62nd Annual North American Meetings of the

Regional Science Association International

Portland, OR

November 11-14, 2015

GeoDaSpace – spreg Workshop
Exercise IPython Notebooks

Luc Anselin

luc.anselin@asu.edu

Contents

Chapter 5 - OLS

November 5, 2015

This notebook contains the PySAL/spreg code for Chapter 5 - OLS
in Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL.
by Luc Anselin and Sergio J. Rey


```
In [1]: __author__ = "Luc Anselin luc.anselin@asu.edu"
```

0.1 Basic Regression Setup

Creating arrays for y and x using the Baltimore example
Preliminaries, importing **numpy** and **PySAL**

```
In [1]: %pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

```
In [2]: import numpy as np
        import pysal
```

Loading the data set and creating the data object
This assumes that the sample data sets are contained in the 'data' subdirectory of the current working directory

```
In [3]: db = pysal.open('data/baltim.dbf','r')
```

Quick check on the number of observations and the contents of the data set in the db object

```
In [4]: len(db)
```

```
Out[4]: 211
```

```
In [5]: db.header
```

```
Out[5]: ['STATION',
         'PRICE',
         'NROOM',
         'DWELL',
         'NBATH',
         'PATIO',
         'FIREPL',
         'AC',
         'BMENT',
         'NSTOR',
         'GAR',
         'AGE',
```

```
           'CITCOU',
           'LOTSZ',
           'SQFT',
           'X',
           'Y']
```

y - dependent variable is PRICE

```
In [6]: y_name = "PRICE"
```

Create the y array as a n by 1 column vector (hence the transpose **T**)

```
In [7]: y = np.array([db.by_col(y_name)]).T
```

Check on the dimensions

```
In [8]: y.shape
```

```
Out[8]: (211, 1)
```

x - the explanatory variables
First create a list with the variable names, then use a list comprehension to create the **x** array

```
In [9]: x_names = ['NROOM','NBATH','PATIO','FIREPL','AC','GAR','AGE',
                   'LOTSZ','SQFT']
        x = np.array([db.by_col(var) for var in x_names]).T
```

Check on dimensions

```
In [10]: x.shape
```

```
Out[10]: (211, 9)
```

Model weights - needed for spatial diagnostics
k nearest neighbor with k=4 constructed from baltim.shp, using STATION as the ID variable

```
In [11]: w = pysal.knnW_from_shapefile('data/baltim.shp',
                                       k=4,idVariable='STATION')
```

Quick check on dimension

```
In [12]: w.n
```

```
Out[12]: 211
```

row-standardize - is **always** necessary

```
In [13]: w.transform = 'r'
```

Quick check on the values of the weights

```
In [14]: w.weights[1]
```

```
Out[14]: [0.25, 0.25, 0.25, 0.25]
```

Kernel weights - needed for HAC standard error option
triangular adaptive bandwidth kernel with k=12, constructed from baltim shape file

```
In [15]: kw = pysal.adaptive_kernelW_from_shapefile('data/baltim.shp',
                                                   k=12,diagonal=True,idVariable='STATION')
```

Check on dimension and actual weights - Note how diagonal element = 1, not 0

This is the case for a triangular kernel function, whether or not the **diagonal** option is set to True or not. Otherwise the latter **must** be set to **True** (see also the example in Chapter 6).

```
In [16]: kw.n

Out[16]: 211

In [17]: kw.weights[1]

Out[17]: [1.0,
          0.4849212978701576,
          0.3611234988877102,
          0.33567595079889523,
          0.31302206649524533,
          0.2768409033403173,
          0.27156871375844516,
          0.14137042299357871,
          0.1179222999771653,
          0.08526805755421585,
          0.0564397505258174,
          0.007682671413889897,
          9.99999900663795e-08]
```

0.2 Basic OLS

Default settings, no variable names

```
In [18]: ols1 = pysal.spreg.OLS(y,x)
```

the regression coefficiens, in the order of the columns of **x**

```
In [19]: ols1

Out[19]: <pysal.spreg.ols.OLS instance at 0x109dcd5f0>

In [20]: dir(ols1)

Out[20]: ['__doc__',
          '__init__',
          '__module__',
          '__summary',
          '_cache',
          'aic',
          'ar2',
          'betas',
          'breusch_pagan',
          'f_stat',
          'jarque_bera',
          'k',
          'koenker_bassett',
          'logll',
          'mean_y',
          'mulColli',
          'n',
          'name_ds',
```

```
          'name_gwk',
          'name_w',
          'name_x',
          'name_y',
          'predy',
          'r2',
          'robust',
          'schwarz',
          'sig2',
          'sig2ML',
          'sig2n',
          'sig2n_k',
          'std_err',
          'std_y',
          'summary',
          't_stat',
          'title',
          'u',
          'utu',
          'vm',
          'x',
          'xtx',
          'xtxi',
          'y']
```

```
In [21]: ols1.betas

Out[21]: array([[ 23.26999634],
                [  0.22249352],
                [  5.64840507],
                [ 10.33587549],
                [ 11.17272765],
                [  7.85416373],
                [  5.40206849],
                [ -0.21345497],
                [  0.09490639],
                [  0.18775621]])
```

```
In [22]: ols1.betas

Out[22]: array([[ 23.26999634],
                [  0.22249352],
                [  5.64840507],
                [ 10.33587549],
                [ 11.17272765],
                [  7.85416373],
                [  5.40206849],
                [ -0.21345497],
                [  0.09490639],
                [  0.18775621]])
```

pretty output, but no variable names (coefficients in order of columns of **x**) or any other descriptive information about the data set, etc.

```
In [23]: print ols1.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :     unknown
Dependent Variable  :     dep_var               Number of Observations:         211
Mean dependent var  :     44.3072               Number of Variables   :          10
S.D. dependent var  :     23.6061               Degrees of Freedom    :         201
R-squared           :      0.6500
Adjusted R-squared  :      0.6343
Sum squared residual:   40960.463               F-statistic           :     41.4718
Sigma-square        :     203.783               Prob(F-statistic)     :    3.24e-41
S.E. of regression  :      14.275               Log likelihood        :    -855.223
Sigma-square ML     :     194.125               Akaike info criterion :    1730.446
S.E of regression ML:     13.9329               Schwarz criterion     :    1763.965

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      23.2699963       5.2241658       4.4542989       0.0000140
               var_1       0.2224935       1.2279277       0.1811943       0.8563976
               var_2       5.6484051       2.0182820       2.7986204       0.0056318
               var_3      10.3358755       3.1202987       3.3124635       0.0010966
               var_4      11.1727277       2.7323132       4.0891094       0.0000626
               var_5       7.8541637       2.7024343       2.9063292       0.0040671
               var_6       5.4020685       1.9636140       2.7510848       0.0064824
               var_7      -0.2134550       0.0573535      -3.7217456       0.0002568
               var_8       0.0949064       0.0165355       5.7395480       0.0000000
               var_9       0.1877562       0.1890598       0.9931047       0.3218530
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER          20.609

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2         429.933           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                9         167.792           0.0000
Koenker-Bassett test              9          39.197           0.0000
================================ END OF REPORT =====================================
```

OLS with variable and data set names

```
In [24]: ols1a = pysal.spreg.OLS(y,x,name_y=y_name,name_x=x_names,
                                 name_ds='baltim.shp')

In [25]: print ols1a.summary

REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
```

```
-----------------------------------------
Data set            :  baltim.shp
Dependent Variable  :       PRICE           Number of Observations:         211
Mean dependent var  :     44.3072           Number of Variables   :          10
S.D. dependent var  :     23.6061           Degrees of Freedom    :         201
R-squared           :      0.6500
Adjusted R-squared  :      0.6343
Sum squared residual:   40960.463           F-statistic           :     41.4718
Sigma-square        :     203.783           Prob(F-statistic)     :    3.24e-41
S.E. of regression  :      14.275           Log likelihood        :    -855.223
Sigma-square ML     :     194.125           Akaike info criterion :    1730.446
S.E of regression ML:     13.9329           Schwarz criterion     :    1763.965

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      23.2699963       5.2241658       4.4542989       0.0000140
                  AC       7.8541637       2.7024343       2.9063292       0.0040671
                 AGE      -0.2134550       0.0573535      -3.7217456       0.0002568
              FIREPL      11.1727277       2.7323132       4.0891094       0.0000626
                 GAR       5.4020685       1.9636140       2.7510848       0.0064824
               LOTSZ       0.0949064       0.0165355       5.7395480       0.0000000
               NBATH       5.6484051       2.0182820       2.7986204       0.0056318
               NROOM       0.2224935       1.2279277       0.1811943       0.8563976
               PATIO      10.3358755       3.1202987       3.3124635       0.0010966
                SQFT       0.1877562       0.1890598       0.9931047       0.3218530
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           20.609

TEST ON NORMALITY OF ERRORS
TEST                             DF         VALUE           PROB
Jarque-Bera                       2          429.933           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF         VALUE           PROB
Breusch-Pagan test                9          167.792           0.0000
Koenker-Bassett test              9           39.197           0.0000
================================ END OF REPORT =====================================
```

note how the order of the coefficients is different from before, it follows the alfabetical order of the variable names

0.3 OLS with White Test

set **white_test = True**

```
In [26]: ols2 = pysal.spreg.OLS(y,x,white_test=True,name_y=y_name,
                                name_x=x_names,name_ds='baltim.shp')

In [27]: print ols2.summary
```

```
REGRESSION
----------
```

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :  baltim.shp
Dependent Variable  :       PRICE                Number of Observations:         211
Mean dependent var  :     44.3072                Number of Variables   :          10
S.D. dependent var  :     23.6061                Degrees of Freedom    :         201
R-squared           :      0.6500
Adjusted R-squared  :      0.6343
Sum squared residual:   40960.463                F-statistic           :     41.4718
Sigma-square        :     203.783                Prob(F-statistic)     :    3.24e-41
S.E. of regression  :      14.275                Log likelihood        :    -855.223
Sigma-square ML     :     194.125                Akaike info criterion :    1730.446
S.E of regression ML:     13.9329                Schwarz criterion     :    1763.965

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      23.2699963       5.2241658       4.4542989       0.0000140
                  AC       7.8541637       2.7024343       2.9063292       0.0040671
                 AGE      -0.2134550       0.0573535      -3.7217456       0.0002568
              FIREPL      11.1727277       2.7323132       4.0891094       0.0000626
                 GAR       5.4020685       1.9636140       2.7510848       0.0064824
               LOTSZ       0.0949064       0.0165355       5.7395480       0.0000000
               NBATH       5.6484051       2.0182820       2.7986204       0.0056318
               NROOM       0.2224935       1.2279277       0.1811943       0.8563976
               PATIO      10.3358755       3.1202987       3.3124635       0.0010966
                SQFT       0.1877562       0.1890598       0.9931047       0.3218530
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER          20.609

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2         429.933           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                9         167.792           0.0000
Koenker-Bassett test              9          39.197           0.0000

SPECIFICATION ROBUST TEST
TEST                             DF        VALUE           PROB
White                            51         164.335           0.0000
================================ END OF REPORT =====================================
```

now the **White** test is included as one of the diagnostics

0.4 OLS with Spatial Diagnostics

specify the weights (**w**) and set **spat_diag = True** and **moran = True**
specify a name for the weights in **name_w**, all the rest is as before
for convenience, white_test is back to default of False

```
In [28]: ols3 = pysal.spreg.OLS(y,x,w=w,spat_diag=True,moran=True,
                                name_y=y_name,name_x=x_names,name_w='baltim_k4',
                                name_ds='baltim.shp')

In [29]: print ols3.summary

REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :  baltim.shp
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE              Number of Observations:         211
Mean dependent var  :     44.3072              Number of Variables   :          10
S.D. dependent var  :     23.6061              Degrees of Freedom    :         201
R-squared           :      0.6500
Adjusted R-squared  :      0.6343
Sum squared residual:   40960.463              F-statistic           :     41.4718
Sigma-square        :     203.783              Prob(F-statistic)     :    3.24e-41
S.E. of regression  :      14.275              Log likelihood        :    -855.223
Sigma-square ML     :     194.125              Akaike info criterion :    1730.446
S.E of regression ML:     13.9329              Schwarz criterion     :    1763.965

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      23.2699963       5.2241658       4.4542989       0.0000140
                  AC       7.8541637       2.7024343       2.9063292       0.0040671
                 AGE      -0.2134550       0.0573535      -3.7217456       0.0002568
              FIREPL      11.1727277       2.7323132       4.0891094       0.0000626
                 GAR       5.4020685       1.9636140       2.7510848       0.0064824
               LOTSZ       0.0949064       0.0165355       5.7395480       0.0000000
               NBATH       5.6484051       2.0182820       2.7986204       0.0056318
               NROOM       0.2224935       1.2279277       0.1811943       0.8563976
               PATIO      10.3358755       3.1202987       3.3124635       0.0010966
                SQFT       0.1877562       0.1890598       0.9931047       0.3218530
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           20.609

TEST ON NORMALITY OF ERRORS
TEST                             DF         VALUE           PROB
Jarque-Bera                       2        429.933          0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF         VALUE           PROB
Breusch-Pagan test                9        167.792          0.0000
Koenker-Bassett test              9         39.197          0.0000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF        VALUE           PROB
Moran's I (error)              0.1095        2.696          0.0070
Lagrange Multiplier (lag)         1         31.896          0.0000
```

```
Robust LM (lag)                   1          29.305           0.0000
Lagrange Multiplier (error)       1           5.658           0.0174
Robust LM (error)                 1           3.067           0.0799
Lagrange Multiplier (SARMA)       2          34.963           0.0000

================================ END OF REPORT =====================================
```

diagnostics for spatial dependence at the bottom of the listing

0.5 OLS with White Standard Errors

set **robust= 'white'** for the default scaling, dividing by $n-k$, the spatial diagnostics have been turned off

```
In [30]: ols4 = pysal.spreg.OLS(y,x,robust='white',
                                name_y=y_name,name_x=x_names,name_ds='baltim.shp')

In [31]: print ols4.summary

REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :  baltim.shp
Dependent Variable  :       PRICE                Number of Observations:         211
Mean dependent var  :     44.3072                Number of Variables   :          10
S.D. dependent var  :     23.6061                Degrees of Freedom    :         201
R-squared           :      0.6500
Adjusted R-squared  :      0.6343
Sum squared residual:   40960.463                F-statistic           :     41.4718
Sigma-square        :     203.783                Prob(F-statistic)     :    3.24e-41
S.E. of regression  :      14.275                Log likelihood        :    -855.223
Sigma-square ML     :     194.125                Akaike info criterion :    1730.446
S.E of regression ML:     13.9329                Schwarz criterion     :    1763.965

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      23.2699963       6.5488415       3.5532997       0.0004741
                  AC       7.8541637       3.1147472       2.5216055       0.0124581
                 AGE      -0.2134550       0.1082717      -1.9714745       0.0500417
              FIREPL      11.1727277       3.0235003       3.6952957       0.0002831
                 GAR       5.4020685       3.2537312       1.6602688       0.0984196
               LOTSZ       0.0949064       0.0260489       3.6433902       0.0003425
               NBATH       5.6484051       2.4995696       2.2597511       0.0249092
               NROOM       0.2224935       1.3934904       0.1596663       0.8733042
               PATIO      10.3358755       4.1978783       2.4621665       0.0146520
                SQFT       0.1877562       0.2099744       0.8941860       0.3722918
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER          20.609

TEST ON NORMALITY OF ERRORS
TEST                             DF         VALUE           PROB
```

```
Jarque-Bera                       2          429.933           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF         VALUE            PROB
Breusch-Pagan test                9          167.792           0.0000
Koenker-Bassett test              9           39.197           0.0000
================================ END OF REPORT =====================================
```

