produce compact Notation
- some programs are matrix oriented
- Excel is a matrix

Dependent Variable

- dependent var is an $n \times 1$ column vector

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

- Subscript = index
- use bold typeface

Independent Variables

- k independent vars and a constant term
- thus, $n \times (k+1)$ matrix size

Linear regression model

- define β as a $(k+1) \times 1$ vector of coefficients
  u as an $n \times 1$ vector of error terms
  - ↳ linear multiple regression in matrix form: $y = x\beta + u$
- keep track of the dimensions

First order condition of Applying RSS

- OLS estimators are residual sum squares (RSS)

$$\frac{dRSS}{d\beta_j} = 0 \Rightarrow \sum_{i=1}^{n} x_{ij}\hat{u}_i = 0, \quad (j = 0, 1, \ldots, k)$$

↳ $\hat{u}$ = residual

- System of $k+1$ equations written as $x'\hat{u} = 0 \rightarrow (k+1) \times 1$ vector of 0
  ↳ transpose of x

OLS Estimators in Matrix form

- $\hat{\beta}$ is a $(k+1) \times 1$ vector of OLS estimates

$$x'u = 0$$
$$x'(y - x\hat{\beta}) = 0$$
$$x'y = (x'x)\hat{\beta}$$
$$\hat{\beta} = (x'x)^{-1}(x'y)$$

An important result

$$\hat{\beta} = (x'x)^{-1}(x'y) = (x'x)^{-1}(x'(x\beta + u)) = \beta + (x'x)^{-1}(x'u)$$

- $\hat{\beta}$ in general differs from β due to the error u
- β is an unknown constant
- distribution of $\hat{\beta}$ is the sampling distribution

Statistical properties of OLS estimator I

- under certain assumptions, the OLS estimator is unbiased

Statistical Properties of OLS estimator II

- most likely $\hat{\beta}$ is biased for two reasons:
  1) data is not independent
  2) $E(u|x) \neq 0$ which can be contributed to an omitted variable, simultaneity, and measurement error

Statistical properties of OLS Estimator III

only valid if homoskedasticity holds

$$E((\hat{\beta} - \beta)(\hat{\beta} - \beta)'|x) = \sigma^2(x'x)^{-1}$$

Heteroskedasticity

$$E((\hat{\beta} - \beta)(\hat{\beta} - \beta)'|x) = (x'x)^{-1}(x'\Omega x)(x'x)^{-1}$$

Ω = diagonal matrix

White sandwich Estimator

$$x'\hat{\Omega}x = \sum_{i=1}^{n} \hat{u}_i^2 x_i x_i'$$

$$(x'x)^{-1}(x'\hat{\Omega}x)(x'x)^{-1}$$

Predicted Values

$$P \equiv x(x'x)^{-1}x'$$

↳ Projection matrix

$$P = P' \qquad PP = P \qquad Px = x$$

Residuals

$$\hat{u} = y - \hat{y} = (I - P)y = My$$

$$M \equiv I - P$$

$$M = M' \qquad MM = M \qquad PM = 0$$

Frisch Waugh Theorem I

1.1 Population Regression Function

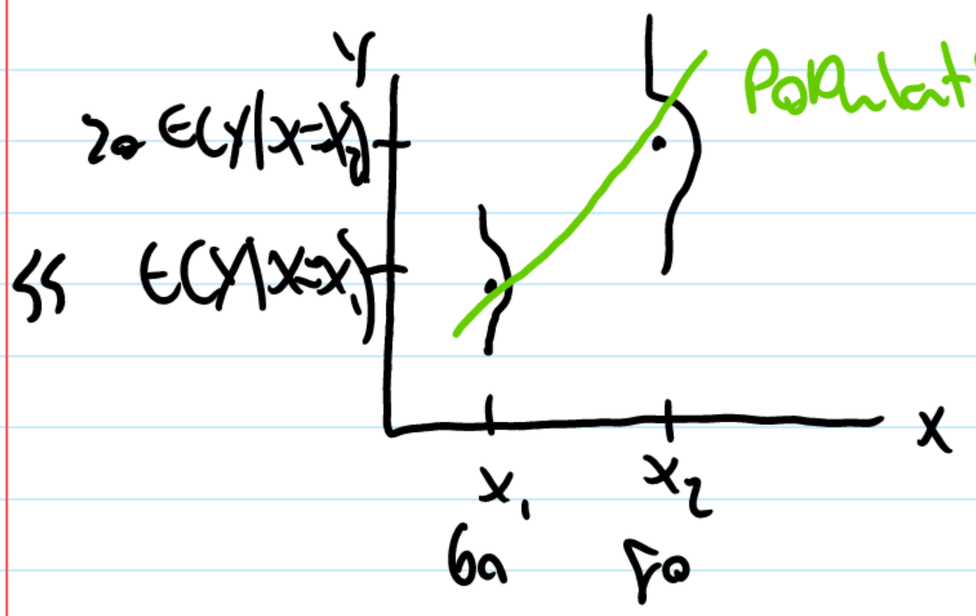
Wednesday, January 13, 2021 9:23 PM

PRF

Expected value $= E(x) = \sum_i f_i x_i = \int_{-\infty}^{\infty} v f(v)\, dv$

Random variable?

Probability density function



Population regression function

$Y_i = E(Y|X) + \text{Residual}$

$E(Y|X) =$ linear function

$E(Y) = \beta_0 + \beta_1 X$

$= \beta_0 + \beta_1 X + \beta_2 X^2$

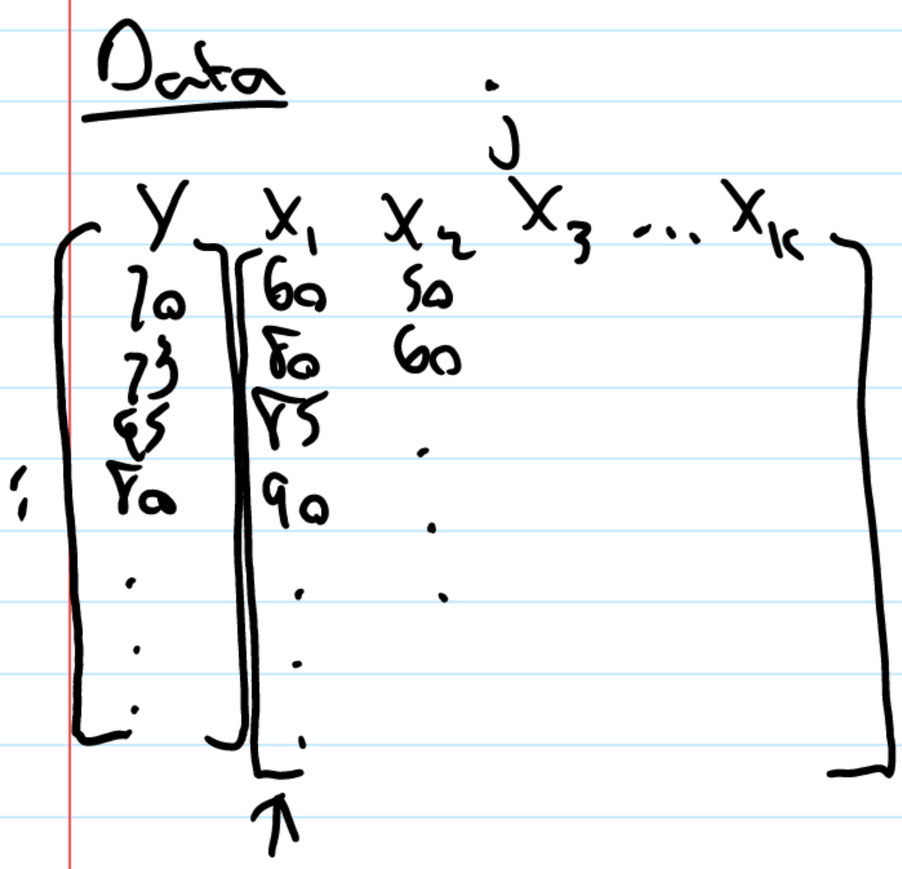
$= \beta_0 + \beta_1 X + \beta_2 1/x$

$E(\ln Y) = \beta_0 + \beta_1 \ln(x)^2$

Specification error: shape of approx $\neq$ shape of PRF

1.2 Data in Matrix Form

Wednesday, January 13, 2021 9:52 PM



$X_0 = 1$ For all

Y X

$E(Y_i|X_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots$

$Y_i = E(Y_i|X_i) + \varepsilon_i$

Predictable

$E(Y|X) = X\beta$

↳ vector of unknowns

↳ x matrix

$Y = X\beta + \varepsilon$

↳ residual matrix

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots \\ 1 & X_{21} & \cdot & \\ 1 & \cdot & \cdot & \\ 1 & \cdot & \cdot & \\ 1 & X_{n1} & \cdot & \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$i \in 1, 2, \dots n$

$Y = X\beta + \varepsilon$

$Y_1 = 1 \cdot \beta_0 + X_{11}\beta_1 + X_{12}\beta_2 + \dots X_{1k}\beta_k + \varepsilon_1$

$\dots Y_n = \dots + \varepsilon_n$

1.3 Estimating the Betas

Wednesday, January 13, 2021 10:03 PM

$$y_i = E(y|x) + \Lambda_i = \sum_j \beta_j x_{ij} + \Lambda_i$$

$$\Lambda_i = y_i - \sum_j \beta_j x_{ij}$$

$$\Lambda_i^2 = (y_i - \sum_j \beta_j x_{ij})^2$$

$$SSR = \sum_i \Lambda_i^2 = \sum_i (y_i - \sum_j \beta_j x_{ij})^2$$

Sample

$$\sum \hat{\Lambda}_i^2 = \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \dots - \hat{\beta}_k x_{ik})^2$$

$\hat{\beta}$ is best estimate of β

$\hat{y}$ is best prediction of y

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots$$

$$\hat{\Lambda}_i = y_i - \hat{y}_i$$

Min sample SSR OLS

1.4 Minimizing the SSR

Wednesday, January 13, 2021 10:13 PM

$$\frac{dSSR}{d\beta_0} = \left(2\sum(y - \sum_j \beta_j x_j)\right)(-1) = 0$$

$$\sum_i y_i = \sum_i \sum_j \beta_j x_{ij}$$

$$\frac{\sum_i y_i}{n} = \sum \bar{y}_i / n$$

$$\bar{y}/n = \bar{y}$$

$$\sum y_i x_{i1} = \sum \bar{y}_i x_{i1}$$

1.5 The Normal Equations

Wednesday, January 13, 2021 10:19 PM

$$\frac{dSSR}{d\beta_j} = -2\sum_i (y_i - \bar{y})\, x_{ij} = 0$$

$$\sum \hat{r}_i \cdot x_{ij} = 0$$

$j = 0 \quad x_{i0} = 1$

$$\sum \hat{r}_i = 0$$

$$\text{Corr}(x_{ij}, r_i) = 0$$

$$\bar{Y}/n = (\sum \tilde{\beta}_0 + \beta_1 x_1 + \tilde{\beta}_2 x_2 + \dots)/n$$

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \dots$$

Choose $\hat{\beta}_0$ so line passes through sample mean

1.6-9 Statistical Properties of the OLS Estimates

Monday, January 18, 2021 7:08 PM

β = slope of population regression function

$\sum_i y_i x_{ij} = \sum \hat{y}_i x_{ij}$

$\sum_i y_i x_{ik} = \sum \hat{y}_i x_{ik}$

$\vdots$

$x'y = x'\hat{y}$

$\hat{y} = x\vec{\beta}$ $\quad x' = x^{\text{Transpose}} = x^T$

$x'y = x'x\hat{\beta}$

$\begin{bmatrix} 1/10 & -1/10 \\ -1/10 & 3/10 \end{bmatrix}\begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ → Identity matrix

$(x^T x)^{-1} x^T y = (x^T x)^{-1} x^T x\vec{\beta}$

$\hat{\beta} = (x^T x)^{-1} x^T y$

$E(\hat{\beta}) = ?$ (unknown pop. values)

$= E[(x^T x)^{-1} x^T (x\beta + \varepsilon)]$

$= E[\beta + (x^T x)^{-1} x^T \varepsilon]$

$= \beta + (x^T x)^{-1} x^T E(\varepsilon | x)$

↳ $= 0$ in population

$E(\hat{\beta}) = \beta$ if no spec error!

$Var(\hat{\beta}) \rightarrow E((\hat{\beta} - E(\hat{\beta}))^2)$

$E((\hat{\beta} - \beta)^2) = E(\beta + (x^T x)^{-1} x\varepsilon - \beta)^2$

$E[(x^T x')\, x^T \varepsilon \varepsilon^T x\, (x^T x')]$

$Var(\hat{\beta}) = (x^T x)^{-1} x^T E(\varepsilon\varepsilon^T | x) x (x^T x)^{-1}$

$\varepsilon\varepsilon^T = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}\begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_n \end{bmatrix}$ = n × n matrix

$E(\varepsilon\varepsilon^T | x) = ?$

Estimating the Sigma Matrix

$E(\varepsilon_i \varepsilon_j | x) = 0$ 1) No spec error 2) Random sample

$\Sigma = E[\text{diagonal matrix}]$

$\hat{\Sigma} = [\text{diagonal hat matrix}]$ → Assume homoskedasticity

$E(\varepsilon_i^2) = E(\varepsilon_j^2)$ $\forall\, i = j$

$E(\varepsilon_i^2) = \sigma_i^2 = \sigma_j^2 = \sigma^2$

I = identity matrix

Homoskedasticity

$Var(\hat{\beta}) = (x^T x)^{-1} x^T \sigma^2 I x (x^T x)^{-1}$

$= \sigma^2 (x^T x)^{-1} x^T x (x^T x)^{-1}$

$= \sigma^2 (x^T x)^{-1}$ → default

= diagonal = variance → OFF diagonal = covariance

Sandwich estimator

Eicker-Huber-White estimator

$\widehat{Var}(\hat{\beta}) = (x^T x)^{-1} x^T \hat{\Sigma} x (x^T x)^{-1}$

(hard to calculate)

Corr factor $= \frac{n}{n-k-1}$

Heteroskedasticity Robust Variance Estimator

Assumptions:

1) Correct specification
2) Random sample

2.1-2 Stata Introduction and Interpreting Regression Results

Thursday, January 21, 2021 9:06 PM

set more off = don't pause after every command

2.3 Understanding Coefficient Variance

Thursday, January 21, 2021 9:54 PM

$$\widehat{Var(\hat{\beta})} = (X^TX)^{-1}X^T\Sigma X(X^TX)^{-1}$$

$$\begin{bmatrix} \widehat{Var(\hat{\beta}_1)} & & \\ & \widehat{Var(\hat{\beta}_2)} & \\ & & \widehat{Var(\hat{\beta}_n)} \end{bmatrix}$$

$\widehat{Var(\hat{\beta}_j)} = s^2/[(1-\tilde{R}_j^2)SST_j]$ → assumes homoscedasticity

↳ $s^2 = [\sum \hat{u}_i^2]/[N-k-1]$

↳ $\tilde{R}_j^2$ = regress x_j on all other x_i $(i \neq j)$ to get R^2

↳ $SST_j = \sum (x_j - \bar{x}_j)^2$ → total variability in x_j

$1/(1-\tilde{R}_j^2)$ → variance inflation factor

clear -> clears everything

set more *off* -> doesn't pause after every command

cd "path" -> changes the directory

log *using* "file name" -> creates a log file with that name

import *delimited* "path" -> imports comma delimited files (csv, tsv, whatever)

/* block comment ***/**
********* single line comment

scatter *y-var x-var* -> creates a scatterplot with the x and y vars specified

regress *y-var x-var* -> creates a simple linear regression with the x and y vars specified

STOP -> isn't a command but will stop the do file because it will throw and error and stop

regress *y-var x-var,* **robust** -> "Robust regression is an alternative to least squares regression when data is contaminated with outliers or influential observations and it can also be used for the purpose of detecting influential observations." (Links to an external site.)

regress *y-var* **c.***x-var***##*i.*x-var2*, **robust** -> same thing as before. c. = continuous variable, i. = indicator/dummy variable, ## = complete interaction between the x variables

testparm i.*y-var* **i.***y-var***#*c.*x-var* <- test parameters associated with y-var and y-var that's interacted with x-var

1. Is there a difference between # and ##?
2. Does it have to be y-var y-var or can it be y-var any-var?

gen *newName* = stuff -> creates a new variable called newName using anything put after the =

corr *var1 var2* -> Find the correlation between the variables

vif -> when run after **corr** it gives the variance inflation factor

predict *var* <- predicts var

predict *var*, **residuals** <- plots against residuals

reg -> same as **regress**

log close -> closes the log

translate *log_file_name new_pdf_name*

2.5 Root Mean Square Error

Thursday, January 21, 2021 10:34 PM

$(Y_i - \hat{Y}_i)^2 / N = MSE = (Y_i - \hat{Y}_i)^2 / (N - k - 1)$

Precision vs accuracy/bias

10 Book Notes

Friday, January 22, 2021 12:05 PM

10-1 The Nature of Time Series Data

Data must go in order

A series of random variables indexed by time is a **stochastic process** or **time series process**
↑ random

10-2 Examples of Time Series Regression Models

10-2a Static models

a static model time has an immediate effect
good for judging tradeoffs between y and z

10-2b Finite Distributed Lag models

FDL model

one or more variables can affect y w/ a lag

δ_0 is the **impact propensity** or **impact multiplier**

Lag distribution summarizes dynamic effect that a temporary increase in z has on y

Long-run propensity/multiplier (LRP)

for any horizon h, we can define the **cumulative effect**

10-2c A convention about the time index

Time starts at t=1

10-3 Finite Sample Properties of OLS under classical assumptions

10-3a Unbiasedness of OLS

Assumptions:
1) Linear in parameters
2) No perfect collinearity
3) Zero conditional mean
4) Homoskedasticity
5) No serial correlation
6) Normality

10-3b The variance of the OLS estimators and the Gauss-Markov Theorem

Homoskedasticity:
Conditional on X, variance of u_t is same for all t
↳ $Var(u_t) = \sigma^2$, $t=1, 2, \ldots, n$

no serial correlation:
Conditional on X, errors in two times are uncorrelated
↳ $Corr(u_t, u_s | X) = 0$, for all $t \neq s$
↳ suffer from serial/auto correlation when false

10-3c Inference under the classical linear model Assumptions

Normality:
errors u_t are independent of X and are independently and identically distributed as Normal$(0, \sigma^2)$

10 Time Series Econometrics for the 21st Century

Friday, January 22, 2021 11:09 AM

1 Introduction

Econometrics started with time series data

2 Overview

Core models to learn:

1) Autoregressive model (AR) $\Rightarrow y_t = \alpha + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + e_t$
2) Regression model $\Rightarrow y_t = \alpha + \delta_0 x_t + e_t$
3) Distributed Lag (DL) model
   $\hookrightarrow y_t = \alpha + \delta_0 x_t + \delta_1 x_{t-1} + \ldots + \delta_q x_{t-q} + e_t$
4) Autoregressive-Distributed Lag (ADL) model
   $\hookrightarrow y_t = \alpha + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \delta_0 x_t + \delta_1 x_{t-1} + \ldots + \delta_q x_{t-q} + e_t$

p = # lags of dependent var y_t
q = # of lags of explanatory var x_t
e_t = mean-zero shock

14! Core insights!

