

Assignment 5: Time Series Models

ECO 321

DUE: Monday, May 14 in class

Instructions: You need to write a Stata do-file to answer question 2. Your do-file must produce a log file that contains your Stata output. You must submit your log file along with your answers to the assignment. Assignments that do not contain log files will be marked down.

1. We want to forecast the unemployment rate in the United States (UrateUS) using quarterly data from the first quarter of 1980 to the fourth quarter of 2016 ($T = 36 \text{ years} \times 4 \text{ quarters} = 144$).

a) The following table presents the first four autocorrelations for the United States unemployment rate and its changes over the time period from Q1:1980 to Q4:2016. Explain briefly what the autocorrelations in the unemployment rate, Y_t , measure.

Table 1: First Four Autocorrelations of the U.S. Unemployment Rate and Its Change, 1980:Q1 – 2016:Q4

Lag	Unemployment Rate Y_t	Change in Unemployment Rate ΔY_t
1	0.97	0.62
2	0.92	0.32
3	0.83	0.12
4	0.75	-0.07

The first four autocorrelations in the unemployment rate measure how the unemployment rate is correlated with itself over time. All of the autocorrelations are positive, so the unemployment rate series exhibits positive autocorrelation. Positive autocorrelation means that if the series is above (below) its mean in the last period (first lag), then it will be above (below) its mean this period. The same can be said for the value of the unemployment rate two, three, and four periods ago and the value today.

b) The accompanying table gives the United States unemployment rate for the period 2015:Q1–2016:Q1 in levels, and the lagged unemployment rate for 2015:Q1. Fill in the blanks for the missing lagged values and the changes in the unemployment rate.

Table 2: Changes in Unemployment Rates in the U.S., Q1:2015 to Q1:2016

Quarter	Unemployment Rate Y_t	First Lag	Change in Unemployment Rate ΔY_t
Q1:2015	4.3	4.4	-0.1
Q2:2015	4.3	4.3	0.0
Q3:2015	4.2	4.3	-0.1
Q4:2015	4.1	4.2	-0.1
Q1:2016	4.0	4.1	-0.1

(c) You decide to test whether the unemployment rate is nonstationary. The result is as follows:

$$\hat{\Delta UrateUS}_t = -0.004 - 0.25 UrateUS_{t-1}$$

$$R^2 = 0.13 \quad SER = 0.55$$

Is the unemployment rate series nonstationary? Why or why not?

This regression represents the Dickey-Fuller test for nonstationarity. The null hypothesis is: $H_0 : \gamma_1 = 0$ against the alternative that $H_1 : \gamma_1 < 0$. The t-statistic on $\gamma_1 = \frac{-0.25}{0.15} = -1.6 > -2.86$, so we fail to reject the null hypothesis of nonstationarity. Therefore, we conclude that the unemployment rate series is nonstationary and violates one of our OLS assumptions for unbiased/consistent forecasts.

$H_0 : \gamma_1 = 0$ means non stat. we assume nonstat

(d) You decide to estimate an AR(1) model using changes (i.e., first-differences) in the US unemployment rate. The result is as follows:

$$\hat{\Delta UrateUS}_t = -0.003 + 0.621 \Delta UrateUS_{t-1}$$

$$R^2 = 0.39 \quad SER = 0.255$$

If most of the forecast error comes from uncertainty about the future error terms instead of sampling error from estimating the coefficients in the AR(1) model, then what is your best guess of the RMSFE here? *root mean square forecast error.*

The best guess of the RMSFE is the $SER = 0.255$ when our sample size is large and/or sampling error associated with estimating β_0 and β_1 is very small.

(e) The actual unemployment rate during the fourth quarter of 2015 is 4.1 percent, and it decreased from the third quarter to the fourth quarter by 0.1 percent. Using the AR(1) model's results, what is your forecast for the unemployment rate level in the first quarter of 2016?

$$\hat{\beta}_0 \quad \hat{\beta}_1 \quad \hat{\Delta Y}_t$$

$$\text{Forecasted Change} = -0.003 + 0.621(-0.10) = -0.003 - 0.0621 = -0.0624$$

Since the unemployment rate was 4.1 percent in Q4:2015, the forecast for the unemployment rate in Q1:2016 is $4.1 - 0.0624 = 4.0386$.

