Time Series Modeling and Forecasting **Spring 2020**

Midterm Exam 2-20-2020

Instructions

- Read these instructions and each question carefully.
- Answer all eight questions below on the answer sheets provided.
- Do not remove the stable holding the answer sheets together.
- Write your name, the date, and the course on each page of the answer sheet as a precaution in case they become separated somehow.
- Write only on the fronts of the answer sheets to keep the work neat and legible.
- Show your work and/or explain your answers.
- Organize your answers logically.
- Be NEAT! If it is too hard to read, it is wrong.
- Be both complete and concise. You will lose points for both errors of omission and commission.
- Use scratch paper to work answers out ahead of time if needed to ensure they are neat, concise, complete, and organized. Use the backs of the pages of questions and output as scratch paper.
- Turn in ONLY the answer sheets. The questions and output are yours to keep or dispose of.

1. (10%) What is meant by seasonality in a time series context? What is the simplest robust way to control for it?

2. (10%) What is covariance stationarity? Why do we care?

3. (10%) We discussed the Wold representation theorem and AR←→ MA invertibility at length. Explain why these things are important to our understanding of how to estimate accurate approximations of time series processes with simple and robust statistical procedures.

4. (10%) What does dynamically complete mean? Why do we care? How do we check whether a model is dynamically complete?

5. (10%) Consider the ARDL model $y_t = \alpha + \beta x_t + r_t$ where x is an exogenous predictor variable and the residual r follows the AR(1) process $r_t = \rho r_{t-1} + \varepsilon_t$ in which ε is a white noise disturbance. Derive the dynamically complete version of the model.

Questions 6-8 are on the following page.

Stata output for questions 6-8

Questions 6-8 refer to the statistical output provided. The output results from an analysis of the relationship of retail sector employment in the Lakeland-Winterhaven Metropolitan Statistical Area (MSA) to five non-seasonally adjusted monthly local time series and one for the U.S., for the period January 1990 through December 2019. The table below presents definitions and relevant context. In the output, the letters "ln" before one of these variables names refers to the natural logarithm of the variable.

Variable	Definition	Relevant Context
retail	Retail employment, thousands	Not a component of the local economic base. Rather, the presence of the economic base creates an induced or derived demand for retail employment
lf	Labor force, thousands	
uer	Unemployment rate	A measure local labor market strength
usuer	US unemployment rate	A measure of national labor market strength
whsto	Warehousing and storage employment, thousands	Part of the local economic base due to Lakeland's location relative to major transportation routes
lh	Leisure and hospitality employment, thousands.	Part of the local economic base because tourism is purchased by those outside the local area
bp	Building permits for single family residences, thousands	Part of the economic base because the need for more living space ties directly to immigrants, whose income comes from out of the local area

6. (10%) Interpret the output provided for question 6. Include an explanation of what that output implies for modeling relationships among these time series.

7. (20%) You are interested in whether the Warehousing and Storage industry or the Leisure and Hospitality industry contribute more to the area's economic base. You estimate Models 1 and 2 to shed light on this question, working under the assumption that the sector with the larger impact on the economic base will have a larger association with induced (or derived) retail employment. Note that variables are not in logarithmic form in these models.

a. What is the difference between models 1 and 2? Why does it matter?

b. What is the cumulative effect of a one unit increase in warehousing and storage employment on retail employment?

