# Econometrics: Multiple Regression and Applications

ECON4004: LAB 5

Duong Trinh

University of Glasgow

March 6, 2024

# Intro

- Duong Trinh
  - PhD Student in Economics (Bayesian Microeconometrics)
  - Email: Duong.Trinh@glasgow.ac.uk
- ECON4004-LB01
  - Wednesday 10am -12 pm
  - 5 sessions (7-Feb, 14-Feb, 21-Feb, 28-Feb, 6-March)
  - ST ANDREWS:357
- ECON4004-LB02
  - Wednesday 12-2 pm
  - 5 sessions (7-Feb, 14-Feb, 21-Feb, 28-Feb, 6-March)
  - ST ANDREWS:357

# Your Attendance & Lab Feedback

# Plan for LAB 5

- Exercise: based on Stock & Watson, E15.1
- We will focus on *"Time series Regression"*

# BRIEF REVIEW

# Time Series Data - What it looks like...

Time series data are data collected on the same observational unit at multiple time periods ($t$).

$$\{y_t\},$$
$$t = 1, \ldots, T$$

Can be of any time frequency - daily, monthly, quarterly, annual, etc.[1]

| freq | GDPC1 | JAPAN_IP | PCECTPI | CPIAUCSL |
|---|---|---|---|---|
| 01/01/1955 | 2683.766 | | 15.755 | 26.793333 |
| 01/04/1955 | 2727.452 | | 15.771 | 26.756667 |
| 01/07/1955 | 2764.128 | | 15.834 | 26.776667 |
| 01/10/1955 | 2780.762 | | 15.878 | 26.856667 |
| 01/01/1956 | 2770.032 | | 15.943 | 26.86 |
| 01/04/1956 | 2792.872 | | 16.051 | 27.036667 |
| 01/07/1956 | 2790.588 | | 16.208 | 27.316667 |
| 01/10/1956 | 2836.238 | | 16.303 | 27.55 |
| 01/01/1957 | 2854.517 | | 16.45 | 27.776667 |
| 01/04/1957 | 2848.186 | | 16.554 | 28.013333 |
| 01/07/1957 | 2875.927 | | 16.687 | 28.263333 |
| 01/10/1957 | 2846.446 | | 16.773 | 28.4 |
| 01/01/1958 | 2772.654 | | 16.979 | 28.736667 |
| 01/04/1958 | 2790.948 | | 17.009 | 28.93 |
| 01/07/1958 | 2855.472 | | 17.023 | 28.913333 |
| 01/10/1958 | 2922.264 | | 17.018 | 28.943333 |
| 01/01/1959 | 2976.629 | | 17.137 | 28.993333 |
| 01/04/1959 | 3049.011 | | 17.204 | 29.043333 |
| 01/07/1959 | 3043.139 | | 17.307 | 29.193333 |
| 01/10/1959 | 3055.104 | | 17.401 | 29.37 |
| 01/01/1960 | 3123.162 | 12.184435 | 17.424 | 29.396667 |
| 01/04/1960 | 3111.31 | 12.676183 | 17.516 | 29.573333 |
| 01/07/1960 | 3119.057 | 13.22257 | 17.583 | 29.59 |
| 01/10/1960 | 3081.3 | 13.850916 | 17.661 | 29.78 |
| 01/01/1961 | 3102.251 | 14.615858 | 17.694 | 29.84 |
| 01/04/1961 | 3159.918 | 15.189564 | 17.692 | 29.83 |

[1] Typical resource: https://fred.stlouisfed.org/

# [SN] Working with dates and times in STATA

Date types in Stata[2]

| Date type | Format | Unit |
|---|---|---|
| Datetime | `%tc` | Milliseconds since 01jan1960 00:00:00.000 |
| Daily date | `%td` | Days since 01jan1960 |
| Weekly date | `%tw` | Weeks since 1960w1 |
| Monthly date | `%tm` | Months since 1960m1 |
| Quarterly date | `%tq` | Quarters since 1960q1 |

[2]See guideline at: https://www.stata.com/bookstore/dtguide.pdf

# [SN] Our Example

```
. * Import Quarterly Data from Excel file
. * first row considered as variable names
. import excel "us_macro_quarterly.xlsx", sheet("Data") firstrow clear
(5 vars, 252 obs)

. describe

Contains data
 Observations:           252
    Variables:             5
------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
------------------------------------------------------------------------
freq            int     %td..                 freq
GDPC1           double  %10.0g                GDPC1
JAPAN_IP        double  %10.0g                JAPAN_IP
PCECTPI         double  %10.0g                PCECTPI
CPIAUCSL        double  %10.0g                CPIAUCSL
------------------------------------------------------------------------
```