turn off scaling (i.e., divide by n) by setting **sig2n_k = False**

```
In [32]: ols5 = pysal.spreg.OLS(y,x,robust='white',sig2n_k=False,
                                name_y=y_name,name_x=x_names,name_ds='baltim.shp')

In [33]: print ols5.summary

REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :  baltim.shp
Dependent Variable  :       PRICE                Number of Observations:         211
Mean dependent var  :     44.3072                Number of Variables   :          10
S.D. dependent var  :     23.6061                Degrees of Freedom    :         201
R-squared           :      0.6500
Adjusted R-squared  :      0.6343
Sum squared residual:   40960.463                F-statistic           :     41.4718
Sigma-square        :     194.125                Prob(F-statistic)     :    3.24e-41
S.E. of regression  :      13.933                Log likelihood        :    -855.223
Sigma-square ML     :     194.125                Akaike info criterion :    1730.446
S.E of regression ML:     13.9329                Schwarz criterion     :    1763.965

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      23.2699963       6.3917721       3.6406174       0.0003459
                  AC       7.8541637       3.0400421       2.5835707       0.0104876
                 AGE      -0.2134550       0.1056749      -2.0199209       0.0447190
              FIREPL      11.1727277       2.9509837       3.7861028       0.0002020
                 GAR       5.4020685       3.1756927       1.7010678       0.0904767
               LOTSZ       0.0949064       0.0254242       3.7329217       0.0002464
               NBATH       5.6484051       2.4396191       2.3152815       0.0216067
               NROOM       0.2224935       1.3600685       0.1635899       0.8702183
               PATIO      10.3358755       4.0971951       2.5226710       0.0124216
                SQFT       0.1877562       0.2049383       0.9161595       0.3606811
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           20.609

TEST ON NORMALITY OF ERRORS
TEST                             DF         VALUE            PROB
Jarque-Bera                       2          429.933           0.0000
```

```
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                9         167.792           0.0000
Koenker-Bassett test              9          39.197           0.0000
================================ END OF REPORT =====================================
```

note how all the estimates remain the same, but the standard errors change (slightly)

0.6 OLS with HAC Standard Errors

set **robust = 'hac'**, specify the kernel weights **gwk = kw** and give a name for the weights in **name_gwk**

```
In [34]: ols6 = pysal.spreg.OLS(y,x,robust='hac',gwk=kw,
                                name_y=y_name,name_x=x_names,
                                name_gwk='baltim_tri_k12',name_ds='baltim.shp')
```

```
In [35]: pysal.spreg.OLS()
```

```
        ---------------------------------------------------------------------------

        TypeError                                 Traceback (most recent call last)

        <ipython-input-35-65d92a8396c9> in <module>()
    ----> 1 pysal.spreg.OLS()


        TypeError: __init__() takes at least 3 arguments (1 given)
```

```
In [36]: print ols6.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :  baltim.shp
Dependent Variable  :       PRICE               Number of Observations:         211
Mean dependent var  :     44.3072               Number of Variables   :          10
S.D. dependent var  :     23.6061               Degrees of Freedom    :         201
R-squared           :      0.6500
Adjusted R-squared  :      0.6343
Sum squared residual:   40960.463               F-statistic           :     41.4718
Sigma-square        :     203.783               Prob(F-statistic)     :    3.24e-41
S.E. of regression  :      14.275               Log likelihood        :    -855.223
Sigma-square ML     :     194.125               Akaike info criterion :    1730.446
S.E of regression ML:     13.9329               Schwarz criterion     :    1763.965

HAC Standard Errors; Kernel Weights: baltim_tri_k12
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      23.2699963       6.7557589       3.4444682       0.0006964
                  AC       7.8541637       3.2786034       2.3955822       0.0175113
                 AGE      -0.2134550       0.1067463      -1.9996471       0.0468849
```

```
          FIREPL      11.1727277       2.9682469       3.7640830       0.0002194
             GAR       5.4020685       3.5732508       1.5118078       0.1321534
           LOTSZ       0.0949064       0.0251785       3.7693418       0.0002151
           NBATH       5.6484051       2.4853389       2.2726901       0.0241027
           NROOM       0.2224935       1.3841624       0.1607424       0.8724578
           PATIO      10.3358755       4.4904057       2.3017687       0.0223729
            SQFT       0.1877562       0.1867878       1.0051843       0.3160167
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER          20.609

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2         429.933           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                9         167.792           0.0000
Koenker-Bassett test              9          39.197           0.0000
================================ END OF REPORT =====================================
```

Again, same estimates as before, but different standard errors

0.7 Accessing Individual Functions

Muticollinearity Diagnostics - Variance inflation factor
use **vif** and pass a regression object, e.g., **ols1**

```
In [37]: v = pysal.spreg.diagnostics.vif(ols1)

In [38]: v

Out[38]: [None,
          (2.1282653585417042, 0.46986622038767001),
          (1.7612737468067756, 0.56777091114485745),
          (1.2635001075481409, 0.79145224763021937),
          (1.4167716982292522, 0.70583002275514606),
          (1.3859551858036305, 0.72152405088059268),
          (1.4320753736980556, 0.69828726781167594),
          (1.3336259007911946, 0.7498354669077244),
          (1.5646798143926102, 0.63910839189050828),
          (2.4833041895651888, 0.40268928961743256)]
```

extract the statistics by skipping over the first element in the list, **v[1:]**

```
In [39]: v[1:]

Out[39]: [(2.1282653585417042, 0.46986622038767001),
          (1.7612737468067756, 0.56777091114485745),
          (1.2635001075481409, 0.79145224763021937),
          (1.4167716982292522, 0.70583002275514606),
          (1.3859551858036305, 0.72152405088059268),
          (1.4320753736980556, 0.69828726781167594),
          (1.3336259007911946, 0.7498354669077244),
          (1.5646798143926102, 0.63910839189050828),
          (2.4833041895651888, 0.40268928961743256)]
```

This list contains a tuple with for each variable (in the order given in **x**) the variance inflation factor (VIF) and its inverse, the tolerance factor. The rule of thumb is to keep the VIF under 5 or 10 and the tolerance factor above 0.2 or 0.1.

Lagrange Multiplier Tests - accessing individual statistics

use **LMtests** with a regression object and a weights object

```
In [40]: vw = pysal.spreg.diagnostics_sp.LMtests(ols1,w)

In [41]: vw

Out[41]: <pysal.spreg.diagnostics_sp.LMtests instance at 0x10ae5a3f8>
```

check the contents of this object using **dir**

```
In [42]: dir(vw)

Out[42]: ['__doc__', '__init__', '__module__', 'lme', 'lml', 'rlme', 'rlml', 'sarma']
```

extract the specific test statistic as attributes, e.g., **lme, lml, rlme, rlml, sarma**

```
In [43]: vw.lme

Out[43]: (5.6577758243448066, 0.017378147300347417)

In [44]: vw.lml

Out[44]: (31.895904502800899, 1.6265975177114666e-08)
```

0.8 Practice

Create a regression object using the classic Harrison-Rubinfeld Boston house price data set (included as Boston.shp). Regression median house value (MEDV) on crime rate (CRIM), Charles river dummy (CHAS), nitric oxides (NOX), number of rooms (RM), age (AGE), weighted distance to five employment centers (DIS) and percent "lower status population" (LSTAT). A full description of the Boston data set is available on the GeoDa Center sample data set site.

Use a k-nearest neighbor spatial weights (k = 4) for the spatial diagnostics, and a triangular adaptive bandwidth kernel weights (k = 12) for the HAC standard errors.

What is the most likely alternative spatial regression model, given the results of a spatial specification search.

Try including the White test and see what happens.

Try any other combinations of explanatory variables, spatial weights and standard error specifications.

```
In [ ]:
```

Chapter 5 - OLS practice solutions

November 5, 2015

0.1 Solutions to Practice Example Chapter 5

```
In [1]: %pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

```
In [2]: import numpy as np
        import pysal
```

```
In [3]: db = pysal.open('data/boston.dbf','r')
```

```
In [4]: len(db)
```

```
Out[4]: 506
```

```
In [5]: db.header
```

```
Out[5]: ['POLY_ID',
         'ID',
         'TOWN',
         'TOWNNO',
         'TRACT',
         'LON',
         'LAT',
         'x',
         'y',
         'MEDV',
         'CMEDV',
         'CRIM',
         'ZN',
         'INDUS',
         'CHAS',
         'NOX',
         'RM',
         'AGE',
         'DIS',
         'RAD',
         'TAX',
         'PTRATIO',
         'B',
         'LSTAT']
```

```
In [6]: y_name = 'MEDV'
        y = np.array([db.by_col(y_name)]).T
```

```
In [7]: x_names = ['CRIM','CHAS','NOX','RM','AGE','DIS','LSTAT']
        x = np.array([db.by_col(var) for var in x_names]).T

In [8]: w = pysal.knnW_from_shapefile('data/boston.shp',
                                       k=4,idVariable='POLY_ID')

In [9]: w.n

Out[9]: 506

In [10]: w[1]

Out[10]: {30: 1.0, 32: 1.0, 33: 1.0, 35: 1.0}

In [11]: w.transform = 'r'

In [12]: kw = pysal.adaptive_kernelW_from_shapefile('data/boston.shp',
                                                    k=12,idVariable='POLY_ID')

In [13]: kw.n

Out[13]: 506

In [14]: kw[1]

Out[14]: {1: 1.0,
          23: 0.017975809698839718,
          24: 0.07789763977511166,
          25: 0.03599595976755843,
          26: 9.99999900663795e-08,
          28: 0.06731962890224807,
          29: 0.1567134489385178,
          30: 0.22098960210545193,
          31: 0.14004706565547476,
          32: 0.24298214626370784,
          33: 0.15996168962289714,
          34: 0.06768918779207955,
          35: 0.1978979513164435}

In [15]: reg1 = pysal.spreg.OLS(y,x,w=w,spat_diag=True,moran=True,
                                name_y=y_name,name_x=x_names,name_w='boston_k4',
                                name_ds='boston.shp')

In [16]: print reg1.summary

REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :  boston.shp
Weights matrix      :   boston_k4
Dependent Variable  :        MEDV                Number of Observations:         506
Mean dependent var  :     22.5328                Number of Variables   :           8
S.D. dependent var  :      9.1971                Degrees of Freedom    :         498
R-squared           :      0.6788
Adjusted R-squared  :      0.6743
Sum squared residual:   13719.651                F-statistic           :    150.3613
```

```
Sigma-square        :      27.550                Prob(F-statistic)     :  1.732e-118
S.E. of regression  :       5.249                Log likelihood        :   -1552.895
Sigma-square ML     :      27.114                Akaike info criterion :    3121.790
S.E of regression ML:      5.2071                Schwarz criterion     :    3155.602

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      11.4725706       3.9048811       2.9380077       0.0034564
                 AGE      -0.0168636       0.0143248      -1.1772306       0.2396658
                CHAS       3.8639095       0.9385306       4.1169776       0.0000449
                CRIM      -0.1074230       0.0315102      -3.4091522       0.0007045
                 DIS      -1.0649851       0.1931115      -5.5148716       0.0000001
               LSTAT      -0.5746656       0.0547460     -10.4969367       0.0000000
                 NOX     -13.2572627       3.5597665      -3.7241945       0.0002183
                  RM       4.9333429       0.4323621      11.4102121       0.0000000
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           47.587

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2         482.230           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                7          97.286           0.0000
Koenker-Bassett test              7          32.560           0.0000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Moran's I (error)              0.4736        16.803           0.0000
Lagrange Multiplier (lag)         1         167.220           0.0000
Robust LM (lag)                   1           4.999           0.0254
Lagrange Multiplier (error)       1         261.062           0.0000
Robust LM (error)                 1          98.841           0.0000
Lagrange Multiplier (SARMA)       2         266.061           0.0000

================================ END OF REPORT =====================================
```

```
In [17]: reg2 = pysal.spreg.OLS(y,x,w=w,robust='hac',gwk=kw,spat_diag=True,moran=True,
                                name_y=y_name,name_x=x_names,name_w='boston_k4',name_gwk='boston_tri_k1
                                name_ds='boston.shp')

In [18]: print reg2.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :  boston.shp
Weights matrix      :   boston_k4
Dependent Variable  :        MEDV                Number of Observations:         506
```

```
Mean dependent var  :     22.5328                Number of Variables   :           8
S.D. dependent var  :      9.1971                Degrees of Freedom    :         498
R-squared           :      0.6788
Adjusted R-squared  :      0.6743
Sum squared residual:   13719.651                F-statistic           :    150.3613
Sigma-square        :      27.550                Prob(F-statistic)     :  1.732e-118
S.E. of regression  :       5.249                Log likelihood        :   -1552.895
Sigma-square ML     :      27.114                Akaike info criterion :    3121.790
S.E of regression ML:      5.2071                Schwarz criterion     :    3155.602

HAC Standard Errors; Kernel Weights: boston_tri_k12
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      11.4725706       9.1535525       1.2533462       0.2106678
                 AGE      -0.0168636       0.0214346      -0.7867480       0.4318037
                CHAS       3.8639095       1.4681591       2.6318056       0.0087567
                CRIM      -0.1074230       0.0390349      -2.7519731       0.0061398
                 DIS      -1.0649851       0.2383633      -4.4679079       0.0000098
               LSTAT      -0.5746656       0.1316164      -4.3662150       0.0000154
                 NOX     -13.2572627       4.9271588      -2.6906506       0.0073709
                  RM       4.9333429       1.1433919       4.3146561       0.0000193
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER          47.587

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2        482.230           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                7         97.286           0.0000
Koenker-Bassett test              7         32.560           0.0000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Moran's I (error)              0.4736       16.803           0.0000
Lagrange Multiplier (lag)         1        167.220           0.0000
Robust LM (lag)                   1          4.999           0.0254
Lagrange Multiplier (error)       1        261.062           0.0000
Robust LM (error)                 1         98.841           0.0000
Lagrange Multiplier (SARMA)       2        266.061           0.0000

================================ END OF REPORT =====================================
```

In [ ]:

In [ ]:

Chapter 6 - 2SLS

November 5, 2015

This notebook contains the PySAL/spreg code for Chapter 6 - 2SLS
in Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL.
by Luc Anselin and Sergio J. Rey