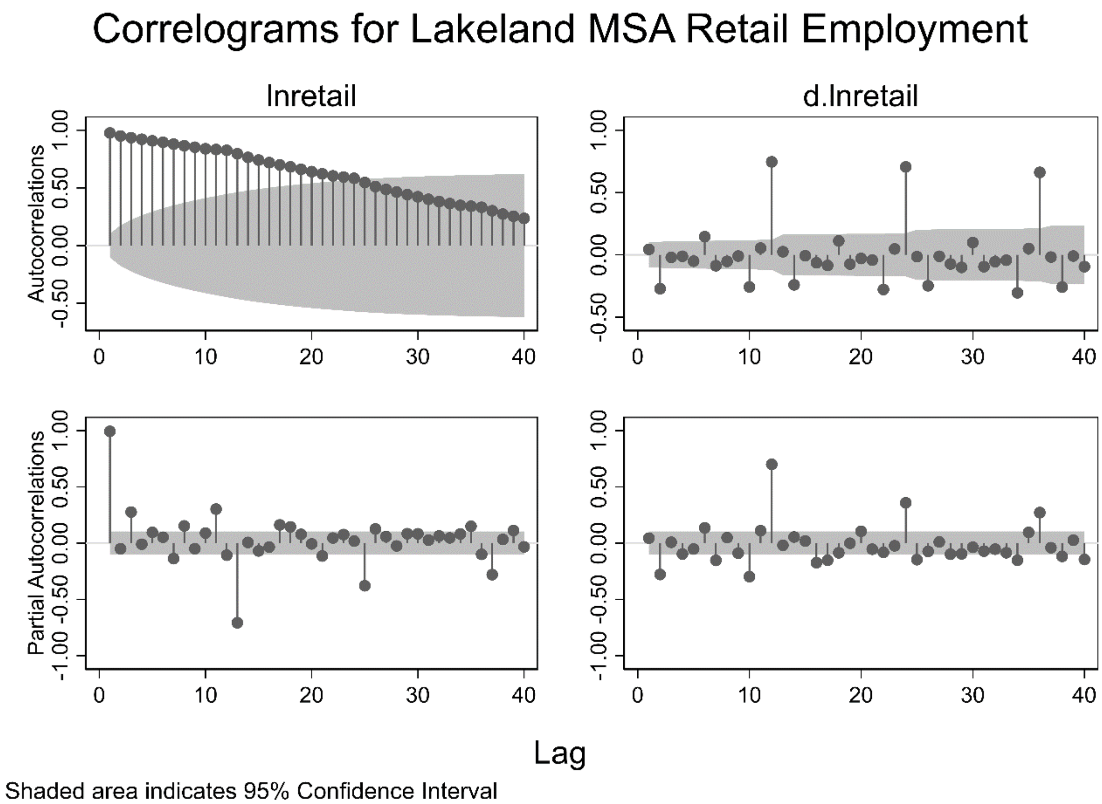
c. What is the cumulative effect of a one unit increase in leisure and hospitality employment on retail employment?

d. Provide and defend an answer to the question of which sector is more important to Lakeland's economic base based on this output.

8. (20%) Consider the output for models 3-6. Which is most promising as a basis for forecasting retail sector employment? Why? Thoroughly support your answer. Note that variables are in logarithmic form in these models.

Stata output for questions 6-8

Output for Question 6




```
. dfuller lnretail,  lag(12)
Augmented Dickey-Fuller test for unit root         Number of obs   =       347

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.598            -3.452            -2.876            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.4845
```

Stata output for questions 6-8

Output for Question 7

Model 1

```
. reg d.retail l(0/2,12)d.whsto l(0/2,12)d.lh i.month

      Source |       SS           df       MS      Number of obs   =       347
-------------+----------------------------------   F(19, 327)      =      8.71
       Model |  28.3297916        19  1.49104166   Prob > F        =    0.0000
    Residual |  55.9887033       327  .171219276   R-squared       =    0.3360
-------------+----------------------------------   Adj R-squared   =    0.2974
       Total |  84.3184949       346  .243695072   Root MSE        =    .41379

------------------------------------------------------------------------------
    D.retail |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       whsto |
         D1. |   1.582245   .3176526     4.98   0.000      .957345    2.207146
         LD. |   .2982512   .3138155     0.95   0.343    -.3191008    .9156031
        L2D. |  -.2752693   .3171678    -0.87   0.386    -.8992162    .3486775
       L12D. |   .9301951   .3256653     2.86   0.005     .2895316    1.570859
             |
          lh |
         D1. |   .3861625   .0842238     4.58   0.000     .2204736    .5518515
         LD. |  -.0936584   .0670806    -1.40   0.164    -.2256223    .0383056
        L2D. |  -.4027094   .0674687    -5.97   0.000     -.535437   -.2699819
       L12D. |   .2053196   .0836569     2.45   0.015      .040746    .3698932
             |
       month |
          2  |    .121886   .1167645     1.04   0.297    -.1078183    .3515903
          3  |   .1393876   .1131004     1.23   0.219    -.0831086    .3618837
          4  |    .199093   .1162354     1.71   0.088    -.0295705    .4277566
          5  |  -.1220647   .1128671    -1.08   0.280    -.3441019    .0999725
          6  |   .1261276   .1138682     1.11   0.269    -.0978792    .3501343
          7  |   .0729668   .1134542     0.64   0.521    -.1502254     .296159
          8  |    .064486   .1168059     0.55   0.581    -.1652999     .294272
          9  |   .1015845   .1148821     0.88   0.377    -.1244167    .3275858
         10  |    .116996   .1134443     1.03   0.303    -.1061768    .3401688
         11  |   .0814679    .114867     0.71   0.479    -.1445036    .3074395
         12  |   .1502732   .1264503     1.19   0.236    -.0984854    .3990319
             |
       _cons |  -.0999964   .0868256    -1.15   0.250    -.2708035    .0708107
------------------------------------------------------------------------------
```