3 Standard Errors and t Statistics

Newey standard error appropriate for simple (non-dynamic) regressions and DL models
- Distributed Lag
  - base used when serial correlation has not been modelled

4 Autoregressive models

5 Distributed Lag models

Estimate impact of one var on another

6 Autoregressive Distributed Lag models

7 Model selection

"Lag selection is inherently a bias-variance trade-off"

8 Spurious Regression

9 Structural Change

Be wary of major shifts within data that throw off estimates

10 Forecasting

10.2 Time Series Basics

Friday, January 22, 2021 1:55 PM

$\{X_t : t = 1, 2, \ldots\}$

Time series data is a random sequential process

observe one outcome at a point in time

1) Past can impact the future
2) Outcomes are random, but obs not independent

Static time series model

$$Y_t = B_0 + B_1 X_t + \Gamma_t$$

$$\underset{\uparrow}{rev}_t = B_0 + B_1 \Gamma GDPC_t + \Gamma_t \longrightarrow \Gamma GDPC = gdp / gdpdef \cdot pop$$

revenue/expenditure

ex_t: depend on what we want income↑ → more depends on available revenue

rev_t: avg tax rate · tax base t

↑

income

10.3 Time Trends

Sunday, January 31, 2021 6:16 PM

past affects future so std error isn't right

Spurious correlations!

linear trends by regressing time and a variable

Time trends

$y_t = \alpha_0 + \alpha_1 t + \varepsilon_t$ $\quad$ $x_t = \gamma_0 + \gamma_1 t + \varepsilon_t$

$\hat{y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 t$

$\hat{\varepsilon}_t = (y_t - \hat{\alpha}_0 - \hat{\alpha}_1 t)$

suppose $Y_t = \beta_0 + \beta_1 x_t + \delta t + \varepsilon_t$

How is y correlated with x?

detrending = put in t trend take residuals afterwards

differencing =

Lags:

$L.y_t = y_{t-1}$ $\quad$ $F.y_t = y_{t+1}$

$L2.y_t = y_{t-2}$

$D.y_t = y_t - y_{t-1}$

↑ differencing

$y_t = \beta_0 + \beta_1 x_t + \delta t + \varepsilon_t$

$y_{t-1} = \beta_0 + \beta_1 x_{t-1} + \delta(t-1) + \varepsilon_{t-1}$

↳ $d.y_t = y_t - y_{t-1} = \delta + \beta_1 (x_t - x_{t-1}) + (\varepsilon_t - \varepsilon_{t-1})$

10.4 Seasonal Trends

Sunday, January 31, 2021 7:29 PM

add quarterly dummy variables to adjust for seasons

$$Y_t = \beta_0 + \beta_1 X_t + \delta t + \underbrace{\gamma_{sp} + \gamma_{sm} + \gamma_{fa}}_{\text{seasons}} + \Lambda_t$$

Why does lag matter with seasons?

10.5-6

Sunday, January 31, 2021 5:27 PM

$\hat{y}_t = -0.76 - 0.22x_t + 0.65x_{t-1} - 0.093t + 0.04sp + .001sm + .012fall$

lag — spring — summer — fall

maybe lag matters more?

$\hat{y} = .15x_t + 1.56x_{t-1} + .67x_{t-2} - .63x_{t-3} - 1.5x_{t-4}$

4 lags is one cycle

spc install estout

lag: (x/y), dot

2 time x to y

estfab → recommended for exporting

different # observations because data is burned to calculate lag

10.7

Sunday, January 31, 2021 6:05 PM

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2}$$

$$= \underset{\uparrow}{\alpha} + \beta_0 X_{t-0} + \beta_1 X_{t-1} + \beta_3 X_{t-3} + \dots$$

intercept

$$= \alpha + \sum_l \beta_l X_{t-l}$$

Y_t = Total N

X_t = Emp by industry of interest

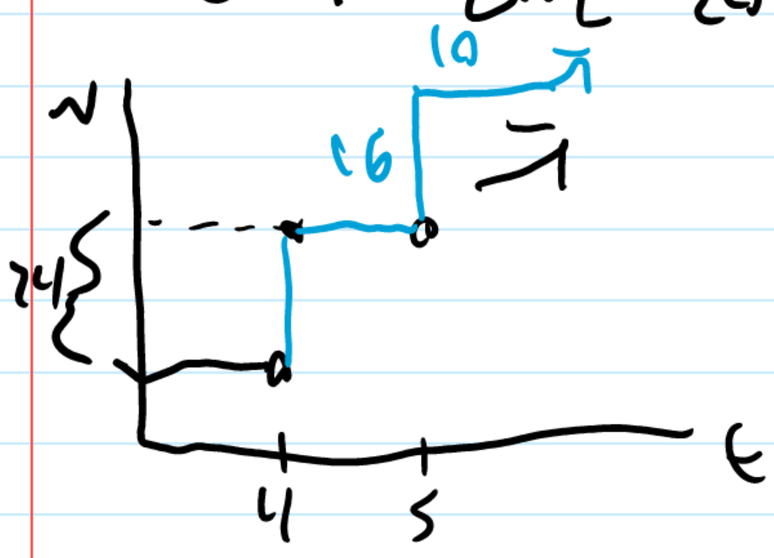
$$N_t = \alpha + \beta_0 N_{It} + \beta_1 N_{It-1} + \beta_2 N_{It-2} + \dots$$

↑ takes 3 years

$$N_t = 100 + 1.2 N_{It} + .8 N_{It-1} + .5 N_{It-2}$$

$$\Delta N_t = 1.2 \Delta N_{It} + .8 \Delta N_{It-1} + .5 \Delta N_{It-2}$$

$t = 4$ $\Delta N_t = 20$



$\rightarrow \Delta N = 50$

Long run effect $= 1.2 + .8 + .5 = 2.5$

11 ARMA Notes

Monday, February 8, 2021 6:40 PM

most econ variables are non-stationary

stationary linear models are building blocks in more complicated non-linear and/or non-stationary models

The Wold Decomposition

Any stationary process $\{z_t\}$ can be expressed as a sum of two components

1) Stochastic: linear combination of a white noise process
2) Deterministic: uncorrelated with stochastic

Importance of Wold

Any stationary process can be written as a linear combination of a lagged value of a white noise process

MA(q) Processes

MA(1) stationarity and Ergodicity

Ergodicity: A point visits all available places

Invertibility: z_t if it admits an autoregressive representation

11 Book Notes

Wednesday, February 3, 2021 7:04 PM

Further Issues in Using OLS with Time Series Data

11-1 Stationary and Weakly Dependent Time Series

11-1a Stationary and nonstationary Time Series

a **stationary process** has stable probabilities over time

a **covariance stationary process** has a constant mean and variance but covariance between x_t and x_{t+h} depends on the size of h

11-1b Weakly Dependent Time series

weakly dependent when correlation between x_t and x_{t+h} trends towards 0 as $x \to \infty$

If instead of x, $h \to \infty \Rightarrow$ **asymptomatically correlated**

11-3 Using highly persistent time series in regression analysis

11-3a Highly Persistent Time series

Y_t depends on Y_{t-h} ← **random walk**

↳ **unit root process:**

random walk w/ drift has a general trend

11-3b Transformations of highly persistent time series

weakly dependent are integrated of order 0 [I(0)]

↳ dont need transform before regression

11-4 Dynamically Complete Models and the absence of Serial Correlation

11-2 Asymptotic Properties of OLS

Linearity and Weak dependence:
$\{(x_t, y_t): t=1, 2, \ldots\}$ is stationary and weakly dependent

No perfect collinearity
Zero conditional mean
Homoskedasticity
No serial correlation

11.1 Temporal Dependence

Wednesday, February 3, 2021 8:41 PM

Spurious correlation & persistence

$\widehat{Var}(\hat{\beta}) = (X'X)^{-1} X' \hat{\Sigma} X (X'X)^{-1}$

$\hat{\sigma}_i^2$ est $E(\sigma_i^2) = \sigma_i^2$

$$\Sigma = \begin{bmatrix} E(\sigma_1 \sigma_1) & E(\sigma_1 \sigma_2) & E(\sigma_1 \sigma_n) \\ \ddots & & \\ & \ddots & \end{bmatrix}$$

Random Walk

$y_t = y_{t-1} + \sigma$

↳ white noise

"highly persistent time series"

↳ $y_t = (y_{t-2} + \sigma_{t-1}) + \sigma_t \rightarrow y_t = (y_{t-3} + \sigma_{t-2}) + \sigma_{t-1} + \sigma_t \rightarrow \dots$

11.2 Testing

Saturday, February 13, 2021 11:43 AM

Dickey Fuller Test

$$y_t = \rho y_{t-1} + \varepsilon_t$$

$\rho = 1$?

$$y_t - y_{t-1} = \rho y_{t-1} - y_{t-1} + \varepsilon_t$$

$$\Delta y_t = (\rho - 1) y_{t-1} + \varepsilon_t$$

IF $\rho = 1$, $\rho - 1 = 0$

$H_0 : \rho - 1 = 0$

$H_a : (\rho - 1) \neq 0$

11.2 More On

Wednesday, February 3, 2021 8:54 PM

$$Y_t = \alpha + Y_{t-1} + \varepsilon_t$$

↳ drift

Y_0

$Y_1 = \alpha + Y_0 + \varepsilon_1$

$Y_2 = 2\alpha + Y_0 + \varepsilon_1 + \varepsilon_2$

$Y_3 = 3\alpha + Y_0 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3$

$Y_t = t\alpha + Y_0 + \text{sum}(j=1:t \mid \varepsilon_j)$

↳ time trend

$\Delta Y_t = Y_t - Y_{t-1} = \alpha t + Y_0 + \varepsilon_t + \varepsilon_{t-1} + \ldots - (\alpha(t-1) + Y_0 + \varepsilon_{t-1} + \varepsilon_{t-2} \ldots)$

$E(\Delta Y_t) = \alpha \qquad \text{Var}(\Delta Y_t) = \sigma^2$

$$Y_t = \alpha t + Y_0 + \sum_{i=1}^{t} \varepsilon_i$$

$E(Y_t) = Y_0 + \alpha t$

$\text{Var}(Y_t) = \sigma_1^2 + \sigma_2^2 + \ldots \sigma_t^2 = t\sigma^2$

$\text{Stdev}(Y_t) = \sqrt{t} \cdot \sigma$

11.4 Stationarity

Wednesday, February 3, 2021 9:27 PM

$(X_t, X_{t+1}, \ldots X_{t+m})$ vs $(X_{t-h}, X_{t-h+1}, \ldots X_{t-h+m})$

X_t is stationary if joint distribution is equal for all t, h, m

Covariance Stationarity

1) $E(X_t)$ $Var(X_t)$ are constant
2) $Cov(X_t, X_{t-h})$ depend on h, not on t

Weak dependence vs high persistence

$\{X_t : t = 1, 2, \ldots\}$ is W.D. if X_t + X_{t-h} are "almost" indep. as h increases for all t

11.6 Autoregressive Processes - 1

Monday, February 8, 2021 6:57 PM

$Y_t = \rho_1 Y_{t-1} + r_t + $

- bMa
- $\hookrightarrow E(r_t) = 0$
- $\hookrightarrow Var(r_t) = \sigma^2$

} AR(1)

$\rho_1 = 1 \Rightarrow$ random walk

Y_0 Fixed Start

$E(Y_t) = Y_0$

$Y_t = \rho(\rho Y_{t-2} + r_{t-1}) + r_t$

$Y_t = \rho^2 Y_{t-2} + \rho r_{t-1} + r_t$

$Y_t = \rho^2(\rho Y_{t-3} + r_{t-2}) + \rho r_{t-1} + r_t$

$= \rho^t Y_0 + \sum_{h=0}^{t} \rho^h r_{t-h}$ $\Rightarrow \rho^0 = 1$ h = lag order

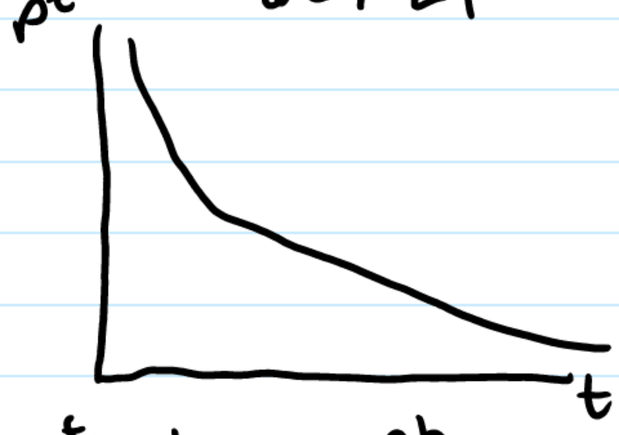
$E(Y_t) = \rho^t Y_0 + \sum_{h=0}^{t} \rho^h E(r_{t-h})$

$\hookrightarrow = 0$

$|\rho| < 1$

$E(Y_t) \to 0$ as $t \to \infty$ if $|\rho_1| < 1$

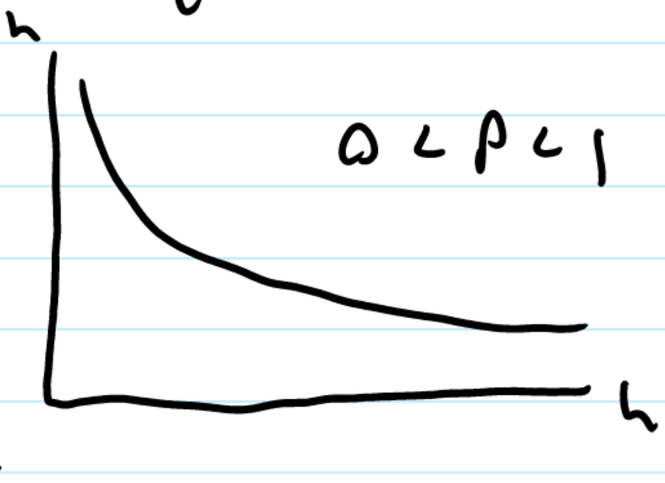
$0 < \rho < 1$



$Y_t = \rho^t Y_0 + \sum_{h=0}^{t} \rho^h r_{t-h}$

$|\rho| < 1$

$Y_t + Y_{t-h} \to$ almost independent if h is large



$Y_t = \rho_1 Y_{t-1} + r_t$

$\Delta Y_t = Y_t - Y_{t-1} = \rho_1 Y_{t-1} - Y_{t-1} + r_t = (\rho_1 - 1) Y_{t-1} + r_t$

$\Delta Y_{t-1} =$

$Y_t = \rho_1 Y_{t-1} + r$

$\hookrightarrow 0 < \rho < 1$

$\Delta Y_t = Y_t - Y_{t-1} = (\rho_1 - 1) Y_t + r_t$

$\rho = .9$	$\rho = .8$
$\rho - 1 = .1$	$(\rho - 1) = .2$
$\rho = .5$	$(\rho - 1) = -.5$

11.7 Autoregressive Processes 2

Monday, February 8, 2021 7:33 PM

AR(p) process
↳ rho

$$y_t = \alpha + \rho_1 y_{t+1} + \rho_2 y_{t-2} + \dots \rho_p y_{t-p} + \varepsilon_t$$

$$= \alpha + \sum_{h=0}^{p} \rho_h y_{t-h} + \varepsilon_t$$

↳ weakly dependent

} stationary + weakly dependent

→ Estimate via OLS

(p|d|1)

daily data

$$\tilde{y} = (y_{t-1} + y_{t-2} + \dots + y_{t-7}) / 7 \quad \rightarrow AR(7)$$

Moving Average → not this!

11.8 Moving Average Processes

Monday, February 8, 2021 7:47 PM

MA(1): $Y_t = \phi + \varepsilon_t + \Theta\varepsilon_{t-1}$ → $\Theta\varepsilon_{t-1}$ = residual

"innovations"
"shocks"
"disturbances"

↳ $E(\varepsilon_t) = 0$
$Var(\varepsilon_t) = \sigma^2$
stationary + weakly dependent

Y_t is corr w/ Y_{t-1} not corr w/ Y_{t-2}

↳ $Y_{t-2} = \phi + \varepsilon_{t-2} + \Theta\varepsilon_{t-3}$

$Y_t = \phi + r_t$, $r_t = \varepsilon_t + \Theta\varepsilon_{t-1}$

MA(q)

$Y_t = \phi + \varepsilon_t + \Theta_1\varepsilon_{t-1} + \Theta_2\varepsilon_{t-2} + \dots + \Theta_q\varepsilon_{t-q}$

$|\Theta_n| < 1$

$Y_t = \phi + \sum_{h=0}^{q} \Theta_h \varepsilon_{t-h}$

ARMA or ARIMA

$Y_t = \alpha + \sum_{h=1}^{p} Y_{t-p} + \sum_{g=0}^{q} \Theta_g \varepsilon_{t-g}$ → ARMA

ARIMA

↳ Integrated

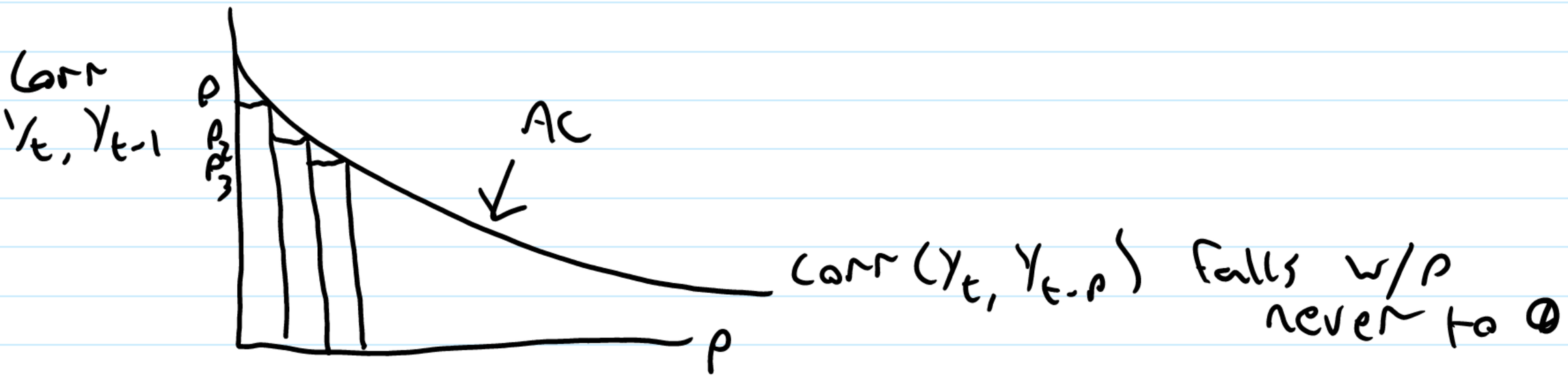
I(1) AR(1) $\rho_1 = 1$ Unit root process Integrated of order 1 process