```
In [1]: __author__ = "Luc Anselin luc.anselin@asu.edu"
```

0.1 Basic Regression Setup

Creating arrays for y, x, the endogenous variables yend and the instruments q
Using the **natregimes.dbf** example
Preliminaries, importing **numpy** and **pysal**

```
In [1]: %pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

```
In [2]: import numpy as np
        import pysal
```

Loading the data set and creating the data object

```
In [3]: db = pysal.open('data/natregimes.dbf','r')
```

```
In [4]: len(db)
```

```
Out[4]: 3085
```

```
In [5]: db.header
```

```
Out[5]: ['REGIONS',
         'NOSOUTH',
         'POLY_ID',
         'NAME',
         'STATE_NAME',
         'STATE_FIPS',
         'CNTY_FIPS',
         'FIPS',
         'STFIPS',
         'COFIPS',
         'FIPSNO',
         'SOUTH',
         'HR60',
         'HR70',
         'HR80',
```

```
'HR90',
'HC60',
'HC70',
'HC80',
'HC90',
'PO60',
'PO70',
'PO80',
'PO90',
'RD60',
'RD70',
'RD80',
'RD90',
'PS60',
'PS70',
'PS80',
'PS90',
'UE60',
'UE70',
'UE80',
'UE90',
'DV60',
'DV70',
'DV80',
'DV90',
'MA60',
'MA70',
'MA80',
'MA90',
'POL60',
'POL70',
'POL80',
'POL90',
'DNL60',
'DNL70',
'DNL80',
'DNL90',
'MFIL59',
'MFIL69',
'MFIL79',
'MFIL89',
'FP59',
'FP69',
'FP79',
'FP89',
'BLK60',
'BLK70',
'BLK80',
'BLK90',
'GI59',
'GI69',
'GI79',
'GI89',
'FH60',
```

```
         'FH70',
         'FH80',
         'FH90',
         'West']
```

y - dependent variable HR90

```
In [6]: y_name = "HR90"
        y = np.array([db.by_col(y_name)]).T
```

```
In [7]: y.shape
```

```
Out[7]: (3085, 1)
```

x - array with observations on explanatory variables

```
In [8]: x_names = ['RD90','MA90','PS90']
        x = np.array([db.by_col(var) for var in x_names]).T
```

```
In [9]: x.shape
```

```
Out[9]: (3085, 3)
```

yend - endogenous explanatory variable, UE90

```
In [10]: yend_names = ['UE90']
         yend = np.array([db.by_col(var) for var in yend_names]).T
```

```
In [11]: yend.shape
```

```
Out[11]: (3085, 1)
```

q - array of instruments

```
In [12]: q_names = ['FH90','FP89','GI89']
         q = np.array([db.by_col(var) for var in q_names]).T
```

```
In [13]: q.shape
```

```
Out[13]: (3085, 3)
```

Creating the model weights, queen contiguity for natregimes.shp, using FIPSNO as the ID variable

```
In [14]: w = pysal.queen_from_shapefile('data/natregimes.shp',idVariable="FIPSNO")
```

```
In [15]: w.n
```

```
Out[15]: 3085
```

row-standardize

```
In [16]: w.transform = 'r'
```

Creating the kernel weights, triangular with k=20
since natregimes.shp coordinates are in lat-lon, use **radius** to get great circle distance
note that **diagonal = True** to ensure that the value of 1 is on the diagonal

```
In [17]: kw = pysal.adaptive_kernelW_from_shapefile('data/natregimes.shp',
                                                     k=20,radius=pysal.cg.RADIUS_EARTH_MILES,
                                                     diagonal=True,idVariable='FIPSNO')
```

0.2 Basic Two Stage Least Squares

default settings including variable names and data set name

```
In [18]: reg1 = pysal.spreg.TSLS(y,x,yend,q,name_y=y_name,name_x=x_names,
                                 name_yend=yend_names,name_q=q_names,name_ds='nat.dbf')
```

regression coefficients, in alphabetical order of the variable names

```
In [19]: reg1.betas

Out[19]: array([[ 15.64555155],
                [  5.72924882],
                [ -0.09837584],
                [  1.8770506 ],
                [ -0.91445539]])
```

pretty listing

```
In [20]: print reg1.summary

REGRESSION
----------
SUMMARY OF OUTPUT: TWO STAGE LEAST SQUARES
------------------------------------------
Data set            :     nat.dbf
Dependent Variable  :        HR90                Number of Observations:        3085
Mean dependent var  :      6.1829                Number of Variables   :           5
S.D. dependent var  :      6.6414                Degrees of Freedom    :        3080
Pseudo R-squared    :      0.3570

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      15.6455516       1.3545018      11.5507796       0.0000000
                MA90      -0.0983758       0.0299492      -3.2847583       0.0010207
                PS90       1.8770506       0.1070934      17.5272273       0.0000000
                RD90       5.7292488       0.2129126      26.9089171       0.0000000
                UE90      -0.9144554       0.0986831      -9.2665854       0.0000000
------------------------------------------------------------------------------------
Instrumented: UE90
Instruments: FH90, FP89, GI89
================================ END OF REPORT =====================================
```

0.3 The Two Stages of 2SLS

create a matrix with all the instruments, i.e., both **x** and **q**

```
In [21]: bigx = np.hstack((x,q))

In [22]: bigx.shape

Out[22]: (3085, 6)
```

OLS regression of endogenous variable on all the instruments (**x** and **q**)

```
In [23]: step1 = pysal.spreg.OLS(yend,bigx)
```

predicted values for endogenous variable

```
In [24]: y2 = step1.predy
```

replace the endogenous variable by its predicted value

```
In [25]: newx = np.hstack((x,y2))
```

second step OLS regression

```
In [26]: step2 = pysal.spreg.OLS(y,newx)
```

```
In [27]: print step2.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :     unknown
Dependent Variable  :     dep_var               Number of Observations:        3085
Mean dependent var  :      6.1829               Number of Variables   :           5
S.D. dependent var  :      6.6414               Degrees of Freedom    :        3080
R-squared           :      0.4027
Adjusted R-squared  :      0.4019
Sum squared residual:   81252.812               F-statistic           :    519.1009
Sigma-square        :      26.381               Prob(F-statistic)     :           0
S.E. of regression  :       5.136               Log likelihood        :   -9422.964
Sigma-square ML     :      26.338               Akaike info criterion :   18855.928
S.E of regression ML:      5.1321               Schwarz criterion     :   18886.100

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      15.6455516       1.2962099      12.0702300       0.0000000
               var_1       5.7292488       0.2037498      28.1190386       0.0000000
               var_2      -0.0983758       0.0286603      -3.4324773       0.0006060
               var_3       1.8770506       0.1024846      18.3154447       0.0000000
               var_4      -0.9144554       0.0944362      -9.6833132       0.0000000
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           30.224

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2      56513.324           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                4       1350.019           0.0000
Koenker-Bassett test              4        120.991           0.0000
================================ END OF REPORT =====================================
```

ignore measure of fit and diagnostics, but the coefficient estimates match, the estimated standard errors do not, because they are based on the wrong residuals

0.4 2SLS with Spatial Diagnostics

set **spat_diag = True** and specify a weights object **w** (and, optionally, its name in **name_w**)

```
In [28]: reg2 = pysal.spreg.TSLS(y,x,yend,q,w=w,spat_diag=True,
                                 name_y=y_name,name_x=x_names,name_yend=yend_names,
                                 name_q=q_names,name_w="nat_queen",name_ds="nat.dbf")
```

```
In [29]: print reg2.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: TWO STAGE LEAST SQUARES
------------------------------------------
Data set            :     nat.dbf
Weights matrix      :   nat_queen
Dependent Variable  :        HR90                Number of Observations:        3085
Mean dependent var  :      6.1829                Number of Variables   :           5
S.D. dependent var  :      6.6414                Degrees of Freedom    :        3080
Pseudo R-squared    :      0.3570

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      15.6455516       1.3545018      11.5507796       0.0000000
                MA90      -0.0983758       0.0299492      -3.2847583       0.0010207
                PS90       1.8770506       0.1070934      17.5272273       0.0000000
                RD90       5.7292488       0.2129126      26.9089171       0.0000000
                UE90      -0.9144554       0.0986831      -9.2665854       0.0000000
------------------------------------------------------------------------------------
Instrumented: UE90
Instruments: FH90, FP89, GI89

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test             1         124.415          0.0000
================================ END OF REPORT =====================================
```

0.5 2SLS with White Standard Errors

set **robust = 'white'**

```
In [30]: reg3 = pysal.spreg.TSLS(y,x,yend,q,robust='white',
                                 name_y=y_name,name_x=x_names,name_yend=yend_names,
                                 name_q=q_names,name_ds="nat.dbf")
```

```
In [31]: print reg3.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: TWO STAGE LEAST SQUARES
------------------------------------------
Data set            :     nat.dbf
Dependent Variable  :        HR90                Number of Observations:        3085
Mean dependent var  :      6.1829                Number of Variables   :           5
S.D. dependent var  :      6.6414                Degrees of Freedom    :        3080
```

```
Pseudo R-squared    :      0.3570

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      15.6455516       1.5393092      10.1640082       0.0000000
                MA90      -0.0983758       0.0316213      -3.1110577       0.0018642
                PS90       1.8770506       0.1688432      11.1171261       0.0000000
                RD90       5.7292488       0.3053397      18.7635242       0.0000000
                UE90      -0.9144554       0.1384631      -6.6043272       0.0000000
------------------------------------------------------------------------------------
Instrumented: UE90
Instruments: FH90, FP89, GI89
================================ END OF REPORT =====================================
```

0.6 2SLS with HAC Standard Errors

set **robust = 'hac'** and specify a kernel weights object as **gkw** (**name_gwk** is optional)

```
In [32]: reg4 = pysal.spreg.TSLS(y,x,yend,q,robust='hac',gwk=kw,
                                 name_y=y_name,name_x=x_names,name_yend=yend_names,
                                 name_q=q_names,name_gwk="nat_k20_triang",
                                 name_ds="nat.dbf")

In [33]: print reg4.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: TWO STAGE LEAST SQUARES
------------------------------------------
Data set            :     nat.dbf
Dependent Variable  :        HR90                Number of Observations:        3085
Mean dependent var  :      6.1829                Number of Variables   :           5
S.D. dependent var  :      6.6414                Degrees of Freedom    :        3080
Pseudo R-squared    :      0.3570

HAC Standard Errors; Kernel Weights: nat_k20_triang
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      15.6455516       1.6405678       9.5366688       0.0000000
                MA90      -0.0983758       0.0341965      -2.8767776       0.0040176
                PS90       1.8770506       0.1982054       9.4702289       0.0000000
                RD90       5.7292488       0.3304847      17.3358973       0.0000000
                UE90      -0.9144554       0.1429221      -6.3982775       0.0000000
------------------------------------------------------------------------------------
Instrumented: UE90
Instruments: FH90, FP89, GI89
================================ END OF REPORT =====================================
```

0.7 Practice

Repeat the 2SLS regression with spatial diagnostics using the natregimes data set, but for a different year, say using HR60.

Check the effect of HAC estimates on the standard errors.

Create a different model weights, using k=6 nearest neighbors and a kernel weights using quadratic (Epanechnicov) kernel.

```
In [ ]:
```

Chapter 6 - 2SLS practice solutions

November 5, 2015

0.1 Solutions to Practice Example Chapter 6

```
In [1]: %pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

```
In [2]: import numpy as np
        import pysal
```

```
In [3]: db = pysal.open('data/natregimes.dbf','r')
```

```
In [4]: len(db)
```

```
Out[4]: 3085
```

```
In [5]: db.header
```

```
Out[5]: ['REGIONS',
         'NOSOUTH',
         'POLY_ID',
         'NAME',
         'STATE_NAME',
         'STATE_FIPS',
         'CNTY_FIPS',
         'FIPS',
         'STFIPS',
         'COFIPS',
         'FIPSNO',
         'SOUTH',
         'HR60',
         'HR70',
         'HR80',
         'HR90',
         'HC60',
         'HC70',
         'HC80',
         'HC90',
         'PO60',
         'PO70',
         'PO80',
         'PO90',
         'RD60',
         'RD70',
         'RD80',
         'RD90',
```

```
 'PS60',
 'PS70',
 'PS80',
 'PS90',
 'UE60',
 'UE70',
 'UE80',
 'UE90',
 'DV60',
 'DV70',
 'DV80',
 'DV90',
 'MA60',
 'MA70',
 'MA80',
 'MA90',
 'POL60',
 'POL70',
 'POL80',
 'POL90',
 'DNL60',
 'DNL70',
 'DNL80',
 'DNL90',
 'MFIL59',
 'MFIL69',
 'MFIL79',
 'MFIL89',
 'FP59',
 'FP69',
 'FP79',
 'FP89',
 'BLK60',
 'BLK70',
 'BLK80',
 'BLK90',
 'GI59',
 'GI69',
 'GI79',
 'GI89',
 'FH60',
 'FH70',
 'FH80',
 'FH90',
 'West']
```

```
In [6]: y_name = 'HR60'
        y = np.array([db.by_col(y_name)]).T
```

```
In [7]: x_names = ['RD60','MA60','PS60']
        x = np.array([db.by_col(var) for var in x_names]).T
```

```
In [8]: yend_names = ['UE60']
        yend = np.array([db.by_col(var) for var in yend_names]).T
```

```
In [9]: q_names = ['FH60','FP59','GI59']
        q = np.array([db.by_col(var) for var in q_names]).T

In [10]: w = pysal.knnW_from_shapefile('data/natregimes.shp',
                                        k=6,radius=pysal.cg.RADIUS_EARTH_MILES,idVariable='FIPSNO')

In [11]: w.n

Out[11]: 3085

In [12]: w.transform = 'r'

In [13]: kw = pysal.adaptive_kernelW_from_shapefile('data/natregimes.shp',
                                                     k=12,function='quadratic',radius=pysal.cg.RADIUS_I
                                                     diagonal=True,idVariable='FIPSNO')

In [14]: kw.n

Out[14]: 3085
```

two stage least squares with White standard errors and diagnostics for spatial dependence

```
In [15]: reg1 = pysal.spreg.TSLS(y,x,yend,q,w=w,robust='white',spat_diag=True,
                                 name_y=y_name,name_x=x_names,name_yend=yend_names,
                                 name_q=q_names,name_w='nat_k6',
                                 name_ds='natregimes.shp')

In [16]: print reg1.summary

REGRESSION
----------
SUMMARY OF OUTPUT: TWO STAGE LEAST SQUARES
------------------------------------------
Data set            :natregimes.shp
Weights matrix      :      nat_k6
Dependent Variable  :        HR60                Number of Observations:        3085
Mean dependent var  :      4.5041                Number of Variables   :           5
S.D. dependent var  :      5.6497                Degrees of Freedom    :        3080
Pseudo R-squared    :      0.2044