Stata output for questions 6-8

Model 2

```
. newey d.retail l(0/2,12)d.whsto l(0/2,12)d.lh i.month , lag(24)

Regression with Newey-West standard errors      Number of obs     =        347
maximum lag: 24                                 F( 19,       327) =      21.78
                                                Prob > F          =     0.0000

------------------------------------------------------------------------------
             |             Newey-West
    D.retail |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       whsto |
         D1. |   1.582245   .3260628     4.85   0.000     .9408001    2.223691
         LD. |   .2982512    .306239     0.97   0.331    -.3041961    .9006984
        L2D. |  -.2752693   .2586356    -1.06   0.288     -.784069    .2335303
       L12D. |   .9301951   .3749445     2.48   0.014     .1925873    1.667803
             |
          lh |
         D1. |   .3861625   .0864389     4.47   0.000     .2161161     .556209
         LD. |  -.0936584   .0587851    -1.59   0.112     -.209303    .0219862
        L2D. |  -.4027094   .0538931    -7.47   0.000    -.5087303   -.2966885
       L12D. |   .2053196   .0644299     3.19   0.002     .0785702     .332069
             |
       month |
          2  |    .121886   .0556603     2.19   0.029     .0123885    .2313835
          3  |   .1393876   .0455228     3.06   0.002     .0498331     .228942
          4  |    .199093   .0519031     3.84   0.000     .0969869    .3011992
          5  |  -.1220647   .0872453    -1.40   0.163    -.2936975    .0495682
          6  |   .1261276   .0667523     1.89   0.060    -.0051905    .2574457
          7  |   .0729668    .050299     1.45   0.148    -.0259837    .1719173
          8  |    .064486   .0615042     1.05   0.295    -.0565077    .1854798
          9  |   .1015845   .0512611     1.98   0.048     .0007414    .2024277
         10  |    .116996   .0481678     2.43   0.016     .0222381     .211754
         11  |   .0814679   .0640806     1.27   0.205    -.0445944    .2075302
         12  |   .1502732   .0556949     2.70   0.007     .0407077    .2598387
             |
       _cons |  -.0999964   .0450287    -2.22   0.027     -.188579   -.0114138
------------------------------------------------------------------------------
```

Stata output for questions 6-8

Output for Question 8

Model 3

```
reg d.lnretail l(1/12)d.lnretail l(1/12)d.lnlf l(1/12)d.lnuer l(1/12)d.lnusuer ///
      l(1/12)d.lnwhsto l(1/12)d.lnlh l(1/12)d.lnbp i.month

(regression output excluded for space)

10-Fold Cross Validation
             |      RMSE
-------------+-----------
        est1 |  .0110013
        est2 |  .0127157
        est3 |  .0132477
        est4 |  .0124585
        est5 |  .0108486
        est6 |  .0129102
        est7 |  .0122873
        est8 |  .0142541
        est9 |  .0134581
       est10 |  .0143277

 Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   .01234727
Mean Absolute Errors     |   .00936199
Pseudo-R2                |   .59541209
-----------------------------------------

Akaike's information criterion and Bayesian information criterion
-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        347  876.3915   1147.103      96   -2102.207  -1732.671
-----------------------------------------------------------------------------
              Note: N=Obs used in calculating BIC; see [R] BIC note.

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |          0.345               1                   0.5571
       2     |          3.997               2                   0.1355
       3     |          6.610               3                   0.0854
       4     |          6.875               4                   0.1427
       5     |          8.304               5                   0.1402
       6     |          8.320               6                   0.2156
       7     |          8.965               7                   0.2552
       8     |          9.397               8                   0.3099
       9     |         20.072               9                   0.0175
      10     |         20.125              10                   0.0281
      11     |         21.396              11                   0.0295
      12     |         31.550              12                   0.0016
---------------------------------------------------------------------------
                        H0: no serial correlation
```