# [SN] Creating dates and times in STATA

(I) Option 1: Building dates and times from components.[3]

| Date type | Format | Pseudofunction | Function |
|---|---|---|---|
| Daily date | `%td` | td(day-month-year) | mdy(M, D, Y) |
| Weekly date | `%tw` | tw(year-week) | yw(Y, W) |
| Monthly date | `%tm` | tm(year-month) | ym(Y, M) |
| Quarterly date | `%tq` | tq(year-quarter) | yq(Y, Q) |

[3] See guideline at: https://www.stata.com/bookstore/dtguide.pdf

# [SN] Our Example

```
. * Create a desired quarterly date, e.g. 1955q1
. display %tq tq(1955q1) //using pseudofunction tq(.)
1955q1

. display %tq yq(1955,1) //using function yq(.)
1955q1

.
. * Generate quarterly date variables, recursively starting from 1955q1
. * By default, tq(1960q1) is defined to be 0 in Stata.
. gen date1 = tq(1955q1) + _n-1

. gen date2 = yq(1955,1) + _n-1

. format %tq date1 date2 // express them in quarterly format, see Data Editor

. list if date1 != date2 // check if both variables are identical
```

# [SN] Creating dates and times in STATA

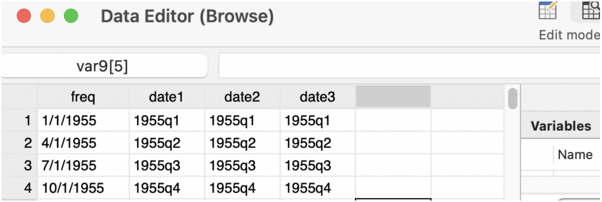
(II) Option 2: Converting dates and times from existing variables.[4]

| From \ To | Daily date | Weekly date | Monthly date | Quarterly date |
|---|---|---|---|---|
| Daily date | | **wofd()** | **mofd()** | **qofd()** |
| Weekly date | **dofw()** | | **mofd(dofw())** | **qofd(dofw())** |
| Monthly date | **dofm()** | **wofd(dofm())** | | **qofd(dofm())** |
| Quarterly date | **dofq()** | **wofd(dofq())** | **mofd(dofq())** | |

[4] See guideline at: https://www.stata.com/bookstore/dtguide.pdf

# [SN] Our Example

```
. * Generate quarterly date variable from existing "freq" var in dataset
. describe freq // note that "freq" is of daily date type (%td)

Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------
freq              int    %td..                freq

. gen date3 = qofd(freq) // convert "freq" to quarterly date type

. format %tq date3 // express it in quarterly format

. br freq date1 date2 date3 // check results, new vars are identical
```





| | freq | date1 | date2 | date3 |
|---|---|---|---|---|
| 1 | 1/1/1955 | 1955q1 | 1955q1 | 1955q1 |
| 2 | 4/1/1955 | 1955q2 | 1955q2 | 1955q2 |
| 3 | 7/1/1955 | 1955q3 | 1955q3 | 1955q3 |
| 4 | 10/1/1955 | 1955q4 | 1955q4 | 1955q4 |

# [SN] STATA command for Setting Data as Time Series

Once having the date variable in a *date format*, we need to declare our data as time series to use the time series operators in Stata.

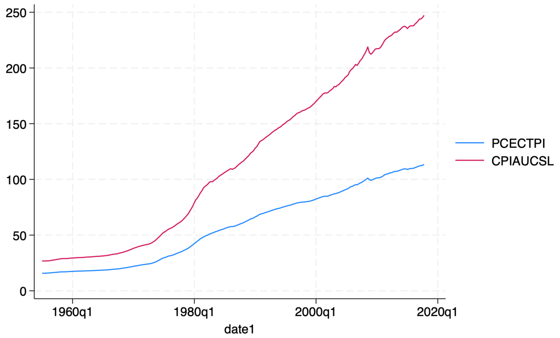
```
*tsset datevar
//choose suitable 'datevar' in the dataset
```

```
. * Declare 'us_macro_quarterly' dataset to be time-series data
. tsset date1

Time variable: date1, 1955q1 to 2017q4
        Delta: 1 quarter
```

# [SN] STATA command for Visualizing Time Series Data

```
*[twoway] tsline varlist_of_inteterest
```

```
. * Line graph of two Price Index series
. twoway tsline PCECTPI CPIAUCSL // create a line graph using declared time-series data

. twoway line PCECTPI CPIAUCSL date1, yline(0) // equivalent graph when using standard command with datevar
```



# [SN] Subsetting tin/twithin in STATA

- `tin` (“times in”, from a to b)
- `twithin` (“times within”, between a and b, it excludes a and b).