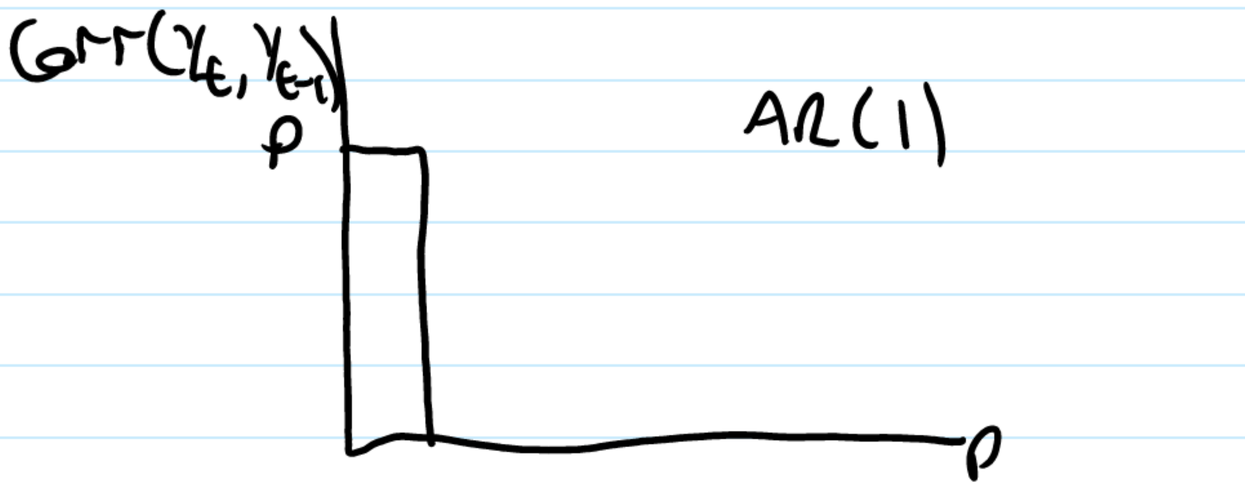
11.9

Saturday, February 13, 2021 10:45 AM

$Y_t = \rho Y_{t-1} + \varepsilon_t$ AC(1)

$Y_t = \rho(\rho Y_{t-2}) + \varepsilon_t + \varepsilon_{t-1}$





Autocorrelation is correlation between now and any given lag in the past

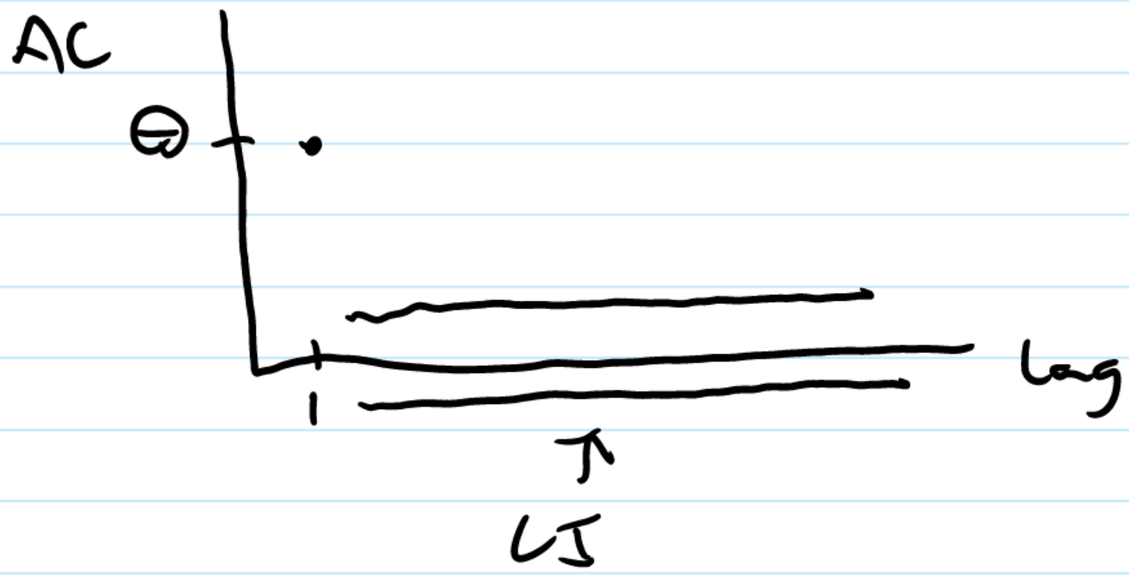
Partial autocorrelation is the same but controlling for all other lags

11.10 ACs

Saturday, February 13, 2021 10:57 AM

MA(1)

$x_t = \varepsilon_t + \theta \varepsilon_{t-1}$



$x_{t-1} = \varepsilon_{t-1} + \theta \varepsilon_{t-1}$ $\rightarrow \varepsilon_{t-1} = x_{t-1} - \theta \varepsilon_{t-2}$

$\hookrightarrow x_t = \varepsilon_t + \theta(x_{t-1} - \theta \varepsilon_{t-2})$

11.11 AR

Saturday, February 13, 2021 11:04 AM

Write an AR(1) with $|\rho|<1$ as infinite MA

$$Y_t = \rho^t Y_0 + \sum_{h=0}^{t} \rho^h \varepsilon_t \qquad Y_t = \sum_{h=0}^{\infty} \rho^h \varepsilon_t$$

Write an MA(1) with $|\theta|<1$ as infinite AR

$$Y_t = \varepsilon_t + \theta \varepsilon_{t-1} \qquad \varepsilon_t = Y_t - \theta \varepsilon_{t-1}$$
$$\varepsilon_{t-1} = Y_{t-1} - \theta \varepsilon_{t-2}$$

$$Y_t = \varepsilon_t + \sum_{h=1}^{\infty} -1^{(h-1)} \theta^h Y_{t-1}$$

$$\rho^h = -1^{(h-1)} \theta^h \qquad |\theta|<1 \qquad |\theta^h| \to 0$$

$$\varepsilon_t = \varepsilon_t$$

$$Y_t = \sum_{h=1}^{\infty} \rho_h Y_{t-1} + \varepsilon_t \qquad AR(\infty)$$

12 Book Notes

Tuesday, February 16, 2021 9:01 PM

12.1 Properties of OLS w/ Serial Correlated Errors

12.1a Unbiasedness and Consistency

When data is weakly dependent, $\hat{\beta}_j$ are still consistent but not necessarily unbiased

12.1b Efficiency and Inference

When $\rho < 0$, ρ^j is negative when j is odd and positive when even

12.1c Goodness of Fit

12.1d Serial Correlation in the Presence of Lagged Dependent Variables

12.2 Testing for Serial Correlation

12.2a A t test for AR(1) Serial Correlation With Strictly Exogenous Regressors

12.2b The Durbin-Watson Test Under Classical Assumptions

12.2c Testing for AR(1) Serial Correlation Without Strictly Exogenous Regressors

12.2d Testing for Higher Order Serial Correlation

$H_0: \rho_1 = 0, \rho_2 = 0, \ldots \rho_n = 0$

12.4 Differencing and Serial Correlation

12.1 Wold Representation Theorem

Tuesday, February 16, 2021 9:57 PM

Any covariance stationary process Y_t can be written (exactly) as $Y_t = \gamma_t + \sum_{n=0}^{\infty} \theta_n \varepsilon_{t-n}$ where $\theta_0 = 1$

$\varepsilon \sim (0, \sigma^2)$

γ_t is completely deterministic

$\sum_{n=0}^{\infty} \theta_n^2 < \infty$

↳ can be written $Y_t = \gamma_t + \sum_{n=1}^{\infty} \rho_n Y_{t-1} + \varepsilon_t$

↓

$X_t \beta$

if $|\rho| < 1$, finite ρ is good enough

$$Y_t = X_t \beta + \sum_{n=1}^{p} \rho_n Y_{t-n} + \varepsilon_t$$

Auto-regressive distributed lag model

12.4-5 Serial

Friday, February 19, 2021 2:16 PM

$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$ where ε_t is correlated with prior ε_t

- $\varepsilon_t = \varepsilon_t + \theta \varepsilon_{t-1}$ MA(1)
- $\varepsilon_{t-1} = \varepsilon_{t-1} + \theta \varepsilon_{t-2}$ MA(q)
- $\varepsilon_t = \rho \varepsilon_{t-1} + \varepsilon_t$ AR(1)
  AR(q)

$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$

$\varepsilon_t = Y_t - \beta_0 - \beta_1 X_t$

$Y_t = \beta_0 + \beta_1 X_t + \rho \varepsilon_{t-1} + \varepsilon_t$

$= \beta_0 + \beta_1 X_t + \rho (Y_{t-1} - \beta_0 - \beta_1 X_{t-1}) + \varepsilon_t$

$= (1-\rho)\beta_0 + \beta_1 X_t - \rho \beta_1 X_{t-1} + \rho Y_{t-1} + \varepsilon_t$

$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + \varepsilon_t$

↳ $\varepsilon_t = \rho \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \varepsilon_t$

$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1}$
$+ \rho_1 (Y_{t-1} - \beta_0 - \beta_1 X_{t-1} - \beta_2 Y_{t-2})$
$+ \rho_2 (Y_{t-2} - \beta_0 - \beta_1 X_{t-2} - \beta_2 Y_{t-3})$

$\Downarrow$

$Y_t = \beta_0 (1 - \rho_1 - \rho_2) + \beta_1 X_t - \rho_1 \beta_1 X_{t-1} + \varepsilon_t$

12.6-7

Friday, February 19, 2021 2:37 PM

$$Var(\hat{\beta}) = (X^TX)^{-1}X^T \; E(\varepsilon\varepsilon^T) \; X(X^TX)^{-1}$$

$$\Sigma = \begin{bmatrix} \sigma_{\varepsilon_1\varepsilon_1} & \cdots & \varepsilon_1\varepsilon_n \\ \varepsilon_n\varepsilon_1 & \ddots & \sigma_{\varepsilon_n\varepsilon_n} \end{bmatrix}$$

$$\begin{bmatrix} \sigma_{11} & & \\ & \ddots & \\ & & \sigma_{nn} \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \hat{\varepsilon}_1\hat{\varepsilon}_1 & \hat{\varepsilon}_1\hat{\varepsilon}_2 & 0 \\ \hat{\varepsilon}_2\hat{\varepsilon}_1 & \hat{\varepsilon}_2\hat{\varepsilon}_2 & \\ 0 & & \hat{\varepsilon}_n\hat{\varepsilon}_n \end{bmatrix} \quad E(\varepsilon_i\varepsilon_j) \neq 0 \qquad n^2 \to \frac{n(n+1)}{2}$$

↓

or fill it all in

$$\Sigma = \hat{\varepsilon}\hat{\varepsilon}^T \begin{bmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \hat{\varepsilon}_n \end{bmatrix} \begin{bmatrix} \hat{\varepsilon}_1\hat{\varepsilon}_2 & \hat{\varepsilon}_n \end{bmatrix}$$

$$\widehat{Var(\hat{\beta})} = (X^TX)^{-1}(X^T\hat{\varepsilon})(\hat{\varepsilon}^TX)(X^TX)^{-1}$$

$$X^T\hat{\varepsilon} = \begin{bmatrix} \sum \hat{\varepsilon}_t \\ \sum x_{t1}\hat{\varepsilon}_t \\ \\ \sum x_{tk}\hat{\varepsilon}_t \end{bmatrix} = 0 \to \text{Normal}$$

$$\hat{\Sigma} = \begin{bmatrix} \hat{\varepsilon}_1\hat{\varepsilon}_1 & \hat{\varepsilon}_1\hat{\varepsilon}_2 & \hat{\varepsilon}_1\hat{\varepsilon}_n \\ \hat{\varepsilon}_2\hat{\varepsilon}_1 & & \\ & & \end{bmatrix}$$

→ weight based on diagonal distance

Noted as w

g is lag

W_g

$w_0 = 1$

$w_g = 1 \qquad g \leq 6$

$w_g = 0 \qquad g > 6$

Newey-west in 1984

↳ $w_g = 1 - \frac{g}{6+1}$

Small sample correction

$\left(\frac{n}{n-k-1}\right)\hat{\Sigma} \to w_g = 1 - \frac{g}{6+1}$

↳ Newey west

$6 = ?$

$6 = \frac{3}{4}T^{1/3}$

12.9

Friday, February 19, 2021 7:02 PM

1) $\{x_t, y_t : t = 1, 2, \ldots, T\}$
   - ↳ covariance stationary
   - ↳ weakly dependent
2) $E(y_t | x_t) = \beta x_t \quad \forall t$
   - ↳ No spec error
   - ↳ $E(r_t | x_t) = 0$
   - -Prediction $E(r_t | x_t) = 0$
   - -Causation by def of PRF
   - $x \perp \tilde{r} = 0$
   - -For causation, need omitted causes uncorrelated with x
     - ↳ Huge Assumption
3) No Perfect Collinearity
4) Homoskedasticity OR Heteroskedasticity robust
5) one of these is true
   1) No serial corr in original model
   2) formulated a dynamically complete transformation
   3) Using serial corr robust std-err (Newey-West)
   4) Explicitly modelled form of serial corr and corrected results

18 Book Notes

Wednesday, February 3, 2021 6:56 PM

Advanced Time Series Topics

17.3 Spurious Regression

There's an extraneous reason for correlation

18.2 Testing For Unit Roots

Unit root process w/ drift very different from one without

Unit root hypothesis
$H_0: \rho = 1$
$H_a: \rho < 1$

$\Theta = \rho - 1$
$\Delta y_t = \alpha + \Theta y_{t-1} + e_t$

asymptotic distribution of t statistic under Ho is
Dickey-Fuller Distribution

Problem Set 1

Gus Lipkin

CAP 4763 Time Series Modelling and Forecasting

> Corrections are underlined

> All uncited quotes are from the Problem Set 1 official solution

Table of Contents

3 Static Model

3a

> Explain why the size of Florida's labor force, the prime age employment to population ratio, and Florida building permits, might be closely related to the number of nonfarm jobs in Florida in a static long run sense. You might want to make some time series plots to give your data context. (Perhaps where one variable is employment and the other, on the other axis, is one of the other variables.)

The size of Florida's labor force can only increase for a few reasons. People either grow up and get a job or people move into the state for one reason or another. These would increase the prime age employment to population ratio but those people need places to work. They could either work in construction or any affiliated field which handles building permits or they could work in a building being constructed by the people handling those permits. In the meantime, as farming becomes more efficient and reliant on technology, not as many people are needed to farm the same parcels of land. This leads to more people employed in non-farm jobs.

"We can think of the number employed as the product of the portion of those in the labor market that are employed and the number that want work and so are in the market. Then the log of total employment is the sum of the logs of those two pieces. From there:

- The number that want to work should closely track labor force in Florida.
- The fraction of those that want to be employed that are employed tracks the strength of the Florida economy, which closely tracks the strength of the national economy, for which the employment to population ratio is a good proxy.
- Construction is a large part of Florida's economic base, due to constant in-migration. So, variations in the strength of the economy may be reflected somewhat in building permits."

3b

> Estimate the static model relating monthly nonfarm employment in Florida to the other three variables (all in logs) without controlling for seasonal impacts or a time trend.

Source	SS	df	MS Number of obs =	396
	F(3, 392) =	5972.65		
Model	10.5356085	3	3.51186951 Prob > F =	0.0000
Residual	.230492978	392	.000587992 R-squared =	0.9786
	Adj R-squared =	0.9784		
Total	10.7661015	395	.027255953 Root MSE =	.02425
ln_fl_nonf~m	Coef.	Std. Err.	t P>t [95% Conf.	Interval]
ln_fl_lf	1.110504	.0092305	120.31 0.000 1.092356	1.128651
ln_us_epr	.6006702	.047797	12.57 0.000 .5066997	.6946407
ln_fl_bp	.0516831	.0028713	18.00 0.000 .0460379	.0573282
_cons	-11.78364	.2925244	-40.28 0.000 -12.35875	-11.20852

3c

> Estimate the static model with month indicators and a time trend.

Source	SS	df	MS Number of obs =	396
	F(15, 380) =	2935.69		
Model	10.6739911	15	.711599408 Prob > F =	0.0000
Residual	.092110398	380	.000242396 R-squared =	0.9914
	Adj R-squared =	0.9911		
Total	10.7661015	395	.027255953 Root MSE =	.01557
ln_fl_nonf~m	Coef.	Std. Err.	t P>t [95% Conf.	Interval]
ln_fl_lf	.9282631	.0413265	22.46 0.000 .8470059	1.00952

ln_us_epr	.9105558	.0514333	17.70 0.000 .8094263	1.011685
ln_fl_bp	.0466812	.0021579	21.63 0.000 .0424382	.0509242
month				
2	.0045623	.0038378	1.19 0.235 -.0029837	.0121084
3	-.001379	.003839	-0.36 0.720 -.0089274	.0061694
4	-.0029373	.0038393	-0.77 0.445 -.0104863	.0046116
5	-.0142748	.0038468	-3.71 0.000 -.0218384	-.0067112
6	-.0356123	.0038709	-9.20 0.000 -.0432234	-.0280012
7	-.0519102	.0038917	-13.34 0.000 -.0595622	-.0442582
8	-.0380965	.0038668	-9.85 0.000 -.0456995	-.0304936
9	-.026004	.0038581	-6.74 0.000 -.0335899	-.0184181
10	-.0215894	.0038763	-5.57 0.000 -.029211	-.0139678
11	-.0014672	.0039082	-0.38 0.708 -.0091517	.0062173
12	.0054514	.0038735	1.41 0.160 -.0021648	.0130675
date	.0003124	.0000637	4.90 0.000 .000187	.0004377
_cons	-10.26323	.498888	-20.57 0.000 -11.24416	-9.282304

3d

> Compare your results from b and c and interpret any differences. What do the seasonal and time trend variables contribute?

Adding the seasonal and time trend variables transform the data into true time series data and give context to the changes. From both you can see that there is a general increase in nonfarm employment. However, by adding the month indicators, you can see that nonfarm employment decreases ever so slightly from March to November, presumably due to prime farming season. "All three coefficients change slightly. The time trend controls for growth at a constant rate over time, while the month indicators control for seasonality. For example, construction employment varies with the weather, employment always varies with holidays, and in Florida employment also varies with tourist season. Presumably, controlling for these effects allows the model to better reveal the underlying relationships between the other variables. (The caveat is we have not checked this data for stationarity or weak dependence, which comes later.)"

3e

> Why should you be cautious using the results of these models for testing any hypotheses about the underlying relationships?

In time series data, the past affects the future and observations are not independent. Standard error and p-value assume that your data is independent which we just established time series data is not.

4 Finite Distributed Lag Model

4a

> Estimate the distributed lag model relating monthly nonfarm employment to lags 0 to 12 of the three predictor variables without month indicators and a time trend.