Stata output for questions 6-8

```
. testparm l(1/12)d.lnretail
 ( 1)  LD.lnretail = 0
 ( 2)  L2D.lnretail = 0
 ( 3)  L3D.lnretail = 0
 ( 4)  L4D.lnretail = 0
 ( 5)  L5D.lnretail = 0
 ( 6)  L6D.lnretail = 0
 ( 7)  L7D.lnretail = 0
 ( 8)  L8D.lnretail = 0
 ( 9)  L9D.lnretail = 0
 (10)  L10D.lnretail = 0
 (11)  L11D.lnretail = 0
 (12)  L12D.lnretail = 0
       F( 12,   251) =    4.42
            Prob > F =    0.0000

. testparm l(1/12)d.lnlf
 ( 1)  LD.lnlf = 0
 ( 2)  L2D.lnlf = 0
 ( 3)  L3D.lnlf = 0
 ( 4)  L4D.lnlf = 0
 ( 5)  L5D.lnlf = 0
 ( 6)  L6D.lnlf = 0
 ( 7)  L7D.lnlf = 0
 ( 8)  L8D.lnlf = 0
 ( 9)  L9D.lnlf = 0
 (10)  L10D.lnlf = 0
 (11)  L11D.lnlf = 0
 (12)  L12D.lnlf = 0
       F( 12,   251) =    1.80
            Prob > F =    0.0485

. testparm l(1/12)d.lnuer
 ( 1)  LD.lnuer = 0
 ( 2)  L2D.lnuer = 0
 ( 3)  L3D.lnuer = 0
 ( 4)  L4D.lnuer = 0
 ( 5)  L5D.lnuer = 0
 ( 6)  L6D.lnuer = 0
 ( 7)  L7D.lnuer = 0
 ( 8)  L8D.lnuer = 0
 ( 9)  L9D.lnuer = 0
 (10)  L10D.lnuer = 0
 (11)  L11D.lnuer = 0
 (12)  L12D.lnuer = 0
       F( 12,   251) =    2.06
            Prob > F =    0.0203

. testparm l(1/12)d.lnusuer
 ( 1)  LD.lnusuer = 0
 ( 2)  L2D.lnusuer = 0
 ( 3)  L3D.lnusuer = 0
 ( 4)  L4D.lnusuer = 0
 ( 5)  L5D.lnusuer = 0
 ( 6)  L6D.lnusuer = 0
 ( 7)  L7D.lnusuer = 0
 ( 8)  L8D.lnusuer = 0
 ( 9)  L9D.lnusuer = 0
 (10)  L10D.lnusuer = 0
 (11)  L11D.lnusuer = 0
 (12)  L12D.lnusuer = 0
       F( 12,   251) =    2.97
            Prob > F =    0.0007
```

```
. testparm l(1/12)d.lnwhsto
 ( 1)  LD.lnwhsto = 0
 ( 2)  L2D.lnwhsto = 0
 ( 3)  L3D.lnwhsto = 0
 ( 4)  L4D.lnwhsto = 0
 ( 5)  L5D.lnwhsto = 0
 ( 6)  L6D.lnwhsto = 0
 ( 7)  L7D.lnwhsto = 0
 ( 8)  L8D.lnwhsto = 0
 ( 9)  L9D.lnwhsto = 0
 (10)  L10D.lnwhsto = 0
 (11)  L11D.lnwhsto = 0
 (12)  L12D.lnwhsto = 0
       F( 12,   251) =    0.79
            Prob > F =    0.6619

. testparm l(1/12)d.lnlh
 ( 1)  LD.lnlh = 0
 ( 2)  L2D.lnlh = 0
 ( 3)  L3D.lnlh = 0
 ( 4)  L4D.lnlh = 0
 ( 5)  L5D.lnlh = 0
 ( 6)  L6D.lnlh = 0
 ( 7)  L7D.lnlh = 0
 ( 8)  L8D.lnlh = 0
 ( 9)  L9D.lnlh = 0
 (10)  L10D.lnlh = 0
 (11)  L11D.lnlh = 0
 (12)  L12D.lnlh = 0
       F( 12,   251) =    2.23
            Prob > F =    0.0109

. testparm l(1/12)d.lnbp

 ( 1)  LD.lnbp = 0
 ( 2)  L2D.lnbp = 0
 ( 3)  L3D.lnbp = 0
 ( 4)  L4D.lnbp = 0
 ( 5)  L5D.lnbp = 0
 ( 6)  L6D.lnbp = 0
 ( 7)  L7D.lnbp = 0
 ( 8)  L8D.lnbp = 0
 ( 9)  L9D.lnbp = 0
 (10)  L10D.lnbp = 0
 (11)  L11D.lnbp = 0
 (12)  L12D.lnbp = 0
       F( 12,   251) =    1.06
            Prob > F =    0.3908

. testparm i.month
 ( 1)  2.month = 0
 ( 2)  3.month = 0
 ( 3)  4.month = 0
 ( 4)  5.month = 0
 ( 5)  6.month = 0
 ( 6)  7.month = 0
 ( 7)  8.month = 0
 ( 8)  9.month = 0
 ( 9)  10.month = 0
 (10)  11.month = 0
 (11)  12.month = 0
       F( 11,   251) =    1.54
            Prob > F =    0.1168
```