The forecast error in Q1:2016 is $4.0 - 4.0386 = -0.0386$.

actual - forecast *we're off by.*

(f) You want to see how sensitive your forecast is to changes in the specification, so you consider an AR(4) model. The results are as follows:

$$\begin{aligned}\Delta UrateUS_t &= -0.005 + 0.663\Delta UrateUS_{t-1} - 0.082\Delta UrateUS_{t-2} \\ &\quad + 0.106\Delta UrateUS_{t-3} - 0.176\Delta UrateUS_{t-4}\end{aligned}$$

$$R^2 = 0.416 \quad SER = 0.253$$

What is your forecast for the unemployment rate level in 2016:Q1? Compare the forecast error of the AR(4) model with the forecast error of the AR(1) model.

$$\text{Change in Forecast} = -0.005 + 0.663(-0.10) - 0.082(-0.10) + 0.106(0) - 0.176(-0.10) = -0.005 - 0.0663 + 0.0082 + 0.0176 = -0.0455$$

Since the unemployment rate was 4.1 percent in Q4:2015, the forecast for the unemployment rate is $4.1 - 0.0455 = 4.0545$ in Q1:2016.

The forecast error is $4.0 - 4.0545 = -0.0545$ in Q1:2016. The forecast error in the AR(4) model is larger than what we obtained from the AR(1) model.

(g) Given the information below, which model should you use for forecasting?

We choose the AR(p) model with the smallest BIC and AIC. We do not pay much attention to the R^2 because it will always be increasing as we add lags to the model. In this example, the BIC and AIC are minimized when $p = 1$. So we choose an AR(1) model.

p	BIC	AIC	R^2
0	0.604	0.624	0.000
1	0.158	0.118	0.393
2	0.185	0.125	0.397
3	0.217	0.138	0.400
4	0.218	0.119	0.416
5	0.277	0.139	0.420

2. The data set macro.dta contains time series data for U.S. real aggregate personal consumption (*cons*), real gross domestic investment (*invest*), interest rates on 3 month Treasury Bills (*intrate*), real M1 money stock (*m1*), and real GDP (*gdp*) for 1947:Q1 - 1988:Q1. *cons*, *invest*, *m1* and *gdp* are all measured in billions of 1982 dollars.

- a) Estimate autocorrelations for $\log(\text{cons})$, $\log(\text{gdp})$, $\log(\text{m1})$ and *intrate* and their first differences. What can you say about the stationarity or non-stationarity of each series?

The autocorrelations for the variables in logs suggest that these variables exhibit nonstationarity because the $\hat{\beta}_1$ s (i.e., first-order autocorrelations) are close to 1 and the 95% confidence intervals include 1 for 3 out of 4 of the variables. If we cannot reject the hypothesis that $\beta_1 = 1$ in an AR(1) model, then that suggests the series is nonstationary.

The autocorrelations for the variables in first differences suggest the first-differenced series might be stationary because the β_1 s are not close to 1 and the 95% confidence intervals do not include 1 for any of the 4 variables. However, we cannot be sure that these variables are stationary until we use the Dickey-Fuller test.

- b) Conduct the Dickey-Fuller test for unit roots in each of the variables listed in part (a). Use 8 lags in each of the Dickey-Fuller regressions. What do you conclude for each series?

We cannot reject the null hypothesis of nonstationarity for all 4 of the variables measured in logs (i.e., levels).

We can reject the null hypothesis of nonstationarity for all 4 of the variables measured in first differences.

- c) The Keynesian consumption function is $\log(\text{cons}) = \beta_0 + \beta_1 \log(\text{gdp}) + u$. Estimate this regression as specified and in first-differenced form. Interpret the effect of *gdp* on *cons*. Does your interpretation change substantively in the two versions of the model (i.e., in levels vs. first differences)? Which estimates are more reliable?