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      13.4510208       0.9883561      13.6094882       0.0000000
                MA60      -0.2485547       0.0212910     -11.6741458       0.0000000
                PS60       0.4393596       0.1330675       3.3017812       0.0009607
                RD60       2.3711959       0.1093345      21.6875296       0.0000000
                UE60      -0.3302298       0.1171311      -2.8193185       0.0048126
------------------------------------------------------------------------------------
Instrumented: UE60
Instruments: FH60, FP59, GI59

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test             1         277.956           0.0000
================================ END OF REPORT =====================================
```

two stage least squares with HAC standard errors and diagnostics for spatial dependence

```
In [17]: reg2 = pysal.spreg.TSLS(y,x,yend,q,w=w,robust='hac',gwk=kw,spat_diag=True,
                                 name_y=y_name,name_x=x_names,name_yend=yend_names,
                                 name_q=q_names,name_w='nat_k6',name_gwk='nat_quadratic_12',
                                 name_ds='natregimes.shp')

In [18]: print reg2.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: TWO STAGE LEAST SQUARES
------------------------------------------
Data set            :natregimes.shp
Weights matrix      :      nat_k6
Dependent Variable  :        HR60                Number of Observations:        3085
Mean dependent var  :      4.5041                Number of Variables   :           5
S.D. dependent var  :      5.6497                Degrees of Freedom    :        3080
Pseudo R-squared    :      0.2044

HAC Standard Errors; Kernel Weights: nat_quadratic_12
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      13.4510208       1.1336621      11.8651063       0.0000000
                MA60      -0.2485547       0.0244053     -10.1844645       0.0000000
                PS60       0.4393596       0.1461997       3.0052022       0.0026540
                RD60       2.3711959       0.1308052      18.1276895       0.0000000
                UE60      -0.3302298       0.1395031      -2.3671862       0.0179239
------------------------------------------------------------------------------------
Instrumented: UE60
Instruments: FH60, FP59, GI59

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test             1         277.956           0.0000
================================ END OF REPORT =====================================
```

```
In [ ]:

In [ ]:
```

Chapter 7 - Spatial 2SLS

November 5, 2015

This notebook contains the PySAL/spreg code for Chapter 7 - Spatial 2SLS
in Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL.
by Luc Anselin and Sergio J. Rey


```
In [1]: __author__ = "Luc Anselin luc.anselin@asu.edu"
```

0.1 Basic Regression Setup

0.2 Spatial Lag without Endogenous Variables

Creating arrays for y and x using the Baltimore example - see also Chapter 5 Notebook

Preliminaries, import **numpy** and **pysal**

```
In [1]: %pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

```
In [2]: import numpy as np
        import pysal
```

the **baltimore** sample data set

```
In [3]: db = pysal.open('data/baltim.dbf','r')
        y_name = "PRICE"
        y = np.array([db.by_col(y_name)]).T
        x_names = ['NROOM','NBATH','PATIO','FIREPL','AC','GAR','AGE',
                   'LOTSZ','SQFT']
        x = np.array([db.by_col(var) for var in x_names]).T
```

model weights - k nearest neighbors with k=4

```
In [4]: w = pysal.knnW_from_shapefile('data/baltim.shp',
                                       k=4,idVariable='STATION')
        w.transform = 'r'
```

kernel weights - triangular with k=12

```
In [5]: kw12 = pysal.adaptive_kernelW_from_shapefile('data/baltim.shp',
                                                  k=12,diagonal=True,idVariable='STATION')
```

0.3 Basic Spatial 2SLS

default settings

```
In [6]: reg1 = pysal.spreg.GM_Lag(y,x,w=w,name_y=y_name,name_x=x_names,
                                  name_w='baltim_k4',name_ds='baltim')
```

coefficients in the order of variables in **x_names** - last one is spatial autoregressive coefficient

```
In [7]: dir(reg1)
```

```
Out[7]: ['__doc__',
         '__init__',
         '__module__',
         '__summary',
         '_cache',
         'betas',
         'e_pred',
         'h',
         'hth',
         'hthi',
         'htz',
         'k',
         'kstar',
         'mean_y',
         'n',
         'name_ds',
         'name_gwk',
         'name_h',
         'name_q',
         'name_w',
         'name_x',
         'name_y',
         'name_yend',
         'name_z',
         'pfora1a2',
         'pr2',
         'pr2_e',
         'predy',
         'predy_e',
         'q',
         'rho',
         'robust',
         'sig2',
         'sig2n',
         'sig2n_k',
         'std_err',
         'std_y',
         'summary',
         'title',
         'u',
         'utu',
         'varb',
         'vm',
         'x',
         'y',
```

```
        'yend',
        'z',
        'z_stat',
        'zthhthi']
```

```
In [8]: reg1.betas
```

```
Out[8]: array([[ 1.17583298],
               [ 0.87624587],
               [ 5.5262249 ],
               [ 6.86878774],
               [ 7.01874585],
               [ 6.43131769],
               [ 3.71781024],
               [-0.09245938],
               [ 0.06707445],
               [ 0.07907289],
               [ 0.48461954]])
```

```
In [9]: print reg1.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :      baltim
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE               Number of Observations:         211
Mean dependent var  :     44.3072               Number of Variables   :          11
S.D. dependent var  :     23.6061               Degrees of Freedom    :         200
Pseudo R-squared    :      0.7080
Spatial Pseudo R-squared:  0.6868

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       1.1758330       5.7402917       0.2048385       0.8376983
                  AC       6.4313177       2.4191155       2.6585410       0.0078480
                 AGE      -0.0924594       0.0543360      -1.7016213       0.0888264
              FIREPL       7.0187459       2.5164274       2.7891708       0.0052843
                 GAR       3.7178102       1.7693095       2.1012775       0.0356166
               LOTSZ       0.0670745       0.0153366       4.3734849       0.0000122
               NBATH       5.5262249       1.7995640       3.0708688       0.0021344
               NROOM       0.8762459       1.0992925       0.7970998       0.4253931
               PATIO       6.8687877       2.8313870       2.4259445       0.0152686
                SQFT       0.0790729       0.1693687       0.4668683       0.6405941
             W_PRICE       0.4846195       0.0735894       6.5854513       0.0000000
------------------------------------------------------------------------------------
Instrumented: W_PRICE
Instruments: W_AC, W_AGE, W_FIREPL, W_GAR, W_LOTSZ, W_NBATH, W_NROOM,
             W_PATIO, W_SQFT
================================ END OF REPORT =====================================
```

using second order spatial lags for the instruments, set **w_lags = 2**

```
In [10]: reg2 = pysal.spreg.GM_Lag(y,x,w=w,w_lags=2,name_y=y_name,
                                    name_x=x_names,name_w='baltim_k4',
                                    name_ds='baltim')

In [11]: print reg2.summary

REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :      baltim
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE                Number of Observations:         211
Mean dependent var  :     44.3072                Number of Variables   :          11
S.D. dependent var  :     23.6061                Degrees of Freedom    :         200
Pseudo R-squared    :      0.7076
Spatial Pseudo R-squared:  0.6856

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       0.6967291       5.6918227       0.1224088       0.9025753
                  AC       6.4004638       2.4204278       2.6443523       0.0081847
                 AGE      -0.0898356       0.0542103      -1.6571679       0.0974855
              FIREPL       6.9286683       2.5140883       2.7559367       0.0058524
                 GAR       3.6812877       1.7696563       2.0802275       0.0375047
               LOTSZ       0.0664709       0.0153170       4.3396943       0.0000143
               NBATH       5.5235755       1.8009199       3.0670855       0.0021616
               NROOM       0.8904222       1.0998836       0.8095604       0.4181929
               PATIO       6.7936052       2.8309029       2.3998016       0.0164040
                SQFT       0.0767161       0.1694537       0.4527261       0.6507460
             W_PRICE       0.4951283       0.0716451       6.9108461       0.0000000
------------------------------------------------------------------------------------
Instrumented: W_PRICE
Instruments: W2_AC, W2_AGE, W2_FIREPL, W2_GAR, W2_LOTSZ, W2_NBATH, W2_NROOM,
             W2_PATIO, W2_SQFT, W_AC, W_AGE, W_FIREPL, W_GAR, W_LOTSZ,
             W_NBATH, W_NROOM, W_PATIO, W_SQFT
================================ END OF REPORT =====================================
```

up to third order spatial lags, set **w_lags=3**

```
In [12]: reg2a = pysal.spreg.GM_Lag(y,x,w=w,w_lags=3,name_y=y_name,
                                    name_x=x_names,name_w='baltim_k4',
                                    name_ds='baltim')

In [13]: print reg2a.summary

REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :      baltim
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE                Number of Observations:         211
Mean dependent var  :     44.3072                Number of Variables   :          11
```

```
S.D. dependent var  :     23.6061                Degrees of Freedom    :         200
Pseudo R-squared    :      0.7078
Spatial Pseudo R-squared:  0.6862

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       0.9254523       5.6633202       0.1634116       0.8701944
                  AC       6.4151934       2.4192610       2.6517161       0.0080084
                 AGE      -0.0910882       0.0541071      -1.6834807       0.0922821
              FIREPL       6.9716711       2.5110371       2.7764111       0.0054963
                 GAR       3.6987235       1.7684991       2.0914478       0.0364879
               LOTSZ       0.0667590       0.0152957       4.3645680       0.0000127
               NBATH       5.5248403       1.8002401       3.0689464       0.0021482
               NROOM       0.8836545       1.0993508       0.8037967       0.4215144
               PATIO       6.8294972       2.8285357       2.4144992       0.0157569
                SQFT       0.0778412       0.1693686       0.4595966       0.6458058
             W_PRICE       0.4901115       0.0706057       6.9415235       0.0000000
------------------------------------------------------------------------------------
Instrumented: W_PRICE
Instruments: W2_AC, W2_AGE, W2_FIREPL, W2_GAR, W2_LOTSZ, W2_NBATH, W2_NROOM,
             W2_PATIO, W2_SQFT, W3_AC, W3_AGE, W3_FIREPL, W3_GAR, W3_LOTSZ,
             W3_NBATH, W3_NROOM, W3_PATIO, W3_SQFT, W_AC, W_AGE, W_FIREPL,
             W_GAR, W_LOTSZ, W_NBATH, W_NROOM, W_PATIO, W_SQFT
================================ END OF REPORT =====================================
```

0.3.1 Direct, Indirect and Total Effects

extract the regression coefficients

```
In [14]: b = reg1.betas[:-1]

In [15]: b

Out[15]: array([[ 1.17583298],
                [ 0.87624587],
                [ 5.5262249 ],
                [ 6.86878774],
                [ 7.01874585],
                [ 6.43131769],
                [ 3.71781024],
                [-0.09245938],
                [ 0.06707445],
                [ 0.07907289]])
```

extract the spatial autoregressive coefficient

```
In [16]: rho = reg1.betas[-1]

In [17]: rho

Out[17]: array([ 0.48461954])
```

total effect using the multiplier

```
In [18]: btot = b / (1.0 - rho)
```

```
In [19]: btot

Out[19]: array([[  2.28148539],
                [  1.70019227],
                [ 10.72261246],
                [ 13.32760616],
                [ 13.61857201],
                [ 12.47877682],
                [  7.21371988],
                [ -0.17940024],
                [  0.13014551],
                [  0.15342625]])
```

indirect effect

```
In [20]: bind = btot - b
```

summary of the results

```
In [21]: varnames = ["CONSTANT"] + x_names
         print "Variable       Direct       Indirect      Total"
         for i in range(len(varnames)):
             print "%10s %12.7f %12.7f %12.7f" % (varnames[i],b[i][0],bind[i][0],btot[i][0])

Variable       Direct       Indirect      Total
  CONSTANT    1.1758330    1.1056524    2.2814854
     NROOM    0.8762459    0.8239464    1.7001923
     NBATH    5.5262249    5.1963876   10.7226125
     PATIO    6.8687877    6.4588184   13.3276062
    FIREPL    7.0187459    6.5998262   13.6185720
        AC    6.4313177    6.0474591   12.4787768
       GAR    3.7178102    3.4959096    7.2137199
       AGE   -0.0924594   -0.0869409   -0.1794002
     LOTSZ    0.0670745    0.0630711    0.1301455
      SQFT    0.0790729    0.0743534    0.1534263
```

0.4 Spatial 2SLS with Spatial Diagnostics

specify the weights as **w=w**, set **spat_diag = True** and optionally specify a name for the weights

```
In [22]: reg3 = pysal.spreg.GM_Lag(y,x,w=w,spat_diag=True,
                                   name_y=y_name,name_x=x_names,
                                   name_w='baltim_k4',name_ds='baltim')

In [23]: print reg3.summary

REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :      baltim
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE                Number of Observations:         211
Mean dependent var  :     44.3072                Number of Variables   :          11
S.D. dependent var  :     23.6061                Degrees of Freedom    :         200
Pseudo R-squared    :      0.7080
```

```
Spatial Pseudo R-squared:  0.6868

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       1.1758330       5.7402917       0.2048385       0.8376983
                  AC       6.4313177       2.4191155       2.6585410       0.0078480
                 AGE      -0.0924594       0.0543360      -1.7016213       0.0888264
              FIREPL       7.0187459       2.5164274       2.7891708       0.0052843
                 GAR       3.7178102       1.7693095       2.1012775       0.0356166
               LOTSZ       0.0670745       0.0153366       4.3734849       0.0000122
               NBATH       5.5262249       1.7995640       3.0708688       0.0021344
               NROOM       0.8762459       1.0992925       0.7970998       0.4253931
               PATIO       6.8687877       2.8313870       2.4259445       0.0152686
                SQFT       0.0790729       0.1693687       0.4668683       0.6405941
             W_PRICE       0.4846195       0.0735894       6.5854513       0.0000000
------------------------------------------------------------------------------------
Instrumented: W_PRICE
Instruments: W_AC, W_AGE, W_FIREPL, W_GAR, W_LOTSZ, W_NBATH, W_NROOM,
             W_PATIO, W_SQFT

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test             1          3.654          0.0559
================================ END OF REPORT =====================================
```