Source	**SS**	**df**	**MS Number of obs =**	**384**
	F(39, 344) =	1506.36		
Model	9.45063897	39	.242324076 Prob > F =	0.0000
Residual	.055338456	344	.000160868 R-squared =	0.9942
	Adj R-squared =	0.9935		
Total	9.50597742	383	.024819784 Root MSE =	.01268
ln_fl_nonf~m	Coef.	Std. Err.	t P>t [95% Conf.	Interval]
ln_fl_lf				
-.	-.3180953	.2192272	-1.45 0.148 -.7492898	.1130992
L1.	-.4936055	.2780395	-1.78 0.077 -1.040477	.0532661
L2.	.3085466	.27846	1.11 0.269 -.239152	.8562452
L3.	1.173922	.2948363	3.98 0.000 .5940134	1.753831
L4.	-.2346487	.2905929	-0.81 0.420 -.8062113	.3369138
L5.	.2808166	.2958343	0.95 0.343 -.3010552	.8626884
L6.	-.2076341	.3372426	-0.62 0.539 -.8709511	.4556829
L7.	.428488	.3391507	1.26 0.207 -.2385821	1.095558

L8.	.4803611	.3332665	1.44 0.150 -.1751354	1.135858
L9.	.2977526	.3112925	0.96 0.339 -.3145235	.9100288
L10.	-.00028	.3217814	-0.00 0.999 -.6331867	.6326267
L11.	-.5860114	.3256137	-1.80 0.073 -1.226456	.0544331
L12.	.0176351	.2499574	0.07 0.944 -.4740021	.5092724
ln_us_epr				
-.	1.180441	.1573579	7.50 0.000 .8709364	1.489946
L1.	.2435207	.202013	1.21 0.229 -.1538155	.6408569
L2.	-.1519264	.2015081	-0.75 0.451 -.5482695	.2444166
L3.	-.719111	.2119425	-3.39 0.001 -1.135977	-.3022447
L4.	.1877102	.2014654	0.93 0.352 -.2085489	.5839692
L5.	-.1596306	.206881	-0.77 0.441 -.5665414	.2472803
L6.	.4937537	.2396216	2.06 0.040 .0224458	.9650615
L7.	-.3031484	.236988	-1.28 0.202 -.7692764	.1629796
L8.	-.2995254	.2312056	-1.30 0.196 -.7542801	.1552293
L9.	.5953076	.2915942	2.04 0.042 .0217756	1.16884
L10.	-.1656984	.352639	-0.47 0.639 -.8592984	.5279015
L11.	.5326939	.3523697	1.51 0.132 -.1603764	1.225764
L12.	-.4280274	.2543508	-1.68 0.093 -.928306	.0722511
ln_fl_bp				
-.	.0177815	.0051888	3.43 0.001 .0075758	.0279872
L1.	.0056999	.0054688	1.04 0.298 -.0050566	.0164565
L2.	.0123023	.0056879	2.16 0.031 .0011149	.0234898
L3.	-.0005041	.0058381	-0.09 0.931 -.0119871	.0109788
L4.	-.0040248	.0058282	-0.69 0.490 -.0154881	.0074385

L5.	.0053648	.0058106	0.92 0.357 -.006064	.0167937
L6.	.0122019	.0057914	2.11 0.036 .0008108	.0235929
L7.	.0146252	.0057698	2.53 0.012 .0032766	.0259737
L8.	.0114715	.0057663	1.99 0.047 .0001299	.0228131
L9.	.0100892	.0057895	1.74 0.082 -.0012981	.0214765
L10.	-.0077443	.0056515	-1.37 0.171 -.0188601	.0033715
L11.	-.0129284	.0055227	-2.34 0.020 -.0237908	-.002066
L12.	-.0156324	.0052843	-2.96 0.003 -.0260261	-.0052388
_cons	-14.00483	.220126	-63.62 0.000 -14.43779	-13.57187

4b

> Estimate the model in (a) but add month indicators and a time trend.

Source	**SS**	**df**	**MS Number of obs =**	**384**
	F(51, 332) =	1880.48		
Model	9.47318331	51	.185748692 Prob > F =	0.0000
Residual	.03279411	332	.000098777 R-squared =	0.9966
	Adj R-squared =	0.9960		
Total	9.50597742	383	.024819784 Root MSE =	.00994
ln_fl_nonf~m	Coef.	Std. Err.	t P>t [95% Conf.	Interval]
ln_fl_lf				
-.	.1395258	.2167149	0.64 0.520 -.2867817	.5658333
L1.	-.0728475	.2909974	-0.25 0.802 -.6452787	.4995837
L2.	-.0401378	.2914261	-0.14 0.891 -.6134123	.5331367
L3.	.4941867	.3004728	1.64 0.101 -.096884	1.085257
L4.	.0243743	.3032608	0.08 0.936 -.5721806	.6209291

L5.	-.0515457	.3007867	-0.17 0.864 -.6432337	.5401424
L6.	.2645611	.3042172	0.87 0.385 -.3338753	.8629975
L7.	.3032209	.3064496	0.99 0.323 -.2996069	.9060486
L8.	.0945934	.3058001	0.31 0.757 -.5069567	.6961435
L9.	-.1097755	.3559336	-0.31 0.758 -.8099451	.590394
L10.	.1539543	.375505	0.41 0.682 -.5847148	.8926234
L11.	-.2776778	.3787638	-0.73 0.464 -1.022757	.4674017
L12.	-.0112724	.279864	-0.04 0.968 -.5618026	.5392579
ln_us_epr				
-.	.8902343	.1499136	5.94 0.000 .595334	1.185135
L1.	.0725186	.1976025	0.37 0.714 -.3161923	.4612294
L2.	.0146862	.1973291	0.07 0.941 -.3734868	.4028593
L3.	-.3099001	.2109421	-1.47 0.143 -.7248517	.1050514
L4.	.137028	.215249	0.64 0.525 -.2863958	.5604519
L5.	-.0073661	.2142714	-0.03 0.973 -.4288668	.4141346
L6.	.0293898	.2200462	0.13 0.894 -.4034709	.4622504
L7.	-.1397223	.2227059	-0.63 0.531 -.5778149	.2983702
L8.	-.0598893	.2228997	-0.27 0.788 -.4983631	.3785844
L9.	.4823653	.4060878	1.19 0.236 -.3164642	1.281195
L10.	.0335197	.4684115	0.07 0.943 -.887909	.9549485
L11.	.4443457	.4733678	0.94 0.349 -.4868327	1.375524
L12.	-.3652099	.3457533	-1.06 0.292 -1.045353	.3149335
ln_fl_bp				
-.	.0174185	.0043812	3.98 0.000 .0088	.0260369
L1.	.0097915	.0047176	2.08 0.039 .0005113	.0190717

L2.	.005989	.0048174	1.24 0.215 -.0034873	.0154654
L3.	.0067099	.0049382	1.36 0.175 -.0030042	.016424
L4.	.0015463	.0049663	0.31 0.756 -.0082232	.0113157
L5.	.0025978	.0049914	0.52 0.603 -.007221	.0124166
L6.	.006001	.0049798	1.21 0.229 -.0037949	.0157968
L7.	.0066017	.0049157	1.34 0.180 -.003068	.0162715
L8.	-.0015491	.0049371	-0.31 0.754 -.011261	.0081628
L9.	.0010036	.0048898	0.21 0.838 -.0086153	.0106225
L10.	-.0004773	.0047767	-0.10 0.920 -.0098737	.008919
L11.	-.0083937	.0046846	-1.79 0.074 -.017609	.0008216
L12.	-.0041455	.0044702	-0.93 0.354 -.0129391	.004648
month				
2	.0077995	.0048077	1.62 0.106 -.001658	.017257
3	.0052085	.0041637	1.25 0.212 -.0029821	.0133991
4	-.0010198	.0053356	-0.19 0.849 -.0115156	.009476
5	-.0012298	.0047478	-0.26 0.796 -.0105694	.0081098
6	-.0122415	.0055844	-2.19 0.029 -.0232267	-.0012563
7	-.0240128	.0047031	-5.11 0.000 -.0332644	-.0147612
8	-.0152756	.0052483	-2.91 0.004 -.0255997	-.0049514
9	-.0111308	.0045365	-2.45 0.015 -.0200548	-.0022068
10	-.0046899	.006722	-0.70 0.486 -.0179129	.0085332
11	.0076979	.0057763	1.33 0.184 -.0036649	.0190607
12	.0151789	.0059337	2.56 0.011 .0035065	.0268514
date	.0003695	.000047	7.86 0.000 .000277	.0004619
_cons	-11.28083	.391293	-28.83 0.000 -12.05055	-10.5111

4d

> Compare your results from a and c and interpret any differences. What do the seasonal and time trend variables contribute?

The model in 4a is accurate to the data it was given but does not make sense and has no practical application because the data is not organized in any way and does not account for the data being time series data. "This is largely the same as it was for question 3. The difference is that since we are controlling for one year ago, the lags themselves may capture some of the seasonal difference in the first model, and that adding seasonal effects purges that, changing the results potentially at all lags. This, though, it just more of the same basic thing."

4e

> Estimate two alternative models that contain month indicators and a time trend but that impose a more parsimonious lag structure for the predictor variables. Explain your choices.

4e Sampling each quarter

Source	**SS**	**df**	**MS Number of obs =**	**384**
	F(24, 359) =	3636.67		
Model	9.46703767	24	.394459903 Prob > F =	0.0000
Residual	.038939751	359	.000108467 R-squared =	0.9959
	Adj R-squared =	0.9956		
Total	9.50597742	383	.024819784 Root MSE =	.01041
ln_fl_nonf~m	Coef.	Std. Err.	t P>t [95% Conf.	Interval]
ln_fl_lf				
-.	.2198644	.118892	1.85 0.065 -.0139479	.4536767
L4.	.3640379	.1628088	2.24 0.026 .0438591	.6842168
L8.	.6241057	.1697337	3.68 0.000 .2903084	.957903
L12.	-.3365352	.1300465	-2.59 0.010 -.592284	-.0807864

ln_us_epr				
-.	.8706823	.0862833	10.09 0.000 .7009981	1.040367
L4.	.0186581	.1180743	0.16 0.875 -.213546	.2508623
L8.	-.1364675	.1363531	-1.00 0.318 -.4046187	.1316838
L12.	.4492816	.1055542	4.26 0.000 .2416993	.6568639
ln_fl_bp				
-.	.0288326	.0033225	8.68 0.000 .0222986	.0353666
L4.	.014784	.0040692	3.63 0.000 .0067816	.0227864
L8.	.0053046	.0040599	1.31 0.192 -.0026795	.0132888
L12.	-.0040886	.0034865	-1.17 0.242 -.0109452	.002768
month				
2	.003724	.0027268	1.37 0.173 -.0016384	.0090864
3	.003428	.0030747	1.11 0.266 -.0026188	.0094747
4	-.0013812	.0030302	-0.46 0.649 -.0073404	.0045779
5	-.0050709	.003101	-1.64 0.103 -.0111693	.0010275
6	-.0215379	.0030889	-6.97 0.000 -.0276125	-.0154633
7	-.0356678	.0033321	-10.70 0.000 -.0422208	-.0291149
8	-.0202856	.0032905	-6.16 0.000 -.0267567	-.0138145
9	-.0118143	.0031977	-3.69 0.000 -.0181028	-.0055257
10	-.0142884	.0031129	-4.59 0.000 -.0204102	-.0081666
11	-.0033333	.0030634	-1.09 0.277 -.0093578	.0026912
12	.0070509	.0028963	2.43 0.015 .001355	.0127468
date	.0004262	.0000476	8.96 0.000 .0003326	.0005197
_cons	-10.60852	.3857432	-27.50 0.000 -11.36712	-9.849916

4e True Quarters

Source	SS	df	MS Number of obs =	392
	F(27, 364) =	2505.01		
Model	10.2740552	27	.380520563 Prob > F =	0.0000
Residual	.055292923	364	.000151904 R-squared =	0.9946
	Adj R-squared =	0.9942		
Total	10.3293481	391	.02641777 Root MSE =	.01232
ln_fl_nonf~m	Coef.	Std. Err.	t P>t [95% Conf.	Interval]
ln_fl_lf				
-.	.2790757	.2536131	1.10 0.272 -.2196552	.7778065
L1.	.2956093	.3348151	0.88 0.378 -.3628055	.9540241
L2.	-.2641756	.3153608	-0.84 0.403 -.8843334	.3559822
L3.	.2832334	.3167687	0.89 0.372 -.339693	.9061598
L4.	.3220421	.2469667	1.30 0.193 -.1636186	.8077029
ln_us_epr				
-.	.869919	.1763402	4.93 0.000 .5231456	1.216693
L1.	-.1508318	.2303364	-0.65 0.513 -.6037889	.3021253
L2.	.1899043	.2170821	0.87 0.382 -.2369882	.6167968
L3.	-.2262386	.2208671	-1.02 0.306 -.6605744	.2080971
L4.	.3389032	.1751932	1.93 0.054 -.0056147	.6834212
ln_fl_bp				
-.	.0204443	.0051347	3.98 0.000 .010347	.0305417
L1.	.0107528	.0054657	1.97 0.050 4.39e-06	.0215012

L2.	.0026867	.0054899	0.49 0.625 -.0081091	.0134826
L3.	.0070439	.0054993	1.28 0.201 -.0037706	.0178583
L4.	.0071123	.0051692	1.38 0.170 -.003053	.0172777
month				
2	.0052225	.0038186	1.37 0.172 -.0022868	.0127318
3	.0086375	.0041006	2.11 0.036 .0005735	.0167014
4	.0012736	.0046541	0.27 0.785 -.0078787	.0104258
5	.0022027	.0038771	0.57 0.570 -.0054216	.0098269
6	-.0193223	.0040672	-4.75 0.000 -.0273206	-.0113241
7	-.0362039	.0038883	-9.31 0.000 -.0438502	-.0285575
8	-.0245188	.0043528	-5.63 0.000 -.0330787	-.015959
9	-.0171602	.0037189	-4.61 0.000 -.0244733	-.0098471
10	-.0193132	.0044175	-4.37 0.000 -.0280001	-.0106262
11	-.004866	.0041178	-1.18 0.238 -.0129637	.0032317
12	.0058531	.0039007	1.50 0.134 -.0018177	.0135238
dateQ	.0010375	.0001584	6.55 0.000 .0007259	.001349
_cons	-10.55687	.4171884	-25.30 0.000 -11.37727	-9.736468

4e Explanation

I was curious to know how sampling lag for a single month from each quarter for a year would compare to generating a new quarter date variable and using that for lag. Unfortunately, I don't think I did it right and I don't know how to get what I want. Instead, what I have for the second chart is quarterly dates but the lagged variables are now only lagged for the first four months of the year.

Based on the MSE of each model, the first one is a little bit better but I don't think either is great.

"The most important lags would seem to be the most recent month, and the same month a year ago."

Appendix A

```
clear
set more off

cd "/Users/guslipkin/Documents/Spring2020/CAP 4763 ~ Time Series/Problem Sets/Problem Set 1"

*2b Load the data
import delimited "Assignment_1_Monthly.txt"

rename lnu02300000 us_epr
rename flnan fl_nonfarm
rename fllfn fl_lf
rename flbppriv fl_bp
rename date datestring

*2c Turn on a log file
log using "Problem Set 1", replace

*2d Generate a monthly date variable (make its display format monthly time, %tm)
gen datec=date(datestring, "YMD")
gen date=mofd(datec)
format date %tm

*2e tsset your data
tsset date

*2f
gen ln_us_epr=log(us_epr)
gen ln_fl_nonfarm=log(fl_nonfarm)
gen ln_fl_lf=log(fl_lf)
gen ln_fl_bp=log(fl_bp)

*3b Estimate the static model relating monthly nonfarm employment in Florida to the other three variables (all in logs) without controlling for seasonal impacts or a time trend.
regress ln_fl_nonfarm ln_fl_lf ln_us_epr ln_fl_bp

*3c Estimate the static model with month indicators and a time trend.
gen month=month(datec)
reg ln_fl_nonfarm ln_fl_lf ln_us_epr ln_fl_bp i.month date

*4a Estimate the distributed lag model relating monthly nonfarm employment to lags 0 to 12 of the three predictor variables without month indicators and a time trend.
regress ln_fl_nonfarm l(0/12).ln_fl_lf l(0/12).ln_us_epr l(0/12).ln_fl_bp

*4b Estimate the model in (a) but add month indicators and a time trend.
```

```
regress ln_fl_nonfarm l(0/12).ln_fl_lf l(0/12).ln_us_epr l(0/12).ln_fl_bp i.month
date

*4e Estimate two alternative models that contain month indicators and a time trend
but that impose a more parsimonious lag structure for the predictor variables.
Explain your choices.
regress ln_fl_nonfarm l(0,4,8,12).ln_fl_lf l(0,4,8,12).ln_us_epr l(0,4,8,12).ln_fl_bp
i.month date
gen dateQ = qofd(datec)
format dateQ %tq
regress ln_fl_nonfarm l(0/4).ln_fl_lf l(0/4).ln_us_epr l(0/4).ln_fl_bp i.month dateQ

log close
```

Appendix B

```
      name:  <unnamed>
       log:  /Users/guslipkin/Documents/Spring2020/CAP 4763 ~ Time Series/Problem Sets/P
> roblem Set 1/Problem Set 1.smcl
  log type:  smcl
 opened on:  11 Feb 2021, 19:36:36

. 
. *2d Generate a monthly date variable (make its display format monthly time, %tm)
. gen datec=date(datestring, "YMD")

. gen date=mofd(datec)

. format date %tm

. 
. *2e tsset your data
. tsset date
        time variable:  date, 1939m1 to 2020m12
                delta:  1 month

. 
. *2f
. gen ln_us_epr=log(us_epr)
(108 missing values generated)

. gen ln_fl_nonfarm=log(fl_nonfarm)

. gen ln_fl_lf=log(fl_lf)
(444 missing values generated)

. gen ln_fl_bp=log(fl_bp)
(588 missing values generated)

. 
. *3b Estimate the static model relating monthly nonfarm employment in Florida to the ot
> her three variables (all in logs) without controlling for seasonal impacts or a time t
> rend.
```

```
. regress ln_fl_nonfarm ln_fl_lf ln_us_epr ln_fl_bp
```

Source	SS	df	MS	Number of obs	=	396
				F(3, 392)	=	5972.65
Model	10.5356085	3	3.51186951	Prob > F	=	0.0000
Residual	.230492978	392	.000587992	R-squared	=	0.9786
				Adj R-squared	=	0.9784
Total	10.7661015	395	.027255953	Root MSE	=	.02425

ln_fl_nonf~m	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_lf	1.110504	.0092305	120.31	0.000	1.092356	1.128651
ln_us_epr	.6006702	.047797	12.57	0.000	.5066997	.6946407
ln_fl_bp	.0516831	.0028713	18.00	0.000	.0460379	.0573282
_cons	-11.78364	.2925244	-40.28	0.000	-12.35875	-11.20852

```
.
. *3c Estimate the static model with month indicators and a time trend.
. gen month=month(datec)

. reg ln_fl_nonfarm ln_fl_lf ln_us_epr ln_fl_bp i.month date
```