Since the variables in levels exhibited nonstationarity, we would prefer the model that regresses the first difference in $\log(\text{consumption})$ on the first difference in $\log(\text{gdp})$. Interpreting the first differences model tells us that as real GDP increases by 1%, consumption increases by 0.37%.

- d) The Keynesian money demand function is $\log(\text{m1}) = \beta_0 + \beta_1 \log(\text{gdp}) + \beta_2 \text{intrate} + u$. Estimate this regression as specified and in first-differenced form. Interpret the effects of *gdp* and *intrate* on *m1*. Does your interpretation change substantively in the two versions? Which estimates are more reliable?

Since the variables in logs exhibited nonstationarity, we would prefer the model that regresses the first difference in $\log(\text{m1})$ on the first difference in $\log(\text{gdp})$ and the first difference in *intrate*. Interpreting the first differences model tells us that as real GDP increases by 1%, the money supply increases by 0.33%. As the interest rate increases by 1%, the money supply decreases by 0.01%, but the effect is just barely statistically significant at the 5% level. The effect of the interest rate on the money supply decreases by a factor of 10 when using the first-differenced model.

(1)

```

    .cnames unnamed>
    logi /Users/jessicavanparys/Dropbox/Working_Docs/teaching/ECO321/Assignments/Assignment_5
    log type: text
    opened on: 15 May 2018, 10:58:41

    .
    *Get a delimiter that allows wrap-around text
    .
    *delimit
    *allow now*
    .
    *macro dta
    .
    *question 2*
    .
    *Open the macro.dta data set;
    .
    *clearar*
    .
    use "/Users/jessicavanparys/Dropbox/Working_Docs/teaching/ECO321/Assignments/Assignment_5.mn"
    > cto.dta";
    .
    .spn

    Variable | Obs   Mean   Std. Dev.   Min   Max
    cons     | 165   1416.371 555.2239 658.1 2527.9
    invest   | 165   385.4519 150.7757 160.2 741.8
    intrate  | 165   4.885115 3.256603 .375667 15.08733
    ml       | 165   492.8857 43.68785 432.8333 634.8
    realgdp | 165   2384.193 843.1918 1012.4 4035.7
    .
    data    | 165   30   47.77552  -52   112
    .
    *sort date;
    .
    foreach x of varlist cons realgdp ml intrate {
    2. gs 1 x~-log("x");
    3. l1
    .
    sort date;
    .
    foreach x of varlist cons realgdp ml intrate {

```

(4)

```

    . reg lrealgdp lag_1_lrealgdp
    Source | SS      df      MS      Number of obs = 164
    Model | 23.797425  1   23.797425 Prob > F      = 0.0000
    Residual | .022025674 162   .000135961 R-squared      = 0.9991
    Total | 23.8094567 163   .14407025  Adj R-squared = 0.9996
    Root MSE = .01166

    lrealgdp | Coef. Std. Err. t P>|t| [95% Conf. Interval]
    lag_1_lrealgdp | .9960486 .0023813 418.28 0.000 .9913462 1.000751
    _cons | .0387302 .0182018 2.12 0.036 .0026289 .0748315

    . reg lml lag_1_lml
    Source | SS      df      MS      Number of obs = 164
    Model | 1.16185283  1   1.16185283 Prob > F      = 0.0000
    Residual | .024762111 162   .000152053 R-squared      = 0.9791
    Total | 1.18661494 163   .007279846  Adj R-squared = 0.9790
    Root MSE = .01236

    lml | Coef. Std. Err. t P>|t| [95% Conf. Interval]
    lag_1_lml | 1.016981 .0116637 87.19 0.000 .9938594 1.039923
    _cons | -.1029566 .0726235 -1.42 0.156 -.2456564 .0397432

    . reg llntrate lag_1_llntrate
    Source | SS      df      MS      Number of obs = 164
    Model | 88.4435339  1   88.4435339 Prob > F      = 0.0000
    Residual | 3.68931222 162   .022717593 R-squared      = 0.9601
    Total | 92.1238471 163   .565176976  Adj R-squared = 0.9598
    Root MSE = .15072

    llntrate | Coef. Std. Err. t P>|t| [95% Conf. Interval]
    lag_1_llntrate | -.9532848 .0152783 62.39 0.000 .9231145 .983455
    _cons | .0787887 .0234835 3.35 0.001 .0324675 .1251538
```