0.5 Spatial 2SLS with White Standard Errors

set **robust = 'white'**

```
In [24]: reg4 = pysal.spreg.GM_Lag(y,x,w=w,robust='white',
                                   spat_diag=True,name_y=y_name,name_x=x_names,
                                   name_w='baltim_k4',name_ds='baltim')

In [25]: print reg4.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :      baltim
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE              Number of Observations:         211
Mean dependent var  :     44.3072              Number of Variables   :          11
S.D. dependent var  :     23.6061              Degrees of Freedom    :         200
Pseudo R-squared    :      0.7080
Spatial Pseudo R-squared:  0.6868

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       1.1758330       7.0047804       0.1678615       0.8666922
                  AC       6.4313177       2.6910698       2.3898739       0.0168542
                 AGE      -0.0924594       0.0983874      -0.9397478       0.3473469
```

```
        FIREPL       7.0187459       2.4195615       2.9008338       0.0037217
           GAR       3.7178102       2.3876143       1.5571234       0.1194412
         LOTSZ       0.0670745       0.0251308       2.6690191       0.0076073
         NBATH       5.5262249       2.1872327       2.5265830       0.0115178
         NROOM       0.8762459       1.3976000       0.6269647       0.5306824
         PATIO       6.8687877       3.1491404       2.1811627       0.0291714
          SQFT       0.0790729       0.2234025       0.3539481       0.7233778
       W_PRICE       0.4846195       0.1259270       3.8484168       0.0001189
------------------------------------------------------------------------------------
Instrumented: W_PRICE
Instruments: W_AC, W_AGE, W_FIREPL, W_GAR, W_LOTSZ, W_NBATH, W_NROOM,
             W_PATIO, W_SQFT

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test             1           3.654           0.0559
================================ END OF REPORT =====================================
```

0.6 Spatial 2SLS with HAC Standard Errors

set **robust = 'hac'** and specify the kernel weights **gwk** and optionally their name **name_gwk**

```
In [26]: reg5 = pysal.spreg.GM_Lag(y,x,w=w,robust='hac',gwk=kw12,
                                   spat_diag=True,name_y=y_name,name_x=x_names,
                                   name_w='baltim_k4',name_gwk='baltim_tri_k12',
                                   name_ds='baltim')
```

```
In [27]: print reg5.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :      baltim
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE                Number of Observations:         211
Mean dependent var  :     44.3072                Number of Variables   :          11
S.D. dependent var  :     23.6061                Degrees of Freedom    :         200
Pseudo R-squared    :      0.7080
Spatial Pseudo R-squared:  0.6868

HAC Standard Errors; Kernel Weights: baltim_tri_k12
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       1.1758330       7.3538150       0.1598943       0.8729644
                  AC       6.4313177       2.9201174       2.2024175       0.0276358
                 AGE      -0.0924594       0.0962413      -0.9607039       0.3367011
              FIREPL       7.0187459       2.4120995       2.9098078       0.0036165
                 GAR       3.7178102       2.4421020       1.5223812       0.1279136
               LOTSZ       0.0670745       0.0229299       2.9251915       0.0034424
               NBATH       5.5262249       2.1861011       2.5278908       0.0114750
               NROOM       0.8762459       1.4166133       0.6185498       0.5362130
               PATIO       6.8687877       3.1066547       2.2109917       0.0270364
                SQFT       0.0790729       0.1980344       0.3992887       0.6896805
```

```
            W_PRICE       0.4846195       0.1132898       4.2776985       0.0000189
------------------------------------------------------------------------------------
Instrumented: W_PRICE
Instruments: W_AC, W_AGE, W_FIREPL, W_GAR, W_LOTSZ, W_NBATH, W_NROOM,
             W_PATIO, W_SQFT

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                              MI/DF       VALUE           PROB
Anselin-Kelejian Test               1           3.654          0.0559
================================ END OF REPORT =====================================
```

0.7 Spatial Lag Model with other Endogenous Variables

create the variable arrays using the **natregimes** sample data set

```
In [28]: db = pysal.open('data/natregimes.dbf','r')
         y_name = "HR90"
         y = np.array([db.by_col(y_name)]).T
         x_names = ['RD90','MA90','PS90']
         x = np.array([db.by_col(var) for var in x_names]).T
         yend_names = ['UE90']
         yend = np.array([db.by_col(var) for var in yend_names]).T
         q_names = ['FH90','FP89','GI89']
         q = np.array([db.by_col(var) for var in q_names]).T
```

model weights

```
In [29]: w = pysal.queen_from_shapefile('data/natregimes.shp',idVariable="FIPSNO")
         w.transform = 'r'
```

0.7.1 Spatial Lag with Endogenous Variables

base case with spatial diagnostics

```
In [30]: reg6 = pysal.spreg.GM_Lag(y,x,yend,q,w=w,spat_diag=True,
                                  name_y=y_name,name_x=x_names,name_yend=yend_names,
                                  name_q=q_names,name_w='natqueen',name_ds='nat')

In [31]: print reg6.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :         nat
Weights matrix      :    natqueen
Dependent Variable  :        HR90                Number of Observations:        3085
Mean dependent var  :      6.1829                Number of Variables   :           6
S.D. dependent var  :      6.6414                Degrees of Freedom    :        3079
Pseudo R-squared    :      0.4186
Spatial Pseudo R-squared:  0.3914

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      10.0338240       1.3616383       7.3689349       0.0000000
```

```
              MA90      -0.0500990       0.0286025      -1.7515613       0.0798493
              PS90       1.5813070       0.1084249      14.5843567       0.0000000
              RD90       4.4092974       0.2400482      18.3683863       0.0000000
              UE90      -0.5182722       0.0882736      -5.8712062       0.0000000
             W_HR90      0.2123364       0.0371805       5.7109639       0.0000000
------------------------------------------------------------------------------------
Instrumented: UE90, W_HR90
Instruments: FH90, FP89, GI89, W_FH90, W_FP89, W_GI89, W_MA90, W_PS90,
             W_RD90

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test            1           2.517           0.1127
================================ END OF REPORT =====================================
```

without spatial lags for the instruments, set **lag_q = False**

```
In [32]: reg7 = pysal.spreg.GM_Lag(y,x,yend,q,w=w,lag_q=False,spat_diag=True,
                                   name_y=y_name,name_x=x_names,name_yend=yend_names,
                                   name_q=q_names,name_w='nat_queen',name_ds='nat')

In [33]: print reg7.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :         nat
Weights matrix      :   nat_queen
Dependent Variable  :        HR90               Number of Observations:        3085
Mean dependent var  :      6.1829               Number of Variables   :           6
S.D. dependent var  :      6.6414               Degrees of Freedom    :        3079
Pseudo R-squared    :      0.4076
Spatial Pseudo R-squared:  0.3802

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      11.2850228       1.4177538       7.9597903       0.0000000
              MA90      -0.0601927       0.0290474      -2.0722259       0.0382444
              PS90       1.6149324       0.1105060      14.6139849       0.0000000
              RD90       4.6642007       0.2537771      18.3791221       0.0000000
              UE90      -0.6580528       0.0951942      -6.9127375       0.0000000
             W_HR90      0.2163835       0.0389967       5.5487653       0.0000000
------------------------------------------------------------------------------------
Instrumented: UE90, W_HR90
Instruments: FH90, FP89, GI89, W_MA90, W_PS90, W_RD90

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test            1           2.442           0.1182
================================ END OF REPORT =====================================
```

0.8 Practice

Replicate the analysis above using a subset of the U.S. counties, i.e., the south data set. Use both k=6 nearest neighbors and queen contiguity as weights and compare the results. Use adaptive bandwidth quadratic kernel weights (k=12) to assess the effect of HAC standard errors.

```
In [ ]:
```

Chapter 7 - S2SLS practice solutions

November 5, 2015

0.1 Solutions to Practice Example Chapter 7

In [1]:
```
%pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

In [2]:
```
import numpy as np
import pysal
```

In [3]:
```
db = pysal.open('data/south.dbf','r')
```

In [4]:
```
len(db)
```

Out[4]:
```
1412
```

In [5]:
```
db.header
```

Out[5]:
```
['FIPSNO',
 'NAME',
 'STATE_NAME',
 'STATE_FIPS',
 'CNTY_FIPS',
 'FIPS',
 'STFIPS',
 'COFIPS',
 'SOUTH',
 'HR60',
 'HR70',
 'HR80',
 'HR90',
 'HC60',
 'HC70',
 'HC80',
 'HC90',
 'PO60',
 'PO70',
 'PO80',
 'PO90',
 'RD60',
 'RD70',
 'RD80',
 'RD90',
 'PS60',
 'PS70',
 'PS80',
```

```
         'PS90',
         'UE60',
         'UE70',
         'UE80',
         'UE90',
         'DV60',
         'DV70',
         'DV80',
         'DV90',
         'MA60',
         'MA70',
         'MA80',
         'MA90',
         'POL60',
         'POL70',
         'POL80',
         'POL90',
         'DNL60',
         'DNL70',
         'DNL80',
         'DNL90',
         'MFIL59',
         'MFIL69',
         'MFIL79',
         'MFIL89',
         'FP59',
         'FP69',
         'FP79',
         'FP89',
         'BLK60',
         'BLK70',
         'BLK80',
         'BLK90',
         'GI59',
         'GI69',
         'GI79',
         'GI89',
         'FH60',
         'FH70',
         'FH80',
         'FH90']
```

```
In [6]: y_name = 'HR90'
        y = np.array([db.by_col(y_name)]).T
```

```
In [7]: x_names = ['RD90','MA90','PS90','UE90']
        x = np.array([db.by_col(var) for var in x_names]).T
```

```
In [8]: xe_names = ['RD90','MA90','PS90']
        xe = np.array([db.by_col(var) for var in xe_names]).T
```

```
In [9]: yend_names = ['UE90']
        yend = np.array([db.by_col(var) for var in yend_names]).T
```

```
In [10]: q_names = ['FH90','FP89','GI89']
         q = np.array([db.by_col(var) for var in q_names]).T
```

```
In [11]: w = pysal.knnW_from_shapefile('data/south.shp',
                                        k=6,radius=pysal.cg.RADIUS_EARTH_MILES,idVariable='FIPSNO')
         w.transform = 'r'
```

```
In [12]: ww = pysal.queen_from_shapefile('data/south.shp',idVariable="FIPSNO")
         ww.transform = 'r'
```

```
In [13]: w.n, ww.n
```

```
Out[13]: (1412, 1412)
```

```
In [14]: kw = pysal.adaptive_kernelW_from_shapefile('data/south.shp',
                                                     k=12,function='quadratic',radius=pysal.cg.RADIUS_
                                                     diagonal=True,idVariable='FIPSNO')
```

```
In [15]: kw.n
```

```
Out[15]: 1412
```

spatial two stage least squares: exogenous variables only, using k=6 nearest neighbors

```
In [16]: reg1 = pysal.spreg.GM_Lag(y,x,w=w,spat_diag=True,
                                   name_y=y_name,name_x=x_names,
                                   name_w='south_k6',name_ds='south')
```

```
In [17]: print reg1.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :       south
Weights matrix      :    south_k6
Dependent Variable  :        HR90                Number of Observations:        1412
Mean dependent var  :      9.5493                Number of Variables   :           6
S.D. dependent var  :      7.0389                Degrees of Freedom    :        1406
Pseudo R-squared    :      0.3170
Spatial Pseudo R-squared:  0.3017

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       8.7095236       2.0183412       4.3151890       0.0000159
                MA90       0.0144354       0.0480777       0.3002507       0.7639859
                PS90       2.0300519       0.2069009       9.8117134       0.0000000
                RD90       4.2657445       0.2694077      15.8337904       0.0000000
                UE90      -0.4609325       0.0741770      -6.2139568       0.0000000
              W_HR90       0.1181498       0.0661620       1.7857661       0.0741371
------------------------------------------------------------------------------------
Instrumented: W_HR90
Instruments: W_MA90, W_PS90, W_RD90, W_UE90

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test            1           2.949           0.0859
================================ END OF REPORT =====================================
```

spatial two stage least squares using k=6 nearest neighbors, with HAC standard errors

```
In [18]: reg1a = pysal.spreg.GM_Lag(y,x,w=w,robust='hac',gwk=kw,spat_diag=True,
                                    name_y=y_name,name_x=x_names,
                                    name_w='south_k6',name_gwk="south_quad_12",name_ds='south')
```

```
In [19]: print reg1a.summary

REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :       south
Weights matrix      :    south_k6
Dependent Variable  :        HR90                Number of Observations:        1412
Mean dependent var  :      9.5493                Number of Variables   :           6
S.D. dependent var  :      7.0389                Degrees of Freedom    :        1406
Pseudo R-squared    :      0.3170
Spatial Pseudo R-squared:  0.3017

HAC Standard Errors; Kernel Weights: south_quad_12
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       8.7095236       2.2063492       3.9474820       0.0000790
                MA90       0.0144354       0.0493446       0.2925419       0.7698723
                PS90       2.0300519       0.3980117       5.1004829       0.0000003
                RD90       4.2657445       0.4770866       8.9412379       0.0000000
                UE90      -0.4609325       0.1019096      -4.5229549       0.0000061
              W_HR90       0.1181498       0.0949992       1.2436927       0.2136127
------------------------------------------------------------------------------------
Instrumented: W_HR90
Instruments: W_MA90, W_PS90, W_RD90, W_UE90

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test             1          2.949           0.0859
================================ END OF REPORT =====================================
```

spatial two stage least squares using queen contiguity

```
In [20]: reg1b = pysal.spreg.GM_Lag(y,x,w=ww,spat_diag=True,
                                    name_y=y_name,name_x=x_names,
                                    name_w='south_queen',name_ds='south')
```

```
In [21]: print reg1b.summary

REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :       south
Weights matrix      : south_queen
Dependent Variable  :        HR90                Number of Observations:        1412
Mean dependent var  :      9.5493                Number of Variables   :           6
S.D. dependent var  :      7.0389                Degrees of Freedom    :        1406
```

```
Pseudo R-squared    :      0.3222
Spatial Pseudo R-squared:  0.3020

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       8.1830219       2.0412020       4.0089231       0.0000610
                MA90       0.0227990       0.0483737       0.4713098       0.6374195
                PS90       2.0297011       0.2037269       9.9628509       0.0000000
                RD90       4.2106113       0.2661952      15.8177558       0.0000000
                UE90      -0.4537805       0.0734397      -6.1789507       0.0000000
              W_HR90       0.1418881       0.0652331       2.1750937       0.0296231
------------------------------------------------------------------------------------
Instrumented: W_HR90
Instruments: W_MA90, W_PS90, W_RD90, W_UE90

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test            1           2.306           0.1289
================================ END OF REPORT =====================================
```

spatial two stage least squares with exogenous and endogenous variables, k=6 nearest neighbors

```
In [22]: reg2 = pysal.spreg.GM_Lag(y,xe,yend,q,w=w,spat_diag=True,
                                   name_y=y_name,name_x=xe_names,name_yend=yend_names,
                                   name_q=q_names,name_w='south_k6',name_ds='south')

In [23]: print reg2.summary

REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :       south
Weights matrix      :    south_k6
Dependent Variable  :        HR90                Number of Observations:        1412
Mean dependent var  :      9.5493                Number of Variables   :           6
S.D. dependent var  :      7.0389                Degrees of Freedom    :        1406
Pseudo R-squared    :      0.2800
Spatial Pseudo R-squared:  0.2774

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      14.4390815       2.2781115       6.3381803       0.0000000
                MA90      -0.0170951       0.0499350      -0.3423476       0.7320893
                PS90       2.2209207       0.2158787      10.2878183       0.0000000
                RD90       5.6815334       0.3619922      15.6951837       0.0000000
                UE90      -1.0886919       0.1299393      -8.3784666       0.0000000
              W_HR90       0.0226703       0.0693006       0.3271294       0.7435700
------------------------------------------------------------------------------------
Instrumented: UE90, W_HR90
Instruments: FH90, FP89, GI89, W_FH90, W_FP89, W_GI89, W_MA90, W_PS90,
             W_RD90
```