# [SN] Time-series operators in STATA

Transforming a time-series y

| | | Stata |
|---|---|---|
| **L operator** | first *lag* $y_{t-1}$ | `L1.y` |
| | 2-period *lag* $y_{t-2}$ | `L2.y` |
| | 3-period *lag* $y_{t-3}$ | `L3.y` |
| **D operator** | first *difference* $\Delta y_t = y_t - y_{t-1}$ | `D1.y` |
| | double *difference* $\Delta y_t - \Delta y_{t-1}$ | `D2.y` |
| | triple *difference* $(\Delta y_t - \Delta y_{t-1}) - (\Delta y_{t-1} - \Delta y_{t-2})$ | `D3.y` |

The first difference of the logarithm of $y_t$ is $\Delta \ln(y_t) = \ln(y_t) - \ln(y_{t-1})$

$$\Delta \ln(y_t) \approx \frac{\text{percentage change of time-series y between } t \text{ and } t-1}{100}$$

(!) approximation is most accurate when the percentage change is small.

# Autocorrelation and Autocovariance

The covariance between $y_t$ and its $j^{\text{th}}$ lag $y_{t-j}$ is called the $j^{\text{th}}$ *autocovariance* of the series $y_t$

$$j^{\text{th}}\text{autocovariance} \equiv Cov(y_t, y_{t-j})$$

The correlation between $y_t$ and its $j^{\text{th}}$ lag $y_{t-j}$ is called the $j^{\text{th}}$ *autocorrelation coefficient*, aka *serial correlation coefficient*

$$j^{\text{th}}\text{autocorrelation} \equiv \rho_j = \rho_{y_t, y_{t-j}} = \frac{Cov(y_t, y_{t-j})}{\sqrt{Var(y_t)Var(y_{t-j})}} \in [-1, 1]$$

$\rightarrow$ Measure how dependent $y_t$ is on its past value $y_{t-j}$.

Autocorrelations of $y$ as a function of its lags $j$ is known as the *autocorrelation function*.

# [SN] Tabulate and Graph Autocorrelations in STATA

Produce a table of the autocorrelations for a time series

```
*corrgram yvar [, lags(p)]
// limit the number of computed autocorrelations to p
```

Plot the autocorrelation function for a time series (correlogram) with pointwise confidence intervals

```
*ac yvar [, lags(p)]
// limit the number of computed autocorrelations to p
```

# [SN] STATA command for Autoregression Model

The autoregressive model of order $p$ - $AR(p)$

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + \epsilon_t$$

To regress $y_t$ against its own lagged values, run one of the followings

```
*regress yvar L1.yvar L2.yvar Lp.yvar [, robust]
```

```
*regress yvar L(1/p).yvar [, robust]
//'p' is the number of lags
```

# [SN] Forecasting in STATA

## Step1. Expanding the Dataset Before Forecasting

Given a set of time-series observations, STATA typically records the dates as running from the first until the last observation.

To forecast a date out-of-sample, these dates need to be in the data set, which requires expanding the dataset to include them.

```
*tsappend, add(k)
//add 'k' dates to the end of the sample
```

## Step2. Point Forecasting Out-of-Sample

To create a series of predicted values both in-sample and out-of-sample, after the `regress` command

```
*predict yvar_hat [if]
//use 'if' condition to restrict the predicted values
```

# Unit Root Test

- Testing for unit roots/stochastic trend is a crucial step in time series analysis. The presence of a unit root in a time series signifies that the series is non-stationary, which can lead to spurious regressions and other potentially serious consequences if no further transformation is made.
  - $H_0$ : $y_t$ has a unit root ($y_t$ is non-stationary).
  - $H_1$ : $y_t$ has no unit root ($y_t$ is stationary).
- The Dickey–Fuller test involves fitting the model

$$y_t = \beta_0 + \rho y_{t-1} + \delta t + u_t$$

  and test $H_0 : \rho = 1$, or, equivalently, that $y_t$ follows a unit root process.
- To control for serial correlation issue, the augmented Dickey–Fuller (ADF) test instead fits

$$\Delta y_t = \beta_0 + \delta y_{t-1} + \alpha t + \gamma_1 \Delta y_{t-1} + \ldots + \gamma_k \Delta y_{t-k} + \epsilon_t$$

  and test $H_0 : \delta = 0$.

Note: Possible restrictions: $\beta_0 = 0$ (without drift); $\alpha = 0$ (no trend term).

# [SN] ADF Test for Unit Roots in STATA

Augmented Dickey–Fuller test using `tsset` data

```
* dfuller yvar
```

```
* dfuller yvar, trend
// include trend term
```

```
* dfuller yvar, lags(k) regress
// include 'k' lagged differences and display the regression table
```

## Excercise: based on Stock & Watson, E15.1

# Picture the Scenario

- **Objective:** Construct forecasting models for the rate of inflation.
- **Dataset:** `us_macro_quaterly.xlsx`.
  - contains quarterly data on several macroeconomic series for the US.
  - use the sample period $1963:Q1 - 2017:Q4$.(where data before 1963 may be used, as necessary, as initial values for lags in regressions)
- **Key variables:** For each country in each year
  - `PCEPI`: the price index for personal consumption expenditures from the U.S. National Income and Product Accounts.