(5)

```

    2. gs lag_1_l`x`~-lag_1`x`[`n-1];
    3. l1
    .
    (1 missing value generated)
    (1 missing value generated)
    (1 missing value generated)
    (1 missing value generated)

    foreach x of varlist cons realgdp ml intrate {
    2. gs fd 1`x`~-lag_1`x`;
    3. l1
    .
    (1 missing value generated)
    (1 missing value generated)
    (1 missing value generated)
    (1 missing value generated)

    foreach x of varlist cons realgdp ml intrate {
    2. gs fd lag_1`x`~-fd 1`x`[`n-1];
    3. l1
    .
    (2 missing values generated)
    (2 missing values generated)
    (2 missing values generated)
    (2 missing values generated)

    .
    .spn

    Variable | Obs   Mean   Std. Dev.   Min   Max
    cons     | 165   1416.371 555.2239 658.1 2527.9
    invest   | 165   385.4519 150.7757 160.2 741.8
    intrate  | 165   4.885115 3.256603 .375667 15.08733
    ml       | 165   492.8857 43.68785 432.8333 634.8
    realgdp | 165   2384.193 843.1918 1012.4 4035.7
    .
    data    | 165   30   47.77552  -52   112
    .
    lcons   | 165   7.37608 1.00000 6.489357 7.831544
    lrealgdp | 165   7.671552 1.005437 6.920679 8.309235
    lml     | 165   6.139584 0.859832 6.070353 6.45331
    llntrate | 165   1.332618 -.7710472 -.9790521 2.713855
    .
    lag_1_lcons | 164   7.172064 1.0016632 6.489357 7.832392
    lag_1_lra-p | 164   7.667703 .38352 6.920679 8.298191
    lag_1_lml | 164   6.195051 .0830247 6.070353 6.45331
    lag_1_llntrate | 164   1.330667 .7727101 -.9790521 2.713855
    fd_lcons | 164   .008206 .0061564 -.0207052 .0391784
    .
    fd_lrealgdp | 164   .008422 .0137228 -.0291042 .0439367
    fd_lml | 164   .0161632 .0124049 -.0493062 .041357
    fd_llntrate | 164   .0156463 .1545468 -.5982903 .4552023
    fd_licons | 163   .0081996 .0061811 -.0207052 .0391784
```

```

    fd_lag_lra-p | 163   .0084547 .0117553 -.0291042 .0439367
    fd_lag_lml | 163   .0016038 .0124431 -.0493002 .041357
    fd_lag_llntrate | 163   .0170023 .1549456 -.5982901 .4552023

    . corrgram lcons, lag(1);
    .
    corrgram lrealgdp, lag(1);
    .
    corrgram lml, lag(1);
    .
    corrgram llntrate, lag(1);
    .
    corrgram lcons lag_1_lcons;
    .
    Source | SS      df      MS      Number of obs = 164
    Model | 26.5757564  1   26.5757564 Prob > F      = 0.0000
    Residual | .00821885 162   .000066802 R-squared      = 0.9996
    Total | 26.5883483 163   .163118701  Adj R-squared = 0.9996
    Root MSE = .00817

    lcons | Coef. Std. Err. t P>|t| [95% Conf. Interval]
    lag_1_lcons | -.990923 .0015084 630.76 0.000 .995965 1.002221
    _cons | .0147122 .0113781 1.29 0.198 -.0077564 .0371867
```