```
DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test             1           7.492           0.0062
================================ END OF REPORT =====================================
```

In [ ]:

In [ ]:

Chapter 8 - ML Spatial Lag

November 5, 2015

This notebook contains the PySAL/spreg code for Chapter 8 - ML Spatial Lag
in Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL.
by Luc Anselin and Sergio J. Rey


```
In [1]: __author__ = "Luc Anselin luc.anselin@asu.edu"
```

0.1 Basic Regression Setup

0.1.1 Model Specification

Creating arrays for y and x using the Baltimore example - see also Chapter 5 Notebook
Preliminaries, import **numpy** and **pysal**

```
In [1]: %pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

```
In [2]: import numpy as np
        import pysal
```

the **baltimore** sample data set

```
In [3]: db = pysal.open(pysal.examples.get_path('baltim.dbf'),'r')
        y_name = "PRICE"
        y = np.array([db.by_col(y_name)]).T
        x_names = ['NROOM','NBATH','PATIO','FIREPL','AC','GAR','AGE',
                   'LOTSZ','SQFT']
        x = np.array([db.by_col(var) for var in x_names]).T
```

model weights - k nearest neighbors with k=4

```
In [4]: w = pysal.knnW_from_shapefile(pysal.examples.get_path('baltim.shp'),
                                        k=4,idVariable='STATION')
        w.transform = 'r'
```

0.2 ML Estimation of Spatial Lag Model

0.2.1 Full Method

```
In [5]: reg1 = pysal.spreg.ML_Lag(y,x,w,name_y=y_name,name_x=x_names,
                                  name_w = "baltim_k4",name_ds = "baltim")
```

```
/Users/luc/anaconda/lib/python2.7/site-packages/scipy/optimize/_minimize.py:573: RuntimeWarning: Method
  "defaulting to absolute tolerance.", RuntimeWarning)
```

contents of the ML_Lag object

```
In [6]: dir(reg1)

Out[6]: ['__doc__',
         '__init__',
         '__module__',
         '__summary',
         '_cache',
         'aic',
         'betas',
         'e_pred',
         'epsilon',
         'k',
         'logll',
         'mean_y',
         'method',
         'n',
         'name_ds',
         'name_w',
         'name_x',
         'name_y',
         'pr2',
         'pr2_e',
         'predy',
         'predy_e',
         'rho',
         'schwarz',
         'sig2',
         'sig2n',
         'sig2n_k',
         'std_err',
         'std_y',
         'summary',
         'title',
         'u',
         'utu',
         'vm',
         'vm1',
         'x',
         'y',
         'z_stat']
```

the regression coefficients, with rho as last

```
In [7]: reg1.betas

Out[7]: array([[ 7.34468119],
               [ 0.69371354],
               [ 5.56033847],
               [ 7.8368234 ],
               [ 8.17856724],
               [ 6.82858647],
               [ 4.18806717],
               [-0.12624221],
```

```
              [ 0.07484533],
              [ 0.10941805],
              [ 0.34931031]])
```

```
In [8]: reg1.betas[-1][0]
```

```
Out[8]: 0.34931030602875046
```

```
In [9]: reg1.rho
```

```
Out[9]: 0.34931030602875046
```

print the full set of results

```
In [10]: print reg1.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: MAXIMUM LIKELIHOOD SPATIAL LAG (METHOD = FULL)
-----------------------------------------------------------------
Data set            :      baltim
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE               Number of Observations:         211
Mean dependent var  :     44.3072               Number of Variables   :          11
S.D. dependent var  :     23.6061               Degrees of Freedom    :         200
Pseudo R-squared    :      0.7073
Spatial Pseudo R-squared:  0.6927
Sigma-square ML     :     162.407               Log likelihood        :    -839.288
S.E of regression   :      12.744               Akaike info criterion :    1700.577
                                                Schwarz criterion     :    1737.447

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       7.3446812       5.0631000       1.4506293       0.1468831
                  AC       6.8285865       2.4153196       2.8271979       0.0046957
                 AGE      -0.1262422       0.0526168      -2.3992737       0.0164276
              FIREPL       8.1785672       2.4835614       3.2930804       0.0009910
                 GAR       4.1880672       1.7562803       2.3846235       0.0170966
               LOTSZ       0.0748453       0.0150251       4.9813401       0.0000006
               NBATH       5.5603385       1.8049519       3.0806020       0.0020658
               NROOM       0.6937135       1.0965606       0.6326267       0.5269774
               PATIO       7.8368234       2.8017708       2.7970966       0.0051564
                SQFT       0.1094180       0.1689413       0.6476690       0.5171991
             W_PRICE       0.3493103       0.0553858       6.3068556       0.0000000
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

0.2.2 Ord Method

```
In [11]: reg2 = pysal.spreg.ML_Lag(y,x,w,method='ord',name_y=y_name,name_x=x_names,
                                   name_w = "baltim_k4",name_ds = "baltim")
```

```
In [12]: print reg2.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: MAXIMUM LIKELIHOOD SPATIAL LAG (METHOD = ORD)
----------------------------------------------------------------
Data set            :      baltim
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE                Number of Observations:         211
Mean dependent var  :     44.3072                Number of Variables   :          11
S.D. dependent var  :     23.6061                Degrees of Freedom    :         200
Pseudo R-squared    :      0.7073
Spatial Pseudo R-squared:  0.6927
Sigma-square ML     :     162.407                Log likelihood        :    -839.288
S.E of regression   :      12.744                Akaike info criterion :    1700.577
                                                 Schwarz criterion     :    1737.447

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       7.3446812       5.0631000       1.4506293       0.1468831
                  AC       6.8285865       2.4153196       2.8271979       0.0046957
                 AGE      -0.1262422       0.0526168      -2.3992737       0.0164276
              FIREPL       8.1785672       2.4835614       3.2930804       0.0009910
                 GAR       4.1880672       1.7562803       2.3846235       0.0170966
               LOTSZ       0.0748453       0.0150251       4.9813401       0.0000006
               NBATH       5.5603385       1.8049519       3.0806020       0.0020658
               NROOM       0.6937135       1.0965606       0.6326267       0.5269774
               PATIO       7.8368234       2.8017708       2.7970966       0.0051564
                SQFT       0.1094180       0.1689413       0.6476690       0.5171991
             W_PRICE       0.3493103       0.0553858       6.3068556       0.0000000
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

In [ ]:

In [ ]:

Chapter 9 - GM_GMM Error

November 5, 2015

This notebook contains the PySAL/spreg code for Chapter 9 - GM/GMM Error in Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL. by Luc Anselin and Sergio J. Rey


```
In [1]: __author__ = "Luc Anselin luc.anselin@asu.edu"
```

0.1 Basic Regression Setup

0.1.1 Exogenous Explanatory Variables Only

Creating arrays for y and x for south.dbf example data set (see previous notebooks)
Preliminaries, import **numpy** and **pysal**

```
In [1]: %pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

```
In [2]: import numpy as np
        import pysal
```

```
In [3]: db = pysal.open('data/south.dbf','r')
        y_name = "HR90"
        y = np.array([db.by_col(y_name)]).T
        x_names = ["RD90","PS90","UE90","DV90"]
        x = np.array([db.by_col(var) for var in x_names]).T
```

0.1.2 Exogenous and Endogenous Explanatory Variables

Creating arrays for yend, q and xe (exogenous only)

```
In [4]: yend_names = ["UE90"]
        yend = np.array([db.by_col(var) for var in yend_names]).T
        q_names = ["FH90","FP89","GI89"]
        q = np.array([db.by_col(var) for var in q_names]).T
        xe_names = ["RD90","PS90","DV90"]
        xe = np.array([db.by_col(var) for var in xe_names]).T
```

0.1.3 Spatial Weights

Queen contiguity, with FIPSNO as the ID variable

```
In [5]: w = pysal.queen_from_shapefile('data/south.shp',idVariable="FIPSNO")
        w.transform = 'r'
```

0.2 GM

0.2.1 Exogenous Variables Only

```
In [6]: gm1 = pysal.spreg.GM_Error(y,x,w,name_y=y_name,name_x=x_names,
                                   name_w="south_q",name_ds="south.dbf")
```

Attributes of the regression object

```
In [7]: dir(gm1)
```

```
Out[7]: ['__doc__',
         '__init__',
         '__module__',
         '__summary',
         '_cache',
         'betas',
         'e_filtered',
         'k',
         'mean_y',
         'n',
         'name_ds',
         'name_w',
         'name_x',
         'name_y',
         'pr2',
         'predy',
         'sig2',
         'std_err',
         'std_y',
         'summary',
         'title',
         'u',
         'vm',
         'x',
         'y',
         'z_stat']
```

The estimated coefficients, including lambda as the last element

```
In [8]: gm1.betas
```

```
Out[8]: array([[ 6.33865368],
               [ 4.43265183],
               [ 1.81335314],
               [-0.3985616 ],
               [ 0.47772164],
               [ 0.26040896]])
```

The spatial autoregressive coefficient

```
In [9]: gm1.betas[-1][0]
```

```
Out[9]: 0.2604089565665465
```

The full listing

```
In [10]: print gm1.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED LEAST SQUARES
---------------------------------------------------
Data set            :   south.dbf
Weights matrix      :     south_q
Dependent Variable  :        HR90               Number of Observations:        1412
Mean dependent var  :      9.5493               Number of Variables   :           5
S.D. dependent var  :      7.0389               Degrees of Freedom    :        1407
Pseudo R-squared    :      0.3066

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       6.3386537       1.0155422       6.2416445       0.0000000
                DV90       0.4777216       0.1203677       3.9688512       0.0000722
                PS90       1.8133531       0.2105237       8.6135328       0.0000000
                RD90       4.4326518       0.2318185      19.1212180       0.0000000
                UE90      -0.3985616       0.0772012      -5.1626346       0.0000002
              lambda       0.2604090
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

0.2.2 Exogenous and Endogenous Variables

```
In [11]: gm2 = pysal.spreg.GM_Endog_Error(y,xe,yend,q,w,name_y=y_name,
                           name_x=xe_names,name_yend=yend_names,name_q=q_names,
                           name_w="south_q",name_ds="south.dbf")
```

Attributes of the regression object

```
In [12]: dir(gm2)
```

```
Out[12]: ['__doc__',
          '__init__',
          '__module__',
          '__summary',
          '_cache',
          'betas',
          'e_filtered',
          'k',
          'mean_y',
          'n',
          'name_ds',
          'name_h',
          'name_q',
          'name_w',
          'name_x',
          'name_y',
          'name_yend',
          'name_z',
          'pr2',
          'predy',
          'sig2',
          'std_err',
```

```
          'std_y',
          'summary',
          'title',
          'u',
          'vm',
          'x',
          'y',
          'yend',
          'z',
          'z_stat']
```

The estimated coefficients, including lambda as the last element

```
In [13]: gm2.betas

Out[13]: array([[ 10.7717841 ],
               [  5.90371303],
               [  2.04553883],
               [  0.49190638],
               [ -1.14071221],
               [  0.23609742]])
```

The spatial autoregressive coefficient

```
In [14]: gm2.betas[-1][0]

Out[14]: 0.23609741823856531
```

The full listing

```
In [15]: print gm2.summary

REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES
-------------------------------------------------------------
Data set            :   south.dbf
Weights matrix      :     south_q
Dependent Variable  :        HR90             Number of Observations:        1412
Mean dependent var  :      9.5493             Number of Variables   :           5
S.D. dependent var  :      7.0389             Degrees of Freedom    :        1407
Pseudo R-squared    :      0.2818

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      10.7717841       1.2771988       8.4339137       0.0000000
                DV90       0.4919064       0.1246483       3.9463541       0.0000794
                PS90       2.0455388       0.2190619       9.3377222       0.0000000
                RD90       5.9037130       0.3473996      16.9940125       0.0000000
                UE90      -1.1407122       0.1483842      -7.6875610       0.0000000
              lambda       0.2360974
------------------------------------------------------------------------------------
Instrumented: UE90
Instruments: FH90, FP89, GI89
================================ END OF REPORT =====================================
```

0.3 GMM Heteroskedastic Case

0.3.1 Exogenous Variables Only

```
In [16]: gm3 = pysal.spreg.GM_Error_Het(y,x,w,name_y=y_name,name_x=x_names,
                             name_w="south_q",name_ds="south.dbf")
```

Attributes of the regression object

```
In [17]: dir(gm3)
```

```
Out[17]: ['__doc__',
          '__init__',
          '__module__',
          '__summary',
          '_cache',
          'betas',
          'e_filtered',
          'iter_stop',
          'iteration',
          'k',
          'mean_y',
          'n',
          'name_ds',
          'name_w',
          'name_x',
          'name_y',
          'pr2',
          'predy',
          'std_err',
          'std_y',
          'step1c',
          'summary',
          'title',
          'u',
          'vm',
          'x',
          'xtx',
          'y',
          'z_stat']
```

The estimated coefficients, including lambda as the last element

```
In [18]: gm3.betas
```

```
Out[18]: array([[ 6.25760366],
                [ 4.41953589],
                [ 1.79832764],
                [-0.38976971],
                [ 0.48116579],
                [ 0.31474155]])
```

The spatial autoregressive coefficient

```
In [19]: gm3.betas[-1][0]
```

```
Out[19]: 0.31474155432811185
```

The full listing

```
In [20]: print gm3.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED LEAST SQUARES (HET)
---------------------------------------------------------
Data set            :   south.dbf
Weights matrix      :     south_q
Dependent Variable  :        HR90                Number of Observations:        1412
Mean dependent var  :      9.5493                Number of Variables   :           5
S.D. dependent var  :      7.0389                Degrees of Freedom    :        1407
Pseudo R-squared    :      0.3062
N. of iterations    :           1                Step1c computed       :          No

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       6.2576037       1.0821873       5.7823668       0.0000000
                DV90       0.4811658       0.1198516       4.0146802       0.0000595
                PS90       1.7983276       0.3359957       5.3522335       0.0000001
                RD90       4.4195359       0.3468537      12.7417874       0.0000000
                UE90      -0.3897697       0.0985644      -3.9544664       0.0000767
              lambda       0.3147416       0.0374883       8.3957227       0.0000000
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

Setting the step1c option

```
In [21]: gm4 = pysal.spreg.GM_Error_Het(y,x,w,step1c=True,name_y=y_name,
                        name_x=x_names,name_w="south_q",name_ds="south.dbf")
```

The full listing

```
In [22]: print gm4.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED LEAST SQUARES (HET)
---------------------------------------------------------
Data set            :   south.dbf
Weights matrix      :     south_q
Dependent Variable  :        HR90                Number of Observations:        1412
Mean dependent var  :      9.5493                Number of Variables   :           5
S.D. dependent var  :      7.0389                Degrees of Freedom    :        1407
Pseudo R-squared    :      0.3059
N. of iterations    :           1                Step1c computed       :         Yes