# Questions

## (a)

i. Compute the inflation rate, $Infl = 400 \times [ln(PCEPI_t) - ln(PCEPI_{t-1})]$. What are the units of `Infl`?

ii. Plot the values of `Infl` from 1963:Q1 through 2017:Q4. Based on the plot, do you think that `Infl` has a stochastic trend? Explain.

## (b) Autocorrelations (»review)

i. Compute the first four autocorrelations of $\Delta Infl$.

ii. Plot the value of $\Delta Infl$ from 1963:Q1 through 2017:Q4. The plot should look choppy or jagged. Explain why this behaviour is consistent with the first autocorrelation that you computed in (i.).

# Questions

## (c) Autoregression Model (»review)

i. Run an OLS regression of $\Delta Infl_t$ on $\Delta Infl_{t-1}$. Does knowing the change in inflation over the current quarter help predict the change in inflation over the next quarter? Explain.

ii. Estimate an $AR(2)$ model of $\Delta Infl_t$ . Is the $AR(2)$ model better than the $AR(1)$ model? Explain.

iii. Use the $AR(2)$ model to predict the change in inflation from 2017:Q4 to 2018:Q1 - that is, to predict the value of $\Delta Infl_{2018:Q1}$.

# Questions

## (d) ADF test (»review)

i. Use the ADF test for $AR(p)$ regression

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \ldots + \gamma_p \Delta Y_{t-p+1} + u_t$$

using 2 lags of $\Delta Infl$ (so that $p = 3$ in the above equation) to test for a stochastic trend in *Infl*.

ii. Is that ADF test based on the above regression preferred to the regression including a determinstic trend

$$\Delta Y_t = \beta_0 + \alpha t + \delta Y_{t-1} + \ldots + \gamma_p \Delta Y_{t-p+1} + u_t$$

for testing for a stochastic trend in *Infl*? Explain.

iii. Based on the ADF tests carried out, does the AR model for *Infl* contain a unit root?

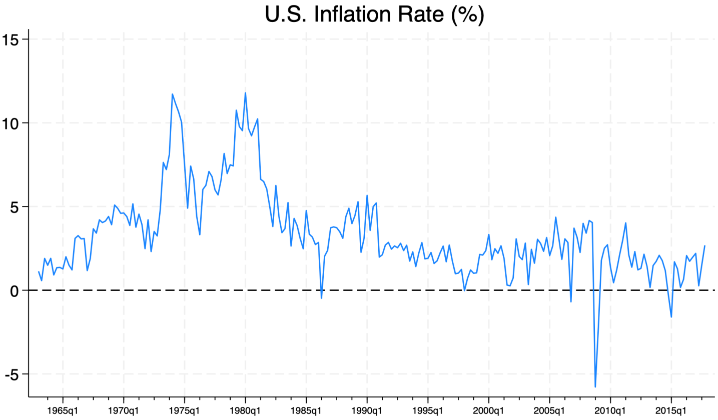
# (a-i) Compute the inflation rate `Infl`. What is its unit?

(»stata)

$$Infl = 400 \times [ln(PCEPI_t) - ln(PCEPI_{t-1})]$$

`Infl` is measured in percentage points at annual rate.

# (a-ii) Plot values of `Infl` from 1963:Q1 through 2017:Q4. Do you think that `Infl` has a stochastic trend?




Inflation increased over the 20-year period 1960-1980, then declined for a decade and has been reasonably stable since then. It appears to have a stochastic trend. (»stata)

# (b-i) Compute the first four autocorrelations of $\Delta Infl$.

```
*corrgram yvar [, lags(p)]
// limit the number of computed autocorrelations to p
```

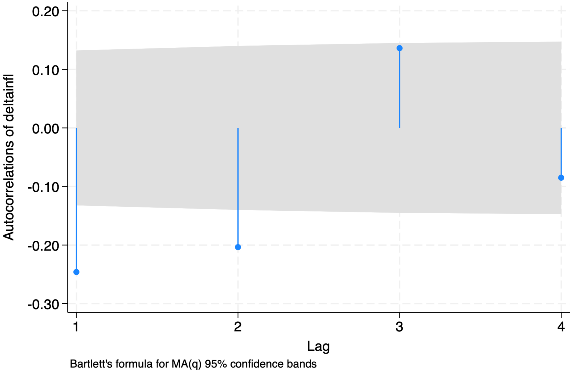
```
. * autocorrelation
. //autocorrelations reported in column named AC
. corrgram deltainfl if tin(1963q1,2017q4), lags(4)
```

| LAG | AC | PAC | Q | Prob>Q | -1 0 1 [Autocorrelation] | -1 0 1 [Partial autocor] |
|---|---|---|---|---|---|---|
| 1 | -0.2460 | -0.2467 | 13.494 | 0.0002 | | |
| 2 | -0.2035 | -0.2843 | 22.772 | 0.0000 | | |
| 3 | 0.1362 | 0.0047 | 26.946 | 0.0000 | | |
| 4 | -0.0850 | -0.1142 | 28.579 | 0.0000 | | |