(6)

```

    . reg fd_llntrate fd_lag_llntrate
    Source | SS      df      MS      Number of obs = 163
    Model | 4.26018534  1   4.26018534 Prob > F      = 0.0000
    Residual | 3.46643685 161   .021530664 R-squared      = 0.3394
    Total | 3.89245936 162   .024027502  Adj R-squared = 0.1039
    Root MSE = .14673

    fd_llntrate | Coef. Std. Err. t P>|t| [95% Conf. Interval]
    fd_lag_llntrate | .3396915 .0744033 4.45 0.000 .1840294 .4778937
    _cons | .0111159 .0115624 0.96 0.338 -.0117177 .0339495

    . "The autocorrelations for the variables in logs look like these time series exhibit
    . nonstationarity because beta_1s are close to 1 and the 95% confidence intervals include
    > 1 in 3 out of 4 cases;
    . "The autocorrelations for the variables in first differences look stationary
    . because the beta_1s are not close to 1 and the 95% confidence intervals do not
    > include 1 in any of the cases; however, we cannot be sure that these variables are
    > stationary until we use the Dickey-Puller test;
    .
    *(B):
    .
    Dickey-Puller test for log variables;
    dfuller lcons, lags(8) regress;

    Augmented Dickey-Puller test for unit root      Number of obs = 156
    Test Statistic          1% Critical Value          5% Critical Value          10% Critical Value
    Z(t)          -0.900           -3.491           -2.886           -2.576
    MacKinnon approximate p-value for Z(t) = 0.7870

    D.icons | Coef. Std. Err. t P>|t| [95% Conf. Interval]
    lcons | .0015012 .0016671 -.090 0.369 -.0047959 .0017936
    LD.1 | -.0479835 .0182157 0.59 -.1125267 .2084937
    LD.2 | -.0214072 .0012042 2.45 0.006 -.1243625 .1666009
    LD.3 | .0215164 .0015698 0.26 0.797 -.143464 .1665793
    LD.4 | -.0570246 .001831788 -.69 0.494 -.2214247 .1073655
    LD.5 | -.107475 .0038548 -1.28 0.202 -.273201 .050251
```

```

L6D. | -.0821621 .0037815 -0.98 0.328 -.2477432 .0034181
L7D. | .1205693 .0021698 1.47 0.144 -.0418265 .2829652
L8D. | -.1582006 .0021927 -1.93 0.056 -.3202989 .0038977
_cons | .0191378 .0121823 1.57 0.118 -.0049385 .0432142

```

`. dfuller lrealgdp, lags(8) regress;`

Augmented Dickey-Fuller test for unit root
Number of obs = 156

Interpolated Dickey-Puller						
Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value			
Z(t)	-1.678	-3.491	-2.886	-2.576		

MacKinnon approximate p-value for Z(t) = 0.4423

D.lrealgdp | Coef. Std. Err. t P>|t| [95% Conf. Interval]

lrealgdp						
L1.	-.0039482	.0034466	-1.68	0.095	-.008554	.0006976
LD.	.0760098	.0779522	4.17	0.000	.2205493	.5355875
L2D.	.1787551	.0950893	1.09	0.337	.1732115	.1981101
L3D.	.0694273	.0861337	-0.81	0.422	-.2396638	.1008093
L4D.	-.1425298	.0857017	-1.66	0.098	-.3119059	.0264963
L5D.	.0655589	.0855592	-1.06	0.296	-.2593133	.0794372
L6D.	.0994947	.0841432	1.17	0.244	-.0678201	.2647472
L7D.	-.1579199	.0788511	-2.13	0.035	-.323757	.0120828
cons	.0366057	.0182873	2.00	0.047	.0004635	.0727478

`. dfuller lml, lags(8) regress;`

Augmented Dickey-Fuller test for unit root
Number of obs = 156

Interpolated Dickey-Puller						
Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value			
Z(t)	-0.876	-3.491	-2.886	-2.576		

MacKinnon approximate p-value for Z(t) = 0.7958

D.lml | Coef. Std. Err. t P>|t| [95% Conf. Interval]

lml						
L1.	-.0167245	.0124688	-0.88	0.383	-.0349265	.0134775
LD.	.5344799	.0974748	6.36	0.000	.3502474	.6086119
L2D.	.008381	.0919475	0.98	0.927	-.1732115	.1891010
L3D.	.1272302	.0923371	1.38	0.170	-.0552897	.3096302
L4D.	-.0185875	.0928957	-0.11	0.909	-.1941815	.1730065
L5D.	.0222311	.0926166	0.62	0.538	-.1253497	.2401313
L6D.	.0322311	.0926446	0.61	0.500	-.1253497	.2135329
L7D.	.1458999	.092612	-1.58	0.117	-.3289323	.0731424
L8D.	.0209589	.0856909	0.34	0.735	-.1402993	.1984131
cons	.067268	.075775	0.89	0.376	-.0824896	.2170257