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       6.1903085       1.0826509       5.7177328       0.0000000
                DV90       0.4840327       0.1199034       4.0368561       0.0000542
                PS90       1.7859187       0.3361596       5.3127111       0.0000001
                RD90       4.4088516       0.3472585      12.6961665       0.0000000
```

```
               UE90      -0.3824992       0.0986013      -3.8792513       0.0001048
             lambda       0.3161445       0.0374169       8.4492406       0.0000000
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

Setting the maximum number of iterations

```
In [23]: gm5 = pysal.spreg.GM_Error_Het(y,x,w,max_iter=10,name_y=y_name,
                name_x=x_names,name_w="south_q",name_ds="south.dbf")
```

The full listing

```
In [24]: print gm5.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED LEAST SQUARES (HET)
---------------------------------------------------------
Data set            :   south.dbf
Weights matrix      :     south_q
Dependent Variable  :        HR90                Number of Observations:        1412
Mean dependent var  :      9.5493                Number of Variables   :           5
S.D. dependent var  :      7.0389                Degrees of Freedom    :        1407
Pseudo R-squared    :      0.3053
N. of iterations    :           5                Step1c computed       :          No

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       6.0484442       1.0837030       5.5812749       0.0000000
                DV90       0.4900931       0.1200190       4.0834631       0.0000444
                PS90       1.7599431       0.3365045       5.2300726       0.0000002
                RD90       4.3869346       0.3481280      12.6014996       0.0000000
                UE90      -0.3672549       0.0986888      -3.7213436       0.0001982
              lambda       0.3191342       0.0372671       8.5634294       0.0000000
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

0.3.2 Exogenous and Endogenous Variables

```
In [25]: gm6 = pysal.spreg.GM_Endog_Error_Het(y,xe,yend,q,w,name_y=y_name,
                           name_x=xe_names,name_yend=yend_names,name_q=q_names,
                           name_w="south_q",name_ds="south.dbf")
```

Attributes of the regression object

```
In [26]: dir(gm6)
```

```
Out[26]: ['__doc__',
          '__init__',
          '__module__',
          '__summary',
          '_cache',
          'betas',
          'e_filtered',
          'h',
```

```
 'hth',
 'iter_stop',
 'iteration',
 'k',
 'mean_y',
 'n',
 'name_ds',
 'name_h',
 'name_q',
 'name_w',
 'name_x',
 'name_y',
 'name_yend',
 'name_z',
 'pr2',
 'predy',
 'q',
 'std_err',
 'std_y',
 'step1c',
 'summary',
 'title',
 'u',
 'vm',
 'x',
 'y',
 'yend',
 'z',
 'z_stat']
```

The estimated coefficients, including lambda as the last element

```
In [27]: gm6.betas
```

```
Out[27]: array([[ 10.74563401],
       [  5.89766591],
       [  2.03579202],
       [  0.49278877],
       [ -1.13750113],
       [  0.26162482]])
```

The spatial autoregressive coefficient

```
In [28]: gm6.betas[-1][0]
```

```
Out[28]: 0.26162482091893224
```

The full listing

```
In [29]: print gm6.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET)
-------------------------------------------------------------------
Data set            :   south.dbf
```

```
Weights matrix      :     south_q
Dependent Variable  :        HR90                Number of Observations:        1412
Mean dependent var  :      9.5493                Number of Variables   :           5
S.D. dependent var  :      7.0389                Degrees of Freedom    :        1407
Pseudo R-squared    :      0.2820
N. of iterations    :           1                Step1c computed       :          No

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      10.7456340       1.5222725       7.0589425       0.0000000
                DV90       0.4927888       0.1266845       3.8898895       0.0001003
                PS90       2.0357920       0.3491174       5.8312538       0.0000000
                RD90       5.8976659       0.5199311      11.3431685       0.0000000
                UE90      -1.1375011       0.2109871      -5.3913304       0.0000001
              lambda       0.2616248       0.0414083       6.3181770       0.0000000
------------------------------------------------------------------------------------
Instrumented: UE90
Instruments: FH90, FP89, GI89
================================ END OF REPORT =====================================
```

0.4 GMM Homoskedastic Case

0.4.1 Exogenous Variables Only

```
In [30]: gm7 = pysal.spreg.GM_Error_Hom(y,x,w,name_y=y_name,name_x=x_names,
                   name_w="south_q",name_ds="south.dbf")
```

Attributes of the regression object

```
In [31]: dir(gm7)
```

```
Out[31]: ['__doc__',
          '__init__',
          '__module__',
          '__summary',
          '_cache',
          'betas',
          'e_filtered',
          'iter_stop',
          'iteration',
          'k',
          'mean_y',
          'n',
          'name_ds',
          'name_w',
          'name_x',
          'name_y',
          'pr2',
          'predy',
          'sig2',
          'std_err',
          'std_y',
          'summary',
          'title',
```

```
          'u',
          'vm',
          'x',
          'xtx',
          'y',
          'z_stat']
```

The estimated coefficients, including lambda as the last element

```
In [32]: gm7.betas
```

```
Out[32]: array([[ 6.33803479],
                [ 4.43255065],
                [ 1.81323806],
                [-0.39849432],
                [ 0.4777479 ],
                [ 0.27985722]])
```

The spatial autoregressive coefficient

```
In [33]: gm7.betas[-1][0]
```

```
Out[33]: 0.2798572154943586
```

The full listing

```
In [34]: print gm7.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED LEAST SQUARES (HOM)
---------------------------------------------------------
Data set            :   south.dbf
Weights matrix      :     south_q
Dependent Variable  :        HR90                Number of Observations:        1412
Mean dependent var  :      9.5493                Number of Variables   :           5
S.D. dependent var  :      7.0389                Degrees of Freedom    :        1407
Pseudo R-squared    :      0.3066
N. of iterations    :           1

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       6.3380348       1.0237066       6.1912612       0.0000000
                DV90       0.4777479       0.1210440       3.9468939       0.0000792
                PS90       1.8132381       0.2118595       8.5586829       0.0000000
                RD90       4.4325506       0.2336868      18.9679144       0.0000000
                UE90      -0.3984943       0.0777957      -5.1223182       0.0000003
              lambda       0.2798572       0.0355242       7.8779334       0.0000000
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

The A1 option
A1 = 'hom'

```
In [35]: gm8a = pysal.spreg.GM_Error_Hom(y,x,w,A1='hom',name_y=y_name,
                             name_x=x_names,name_w="south_q",
                             name_ds="south.dbf")
```

Full listing

```
In [36]: print gm8a.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED LEAST SQUARES (HOM)
---------------------------------------------------------
Data set            :   south.dbf
Weights matrix      :     south_q
Dependent Variable  :        HR90                Number of Observations:        1412
Mean dependent var  :      9.5493                Number of Variables   :           5
S.D. dependent var  :      7.0389                Degrees of Freedom    :        1407
Pseudo R-squared    :      0.3066
N. of iterations    :           1

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       6.3392818       1.0236978       6.1925324       0.0000000
                DV90       0.4776950       0.1210433       3.9464801       0.0000793
                PS90       1.8134699       0.2118581       8.5598346       0.0000000
                RD90       4.4327545       0.2336848      18.9689495       0.0000000
                UE90      -0.3986299       0.0777951      -5.1241028       0.0000003
              lambda       0.2798359       0.0355255       7.8770510       0.0000000
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

A1 = 'het'

```
In [37]: gm8b = pysal.spreg.GM_Error_Hom(y,x,w,A1='het',name_y=y_name,
                    name_x=x_names,name_w="south_q",
                    name_ds="south.dbf")
```

Full listing

```
In [38]: print gm8b.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED LEAST SQUARES (HOM)
---------------------------------------------------------
Data set            :   south.dbf
Weights matrix      :     south_q
Dependent Variable  :        HR90                Number of Observations:        1412
Mean dependent var  :      9.5493                Number of Variables   :           5
S.D. dependent var  :      7.0389                Degrees of Freedom    :        1407
Pseudo R-squared    :      0.3062
N. of iterations    :           1

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       6.2576037       1.0361188       6.0394656       0.0000000
                DV90       0.4811658       0.1220549       3.9422070       0.0000807
                PS90       1.7983276       0.2138587       8.4089517       0.0000000
```

```
            RD90       4.4195359       0.2365100      18.6864665       0.0000000
            UE90      -0.3897697       0.0786895      -4.9532615       0.0000007
          lambda       0.3088889       0.0349241       8.8445872       0.0000000
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

0.4.2 Exogenous and Endogenous Variables

```
In [39]: gm9 = pysal.spreg.GM_Endog_Error_Hom(y,xe,yend,q,w,name_y=y_name,
                           name_x=xe_names,name_yend=yend_names,name_q=q_names,
                           name_w="south_q",name_ds="south.dbf")
```

Full listing

```
In [40]: print gm9.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HOM)
-------------------------------------------------------------------
Data set            :   south.dbf
Weights matrix      :     south_q
Dependent Variable  :        HR90                Number of Observations:        1412
Mean dependent var  :      9.5493                Number of Variables   :           5
S.D. dependent var  :      7.0389                Degrees of Freedom    :        1407
Pseudo R-squared    :      0.2818
N. of iterations    :           1

------------------------------------------------------------------------------------
        Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
        CONSTANT      10.7713463       1.2834619       8.3924158       0.0000000
            DV90       0.4919212       0.1249985       3.9354165       0.0000831
            PS90       2.0453736       0.2197377       9.3082522       0.0000000
            RD90       5.9036116       0.3489835      16.9165947       0.0000000
            UE90      -1.1406585       0.1491507      -7.6476928       0.0000000
          lambda       0.2431636       0.0389702       6.2397251       0.0000000
------------------------------------------------------------------------------------
Instrumented: UE90
Instruments: FH90, FP89, GI89
================================ END OF REPORT =====================================
```

The A1 option
A1 = 'hom'

```
In [41]: gm10a = pysal.spreg.GM_Endog_Error_Hom(y,xe,yend,q,w,A1='hom',
                               name_y=y_name,name_x=xe_names,name_yend=yend_names,
                               name_q=q_names,name_w="south_q",
                               name_ds="south.dbf")
```

Full listing

```
In [42]: print gm10a.summary
```

```
REGRESSION
----------
```

```
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HOM)
-------------------------------------------------------------------
Data set            :   south.dbf
Weights matrix      :     south_q
Dependent Variable  :        HR90                Number of Observations:        1412
Mean dependent var  :      9.5493                Number of Variables   :           5
S.D. dependent var  :      7.0389                Degrees of Freedom    :        1407
Pseudo R-squared    :      0.2818
N. of iterations    :           1

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      10.7722363       1.2834693       8.3930611       0.0000000
                DV90       0.4918910       0.1249994       3.9351462       0.0000831
                PS90       2.0457096       0.2197392       9.3097152       0.0000000
                RD90       5.9038177       0.3489855      16.9170837       0.0000000
                UE90      -1.1407677       0.1491515      -7.6483826       0.0000000
              lambda       0.2431583       0.0389707       6.2395091       0.0000000
------------------------------------------------------------------------------------
Instrumented: UE90
Instruments: FH90, FP89, GI89
================================ END OF REPORT =====================================
```

A1 = 'het'

```
In [43]: gm10b = pysal.spreg.GM_Endog_Error_Hom(y,xe,yend,q,w,A1='het',
                              name_y=y_name,name_x=xe_names,name_yend=yend_names,
                              name_q=q_names,name_w="south_q",
                              name_ds="south.dbf")
```

Full listing

```
In [44]: print gm10b.summary

REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HOM)
-------------------------------------------------------------------
Data set            :   south.dbf
Weights matrix      :     south_q
Dependent Variable  :        HR90                Number of Observations:        1412
Mean dependent var  :      9.5493                Number of Variables   :           5
S.D. dependent var  :      7.0389                Degrees of Freedom    :        1407
Pseudo R-squared    :      0.2820
N. of iterations    :           1

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      10.7456340       1.2965714       8.2877305       0.0000000
                DV90       0.4927888       0.1257156       3.9198710       0.0000886
                PS90       2.0357920       0.2211232       9.2065972       0.0000000
                RD90       5.8976659       0.3522831      16.7412669       0.0000000
                UE90      -1.1375011       0.1507539      -7.5454198       0.0000000
```

```
              lambda       0.2580216       0.0387024       6.6668084       0.0000000
------------------------------------------------------------------------------------
Instrumented: UE90
Instruments: FH90, FP89, GI89
================================ END OF REPORT =====================================
```

0.5 Practice

Since the spatial diagnostics for the Boston house price example (Chapter 5 practice) pointed to a spatial error alternative, estimate this specification by means of GM, GMM-het and GMM-hom. Compare the results and the inference. Feel free to experiment with the various options (number of iterations, etc.). To assess the effect of endogenous variables, use the south or natregimes data sets for one of the HR specifications (see Chapter 7 practice).

```
In [ ]:
```

Chapter 9 - GM,GMM practice solutions

November 5, 2015

0.1 Solutions to Practice Example Chapter 9

In [1]:
```
%pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

In [2]:
```
import numpy as np
import pysal
```

In [3]:
```
db = pysal.open('data/boston.dbf','r')
```

In [4]:
```
len(db)
```

Out[4]:
```
506
```

In [5]:
```
db.header
```

Out[5]:
```
['POLY_ID',
 'ID',
 'TOWN',
 'TOWNNO',
 'TRACT',
 'LON',
 'LAT',
 'x',
 'y',
 'MEDV',
 'CMEDV',
 'CRIM',
 'ZN',
 'INDUS',
 'CHAS',
 'NOX',
 'RM',
 'AGE',
 'DIS',
 'RAD',
 'TAX',
 'PTRATIO',
 'B',
 'LSTAT']
```

In [6]:
```
y_name = 'MEDV'
y = np.array([db.by_col(y_name)]).T
```

```
In [7]: x_names = ['CRIM','CHAS','NOX','RM','AGE','DIS','LSTAT']
        x = np.array([db.by_col(var) for var in x_names]).T

In [8]: w = pysal.knnW_from_shapefile('data/boston.shp',
                                        k=4,idVariable='POLY_ID')
        w.transform = 'r'

In [9]: w.n

Out[9]: 506
```

0.2 GM

GM estimation (no inference for λ

```
In [10]: reg1 = pysal.spreg.GM_Error(y,x,w,name_y=y_name,name_x=x_names,
                                  name_w="boston_k4",name_ds="boston.dbf")

In [11]: print reg1.summary

REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED LEAST SQUARES
---------------------------------------------------
Data set            :  boston.dbf
Weights matrix      :   boston_k4
Dependent Variable  :        MEDV                Number of Observations:         506
Mean dependent var  :     22.5328                Number of Variables   :           8
S.D. dependent var  :      9.1971                Degrees of Freedom    :         498
Pseudo R-squared    :      0.6563