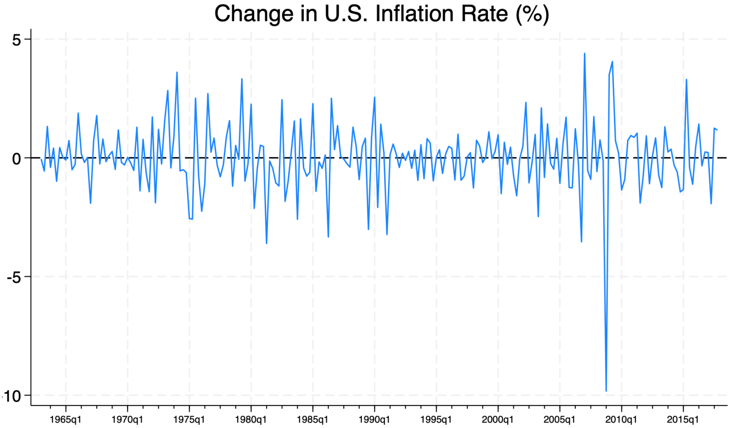
The first four autocorrelations are (rounded to 2 decimal places):
$-0.25$, $-0.20$, $0.14$, and $-0.08$.

```
*ac yvar [, lags(p)]
// limit the number of computed autocorrelations to p
```

```
. * plot of autocorrelations with 95% confidence bands (optional)
. ac deltainfl if tin(1963q1,2017q4), lags(4)
```



# (b-ii) Plot values of $\Delta \mathit{Infl}$ from 1963:Q1 through 2017:Q4.




The change in inflation is slightly negatively serially correlated (the first autocorrelation is $\hat{\rho}_1 = -0.25$) so that values above the mean tend to be followed by values below the mean. (»stata)

## (c-i) Run an OLS regression of $\Delta Infl_t$ on $\Delta Infl_{t-1}$.

Autoregressive model of order one - AR(1)

$$\Delta Infl_t = \beta_0 + \beta_1 \cdot \Delta Infl_{t-1} + \epsilon_t$$

# (c-i) Run an OLS regression of $\Delta Infl_t$ on $\Delta Infl_{t-1}$.

```
*regress yvar L1.yvar [, robust]
```

```
. * AR(1) regression - OLS estimates
. * L. takes the lag
. //ldeltainfl = deltainfl[_n] - deltainfl[_n-1]
. regress deltainfl L1.deltainfl if tin(1963q1,2017q4), robust

Linear regression                               Number of obs     =        220
                                                F(1, 218)         =      13.03
                                                Prob > F          =     0.0004
                                                R-squared         =     0.0607
                                                Root MSE          =     1.4383
```

| deltainfl | Coefficient | Robust std. err. | t | P>\|t\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| deltainfl | | | | | | |
| L1. | -.2466842 | .0683325 | -3.61 | 0.000 | -.3813611 | -.1120074 |
| _cons | .0073719 | .0969709 | 0.08 | 0.939 | -.1837485 | .1984923 |

## (c-i) Run an OLS regression of $\Delta Infl_t$ on $\Delta Infl_{t-1}$. Does knowing the change in inflation over the current quarter help predict the change in inflation over the next quarter?

Autoregressive model of order one - AR(1)

$$\Delta Infl_t = \beta_0 + \beta_1 \cdot \Delta Infl_{t-1} + \epsilon_t$$

OLS estimation results (»stata)

$$\underset{(se)}{\widehat{\Delta Infl_t}} = \underset{(0.096)}{0.007} - \underset{(0.068)}{0.247} \cdot \Delta Infl_{t-1}, \qquad R^2 = 0.061$$

- ◇ The coefficient on lagged inflation is statistically significant, so that lagged inflation helps predict current inflation.

## (c-ii) Estimate an $AR(2)$ model of $\Delta Infl_t$. Is the $AR(2)$ model better than the $AR(1)$ model?

AR(2) model

$$\Delta Infl_t = \beta_0 + \beta_1 \cdot \Delta Infl_{t-1} + \beta_2 \cdot \Delta Infl_{t-2} + \epsilon_t$$

# (c-ii) Estimate an $AR(2)$ model of $\Delta Infl_t$. Is the $AR(2)$ model better than the $AR(1)$ model?

AR(2) model

$$\Delta Infl_t = \beta_0 + \beta_1 \cdot \Delta Infl_{t-1} + \beta_2 \cdot \Delta Infl_{t-2} + \epsilon_t$$

OLS estimation results (»stata)

$$\underset{(se)}{\widehat{\Delta Infl_t}} = \underset{(0.093)}{0.006} - \underset{(0.064)}{0.138} \cdot \Delta Infl_{t-1} - \underset{(0.076)}{0.284} \cdot \Delta Infl_{t-2}, \qquad \text{Adj-}R^2 = 0.128$$

- The estimated coefficient on $\Delta Infl_{t-2}$ is statistically significant, so the AR(2) model is preferred to the AR(1) model.
- Note also that adjusted R-squared increases from 0.061 in the AR(1) model to 0.128 in the AR(2) model, showing better goodness-of-fit.