Source	SS	df	MS	Number of obs	=	396
				F(15, 380)	=	2935.69
Model	10.6739911	15	.711599408	Prob > F	=	0.0000
Residual	.092110398	380	.000242396	R-squared	=	0.9914
				Adj R-squared	=	0.9911
Total	10.7661015	395	.027255953	Root MSE	=	.01557

ln_fl_nonf~m	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_lf	.9282631	.0413265	22.46	0.000	.8470059	1.00952
ln_us_epr	.9105558	.0514333	17.70	0.000	.8094263	1.011685
ln_fl_bp	.0466812	.0021579	21.63	0.000	.0424382	.0509242
month						
2	.0045623	.0038378	1.19	0.235	-.0029837	.0121084
3	-.001379	.003839	-0.36	0.720	-.0089274	.0061694
4	-.0029373	.0038393	-0.77	0.445	-.0104863	.0046116
5	-.0142748	.0038468	-3.71	0.000	-.0218384	-.0067112
6	-.0356123	.0038709	-9.20	0.000	-.0432234	-.0280012
7	-.0519102	.0038917	-13.34	0.000	-.0595622	-.0442582
8	-.0380965	.0038668	-9.85	0.000	-.0456995	-.0304936
9	-.026004	.0038581	-6.74	0.000	-.0335899	-.0184181
10	-.0215894	.0038763	-5.57	0.000	-.029211	-.0139678
11	-.0014672	.0039082	-0.38	0.708	-.0091517	.0062173
12	.0054514	.0038735	1.41	0.160	-.0021648	.0130675
date	.0003124	.0000637	4.90	0.000	.000187	.0004377
_cons	-10.26323	.498888	-20.57	0.000	-11.24416	-9.282304

```
.
. *4a Estimate the distributed lag model relating monthly nonfarm employment to lags 0 t
```

```
> o 12 of the three predictor variables without month indicators and a time trend.
. regress ln_fl_nonfarm l(0/12).ln_fl_lf l(0/12).ln_us_epr l(0/12).ln_fl_bp
```

Source	SS	df	MS	Number of obs	=	384
				F(39, 344)	=	1506.36
Model	9.45063897	39	.242324076	Prob > F	=	0.0000
Residual	.055338456	344	.000160868	R-squared	=	0.9942
				Adj R-squared	=	0.9935
Total	9.50597742	383	.024819784	Root MSE	=	.01268

ln_fl_nonf~m	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_lf						
--.	-.3180953	.2192272	-1.45	0.148	-.7492898	.1130992
L1.	-.4936055	.2780395	-1.78	0.077	-1.040477	.0532661
L2.	.3085466	.27846	1.11	0.269	-.239152	.8562452
L3.	1.173922	.2948363	3.98	0.000	.5940134	1.753831
L4.	-.2346487	.2905929	-0.81	0.420	-.8062113	.3369138
L5.	.2808166	.2958343	0.95	0.343	-.3010552	.8626884
L6.	-.2076341	.3372426	-0.62	0.539	-.8709511	.4556829
L7.	.428488	.3391507	1.26	0.207	-.2385821	1.095558
L8.	.4803611	.3332665	1.44	0.150	-.1751354	1.135858
L9.	.2977526	.3112925	0.96	0.339	-.3145235	.9100288
L10.	-.00028	.3217814	-0.00	0.999	-.6331867	.6326267
L11.	-.5860114	.3256137	-1.80	0.073	-1.226456	.0544331
L12.	.0176351	.2499574	0.07	0.944	-.4740021	.5092724
ln_us_epr						
--.	1.180441	.1573579	7.50	0.000	.8709364	1.489946
L1.	.2435207	.202013	1.21	0.229	-.1538155	.6408569
L2.	-.1519264	.2015081	-0.75	0.451	-.5482695	.2444166
L3.	-.719111	.2119425	-3.39	0.001	-1.135977	-.3022447
L4.	.1877102	.2014654	0.93	0.352	-.2085489	.5839692
L5.	-.1596306	.206881	-0.77	0.441	-.5665414	.2472803
L6.	.4937537	.2396216	2.06	0.040	.0224458	.9650615
L7.	-.3031484	.236988	-1.28	0.202	-.7692764	.1629796
L8.	-.2995254	.2312056	-1.30	0.196	-.7542801	.1552293
L9.	.5953076	.2915942	2.04	0.042	.0217756	1.16884
L10.	-.1656984	.352639	-0.47	0.639	-.8592984	.5279015
L11.	.5326939	.3523697	1.51	0.132	-.1603764	1.225764
L12.	-.4280274	.2543508	-1.68	0.093	-.928306	.0722511
ln_fl_bp						
--.	.0177815	.0051888	3.43	0.001	.0075758	.0279872
L1.	.0056999	.0054688	1.04	0.298	-.0050566	.0164565
L2.	.0123023	.0056879	2.16	0.031	.0011149	.0234898
L3.	-.0005041	.0058381	-0.09	0.931	-.0119871	.0109788
L4.	-.0040248	.0058282	-0.69	0.490	-.0154881	.0074385
L5.	.0053648	.0058106	0.92	0.357	-.006064	.0167937
L6.	.0122019	.0057914	2.11	0.036	.0008108	.0235929
L7.	.0146252	.0057698	2.53	0.012	.0032766	.0259737
L8.	.0114715	.0057663	1.99	0.047	.0001299	.0228131
L9.	.0100892	.0057895	1.74	0.082	-.0012981	.0214765
L10.	-.0077443	.0056515	-1.37	0.171	-.0188601	.0033715
L11.	-.0129284	.0055227	-2.34	0.020	-.0237908	-.002066
L12.	-.0156324	.0052843	-2.96	0.003	-.0260261	-.0052388

_cons	-14.00483	.220126	-63.62	0.000	-14.43779	-13.57187

```
.
. *4b Estimate the model in (a) but add month indicators and a time trend.
. regress ln_fl_nonfarm l(0/12).ln_fl_lf l(0/12).ln_us_epr l(0/12).ln_fl_bp i.month date
```

Source	SS	df	MS	Number of obs	=	384
				F(51, 332)	=	1880.48
Model	9.47318331	51	.185748692	Prob > F	=	0.0000
Residual	.03279411	332	.000098777	R-squared	=	0.9966
				Adj R-squared	=	0.9960
Total	9.50597742	383	.024819784	Root MSE	=	.00994

ln_fl_nonf~m	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_lf						
--.	.1395258	.2167149	0.64	0.520	-.2867817	.5658333
L1.	-.0728475	.2909974	-0.25	0.802	-.6452787	.4995837
L2.	-.0401378	.2914261	-0.14	0.891	-.6134123	.5331367
L3.	.4941867	.3004728	1.64	0.101	-.096884	1.085257
L4.	.0243743	.3032608	0.08	0.936	-.5721806	.6209291
L5.	-.0515457	.3007867	-0.17	0.864	-.6432337	.5401424
L6.	.2645611	.3042172	0.87	0.385	-.3338753	.8629975
L7.	.3032209	.3064496	0.99	0.323	-.2996069	.9060486
L8.	.0945934	.3058001	0.31	0.757	-.5069567	.6961435
L9.	-.1097755	.3559336	-0.31	0.758	-.8099451	.590394
L10.	.1539543	.375505	0.41	0.682	-.5847148	.8926234
L11.	-.2776778	.3787638	-0.73	0.464	-1.022757	.4674017
L12.	-.0112724	.279864	-0.04	0.968	-.5618026	.5392579
ln_us_epr						
--.	.8902343	.1499136	5.94	0.000	.595334	1.185135
L1.	.0725186	.1976025	0.37	0.714	-.3161923	.4612294
L2.	.0146862	.1973291	0.07	0.941	-.3734868	.4028593
L3.	-.3099001	.2109421	-1.47	0.143	-.7248517	.1050514
L4.	.137028	.215249	0.64	0.525	-.2863958	.5604519
L5.	-.0073661	.2142714	-0.03	0.973	-.4288668	.4141346
L6.	.0293898	.2200462	0.13	0.894	-.4034709	.4622504
L7.	-.1397223	.2227059	-0.63	0.531	-.5778149	.2983702
L8.	-.0598893	.2228997	-0.27	0.788	-.4983631	.3785844
L9.	.4823653	.4060878	1.19	0.236	-.3164642	1.281195
L10.	.0335197	.4684115	0.07	0.943	-.887909	.9549485
L11.	.4443457	.4733678	0.94	0.349	-.4868327	1.375524
L12.	-.3652099	.3457533	-1.06	0.292	-1.045353	.3149335
ln_fl_bp						
--.	.0174185	.0043812	3.98	0.000	.0088	.0260369
L1.	.0097915	.0047176	2.08	0.039	.0005113	.0190717
L2.	.005989	.0048174	1.24	0.215	-.0034873	.0154654
L3.	.0067099	.0049382	1.36	0.175	-.0030042	.016424
L4.	.0015463	.0049663	0.31	0.756	-.0082232	.0113157
L5.	.0025978	.0049914	0.52	0.603	-.007221	.0124166
L6.	.006001	.0049798	1.21	0.229	-.0037949	.0157968
L7.	.0066017	.0049157	1.34	0.180	-.003068	.0162715
L8.	-.0015491	.0049371	-0.31	0.754	-.011261	.0081628

L8.	-.0015491	.0049571	-0.31	0.754	-.011261	.0081628
L9.	.0010036	.0048898	0.21	0.838	-.0086153	.0106225
L10.	-.0004773	.0047767	-0.10	0.920	-.0098737	.008919
L11.	-.0083937	.0046846	-1.79	0.074	-.017609	.0008216
L12.	-.0041455	.0044702	-0.93	0.354	-.0129391	.004648
month						
2	.0077995	.0048077	1.62	0.106	-.001658	.017257
3	.0052085	.0041637	1.25	0.212	-.0029821	.0133991
4	-.0010198	.0053356	-0.19	0.849	-.0115156	.009476
5	-.0012298	.0047478	-0.26	0.796	-.0105694	.0081098
6	-.0122415	.0055844	-2.19	0.029	-.0232267	-.0012563
7	-.0240128	.0047031	-5.11	0.000	-.0332644	-.0147612
8	-.0152756	.0052483	-2.91	0.004	-.0255997	-.0049514
9	-.0111308	.0045365	-2.45	0.015	-.0200548	-.0022068
10	-.0046899	.006722	-0.70	0.486	-.0179129	.0085332
11	.0076979	.0057763	1.33	0.184	-.0036649	.0190607
12	.0151789	.0059337	2.56	0.011	.0035065	.0268514
date	.0003695	.000047	7.86	0.000	.000277	.0004619
_cons	-11.28083	.391293	-28.83	0.000	-12.05055	-10.5111

```
.
. *4e Estimate two alternative models that contain month indicators and a time trend but
>  that impose a more parsimonious lag structure for the predictor variables. Explain yo
> ur choices.
. regress ln_fl_nonfarm l(0,4,8,12).ln_fl_lf l(0,4,8,12).ln_us_epr l(0,4,8,12).ln_fl_bp
> i.month date
```

Source	SS	df	MS		
Model	9.46703767	24	.394459903	Number of obs	= 384
Residual	.038939751	359	.000108467	F(24, 359)	= 3636.67
				Prob > F	= 0.0000
Total	9.50597742	383	.024819784	R-squared	= 0.9959
				Adj R-squared	= 0.9956
				Root MSE	= .01041

ln_fl_nonf~m	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_lf						
--.	.2198644	.118892	1.85	0.065	-.0139479	.4536767
L4.	.3640379	.1628088	2.24	0.026	.0438591	.6842168
L8.	.6241057	.1697337	3.68	0.000	.2903084	.957903
L12.	-.3365352	.1300465	-2.59	0.010	-.592284	-.0807864
ln_us_epr						
--.	.8706823	.0862833	10.09	0.000	.7009981	1.040367
L4.	.0186581	.1180743	0.16	0.875	-.213546	.2508623
L8.	-.1364675	.1363531	-1.00	0.318	-.4046187	.1316838
L12.	.4492816	.1055542	4.26	0.000	.2416993	.6568639
ln_fl_bp						
--.	.0288326	.0033225	8.68	0.000	.0222986	.0353666
L4.	.014784	.0040692	3.63	0.000	.0067816	.0227864
L8.	.0053046	.0040599	1.31	0.192	-.0026795	.0132888
L12.	-.0040886	.0034865	-1.17	0.242	-.0109452	.002768

month						
2	.003724	.0027268	1.37	0.173	-.0016384	.0090864
3	.003428	.0030747	1.11	0.266	-.0026188	.0094747
4	-.0013812	.0030302	-0.46	0.649	-.0073404	.0045779
5	-.0050709	.003101	-1.64	0.103	-.0111693	.0010275
6	-.0215379	.0030889	-6.97	0.000	-.0276125	-.0154633
7	-.0356678	.0033321	-10.70	0.000	-.0422208	-.0291149
8	-.0202856	.0032905	-6.16	0.000	-.0267567	-.0138145
9	-.0118143	.0031977	-3.69	0.000	-.0181028	-.0055257
10	-.0142884	.0031129	-4.59	0.000	-.0204102	-.0081666
11	-.0033333	.0030634	-1.09	0.277	-.0093578	.0026912
12	.0070509	.0028963	2.43	0.015	.001355	.0127468
date	.0004262	.0000476	8.96	0.000	.0003326	.0005197
_cons	-10.60852	.3857432	-27.50	0.000	-11.36712	-9.849916

```
. gen dateQ = qofd(datec)

. format dateQ %tq

. regress ln_fl_nonfarm l(0/4).ln_fl_lf l(0/4).ln_us_epr l(0/4).ln_fl_bp i.month dateQ
```

Source	SS	df	MS		
Model	10.2740552	27	.380520563	Number of obs	= 392
				F(27, 364)	= 2505.01
Residual	.055292923	364	.000151904	Prob > F	= 0.0000
				R-squared	= 0.9946
				Adj R-squared	= 0.9942
Total	10.3293481	391	.02641777	Root MSE	= .01232

ln_fl_nonf~m	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_lf						
--.	.2790757	.2536131	1.10	0.272	-.2196552	.7778065
L1.	.2956093	.3348151	0.88	0.378	-.3628055	.9540241
L2.	-.2641756	.3153608	-0.84	0.403	-.8843334	.3559822
L3.	.2832334	.3167687	0.89	0.372	-.339693	.9061598
L4.	.3220421	.2469667	1.30	0.193	-.1636186	.8077029
ln_us_epr						
--.	.869919	.1763402	4.93	0.000	.5231456	1.216693
L1.	-.1508318	.2303364	-0.65	0.513	-.6037889	.3021253
L2.	.1899043	.2170821	0.87	0.382	-.2369882	.6167968
L3.	-.2262386	.2208671	-1.02	0.306	-.6605744	.2080971
L4.	.3389032	.1751932	1.93	0.054	-.0056147	.6834212
ln_fl_bp						
--.	.0204443	.0051347	3.98	0.000	.010347	.0305417
L1.	.0107528	.0054657	1.97	0.050	4.39e-06	.0215012
L2.	.0026867	.0054899	0.49	0.625	-.0081091	.0134826
L3.	.0070439	.0054993	1.28	0.201	-.0037706	.0178583
L4.	.0071123	.0051692	1.38	0.170	-.003053	.0172777
month						
2	.0052225	.0038186	1.37	0.172	-.0022868	.0127318
3	.0086375	.0041006	2.11	0.036	.0005735	.0167014

4	.0012736	.0046541	0.27	0.785	-.0078787	.0104258
5	.0022027	.0038771	0.57	0.570	-.0054216	.0098269
6	-.0193223	.0040672	-4.75	0.000	-.0273206	-.0113241
7	-.0362039	.0038883	-9.31	0.000	-.0438502	-.0285575
8	-.0245188	.0043528	-5.63	0.000	-.0330787	-.015959
9	-.0171602	.0037189	-4.61	0.000	-.0244733	-.0098471
10	-.0193132	.0044175	-4.37	0.000	-.0280001	-.0106262
11	-.004866	.0041178	-1.18	0.238	-.0129637	.0032317
12	.0058531	.0039007	1.50	0.134	-.0018177	.0135238
dateQ	.0010375	.0001584	6.55	0.000	.0007259	.001349
_cons	-10.55687	.4171884	-25.30	0.000	-11.37727	-9.736468

```
.
.
. log close
      name:  <unnamed>
       log:  /Users/guslipkin/Documents/Spring2020/CAP 4763 ~ Time Series/Problem Sets/P
> roblem Set 1/Problem Set 1.smcl
  log type:  smcl
 closed on:  11 Feb 2021, 19:36:37
```

Problem Set 2

Gus Lipkin

CAP 4763 Time Series Modelling and Forecasting

> All underlined portions are the corrections

> All uncited quotes are from the problem set

Table of Contents

Part A

1. Writethemodel $y_t = \alpha + \delta t + \rho y_{t-1} + \beta x_{t-1} + r$ in first differences.

- $\Delta y_t = \delta + \rho \Delta y_{t-1} + \beta \Delta x_{t-1} + \Delta r_t$

2. Suppose after first differencing a model is $\Delta y_t = \delta - \varphi - 2\varphi t + \rho\, \Delta\, y_{t-1} + \beta\, \Delta\, x_{t-1} + \Delta r_t$. What was it before the first difference was taken? (Hint: both t and t^2 are in it.)

- $y_t = \delta t + \varphi t^2 + \varphi t - \varphi + \rho y_{t-1} + \beta x_{t-1} + r_t$ <- WRONG
- $\Delta y_t = \delta - \phi + 2\phi t + \rho \Delta t_{t-1} + \beta \Delta x_{t-1} + \Delta r_t$ <- RIGHT

3. Suppose you are originally interested in the model $y_t = \alpha + \delta t + \rho y_{t-1} + \beta x_{t-1} + r_t$, where $r_t = \gamma r_{t-1} + \varepsilon_t$ and ε_t is an independent random disturbance. Write the dynamically complete model in first differences. Hint: first substitute to make the model dynamically complete, and then take the first difference.