Stata output for questions 6-8

Model 4

```
reg d.lnretail l(1/12)d.lnretail l(1/12)d.lnlf l(1/12)d.lnuer l(1/12)d.lnusuer ///
      l(1/12)d.lnlh i.month

(regression output excluded for space)

10-Fold Cross Validation
             |      RMSE
-------------+-----------
        est1 |  .0121968
        est2 |    .01538
        est3 |  .0111407
        est4 |  .0111009
        est5 |  .0161237
        est6 |  .0088344
        est7 |  .0116765
        est8 |  .0122955
        est9 |  .0108636
       est10 |  .0120957

Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   .0117591
Mean Absolute Errors     |   .00888454
Pseudo-R2                |   .62384824
-----------------------------------------

Akaike's information criterion and Bayesian information criterion
-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        347  876.3915   1131.786      72   -2119.572  -1842.421
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |          0.826               1                   0.3635
       2     |          4.449               2                   0.1081
       3     |          6.714               3                   0.0816
       4     |          6.976               4                   0.1372
       5     |          8.601               5                   0.1261
       6     |          8.611               6                   0.1967
       7     |          8.861               7                   0.2628
       8     |          9.156               8                   0.3293
       9     |         19.362               9                   0.0223
      10     |         20.354              10                   0.0261
      11     |         21.420              11                   0.0293
      12     |         33.047              12                   0.0010
---------------------------------------------------------------------------
                        H0: no serial correlation
```

Stata output for questions 6-8

Model 5

```
reg d.lnretail l(1,2,12,24)d.lnretail l(1,2,12)d.lnlf l(1,2,12)d.lnusuer ///
      l(1,2,12)d.lnlh i.month

(regression output excluded for space)

10-Fold Cross Validation
             |      RMSE
-------------+-----------
        est1 |  .0109601
        est2 |   .011824
        est3 |  .0098204
        est4 |  .0107243
        est5 |  .0080266
        est6 |  .0114437
        est7 |  .0099501
        est8 |  .0101516
        est9 |  .0105282
       est10 |  .0132597

Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   .01072511
Mean Absolute Errors     |   .00794259
Pseudo-R2                |   .67542327
-----------------------------------------

Akaike's information criterion and Bayesian information criterion
-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        335   855.331   1073.232      25   -2096.464  -2001.111
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |          0.241               1                   0.6236
       2     |          1.913               2                   0.3843
       3     |          5.381               3                   0.1459
       4     |          7.792               4                   0.0995
       5     |          8.714               5                   0.1210
       6     |          8.786               6                   0.1860
       7     |          9.698               7                   0.2064
       8     |         11.333               8                   0.1835
       9     |         11.583               9                   0.2379
      10     |         11.583              10                   0.3139
      11     |         12.216              11                   0.3476
      12     |         30.167              12                   0.0026
---------------------------------------------------------------------------
                        H0: no serial correlation
```

Stata output for questions 6-8

Model 6

```
reg d.lnretail l(1/4,12)d.lnretail i.month
```

(regression output excluded for space)

	RMSE
est1	.0174193
est2	.0107302
est3	.0119368
est4	.0112553
est5	.0087022
est6	.0122575
est7	.0090112
est8	.0109853
est9	.0170309
est10	.0153709

Leave-One-Out Cross-Validation Results

Method	Value
Root Mean Squared Errors	.01217818
Mean Absolute Errors	.00887577
Pseudo-R2	.58433196

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	347	876.3915	1041.371	17	-2048.743	-1983.304

Note: N=Obs used in calculating BIC; see [R] BIC note.

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	0.053	1	0.8173
2	3.794	2	0.1500
3	6.354	3	0.0956
4	12.159	4	0.0162
5	13.370	5	0.0201
6	13.530	6	0.0353
7	16.770	7	0.0189
8	25.818	8	0.0011
9	25.853	9	0.0022
10	25.919	10	0.0039
11	26.303	11	0.0058
12	84.295	12	0.0000

H0: no serial correlation