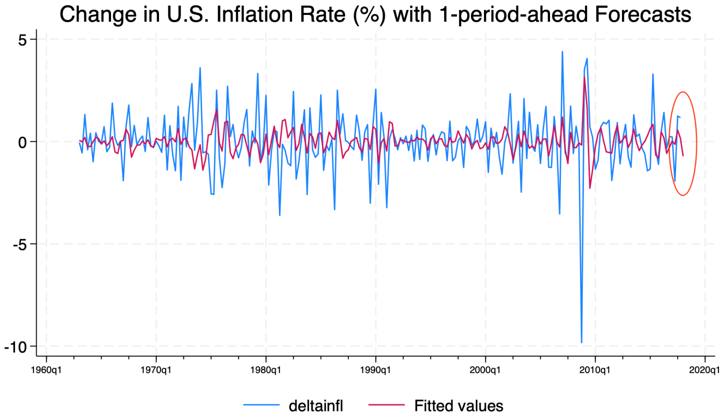
## (c-iii) Use the $AR(2)$ model to predict change in inflation from 2017:Q4 to 2018:Q1 - that is, value of $\Delta Infl_{2018:Q1}$.

```
. * Forecasting 1 period ahead deltainflation
. * add the extra period2 (next quarter)
. tsappend, add(1)

. * estimate AR(2)
. quiet regress deltainfl L.deltainfl L2.deltainfl if tin(1963q1,2017q4), r

. * predict for next quarter
. predict deltainflhat
(option xb assumed; fitted values)
(4 missing values generated)

. list deltainflhat if tin(2017q4,2018q1)
```

| | deltain~t |
|---|---|
| 252. | .1589877 |
| 253. | -.7205634 |

```
. // the last observation is the predicted inflation change
```

(c-iii) Use the $AR(2)$ model to predict change in inflation from 2017:Q4 to 2018:Q1 - that is, value of $\Delta Infl_{2018:Q1}$.




The predicted change in inflation from 2017Q4 to 2018Q1 is $-\mathbf{0.72}$ (a drop).(»stata) (d-i) Use the ADF test for $AR(p)$ regression using 2 lags of $\Delta Infl$ (so that $p = 3$ in the above equation) to test for a stochastic trend in $Infl$.

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \ldots + \gamma_p \Delta Y_{t-p+1} + u_t$$

$$\Delta Infl_t = \beta_0 + \delta Infl_{t-1} + \gamma_1 \Delta Infl_{t-1} + \gamma_2 \Delta Infl_{t-2} + u_t$$

```
* dfuller yvar, lags(k) regress
// include 'k' lagged differences and display the regression table
```

```
. dfuller infl if tin(1963q1,2017q4), lags(2) regress

Augmented Dickey-Fuller test for unit root

Variable: infl                          Number of obs  = 220
                                        Number of lags =   2

H0: Random walk without drift, d = 0
```

| | Test statistic | Dickey-Fuller critical value 1% | 5% | 10% |
|---|---|---|---|---|
| Z(t) | **-2.740** | -3.470 | -2.882 | -2.572 |

MacKinnon approximate *p*-value for Z(t) = **0.0674**.

Regression table

| D.infl | Coefficient | Std. err. | t | P>\|t\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| infl | | | | | | |
| L1. | -.1057272 | .0385888 | -2.74 | 0.007 | -.1817859 | -.0296684 |
| LD. | -.2543744 | .0683361 | -3.72 | 0.000 | -.3890653 | -.1196835 |
| L2D. | -.2415464 | .066275 | -3.64 | 0.000 | -.3721749 | -.1109179 |
| _cons | .3591525 | .1582484 | 2.27 | 0.024 | .0472437 | .6710613 |

(d-i) Use the ADF test for $AR(p)$ regression using 2 lags of $\Delta Infl$ (so that $p = 3$ in the above equation) to test for a stochastic trend in $Infl$.

The ADF t-statistic is $-2.74$. The 10% critical vale is $-2.57$ and the 5% critical value of $-2.88$; thus the unit root null hypothesis can be rejected at the 10% but not the 5% significance level.