- $y_t = \alpha + \delta t + \rho y_{t-1} + \beta x_{t-1} + \gamma r_{t-1} + \varepsilon_t$ <- WRONG
- $\Delta y_t = \delta + \rho \Delta y_{t-1} + \beta \Delta x_{t-1} + \gamma \Delta r_{t-1} + \Delta \varepsilon_t$ <- WRONG
- $\Delta y_t = \delta(1-\gamma) + (\rho+\gamma)\Delta y_{t-1} - \gamma\rho \Delta y_{t-2} + \beta \Delta x_{t-2} + \varepsilon_t - \varepsilon_{t-1}$ <- RIGHT

Part B

3. Autocorrelation and Weak Dependence

1. Obtain the correlation of each variable with its one period lag.

Variable	Correlation with Lag
lnflnonfarm	.9981
lnfllf	.9994
lnusepr	.9821
lnflbp	.9477

(obs=875)	corr ln_us_epr l1.ln_us_epr
	L.
	ln_us~r *ln_us*~r
ln_us_epr	
-.	1.0000
L1.	0.9758 1.0000

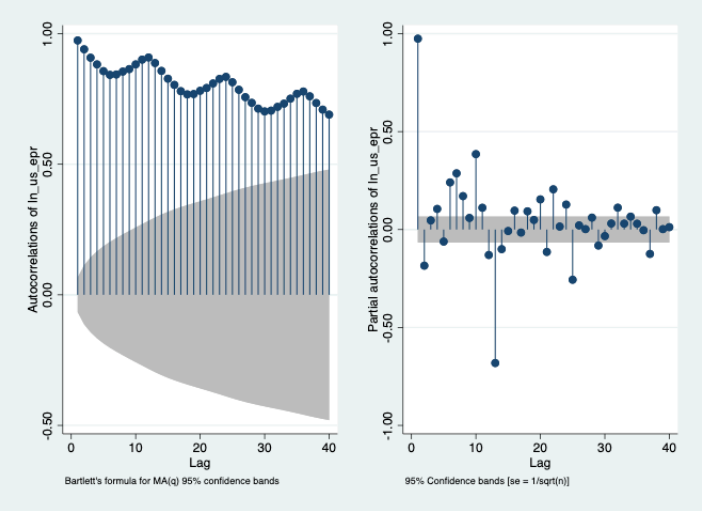
(obs=983)	corr ln_fl_nonfarm l1.ln_fl_nonfarm
	L.
	ln_fl~*m ln_fl*~m
ln_fl_nonf~m	
-.	1.0000
L1.	0.9999 1.0000

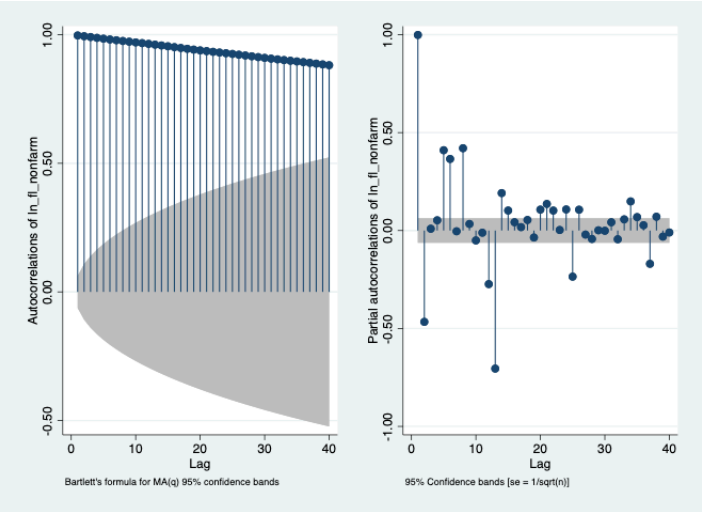
(obs=539)	corr ln_fl_lf l1.ln_fl_lf
	L.
	ln_fl_lf ln_fl_lf
ln_fl_lf	
-.	1.0000
L1.	0.9997 1.0000

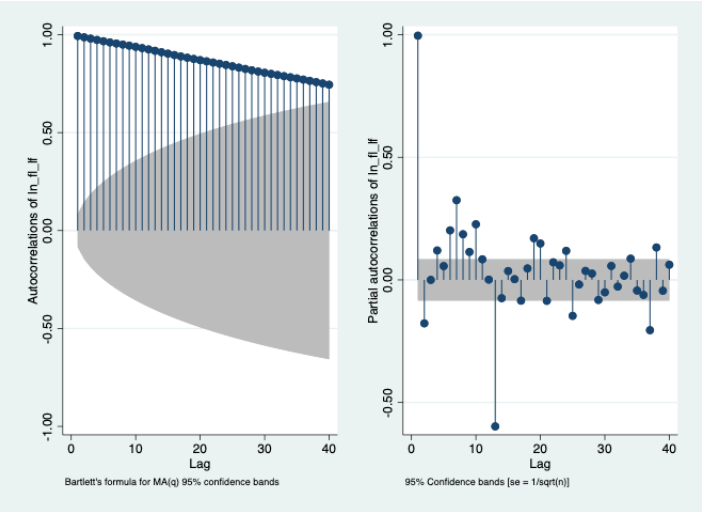
(obs=395)	corr ln_fl_bp l1.ln_fl_bp
	L.
	ln_fl_bp ln_fl_bp
ln_fl_bp	
-.	1.0000
L1.	0.9470 1.0000

- There appears to be very high correlation between the log form of each variable and its first lag. The highest is ln_fl_nonfarm with a correlation of .9999, followed by ln_fl_lf, ln_us_epr, and ln_fl_bp with .9997, .9758, and .9470 respectively.

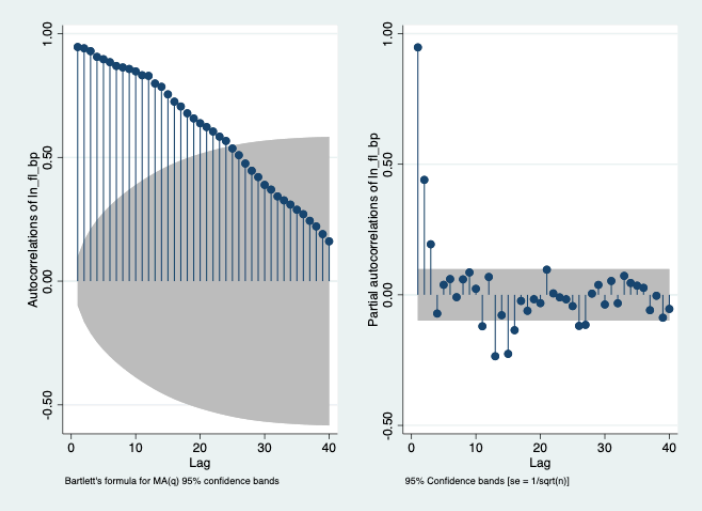
2. Obtain the autocorrelogram and partial autocorrelagram for each variable.







For the above three graphs, because all of the points are outside and above the cone, we can conclude that there is an autoregressive term in the data and should consult the partial autocorrelation graph. The PAC suggests that this is a higher order moving average.



For the last graph, the autocorrelation is not all outside of the confidence interval. When we look at the PAC we see that there are significant correlations in the first few terms followed by insignificant correlations in the rest. This suggests the order of the autoregressive term.

3. Conduct the Dickey-Fuller unit root rest for each variable.

Variable	**Dickey-Fuller p-value**
lnflnonfarm	.0328
lnfllf	.6285
lnusepr	.2246
lnflbp	.7774

```
. dfuller ln_us_epr, trend regress

Dickey-Fuller test for unit root                    Number of obs   =       875

                                 ---------- Interpolated Dickey-Fuller ----------
                    Test         1% Critical       5% Critical      10% Critical
                 Statistic           Value             Value            Value
------------------------------------------------------------------------------
 Z(t)             -4.020            -3.960            -3.410           -3.120
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0082

------------------------------------------------------------------------------
  D.ln_us_epr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
    ln_us_epr |
          L1. |  -.0392314   .0097585    -4.02   0.000    -.0583843   -.0200784
              |
       _trend |   4.02e-06   1.84e-06     2.18   0.030     3.99e-07    7.63e-06
        _cons |   .1583652   .0392952     4.03   0.000     .0812411    .2354894
------------------------------------------------------------------------------
```

```
. dfuller ln_fl_nonfarm, trend regress

Dickey-Fuller test for unit root                   Number of obs   =       983

                         ------------ Interpolated Dickey-Fuller ------------
               Test         1% Critical       5% Critical      10% Critical
             Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)          -0.653            -3.960            -3.410            -3.120
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.9761
```

D. ln_fl_nonfarm	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_nonfarm L1.	-.001659	.0025399	-0.65	0.514	-.0066433	.0033253
_trend	6.93e-07	8.56e-06	0.08	0.935	-.0000161	.0000175
_cons	.0159216	.0160282	0.99	0.321	-.0155318	.0473751

```
. dfuller ln_fl_lf, trend regress

Dickey-Fuller test for unit root                   Number of obs   =       539

                         ------------ Interpolated Dickey-Fuller ------------
               Test         1% Critical       5% Critical      10% Critical
             Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)          -1.724            -3.960            -3.410            -3.120
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.7400
```

D.ln_fl_lf	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_lf L1.	-.0076457	.0044337	-1.72	0.085	-.0163552	.0010639
_trend	7.40e-06	8.61e-06	0.86	0.391	-9.52e-06	.0000243
_cons	.120517	.0676628	1.78	0.075	-.0123997	.2534337

```
. dfuller ln_fl_bp, trend regress

Dickey-Fuller test for unit root                   Number of obs   =       395

                         ------------ Interpolated Dickey-Fuller ------------
               Test         1% Critical       5% Critical      10% Critical
             Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)          -3.256            -3.984            -3.424            -3.130
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0738
```

D.ln_fl_bp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_bp L1.	-.0545463	.0167509	-3.26	0.001	-.0874792	-.0216134
_trend	-.0000375	.0000734	-0.51	0.609	-.0001817	.0001067
_cons	.5091679	.1583766	3.21	0.001	.1977942	.8205417

For both Dickey-Fuller of the ln_us_epr, the p-value is extremely low at .0082 and so we accept the null hypothesis. For all others, we fail to reject the null hypothesis. Especially ln_fl_nonfarm and ln_fl_lf.

4. "Looking at the AC and PAC, all four show strong enough first order autoregressive relationships to merit differencing. We can reject the null of an I(1) process for the log of non-farm employment. But, the partial autocorrelation coefficient is so close to one that we should difference anyway. The AC and PAC for the log difference of non-farm employment are bwlow, illustrating the differences are clearly not I(1)."

4. ARDL Model and Breusch-Godfrey Test

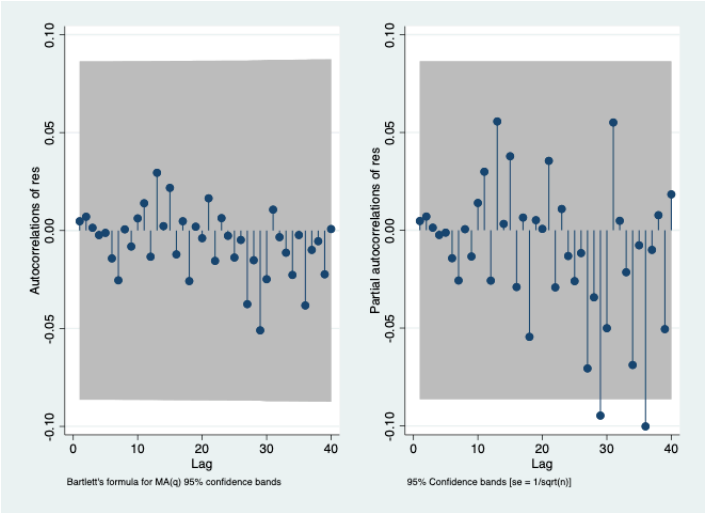
Given the results of the previous question, transform the data as needed and estimate a dynamically complete ARDL model for non-farm employment. Include at least one lag of the relevant dependent variable. How many additional lags of the dependent variable, and how many lags of which independent variables you include, are up to you. Looking back at what you did for Problem Set 1 might be informative, but don't be limited by it. Produce and interpret the AC and PAC for the residuals and the results of a Breusch-Godfrey test. In your write up, justify your specification and interpret the results.

.	**regress d.ln_fl_nonfarm l(1/48)d.ln_fl_nonfarm**	**l(12/24)d.ln_us_epr**	**l(1/18, 24)d.ln_fl_lf**	**date**
	Source SS df MS	Number of obs =	515	
	F(81, 433) =	15.48		
Model .050091055 81 .000618408	Prob > F =	0.0000		
Residual .01729335 433 .000039938	R-squared =	0.7434		
	Adj R-squared =	0.6954		
Total .067384405 514 .000131098	Root MSE =	.00632		
----------	---------	---------	-----	-----
D.				
ln_fl_nonfarm	Coef.	Std. Err.	t	P>t
ln_fl_nonfarm				
LD.	-.1441103	.059346	-2.43	0.016
L2D.	-.1332106	.060728	-2.19	0.029
L3D.	.0520745	.060831	0.86	0.392
L4D.	.1139409	.0609067	1.87	0.062
L5D.	.066288	.0611891	1.08	0.279
L6D.	.1944856	.0614959	3.16	0.002
L7D.	.0759452	.0622902	1.22	0.223
L8D.	.0829208	.0631492	1.31	0.190
L9D.	.2532911	.0930772	2.72	0.007
L10D.	.1403499	.0960901	1.46	0.145
L11D.	.1893271	.0946093	2.00	0.046
L12D.	.4685154	.0957577	4.89	0.000

L13D.	.0758492	.1003991	0.76	0.450
L14D.	.0089228	.1008964	0.09	0.930
L15D.	.0490602	.1006788	0.49	0.626
L16D.	-.0187785	.1013922	-0.19	0.853
L17D.	.0547956	.1017669	0.54	0.591
L18D.	.0863921	.1011552	0.85	0.394
L19D.	-.25835	.1016689	-2.54	0.011
L20D.	-.1621826	.1009034	-1.61	0.109
L21D.	-.0839614	.1033319	-0.81	0.417
L22D.	-.1719582	.1017154	-1.69	0.092
L23D.	.0347504	.1011416	0.34	0.731
L24D.	.2927769	.0998811	2.93	0.004
L25D.	.1178616	.098203	1.20	0.231
L26D.	.0999885	.0980021	1.02	0.308
L27D.	-.1283723	.0980801	-1.31	0.191
L28D.	-.2031139	.0980964	-2.07	0.039
L29D.	-.2892074	.097907	-2.95	0.003
L30D.	-.5772115	.0991658	-5.82	0.000
L31D.	.6236058	.1020615	6.11	0.000
L32D.	.1870999	.1073141	1.74	0.082
L33D.	.1426809	.1091241	1.31	0.192
L34D.	.1068341	.1078243	0.99	0.322
L35D.	-.0794067	.1078368	-0.74	0.462
L36D.	.1327386	.1064489	1.25	0.213
L37D.	-.0639028	.099194	-0.64	0.520
L38D.	-.048562	.0984536	-0.49	0.622
L39D.	.0871388	.0975069	0.89	0.372
L40D.	-.1442082	.0974565	-1.48	0.140
L41D.	-.0032331	.0966638	-0.03	0.973
L42D.	.0938246	.0970599	0.97	0.334
L43D.	-.3559573	.0966539	-3.68	0.000
L44D.	-.0089124	.0978207	-0.09	0.927

L45D.	-.0882528	.0966085	-0.91	0.361
L46D.	.1086727	.091884	1.18	0.238
L47D.	.0313382	.091654	0.34	0.733
L48D.	.0609195	.091323	0.67	0.505
ln_us_epr				
L12D.	-.0155085	.1744885	-0.09	0.929
L13D.	-.3056076	.153451	-1.99	0.047
L14D.	-.5608006	.1545155	-3.63	0.000
L15D.	-.3645519	.1519838	-2.40	0.017
L16D.	.0029936	.1580302	0.02	0.985
L17D.	.0422232	.1559561	0.27	0.787
L18D.	.3199335	.1565006	2.04	0.042
L19D.	-.07463	.0988972	-0.75	0.451
L20D.	.0625226	.0999685	0.63	0.532
L21D.	-.0436852	.1002131	-0.44	0.663
L22D.	.2231831	.0985078	2.27	0.024
L23D.	-.0081188	.0960409	-0.08	0.933
L24D.	-.2688582	.1616447	-1.66	0.097
ln_fl_lf				
LD.	.1762398	.0704433	2.50	0.013
L2D.	-.1356975	.0715783	-1.90	0.059
L3D.	-.1659446	.0715828	-2.32	0.021
L4D.	-.0977864	.0709175	-1.38	0.169
L5D.	-.1364495	.0722069	-1.89	0.059
L6D.	-.2270642	.0723796	-3.14	0.002
L7D.	-.1332104	.0724525	-1.84	0.067
L8D.	-.2396185	.0727056	-3.30	0.001
L9D.	-.1256755	.079465	-1.58	0.114
L10D.	-.180737	.0797732	-2.27	0.024
L11D.	-.005726	.0808095	-0.07	0.944

L12D.	.0558537	.1334055	0.42	0.676
L13D.	.0173463	.1262683	0.14	0.891
L14D.	.2969825	.1275491	2.33	0.020
L15D.	.125207	.1266497	0.99	0.323
L16D.	-.0665773	.1288379	-0.52	0.606
L17D.	-.1292395	.1273895	-1.01	0.311
L18D.	-.2883037	.1278108	-2.26	0.025
L24D.	.2278015	.1255369	1.81	0.070
date	-8.65e-06	3.19e-06	-2.72	0.007
_cons	.0058495	.0022338	2.62	0.009



I don't think there's any correlation because almost everything is inside the interval.

. estat bgodfrey, lag(1/48)	
Breusch-Godfrey LM test for	autocorrelation

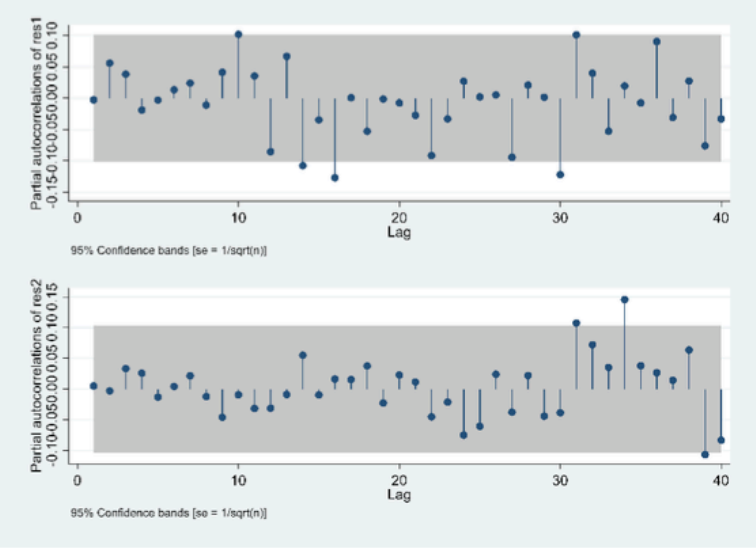
lags(p)	chi2	df	Prob > chi2
1	0.617	1	0.4321
2	1.630	2	0.4427

3	1.639	3	0.6506
4	1.665	4	0.7970
5	1.730	5	0.8850
6	2.757	6	0.8387
7	8.252	7	0.3109
8	8.536	8	0.3830
9	8.707	9	0.4648
10	8.803	10	0.5509
11	9.015	11	0.6205
12	10.913	12	0.5364
13	12.697	13	0.4715
14	12.775	14	0.5443
15	14.075	15	0.5198
16	15.212	16	0.5091
17	15.284	17	0.5751
18	18.315	18	0.4351
19	18.317	19	0.5014
20	19.893	20	0.4647
21	19.920	21	0.5263
22	20.203	22	0.5704
23	20.218	23	0.6287
24	20.362	24	0.6760
25	21.112	25	0.6864
26	21.381	26	0.7221
27	23.290	27	0.6693
28	24.359	28	0.6624
29	25.322	29	0.6615

30	27.716	30	0.5855
31	28.706	31	0.5846
32	28.728	32	0.6330
33	29.272	33	0.6533
34	30.894	34	0.6207
35	30.897	35	0.6666
36	33.834	36	0.5720
37	35.071	37	0.5597
38	35.519	38	0.5847
39	38.229	39	0.5049
40	38.448	40	0.5402
41	38.548	41	0.5801
42	39.001	42	0.6034
43	39.107	43	0.6408
44	39.122	44	0.6804
45	39.431	45	0.7061
46	39.812	46	0.7278
47	40.011	47	0.7550
48	40.617	48	0.7664
H0: no serial correlation			

4) Given the results of the previous question, transform the data as needed and estimate a dynamically complete ARDL model for non-farm employment. Include at least one lag of the relevant dependent variable. How many additional lags of the dependent variable, and how many lags of which independent variables you include, are up to you. Looking back at what you did for Problem Set 1 might be informative, but don't be limited by it. Produce and interpret the AC and PAC for the residuals and the results of a Breusch-Godfrey test. In your write up, justify your specification and interpret the results.