`. dfuller llntrate, lags(8) regress;`

Augmented Dickey-Fuller test for unit root
Number of obs = 156

Interpolated Dickey-Puller						
Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value			
Z(t)	-2.001	-3.491	-2.886	-2.576		

MacKinnon approximate p-value for Z(t) = 0.2863

D.llntrate | Coef. Std. Err. t P>|t| [95% Conf. Interval]

llntrate						
L1.	-.0984311	.0148111	-2.00	0.047	-.058905	-.0003612
LD.	.2389852	.0857017	6.36	0.000	.3502475	.6082229
L2D.	-.4054130	.0918763	-4.00	0.000	-.4469179	-.101444
L3D.	.2601602	.0953931	2.61	0.019	-.0634365	.4550384
L4D.	-.1265994	.1917053	-1.24	0.218	-.3267043	.0733055
L5D.	.0226333	.1904397	-0.03	0.979	-.2011369	.1987004
L6D.	.1446646	.0974748	0.145	0.524	-.3302024	.0580015
L7D.	-.0877361	.0848281	-1.03	0.373	-.1732115	.1093335
L8D.	-.0077366	.0756629	-0.10	0.919	-.1572726	.1417994
cons	.0541495	.0232291	2.33	0.021	.0082409	.1005082

*We cannot reject the null hypothesis of nonstationarity for all 4 of the log variables;
*Dickey-Fuller test for first-difference variables;
`. dfuller fd_llcons, lags(8) regress;`

`. reg fd_llcons r;`

Linear regression
Number of obs = 156
F(1, 163) = 10963.35
Prob > F = 0.0000
R-squared = 0.9921
Root MSE = .03632

fd_llcons	Coef. Std. Err. t P> t [95% Conf. Interval]
lrealgdp	1.049442 .0100227 104.71 0.000 1.029651 1.069233
cons	-.8747651 .0766494 -11.12 0.000 -1.030068 -.719462

`. reg fd_llcons fd_lrealgdp, r;`

Linear regression
Number of obs = 156
F(2, 162) = 35.51
Prob > F = 0.0000
R-squared = 0.2920
Root MSE = .05688

fd_llcons	Coef. Std. Err. t P> t [95% Conf. Interval]
fd_lrealgdp	.3759713 .0630969 5.96 0.000 .2513729 .5095698
cons	.0050358 .0066812 7.39 0.000 .0036907 .0063809

`. reg fd_llcons fd_llntrate, r;`

Linear regression
Number of obs = 156
F(2, 162) = 133.36
Prob > F = 0.0000
R-squared = 0.6164
Root MSE = .05302

fd_llcons	Coef. Std. Err. t P> t [95% Conf. Interval]
fd_llntrate	.3279549 .0299266 10.96 0.000 .2688584 .3970514
lrealgdp	-.0960165 .0159211 -6.03 0.000 -.1274562 -.0645769
cons	3.008614 .0209245 18.20 0.000 3.395422 4.221807

`. reg fd_llcons fd_lrealgdp fd_llntrate, r;`

Linear regression
Number of obs = 164
F(3, 161) = 7.14
Prob > F = 0.0006
R-squared = 0.0835
Root MSE = .01195

fd_llcons	Coef. Std. Err. t P> t [95% Conf. Interval]				
fd_lrealgdp	.3252979 .091851 3.31 0.001 -.1314012 .5193947				
fd_llntrate	-.0108141 .0053355 -2.03 0.044 -.0213507 -.0022775				
lrealgdp	-.00608797 .0013299 -0.66 0.569 -.003506 .0017466				
cons					

`. clear;`