(d-ii) Is that ADF test based on the above regression preferred to the regression including a determinstic trend for testing for a stochastic trend in *Infl*?

$$\Delta Y_t = \beta_0 + \alpha t + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \ldots + \gamma_p \Delta Y_{t-p+1} + u_t$$

$$\Delta Infl_t = \beta_0 + \alpha t + \delta Infl_{t-1} + \gamma_1 \Delta Infl_{t-1} + \gamma_2 \Delta Infl_{t-2} + u_t$$

```
* dfuller yvar, lags(k) trend regress
// include 'k' lagged differences and trend term
```

```
. * ADF on Infl with a deterministic trend
. dfuller infl if tin(1963q1,2017q4), lags(2) trend regress

Augmented Dickey-Fuller test for unit root

Variable: infl                          Number of obs  = 220
                                        Number of lags =   2

H0: Random walk with or without drift

                                       Dickey-Fuller
                   Test      ------- critical value ---------
              statistic           1%           5%          10%
--------------------------------------------------------------
 Z(t)            -3.392       -4.000       -3.434       -3.134
--------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0525.

Regression table
------------------------------------------------------------------------------
      D.infl | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        infl |
         L1. |  -.1475798   .0435074    -3.39   0.001    -.2333354   -.0618242
         LD. |  -.2336542    .068611    -3.41   0.001    -.3688904   -.0984179
        L2D. |  -.2295046   .0660681    -3.47   0.001    -.3597287   -.0992806
             |
      _trend |  -.0033129   .0016318    -2.03   0.044    -.0065293   -.0000964
       _cons |   .8616621   .2931807     2.94   0.004     .2837856    1.439539
------------------------------------------------------------------------------

. // regress option to produce the regression results
. // trend option to include the linear trend in AR(p) regression
```

## (d-ii) Is that ADF test based on the above regression preferred to the regression including a determinstic trend for testing for a stochastic trend in *Infl*?

- ◇ Based on the t-statistic and the critical values, the results are similar to the previous regression.
- ◇ The coefficient on trend is quite close to zero so the inflation rate does not exhibit a linear trend. Thus, the specification that includes an intercept, but no time trend is appropriate.

# (d-iii) Based on the ADF tests carried out, does the AR model for *Infl* contain a unit root?

$$\underset{(se)}{\widehat{\Delta Infl_t}} = \underset{(.158)}{.359} - \underset{(.039)}{.106} Infl_{t-1} - \underset{(.068)}{.254} \Delta Infl_{t-1} - \underset{(.066)}{.242} \Delta Infl_{t-2}$$

From the results in (d-i), it is clear that inflation contains a unit root, thereby being highly persistent. The null hypothesis that $\delta = 0$ or, equivalently, $\rho = 1.0$ cannot be rejected at the 5% significance level.

# STATA CODES & RESULTS

# Q(a-i,ii) (»back(a-i)) (»back(a-ii))

```
. * log() is the same as ln() in Stata
. gen logpce = log(PCECTPI)

. * using stata's first difference operator
. gen infl = 400*D.logpce
(1 missing value generated)

. * alternatively
. * gen infl = 400*(logpce[_n] - logpce[_n-1])
```

```
. * Plot 'infl' for defined time range
. twoway (tsline infl if tin(1963q1,2017q4), yline(0)), title(" U.S. Inflation Rate (%)") ///
>                                                       ytitle("") xtitle("") ///
>                                                       tlabel(#12, labsize(vsmall)) tmtick(##4)
```

# Q(b-i,ii) (»back(b-ii))

```
. * Create Delta Infl as first difference of inflation rate
. gen deltainfl = D.infl
(2 missing values generated)
```

```
. * Plot 'deltainfl' for defined time range
. twoway (tsline deltainfl if tin(1963q1,2017q4), yline(0)), title(" Change in U.S. Inflation Rate (%)") ///
>                                                                 ytitle("") xtitle("") ///
>                                                                 tlabel(#12, labsize(vsmall)) tmtick(##4)
```

# Q(b-i)

```
*corrgram yvar [, lags(p)]
// limit the number of computed autocorrelations to p
```

```
. * autocorrelation
. //autocorrelations reported in column named AC
. corrgram deltainfl if tin(1963q1,2017q4), lags(4)
```

| LAG | AC | PAC | Q | Prob>Q | -1 0 1 [Autocorrelation] | -1 0 1 [Partial autocor] |
|---|---|---|---|---|---|---|
| 1 | -0.2460 | -0.2467 | 13.494 | 0.0002 | | |
| 2 | -0.2035 | -0.2843 | 22.772 | 0.0000 | | |
| 3 | 0.1362 | 0.0047 | 26.946 | 0.0000 | | |
| 4 | -0.0850 | -0.1142 | 28.579 | 0.0000 | | |

```
*ac yvar [, lags(p)]
// limit the number of computed autocorrelations to p
```

```
. * plot of autocorrelations with 95% confidence bands (optional)
. ac deltainfl if tin(1963q1,2017q4), lags(4)
```