I estimated two models, one with all lags back 12 months and one going back 24 months. Breush-Godfrey test results are in the table below. In the first case, the null of no serial correlation is rejected. For the second, the null can't be rejected at 24 lags, but it neither is it convincingly rejected (p=0.16). However, examining the PACs for the residuals in the figure below gives a bit more confidence in the second model. It also suggests some lags from year 3 and 4 may be worth including.



Breush-Godfrey tests for question 4

	p > chi	
Lags	Model 1	Model 2
1	0.8812	0.4861
2	0.0332	0.778
3	0.0074	0.0585
4	0.0129	0.0386
5	0.0266	0.0709
6	0.0475	0.0774
7	0.0787	0.1049
8	0.0453	0.1426
9	0.068	0.1042
10	0.0467	0.1464
11	0.0688	0.1728
12	0.005	0.2121
13	0.0035	0.2321
14	0.0047	0.1206
15	0.0064	0.1304
16	0.0007	0.1483
17	0.0012	0.1816
18	0.0019	0.2039
19	0.0028	0.2119
20	0.0037	0.2138
21	0.0056	0.255
22	0.006	0.2065
23	0.0066	0.2485
24	0.0079	0.1618

A more parsimonious model, possibly with selected lags out further, might be a good idea. However, with some careful thought and exploration, I still have not come up with one that passed a Breusch-Godfrey test. Perhaps you did… Really, we will need more model selection tools to help us choose if we want to forecast. If we need to estimate parameters, we need to choose the appropriate model for the purpose, even if not dynamically complete, and then use appropriately adjusted standard errors. That is the point of the next problem.

5. Dynamically Complete Models and Newey-West Standard Errors

```
. reg d.ln_fl_nonfarm l(0/4)d.ln_fl_bp if tin(1948m1,2020m1)
```

Source	SS	df	MS	Number of obs	=	380
				F(5, 374)	=	2.97
Model	.00146591	5	.000293182	Prob > F	=	0.0122
Residual	.036972226	374	.000098856	R-squared	=	0.0381
				Adj R-squared	=	0.0253
Total	.038438136	379	.00010142	Root MSE	=	.00994

D. ln_fl_nonf~m	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_bp						
D1.	-.0043445	.0035864	-1.21	0.227	-.0113965	.0027075
LD.	-.0115113	.0040594	-2.84	0.005	-.0194935	-.0035291
L2D.	.0019871	.0041056	0.48	0.629	-.0060858	.01006
L3D.	-.0011778	.0040768	-0.29	0.773	-.0091941	.0068385
L4D.	-.0028262	.0036121	-0.78	0.434	-.0099287	.0042763
_cons	.0015358	.0005101	3.01	0.003	.0005328	.0025387

```
. newey d.ln_fl_nonfarm l(0/4)d.ln_fl_bp if tin(1948m1,2020m1), lag(4)
```

```
Regression with Newey-West standard errors      Number of obs     =        380
maximum lag: 4                                  F(  5,       374) =       4.01
                                                Prob > F          =     0.0015
```

D. ln_fl_nonf~m	Coef.	Newey-West Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_bp						
D1.	-.0043445	.003622	-1.20	0.231	-.0114665	.0027776
LD.	-.0115113	.0036606	-3.14	0.002	-.0187093	-.0043133
L2D.	.0019871	.0043475	0.46	0.648	-.0065616	.0105358
L3D.	-.0011778	.004813	-0.24	0.807	-.0106416	.008286
L4D.	-.0028262	.003664	-0.77	0.441	-.0100308	.0043783
_cons	.0015358	.0004154	3.70	0.000	.0007189	.0023526

if fuller high, can't reject

5) Suppose you are interested in the relationship between the first difference in non-farm employment and the lags 0 to 4 of the differences of Florida building permits, controlling for seasonal impacts, but not controlling for any other variables or lags, including lags of employment. That is, you explicitly do not want to a dynamically complete model. (Don't worry about why, for this purpose.) Estimate the model both with and without Newey-West standard errors and discuss the difference that makes.

The results of interest are in the table at right. Note that the Newey-West standard errors are larger for the first three coefficients and smaller for the last. The regular standard errors are misleading regarding the precision of the estimates.

Models for question 5

Std Err	Regular	Newey-West
D.lnflbp	0.00820***	0.00820**
	(0.00203)	(0.00250)
LD.lnflbp	0.00793***	0.00793**
	(0.00236)	(0.00294)
L2D.lnflbp	0.00627*	0.00627
	(0.00244)	(0.00348)
L3D.lnflbp	0.00730**	0.00730*
	(0.00237)	(0.00300)
L4D.lnflbp	0.00430*	0.00430*
	(0.00204)	(0.00199)
N	379	379
R^2	0.764	

Standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$
Constant, trend, and month coefficients omitted for space

Appendix A

```
clear
set more off

cd "/Users/guslipkin/Documents/Spring2020/CAP 4763 ~ Time Series/Problem Sets/Problem Set 2"

*2a
*Done

*2b Load the data
import delimited "Assignment_1_Monthly.txt"

rename lnu02300000 us_epr
rename flnan fl_nonfarm
rename fllfn fl_lf
rename flbppriv fl_bp
rename date datestring
```

```
17
18  *2c Turn on a log file
19  log using "Problem Set 1", replace
20
21  *2d Generate a monthly date variable (make its display format monthly time, %tm)
22  gen datec=date(datestring, "YMD")
23  gen date=mofd(datec)
24  format date %tm
25
26  *2e tsset your data
27  tsset date
28
29  *2f
30  gen ln_us_epr=log(us_epr)
31  gen ln_fl_nonfarm=log(fl_nonfarm)
32  gen ln_fl_lf=log(fl_lf)
33  gen ln_fl_bp=log(fl_bp)
34
35  *3a
36  corr ln_us_epr l1.ln_us_epr
37  corr ln_fl_nonfarm l1.ln_fl_nonfarm
38  corr ln_fl_lf l1.ln_fl_lf
39  corr ln_fl_bp l1.ln_fl_bp
40
41  *3b
42  ac ln_us_epr, saving(ac_ln_us_epr.gph, replace)
43  pac ln_us_epr, saving(pac_ln_us_epr.gph, replace)
44  graph combine ac_ln_us_epr.gph pac_ln_us_epr.gph, saving(combo_ln_us_epr.gph,
    replace)
45
46  ac ln_fl_nonfarm, saving(ac_ln_fl_nonfarm.gph, replace)
47  pac ln_fl_nonfarm, saving(pac_ln_fl_nonfarm.gph, replace)
48  graph combine ac_ln_fl_nonfarm.gph pac_ln_fl_nonfarm.gph,
    saving(combo_ln_fl_nonfarm.gph, replace)
49
50  ac ln_fl_lf, saving(ac_ln_fl_lf.gph, replace)
51  pac ln_fl_lf, saving(pac_ln_fl_lf.gph, replace)
52  graph combine ac_ln_fl_lf.gph pac_ln_fl_lf.gph, saving(combo_ln_fl_lf.gph, replace)
53
54  ac ln_fl_bp, saving(ac_ln_fl_bp.gph, replace)
55  pac ln_fl_bp, saving(pac_ln_fl_bp.gph, replace)
56  graph combine ac_ln_fl_bp.gph pac_ln_fl_bp.gph, saving(combo_ln_fl_bp.gph, replace)
57
58  *3c
59  dfuller ln_us_epr, trend regress
60  dfuller ln_fl_nonfarm, trend regress
61  dfuller ln_fl_lf, trend regress
62  dfuller ln_fl_bp, trend regress
63
64  *4
```

```
regress d.ln_fl_nonfarm l(1/48)d.ln_fl_nonfarm l(12/24)d.ln_us_epr l(1/18,
24)d.ln_fl_lf date
predict res, residual
ac res, saving(p4_ac.gph, replace)
pac res, saving(p4_pac.gph, replace)
graph combine p4_ac.gph p4_pac.gph, saving(p4_combo.gph, replace)
estat bgodfrey, lag(1/48)

*5
reg d.ln_fl_nonfarm l(0/4)d.ln_fl_bp if tin(1948m1,2020m1)
newey d.ln_fl_nonfarm l(0/4)d.ln_fl_bp if tin(1948m1,2020m1), lag(4)

log close
```

Appendix B

```
      name:  <unnamed>
       log:  /Users/guslipkin/Documents/Spring2020/CAP 4763 ~ Time Series/Problem S
> ets/Problem Set 2/Problem Set 1.smcl
  log type:  smcl
 opened on:  26 Feb 2021, 18:08:56

. 
. *2d Generate a monthly date variable (make its display format monthly time, %tm)
. gen datec=date(datestring, "YMD")

. gen date=mofd(datec)

. format date %tm

. 
. *2e tsset your data
. tsset date
        time variable:  date, 1939m1 to 2020m12
                delta:  1 month

. 
. *2f
. gen ln_us_epr=log(us_epr)
(108 missing values generated)

. gen ln_fl_nonfarm=log(fl_nonfarm)

. gen ln_fl_lf=log(fl_lf)
(444 missing values generated)

. gen ln_fl_bp=log(fl_bp)
(588 missing values generated)

. 
. *3a
. corr ln_us_epr l1.ln_us_epr
(obs=875)

             |                 L.
             | ln_us_~r ln_us_~r
-------------+------------------
   ln_us_epr |
          --.|   1.0000
          L1.|   0.9758   1.0000


. corr ln_fl_nonfarm l1.ln_fl_nonfarm
(obs=983)

             |                 L.
             | ln_fl_~m ln_fl_~m
-------------+------------------
ln_fl_nonf~m |
          --.|   1.0000
          L1.|   0.9999   1.0000


. corr ln_fl_lf l1.ln_fl_lf
(obs=539)

             |                 L.
             | ln_fl_lf ln_fl_lf
-------------+------------------
    ln_fl_lf |
```

```
        --. |   1.0000
        L1. |   0.9997   1.0000


. corr ln_fl_bp l1.ln_fl_bp
(obs=395)

             |               L.
             | ln_fl_bp ln_fl_bp
-------------+------------------
    ln_fl_bp |
         --. |   1.0000
         L1. |   0.9470   1.0000


.
. *3b
. ac ln_us_epr, saving(ac_ln_us_epr.gph, replace)
(file ac_ln_us_epr.gph saved)

. pac ln_us_epr, saving(pac_ln_us_epr.gph, replace)
(file pac_ln_us_epr.gph saved)

. graph combine ac_ln_us_epr.gph pac_ln_us_epr.gph, saving(combo_ln_us_epr.gph, rep
> lace)
(file combo_ln_us_epr.gph saved)

.
. ac ln_fl_nonfarm, saving(ac_ln_fl_nonfarm.gph, replace)
(file ac_ln_fl_nonfarm.gph saved)

. pac ln_fl_nonfarm, saving(pac_ln_fl_nonfarm.gph, replace)
(file pac_ln_fl_nonfarm.gph saved)

. graph combine ac_ln_fl_nonfarm.gph pac_ln_fl_nonfarm.gph, saving(combo_ln_fl_nonf
> arm.gph, replace)
(file combo_ln_fl_nonfarm.gph saved)

.
. ac ln_fl_lf, saving(ac_ln_fl_lf.gph, replace)
(file ac_ln_fl_lf.gph saved)

. pac ln_fl_lf, saving(pac_ln_fl_lf.gph, replace)
(file pac_ln_fl_lf.gph saved)

. graph combine ac_ln_fl_lf.gph pac_ln_fl_lf.gph, saving(combo_ln_fl_lf.gph, replac
> e)
(file combo_ln_fl_lf.gph saved)

.
. ac ln_fl_bp, saving(ac_ln_fl_bp.gph, replace)
(file ac_ln_fl_bp.gph saved)

. pac ln_fl_bp, saving(pac_ln_fl_bp.gph, replace)
(file pac_ln_fl_bp.gph saved)

. graph combine ac_ln_fl_bp.gph pac_ln_fl_bp.gph, saving(combo_ln_fl_bp.gph, replac
> e)
(file combo_ln_fl_bp.gph saved)

.
. *3c
. dfuller ln_us_epr, trend regress

Dickey-Fuller test for unit root                    Number of obs   =       875

                         ------------ Interpolated Dickey-Fuller ------------
               Test         1% Critical       5% Critical      10% Critical
             Statistic           Value             Value            Value
------------------------------------------------------------------------------
 Z(t)            -4.020            -3.960            -3.410           -3.120
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0082

------------------------------------------------------------------------------
 D.ln_us_epr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   ln_us_epr |
         L1. |  -.0392314   .0097585    -4.02   0.000    -.0583843   -.0200784
             |
       _trend |   4.02e-06   1.84e-06     2.18   0.030     3.99e-07    7.63e-06
        _cons |   .1583652   .0392952     4.03   0.000     .0812411    .2354894
------------------------------------------------------------------------------

. dfuller ln_fl_nonfarm, trend regress

Dickey-Fuller test for unit root                    Number of obs   =       983

                         ------------ Interpolated Dickey-Fuller ------------
               Test         1% Critical       5% Critical      10% Critical
             Statistic           Value             Value            Value
------------------------------------------------------------------------------
 Z(t)            -0.653            -3.960            -3.410           -3.120
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.9761
```

D. ln_fl_nonfarm	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_nonfarm						
L1.	-.001659	.0025399	-0.65	0.514	-.0066433	.0033253
_trend	6.93e-07	8.56e-06	0.08	0.935	-.0000161	.0000175
_cons	.0159216	.0160282	0.99	0.321	-.0155318	.0473751

```
. dfuller ln_fl_lf, trend regress

Dickey-Fuller test for unit root                    Number of obs   =       539

                                 ---------- Interpolated Dickey-Fuller ----------
                 Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value            Value
------------------------------------------------------------------------------
 Z(t)            -1.724            -3.960            -3.410           -3.120
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.7400
```

D.ln_fl_lf	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_lf						
L1.	-.0076457	.0044337	-1.72	0.085	-.0163552	.0010639
_trend	7.40e-06	8.61e-06	0.86	0.391	-9.52e-06	.0000243
_cons	.120517	.0676628	1.78	0.075	-.0123997	.2534337

```
. dfuller ln_fl_bp, trend regress

Dickey-Fuller test for unit root                    Number of obs   =       395

                                 ---------- Interpolated Dickey-Fuller ----------
                 Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value            Value
------------------------------------------------------------------------------
 Z(t)            -3.256            -3.984            -3.424           -3.130
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0738
```

D.ln_fl_bp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_bp						
L1.	-.0545463	.0167509	-3.26	0.001	-.0874792	-.0216134
_trend	-.0000375	.0000734	-0.51	0.609	-.0001817	.0001067
_cons	.5091679	.1583766	3.21	0.001	.1977942	.8205417

```
.
. *4
. regress d.ln_fl_nonfarm l(1/48)d.ln_fl_nonfarm l(12/24)d.ln_us_epr l(1/18, 24)d.l
> n_fl_lf date
```

Source	SS	df	MS		
Model	.050091055	81	.000618408	Number of obs	= 515
Residual	.01729335	433	.000039938	F(81, 433)	= 15.48
				Prob > F	= 0.0000
Total	.067384405	514	.000131098	R-squared	= 0.7434
				Adj R-squared	= 0.6954
				Root MSE	= .00632

D. ln_fl_nonfarm	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_nonfarm						
LD.	-.1441103	.059346	-2.43	0.016	-.2607523	-.0274683
L2D.	-.1332106	.060728	-2.19	0.029	-.2525689	-.0138524
L3D.	.0520745	.060831	0.86	0.392	-.0674863	.1716354
L4D.	.1139409	.0609067	1.87	0.062	-.0057687	.2336504
L5D.	.066288	.0611891	1.08	0.279	-.0539766	.1865526
L6D.	.1944856	.0614959	3.16	0.002	.073618	.3153532
L7D.	.0759452	.0622902	1.22	0.223	-.0464836	.1983741
L8D.	.0829208	.0631492	1.31	0.190	-.0411963	.207038
L9D.	.2532911	.0930772	2.72	0.007	.0703519	.4362303
L10D.	.1403499	.0960901	1.46	0.145	-.0485112	.329211
L11D.	.1893271	.0946093	2.00	0.046	.0033766	.3752776
L12D.	.4685154	.0957577	4.89	0.000	.2803077	.6567232
L13D.	.0758492	.1003991	0.76	0.450	-.1214811	.2731795
L14D.	.0089228	.1008964	0.09	0.930	-.1893847	.2072303
L15D.	.0490602	.1006788	0.49	0.626	-.1488197	.24694
L16D.	-.0187785	.1013922	-0.19	0.853	-.2180605	.1805035
L17D.	.0547956	.1017669	0.54	0.591	-.1452228	.2548141
L18D.	.0863921	.1011552	0.85	0.394	-.1124241	.2852084
L19D.	-.25835	.1016689	-2.54	0.011	-.4581759	-.0585241
L20D.	-.1621826	.1009034	-1.61	0.109	-.360504	.0361389
L21D.	-.0839614	.1033319	-0.81	0.417	-.2870559	.1191331
L22D.	-.1719582	.1017154	-1.69	0.092	-.3718755	.0279592