# Q(c-i) (»back(c-i))

```
*regress yvar L1.yvar [, robust]
```

```
. * AR(1) regression - OLS estimates
. * L. takes the lag
. //ldeltainfl = deltainfl[_n] - deltainfl[_n-1]
. regress deltainfl L1.deltainfl if tin(1963q1,2017q4), robust

Linear regression                               Number of obs     =        220
                                                F(1, 218)         =      13.03
                                                Prob > F          =     0.0004
                                                R-squared         =     0.0607
                                                Root MSE          =     1.4383
```

| deltainfl | Coefficient | Robust std. err. | t | P>\|t\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| deltainfl L1. | -.2466842 | .0683325 | -3.61 | 0.000 | -.3813611 | -.1120074 |
| _cons | .0073719 | .0969709 | 0.08 | 0.939 | -.1837485 | .1984923 |

## Q(c-ii) (»back(c-ii))

```
*regress yvar L1.yvar L2.yvar [, robust]
```

```
. * AR(2) regression
. regress deltainfl L1.deltainfl L2.deltainfl if tin(1963q1,2017q4), robust

Linear regression                               Number of obs     =        220
                                                F(2, 217)         =      13.98
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1361
                                                Root MSE          =     1.3826

------------------------------------------------------------------------------
             |               Robust
   deltainfl | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   deltainfl |
         L1. |  -.3177357   .0644048    -4.93   0.000    -.4446747   -.1907966
         L2. |  -.2844172   .0763383    -3.73   0.000    -.4348766   -.1339578
             |
       _cons |   .0060757   .0932463     0.07   0.948    -.1777087      .18986
------------------------------------------------------------------------------

. display e(r2_a) // Adjusted R-squared
.12810942
```

# Q(c-iii-I)

```
. * Forecasting 1 period ahead deltainflation
. * add the extra period2 (next quarter)
. tsappend, add(1)

. * estimate AR(2)
. quiet regress deltainfl L.deltainfl L2.deltainfl if tin(1963q1,2017q4), r

. * predict for next quarter
. predict deltainflhat
(option xb assumed; fitted values)
(4 missing values generated)

. list deltainflhat if tin(2017q4,2018q1)

      +-----------+
      | deltain~t |
      |-----------|
252.  |  .1589877 |
253.  | -.7205634 |
      +-----------+

. // the last observation is the predicted inflation change
```

# Q(c-iii-II) (»back(c-iii))

```
. * plot change in inflation vs predicted change
. twoway (tsline deltainfl) (tsline deltainflhat) if tin(1963q1,2018q1), ///
>         ytitle("") xtitle("") ///
>         tlabel(#8, labsize(vsmall)) tmtick(##4) ///
>         legend(pos(6) cols(2)) ///
>         title("Change in U.S. Inflation Rate (%) with 1-period-ahead Forecasts")
```

# Q(d-i) (»back(d-i))

```
. dfuller infl if tin(1963q1,2017q4), lags(2) regress

Augmented Dickey–Fuller test for unit root

Variable: infl                              Number of obs  = 220
                                            Number of lags =   2

H0: Random walk without drift, d = 0
```

| | Test statistic | Dickey–Fuller critical value 1% | 5% | 10% |
|---|---|---|---|---|
| Z(t) | -2.740 | -3.470 | -2.882 | -2.572 |

MacKinnon approximate *p*-value for Z(t) = **0.0674.**

Regression table

| D.infl | Coefficient | Std. err. | t | P>\|t\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| infl | | | | | | |
| L1. | -.1057272 | .0385888 | -2.74 | 0.007 | -.1817859 | -.0296684 |
| LD. | -.2543744 | .0683361 | -3.72 | 0.000 | -.3890653 | -.1196835 |
| L2D. | -.2415464 | .066275 | -3.64 | 0.000 | -.3721749 | -.1109179 |
| _cons | .3591525 | .1582484 | 2.27 | 0.024 | .0472437 | .6710613 |

# Q(d-ii) (»back(d-ii))

```
. * ADF on Infl with a deterministic trend
. dfuller infl if tin(1963q1,2017q4), lags(2) trend regress

Augmented Dickey-Fuller test for unit root

Variable: infl                          Number of obs  = 220
                                        Number of lags =   2

H0: Random walk with or without drift
```

| | Test statistic | Dickey-Fuller critical value 1% | 5% | 10% |
|---|---|---|---|---|
| Z(t) | -3.392 | -4.000 | -3.434 | -3.134 |

MacKinnon approximate *p*-value for Z(t) = **0.0525.**

Regression table

| D.infl | Coefficient | Std. err. | t | P>\|t\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| infl | | | | | | |
| L1. | -.1475798 | .0435074 | -3.39 | 0.001 | -.2333354 | -.0618242 |
| LD. | -.2336542 | .068611 | -3.41 | 0.001 | -.3688904 | -.0984179 |
| L2D. | -.2295046 | .0660681 | -3.47 | 0.001 | -.3597287 | -.0992806 |
| _trend | -.0033129 | .0016318 | -2.03 | 0.044 | -.0065293 | -.0000964 |
| _cons | .8616621 | .2931807 | 2.94 | 0.004 | .2837856 | 1.439539 |

```
. // regress option to produce the regression results
. // trend option to include the linear trend in AR(p) regression
```