L23D.	.0347504	.1011416	0.34	0.731	-.1640391	.2335399
L24D.	.2927769	.0998811	2.93	0.004	.0964647	.489089
L25D.	.1178616	.098203	1.20	0.231	-.0751523	.3108754
L26D.	.0999885	.0980021	1.02	0.308	-.0926304	.2926074
L27D.	-.1283723	.0980801	-1.31	0.191	-.3211445	.0644
L28D.	-.2031139	.0980964	-2.07	0.039	-.3959182	-.0103096
L29D.	-.2892074	.097907	-2.95	0.003	-.4816395	-.0967753
L30D.	-.5772115	.0991658	-5.82	0.000	-.7721176	-.3823054
L31D.	.6236058	.1020615	6.11	0.000	.4230083	.8242034
L32D.	.1870999	.1073141	1.74	0.082	-.0238215	.3980212
L33D.	.1426809	.1091241	1.31	0.192	-.0717978	.3571596
L34D.	.1068341	.1078243	0.99	0.322	-.1050899	.3187581
L35D.	-.0794067	.1078368	-0.74	0.462	-.2913554	.1325421
L36D.	.1327386	.1064489	1.25	0.213	-.0764823	.3419594
L37D.	-.0639028	.099194	-0.64	0.520	-.2588645	.1310589
L38D.	-.048562	.0984536	-0.49	0.622	-.2420684	.1449445
L39D.	.0871388	.0975069	0.89	0.372	-.104507	.2787845
L40D.	-.1442082	.0974565	-1.48	0.140	-.3357548	.0473384
L41D.	-.0032331	.0966638	-0.03	0.973	-.1932218	.1867555
L42D.	.0938246	.0970599	0.97	0.334	-.0969425	.2845917
L43D.	-.3559573	.0966539	-3.68	0.000	-.5459264	-.1659882
L44D.	-.0089124	.0978207	-0.09	0.927	-.2011749	.1833501
L45D.	-.0882528	.0966085	-0.91	0.361	-.2781327	.1016272
L46D.	.1086727	.091884	1.18	0.238	-.0719214	.2892668
L47D.	.0313382	.091654	0.34	0.733	-.1488038	.2114803
L48D.	.0609195	.091323	0.67	0.505	-.118572	.240411
ln_us_epr						
L12D.	-.0155085	.1744885	-0.09	0.929	-.3584584	.3274413
L13D.	-.3056076	.153451	-1.99	0.047	-.607209	-.0040062
L14D.	-.5608006	.1545155	-3.63	0.000	-.8644942	-.257107
L15D.	-.3645519	.1519838	-2.40	0.017	-.6632696	-.0658341
L16D.	.0029936	.1580302	0.02	0.985	-.3076081	.3135954
L17D.	.0422232	.1559561	0.27	0.787	-.264302	.3487483
L18D.	.3199335	.1565006	2.04	0.042	.0123381	.6275288
L19D.	-.07463	.0988972	-0.75	0.451	-.2690083	.1197484
L20D.	.0625226	.0999685	0.63	0.532	-.1339614	.2590065
L21D.	-.0436852	.1002131	-0.44	0.663	-.2406498	.1532795
L22D.	.2231831	.0985078	2.27	0.024	.0295703	.416796
L23D.	-.0081188	.0960409	-0.08	0.933	-.1968832	.1806456
L24D.	-.2688582	.1616447	-1.66	0.097	-.5865639	.0488476
ln_fl_lf						
LD.	.1762398	.0704433	2.50	0.013	.0377865	.3146932
L2D.	-.1356975	.0715783	-1.90	0.059	-.2763815	.0049866
L3D.	-.1659446	.0715828	-2.32	0.021	-.3066375	-.0252517
L4D.	-.0977864	.0709175	-1.38	0.169	-.2371718	.041599
L5D.	-.1364495	.0722069	-1.89	0.059	-.278369	.00547
L6D.	-.2270642	.0723796	-3.14	0.002	-.3693234	-.0848051
L7D.	-.1332104	.0724525	-1.84	0.067	-.2756127	.0091919
L8D.	-.2396185	.0727056	-3.30	0.001	-.3825182	-.0967187
L9D.	-.1256755	.079465	-1.58	0.114	-.2818605	.0305095
L10D.	-.180737	.0797732	-2.27	0.024	-.3375278	-.0239463
L11D.	-.005726	.0808095	-0.07	0.944	-.1645537	.1531017
L12D.	.0558537	.1334055	0.42	0.676	-.2063492	.3180565
L13D.	.0173463	.1262683	0.14	0.891	-.2308286	.2655213
L14D.	.2969825	.1275491	2.33	0.020	.0462901	.547675
L15D.	.125207	.1266497	0.99	0.323	-.1237177	.3741316
L16D.	-.0665773	.1288379	-0.52	0.606	-.3198027	.186648
L17D.	-.1292395	.1273895	-1.01	0.311	-.3796182	.1211391
L18D.	-.2883037	.1278108	-2.26	0.025	-.5395104	-.037097
L24D.	.2278015	.1255369	1.81	0.070	-.018936	.4745389
date	-8.65e-06	3.19e-06	-2.72	0.007	-.0000149	-2.39e-06
_cons	.0058495	.0022338	2.62	0.009	.0014592	.0102399

```
. predict res, residual
(469 missing values generated)

. ac res, saving(p4_ac.gph, replace)
(file p4_ac.gph saved)

. pac res, saving(p4_pac.gph, replace)
(file p4_pac.gph saved)

. graph combine p4_ac.gph p4_pac.gph, saving(p4_combo.gph, replace)
(file p4_combo.gph saved)

. estat bgodfrey, lag(1/48)

Breusch-Godfrey LM test for autocorrelation
```

lags(p)	chi2	df	Prob > chi2
1	0.617	1	0.4321
2	1.630	2	0.4427
3	1.639	3	0.6506
4	1.665	4	0.7970
5	1.730	5	0.8850
6	2.757	6	0.8387
7	8.252	7	0.3109
8	8.536	8	0.3830
9	8.707	9	0.4648
10	8.803	10	0.5509

10	8.008	10	0.6307
11	9.015	11	0.6205
12	10.913	12	0.5364
13	12.697	13	0.4715
14	12.775	14	0.5443
15	14.075	15	0.5198
16	15.212	16	0.5091
17	15.284	17	0.5751
18	18.315	18	0.4351
19	18.317	19	0.5014
20	19.893	20	0.4647
21	19.920	21	0.5263
22	20.203	22	0.5704
23	20.218	23	0.6287
24	20.362	24	0.6760
25	21.112	25	0.6864
26	21.381	26	0.7221
27	23.290	27	0.6693
28	24.359	28	0.6624
29	25.322	29	0.6615
30	27.716	30	0.5855
31	28.706	31	0.5846
32	28.728	32	0.6330
33	29.272	33	0.6533
34	30.894	34	0.6207
35	30.897	35	0.6666
36	33.834	36	0.5720
37	35.071	37	0.5597
38	35.519	38	0.5847
39	38.229	39	0.5049
40	38.448	40	0.5402
41	38.548	41	0.5801
42	39.001	42	0.6034
43	39.107	43	0.6408
44	39.122	44	0.6804
45	39.431	45	0.7061
46	39.812	46	0.7278
47	40.011	47	0.7550
48	40.617	48	0.7664

H0: no serial correlation

```
.
. *5
. reg d.ln_fl_nonfarm l(0/4)d.ln_fl_bp if tin(1948m1,2020m1)
```

Source	SS	df	MS	Number of obs	=	380
				F(5, 374)	=	2.97
Model	.00146591	5	.000293182	Prob > F	=	0.0122
Residual	.036972226	374	.000098856	R-squared	=	0.0381
				Adj R-squared	=	0.0253
Total	.038438136	379	.00010142	Root MSE	=	.00994

D. ln_fl_nonf~m	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_bp						
D1.	-.0043445	.0035864	-1.21	0.227	-.0113965	.0027075
LD.	-.0115113	.0040594	-2.84	0.005	-.0194935	-.0035291
L2D.	.0019871	.0041056	0.48	0.629	-.0060858	.01006
L3D.	-.0011778	.0040768	-0.29	0.773	-.0091941	.0068385
L4D.	-.0028262	.0036121	-0.78	0.434	-.0099287	.0042763
_cons	.0015358	.0005101	3.01	0.003	.0005328	.0025387

```
. newey d.ln_fl_nonfarm l(0/4)d.ln_fl_bp if tin(1948m1,2020m1), lag(4)

Regression with Newey-West standard errors      Number of obs     =        380
maximum lag: 4                                  F(  5,       374) =       4.01
                                                Prob > F          =     0.0015
```

D. ln_fl_nonf~m	Coef.	Newey-West Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_fl_bp						
D1.	-.0043445	.003622	-1.20	0.231	-.0114665	.0027776
LD.	-.0115113	.0036606	-3.14	0.002	-.0187093	-.0043133
L2D.	.0019871	.0043475	0.46	0.648	-.0065616	.0105358
L3D.	-.0011778	.004813	-0.24	0.807	-.0106416	.008286
L4D.	-.0028262	.003664	-0.77	0.441	-.0100308	.0043783
_cons	.0015358	.0004154	3.70	0.000	.0007189	.0023526

```
.
. log close
      name:  <unnamed>
       log:  /Users/guslipkin/Documents/Spring2020/CAP 4763 ~ Time Series/Problem S
> ets/Problem Set 2/Problem Set 1.smcl
  log type:  smcl
 closed on:  26 Feb 2021, 18:09:11
```

Gus Lipkin - CAP 4763 Time Series Midterm

> **Corrections are marked by underlines**

> 1. Express the model above in first differences. Under what conditions would you need to work with the differenced model instead of the original?

- $List_t = \beta_0 + \beta_P lnPermits_{t-1} + \beta_{Int} lnInterest_{t-1} + \beta_{Inf} lnInflation_{t-1} + r_t$
- $\Delta List_t = \beta_0 + \beta_P \Delta lnPermits_{t-1} + \beta_{Int} \Delta lnInterest_{t-1} + \beta_{Inf} \Delta lnInflation_{t-1} + \Delta r_t$
- $\underline{\Delta List_t = \beta_0 + \beta_P \Delta lnPermits_{t-1} + \beta_{Int} \Delta lnInterest_{t-1} + \beta_{Inf} \Delta lnInflation_{t-1} + \Delta r_t}$

> 2. Suppose you think the residual, r_t, follows an AR(1) process with parameter ρ. Write the dynamically complete version of the original model ***and*** of the model in first differences.

- $List_t = \beta_0 + \beta_P lnPermits_{t-1} + \beta_{Int} lnInterest_{t-1} + \beta_{Inf} lnInflation_{t-1} + r_t$
- $\Delta List_t = \underline{\beta_0} + \beta_P lnPermits_{t-1} + \beta_{Int} lnInterest_{t-1} + \beta_{Inf} lnInflation_{t-1} + \rho(List_{t-1} - \beta_0 - \beta_P lnPermits_{t-1} - \beta_{Int} lnInterest_{t-1} - \beta_{Inf} lnInflation_{t-1}) + \varepsilon_t$
- $\underline{\Delta List_t = \beta_0 + \beta_P lnPermits_{t-1} + \beta_{Int} lnInterest_{t-1} + \beta_{Inf} lnInflation_{t-1} + \rho(List_{t-1} - \beta_0 - \beta_P lnPermits_{t-1} - \beta_{Int} lnInterest_{t-1} - \beta_{Inf} lnInflation_{t-1}) + \varepsilon_t}$
- That second equation can then be distributed out and further reduced.

> 3. What is the purpose of the commands on page 1 lines 25-32?

- In order the functions do as follows, renaming the *date* variable to *datestring*, generating a new variable called *datec* that is the *datestring* variable in YMD format, generating another variable, *date* that is *datec* in monthly data form, then formatting the *date* variable as a timeseries data-type, then setting the beginning of the time-series data, and finally generating a new *month* variable that is the month of the date.
- We do all of this to make sure that our data is in a format that STATA can work with and so that we have all of the variables we will need later ready to go at the beginning of the analysis.
- "<u>These lines make sure Stata properly recognizes time in the data. A monthly date is set for time related calculations such as lags and differences and a categorical variable is created for month of year (1-12) for capturing seasonal effects.</u>"

> 4. What is the purpose of the commands and results from page 2 line 4 through page 3 line 30, and what conclusion should be drawn from the results of these commands?

- The *ac* command generates an autocorrelogram graph while the *pac* chart generates a partial autocorrelogram. *dfuller* runs a Dickey-Fuller test on the data. The Dickey-Fuller test has an option for *lag(12)* which lets us lag for 12 months (an entire year) of data.
- AC and PAC graphs can be used in conjunction to identify ARIMA models. If a point is significant, it extends beyond the shaded boundary. For the non-differenced models, we see that significance decreases as time passes. We can then look to the PAC chart and see that there is significant correlation in the first lag and correlations that are not significant. This suggests a higher order autorgressive term. In the differenced models, the AC graph spikes are right at the edges of the range. If they are significant, it suggests an autoregressive term, if they are not significant, it suggests a moving average term. The same goes for the PAC but instead the insignificant values suggest autoregressive while the significant

suggest moving average.

- The Dickey-Fuller test is interpreted by its p-value. I don't remember what it does and everything I have says that it tests to see if there is a unit root but I have no clue what a unit root is and can't figure it out. I found this article https://stats.stackexchange.com/questions/29121/intuitive-explanation-of-unit-root which has a very funny joke at the bottom of the accepted answer. If I'm understanding what A.A. Milne 2.0 is saying, the model does not have a unit root because the p-value is less than one, the data will converge back to the same spot.
- "The purpose is to determine whether lnList is exhibits high persistence or only weak dependence, since further analysis required the time series be stationary and weakly dependent. The AC and PAC are consistent with an I(1) process and indicate very high persistence, and the Dickey Fuller test cannot reject the null hypothesis of an I(1) process. The conclusion is that this series demonstrates high persistence and should be differenced before further analysis."

> 5. Four sets of models are estimates. What are the differences between the sets (**not** between the models in a given set)? Which set is better for the purpose at hand? Why?

- Model 1 is only lagged, model 2 is lagged with the date and month indicators, model three is lagged and differenced, model 4 is lagged and differenced with the date and month indicators.
- It's possible that these are AR/DL models in all four forms. None, autoregressive, distributed lag, and autoregressive distributed lag
- I'm going to choose Model Set 3. There's something bothering me suggesting maybe I should choose 4 instead but I'm going to stick with 3. I'm not super confident with why, but I feel okay with it and I'm running out of time. It's the only one where there's ever any enough evidence to reject the null hypothesis of the Breusch-Godfrey test.
- "The differences are a) whether month dummies and a time trend are included and b) whether the data is differenced before estimating the model. The PAC and AC and Dickey Fuller test discussed previously indicate differencing is necessary. Seasonal indicator variables should be used to deal with purely seasonal effects (e.g. weather impacts on home building) unless there is a clear reason not to. Hence, set 4, which does both, is the best of these."

> 6. There are three models within the set you chose. Each of those is estimated twice. What is the difference between the two sets of estimates? Does the difference matter? Why? Which is better? Why?

- The first set of models is estimated only using the first lag. The second set of models uses the first and second lags.
- The difference does matter because it changes how far back the model looks when making its predictions.
- I think the first option is better because it is taking all three variables into account rather than just two. Like I say below, it is best to include all hypothesis variables when testing.
- "For each model in the set, the first estimate uses the command **regress** which calculated default standard errors that assume no serial correlation and no heteroskedasticity in the residuals, while the second uses the command **newey** which calculates standard errors robust to autocorrelation and heteroskedasticity in the residuals. Thus, unless you are very confident your model has no autocorrelation or heteroskedasticity in the residuals, use the second estimate with Newey-West standard errors. Note the Breusch-Godfrey tests suggest autocorrelation remains."

> 7. For the set you chose as best, interpret the F-test for the first model in that set. That is, if set X is

> best, interpret the F-test that follows one of the two estimates of Model X.1. Again, there are two versions. Use the better one. Your answer to 6 should have made it clear what the difference is, which is better, and why.

- (I'm not entirely sure if *testparm* or *estat bgodfrey* is the F-test and I don't have enough time to figure it out. I'm going to assume it's *testparm*. I also don't know what *test* is either...)
  Although the p-value is very close to .05, it is still just above it at .0566. This suggests that we our model does not fit the data as well as it could.
- "This is a test (using the **testparm** command for testing sets of parameters) of the null hypothesis that neither the first lag of inflation nor the first lag of the interest rate have predictive power. The second version is best because it uses calculations robust to autocorrelation and heteroskedasticity. The p-value of 0.3545 indicates there is very little evidence upon which to conclude these variables are predictive of list prices."

> 8. How do the three models in your chosen set relate to the model set out on the previous page and to questions 1 and 2?

- One model is like the original model given, one of them is like the first difference model from problem 1, and one of them is like the autoregressive dynamically complete model from question 2.
- "All are derived from the model set out in the background section. Model 4.1 is in first differences, like in question 1. Model 4.2 is the second equation in question 2, which would be dynamically complete if the residuals of the baseline model were a simple AR(1) process. The third model simply deleted inflation and interest as predictors from Model 4.2 following the test that shows no evidence they contribute predictive power. (If you incorrectly chose a non-differenced set, this would differ slightly)"

> 9. What assumption must be defended to apply a causal interpretation to the results of this model, as opposed to a purely predictive one?

- You must assume that all relevant variables are accounted for and that there are no *omitted variables*. You must also assume that there is no *multicollinearity* in the data. The first means you shouldn't leave out important data and the second means you shouldn't include two pieces of data that are correlated with each other such as the amount of cereal consumed and the amount of milk consumed. Most people consume those items together, so using them both can be redundant and detrimental to the model.
- "No omitted causes of list price are contemporaneously correlated with permits, interest, or inflation."

> 10. Within the set of models you chose as best, X, which model is best for predicting *List*? That is, X.1, X.2, or X.3? Why? Which is best for testing they hypotheses of interest? Why? If the two are different, why?

- I think model 3.3 is the best predictor but 3.2 also looks pretty good. The best for testing the hypothesis is 3.2 because it is the only one that takes the number of building permits, the interest rate, and the inflation rate into account. They could be different because when testing a hypothesis, it is important to test all of the variables discussed in your hypothesis. If I say "high consumption of pizza and beers leads to heart disease", I can't only test if pizza leads to heart disease, I have to test both. That said, it might turn out that pizza is a much better indicator than beers and that beers doesn't add much. In that case, the pizza only model would be better because there is less room for error.

- "For prediction, one could make an argument for either model 4.2 or model 4.3. Model 4.3 is more parsimonious, dropping variables that seem not to have predictive power. But if the content knowledge indicating they should be controlled for is strong, Model 4.2 is better and 4.3 is simply overfit to the data. Model 4.1 is not as useful because it lacks lagged variables that we see have predictive power.

  For causation, the three null hypotheses are that the three coefficients in the original model are zero. The alternative hypotheses are that the coefficient on permits is negative, the coefficient on interest rates is negative, and that the coefficient on inflation is positive. Model 4.3 does not contain all three coefficients, so it simply cannot test these hypotheses. As long as we use the Newey-West standard errors, and can defend the assumption discussed in question 9, we can make an argument for either model 4.1 or model 4.2. The coefficients of the first model are a bit more precisely estimated because the second lags of the three predictors in Model 4.2 don't appear to add anything useful, so I would probably take Model 4.1, which is the direct application of the hypothesized model after differencing. But, if you chose 4.2, that is not a bad answer."