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      13.2030606       4.3315934       3.0480840       0.0023031
                 AGE      -0.0439815       0.0137959      -3.1880025       0.0014326
                CHAS       0.3561344       0.8493020       0.4193260       0.6749779
                CRIM      -0.0997814       0.0278667      -3.5806686       0.0003427
                 DIS      -1.2060859       0.2713816      -4.4442426       0.0000088
               LSTAT      -0.3884304       0.0511833      -7.5890133       0.0000000
                 NOX     -19.2017960       4.6625597      -4.1182949       0.0000382
                  RM       5.2655604       0.3780721      13.9273967       0.0000000
              lambda       0.6194043
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

0.3 GMM

GMM estimation, heteroskedasticity

```
In [12]: reg2 = pysal.spreg.GM_Error_Het(y,x,w,name_y=y_name,name_x=x_names,
                                  name_w="boston_k4",name_ds="boston.dbf")

In [13]: print reg2.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED LEAST SQUARES (HET)
---------------------------------------------------------
Data set            :  boston.dbf
Weights matrix      :   boston_k4
Dependent Variable  :        MEDV               Number of Observations:         506
Mean dependent var  :     22.5328               Number of Variables   :           8
S.D. dependent var  :      9.1971               Degrees of Freedom    :         498
Pseudo R-squared    :      0.6546
N. of iterations    :           1               Step1c computed       :          No

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      13.3293495       7.9502765       1.6765894       0.0936228
                 AGE      -0.0447930       0.0143511      -3.1212345       0.0018009
                CHAS       0.2680795       1.3168505       0.2035763       0.8386847
                CRIM      -0.0986761       0.0378155      -2.6094115       0.0090698
                 DIS      -1.2035485       0.3259144      -3.6928361       0.0002218
               LSTAT      -0.3824789       0.0897396      -4.2620956       0.0000203
                 NOX     -19.4566834       5.1702338      -3.7632115       0.0001677
                  RM       5.2668820       0.9616157       5.4771174       0.0000000
              lambda       0.7319953       0.0516632      14.1685990       0.0000000
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

GMM estimation, homoskedasticity

```
In [14]: reg3 = pysal.spreg.GM_Error_Hom(y,x,w,name_y=y_name,name_x=x_names,
                    name_w="boston_k4",name_ds="boston.dbf")

In [15]: print reg3.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED LEAST SQUARES (HOM)
---------------------------------------------------------
Data set            :  boston.dbf
Weights matrix      :   boston_k4
Dependent Variable  :        MEDV               Number of Observations:         506
Mean dependent var  :     22.5328               Number of Variables   :           8
S.D. dependent var  :      9.1971               Degrees of Freedom    :         498
Pseudo R-squared    :      0.6563
N. of iterations    :           1

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      13.2024291       4.4531626       2.9647310       0.0030295
                 AGE      -0.0439774       0.0136983      -3.2104366       0.0013253
                CHAS       0.3565845       0.8388698       0.4250772       0.6707804
                CRIM      -0.0997870       0.0275156      -3.6265588       0.0002872
                 DIS      -1.2060960       0.2933683      -4.1112010       0.0000394
               LSTAT      -0.3884606       0.0509103      -7.6303011       0.0000000
```

```
               NOX     -19.2005124       4.8708867      -3.9418926       0.0000808
                RM       5.2655520       0.3725242      14.1347915       0.0000000
            lambda       0.6914855       0.0311299      22.2129044       0.0000000
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

In [ ]:

In [ ]:

Chapter 10 - ML Error

November 5, 2015

This notebook contains the PySAL/spreg code for Chapter 10 - ML Error
in Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL.
by Luc Anselin and Sergio J. Rey


```
In [1]: __author__ = "Luc Anselin luc.anselin@asu.edu"
```

0.1 Basic Regression Setup

0.1.1 Model Specification

Creating arrays for y and x for south.dbf example data set

Preliminaries, import **numpy** and **pysal**

```
In [1]: %pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

```
In [2]: import numpy as np
        import pysal
```

Basic Model

```
In [3]: db = pysal.open(pysal.examples.get_path('south.dbf'),'r')
        y_name = "HR90"
        y = np.array([db.by_col(y_name)]).T
        x_names = ["RD90","PS90","UE90","DV90"]
        x = np.array([db.by_col(var) for var in x_names]).T
```

Spatial Weights

Reading in the spatial weights

```
In [4]: galw = pysal.open(pysal.examples.get_path("south_q.gal"),'r')
        w = galw.read()
        galw.close
        w.transform = 'r'
```

Alternatively, creating from scratch

```
In [5]: w = pysal.queen_from_shapefile(pysal.examples.get_path("south.shp"),idVariable="FIPSNO")
        w.transform = 'r'
```

0.2 ML Error - Method Full

```
In [10]: reg1 = pysal.spreg.ML_Error(y,x,w,name_y=y_name,name_x=x_names,name_w = "south_q",name_ds = "s
```

Contents of the ML_Error object

```
In [11]: dir(reg1)
```

```
Out[11]: ['__doc__',
          '__init__',
          '__module__',
          '__summary',
          '_cache',
          'aic',
          'betas',
          'e_filtered',
          'epsilon',
          'get_x_lag',
          'k',
          'lam',
          'logll',
          'mean_y',
          'method',
          'n',
          'name_ds',
          'name_w',
          'name_x',
          'name_y',
          'pr2',
          'predy',
          'schwarz',
          'sig2',
          'sig2n',
          'sig2n_k',
          'std_err',
          'std_y',
          'summary',
          'title',
          'u',
          'utu',
          'vm',
          'vm1',
          'x',
          'y',
          'z_stat']
```

Regression coefficients, with lambda as last one

```
In [12]: reg1.betas
```

```
Out[12]: array([[ 6.14922463],
                [ 4.40242011],
                [ 1.77837122],
                [-0.3780731 ],
                [ 0.48578576],
                [ 0.29907787]])
```

Lambda extracted

```
In [13]: reg1.betas[-1][0]

Out[13]: 0.29907786543300452
```

Lambda explicitly

```
In [14]: reg1.lam

Out[14]: 0.29907786543300452
```

Full listing

```
In [15]: print reg1.summary

REGRESSION
----------
SUMMARY OF OUTPUT: MAXIMUM LIKELIHOOD SPATIAL ERROR (METHOD = FULL)
-------------------------------------------------------------------
Data set            :       south
Weights matrix      :     south_q
Dependent Variable  :        HR90                Number of Observations:        1412
Mean dependent var  :      9.5493                Number of Variables   :           5
S.D. dependent var  :      7.0389                Degrees of Freedom    :        1407
Pseudo R-squared    :      0.3058
Sigma-square ML     :      32.407                Log likelihood        :   -4471.407
S.E of regression   :       5.693                Akaike info criterion :    8952.814
                                                 Schwarz criterion     :    8979.078

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       6.1492246       1.0318747       5.9592748       0.0000000
                DV90       0.4857858       0.1217110       3.9913061       0.0000657
                PS90       1.7783712       0.2131787       8.3421630       0.0000000
                RD90       4.4024201       0.2355472      18.6901825       0.0000000
                UE90      -0.3780731       0.0783853      -4.8232683       0.0000014
              lambda       0.2990779       0.0378155       7.9088793       0.0000000
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

0.3 ML Error - Method Ord

```
In [17]: reg2 = pysal.spreg.ML_Error(y,x,w,method='ord',\
                                             name_y=y_name,name_x=x_names,name_w = "south_q",name_ds =
```

Full listing of results

```
In [18]: print reg2.summary

REGRESSION
----------
SUMMARY OF OUTPUT: MAXIMUM LIKELIHOOD SPATIAL ERROR (METHOD = ORD)
------------------------------------------------------------------
Data set            :       south
Weights matrix      :     south_q
```

```
Dependent Variable  :       HR90          Number of Observations:      1412
Mean dependent var  :     9.5493          Number of Variables   :         5
S.D. dependent var  :     7.0389          Degrees of Freedom    :      1407
Pseudo R-squared    :     0.3058
Sigma-square ML     :     32.407          Log likelihood        :  -4471.407
S.E of regression   :      5.693          Akaike info criterion :   8952.814
                                          Schwarz criterion     :   8979.078

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       6.1492248       1.0318746       5.9592751       0.0000000
                DV90       0.4857858       0.1217110       3.9913061       0.0000657
                PS90       1.7783713       0.2131787       8.3421633       0.0000000
                RD90       4.4024201       0.2355472      18.6901829       0.0000000
                UE90      -0.3780731       0.0783853      -4.8232686       0.0000014
              lambda       0.2990778       0.0378155       7.9088781       0.0000000
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

In [ ]:

In [ ]:

Chapter 11 - Combo Model

November 5, 2015

This notebook contains the PySAL/spreg code for Chapter 11 - Combo Model
in Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL.
by Luc Anselin and Sergio J. Rey


```
In [1]: __author__ = "Luc Anselin luc.anselin@asu.edu"
```

0.1 Basic Regression Setup

0.1.1 Exogenous Explanatory Variables Only

```
In [1]: %pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

```
In [2]: import numpy as np
        import pysal
```

Creating arrays for y and x for nat.dbf example data set

```
In [7]: db = pysal.open('data/natregimes.dbf','r')
        y_name = "HR60"
        y = np.array([db.by_col(y_name)]).T
        x_names = ["RD60","PS60","UE60","DV60","BLK60"]
        x = np.array([db.by_col(var) for var in x_names]).T
```

0.1.2 Exogenous and Endogenous Explanatory Variables

Creating arrays for yend, q and xe (exogenous only)

```
In [8]: yend_names = ["UE60"]
        yend = np.array([db.by_col(var) for var in yend_names]).T
        q_names = ["FH60","FP59","GI59"]
        q = np.array([db.by_col(var) for var in q_names]).T
        xe_names = ["RD60","PS60","DV60","BLK60"]
        xe = np.array([db.by_col(var) for var in xe_names]).T
```

0.1.3 Spatial Weights

Reading in the weights file

```
In [5]: galw = pysal.open(pysal.examples.get_path("nat_queen.gal"),'r')
        w = galw.read()
        galw.close
        w.transform = 'r'
```

Alternatively, creating from scratch

```
In [12]: w = pysal.queen_from_shapefile('data/natregimes.shp',idVariable="FIPSNO")
         w.transform = 'r'
```

0.2 GM Combo

Exogenous Only

```
In [13]: combo1 = pysal.spreg.GM_Combo(y,x,w=w,name_y=y_name,
                                  name_x=x_names,name_w="nat_queen",
                                  name_ds="NAT")
```

```
In [14]: dir(combo1)
```

```
Out[14]: ['__doc__',
          '__init__',
          '__module__',
          '__summary',
          '_cache',
          'betas',
          'e_filtered',
          'e_pred',
          'k',
          'mean_y',
          'n',
          'name_ds',
          'name_h',
          'name_q',
          'name_w',
          'name_x',
          'name_y',
          'name_yend',
          'name_z',
          'pr2',
          'pr2_e',
          'predy',
          'predy_e',
          'rho',
          'sig2',
          'std_err',
          'std_y',
          'summary',
          'title',
          'u',
          'vm',
          'x',
          'y',
          'yend',
          'z',
          'z_stat']
```

The coefficient estimates

```
In [15]: combo1.betas
```

```
Out[15]: array([[ 0.32411091],
               [ 0.80866252],
               [ 0.1056478 ],
               [ 0.05279337],
               [ 0.61577086],
```

```
                [ 0.07096489],
                [ 0.44950309],
                [-0.1884266 ]])
```

The spatial autoregressive (lag) coefficient

```
In [16]: combo1.rho

Out[16]: array([ 0.44950309])

In [17]: combo1.betas[-2][0]

Out[17]: 0.44950309137162137
```

The spatial autoregressive error coefficient

```
In [18]: combo1.betas[-1][0]

Out[18]: -0.1884265986340069
```

Full listing

```
In [19]: print combo1.summary

REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES
-------------------------------------------------------------
Data set            :         NAT
Weights matrix      :   nat_queen
Dependent Variable  :        HR60                Number of Observations:        3085
Mean dependent var  :      4.5041                Number of Variables   :           7
S.D. dependent var  :      5.6497                Degrees of Freedom    :        3078
Pseudo R-squared    :      0.3333
Spatial Pseudo R-squared:  0.2854

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       0.3241109       0.2492294       1.3004522       0.1934460
               BLK60       0.0709649       0.0105437       6.7305432       0.0000000
                DV60       0.6157709       0.0980317       6.2813471       0.0000000
                PS60       0.1056478       0.0843697       1.2522008       0.2104967
                RD60       0.8086625       0.1400977       5.7721323       0.0000000
                UE60       0.0527934       0.0308956       1.7087671       0.0874941
               W_HR60      0.4495031       0.0652376       6.8902409       0.0000000
               lambda     -0.1884266
------------------------------------------------------------------------------------
Instrumented: W_HR60
Instruments: W_BLK60, W_DV60, W_PS60, W_RD60, W_UE60
================================ END OF REPORT =====================================
```

Exogenous and Endogenous explanatory variables

```
In [20]: combo2 = pysal.spreg.GM_Combo(y,xe,yend=yend,q=q,w=w,
                                name_y=y_name,name_x=xe_names,
                                name_yend=yend_names,name_q=q_names,
                                name_w="nat_queen",name_ds="NAT")
```

```
In [21]: print combo2.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES
-------------------------------------------------------------
Data set            :         NAT
Weights matrix      :   nat_queen
Dependent Variable  :        HR60                Number of Observations:        3085
Mean dependent var  :      4.5041                Number of Variables   :           7
S.D. dependent var  :      5.6497                Degrees of Freedom    :        3078
Pseudo R-squared    :      0.3328
Spatial Pseudo R-squared:  0.2812

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      -0.0618162       0.3505032      -0.1763642       0.8600079
               BLK60       0.0703360       0.0108268       6.4964399       0.0000000
                DV60       0.5411868       0.1009337       5.3618036       0.0000001
                PS60       0.0636171       0.0911780       0.6977245       0.4853495
                RD60       0.6903681       0.1485215       4.6482704       0.0000033
                UE60       0.1276860       0.0688647       1.8541563       0.0637168
              W_HR60       0.4818021       0.0610440       7.8926964       0.0000000
              lambda      -0.1901571
------------------------------------------------------------------------------------
Instrumented: UE60, W_HR60
Instruments: FH60, FP59, GI59, W_BLK60, W_DV60, W_FH60, W_FP59, W_GI59,
             W_PS60, W_RD60
================================ END OF REPORT =====================================
```

0.3 GM Combo with Homoskedastic Errors

Exogenous only

```
In [22]: combo3 = pysal.spreg.GM_Combo_Hom(y,x,w=w,name_y=y_name,
                      name_x=x_names,name_w="nat_queen",
                      name_ds="NAT")
```

```
In [23]: dir(combo3)
```

```
Out[23]: ['__doc__',
          '__init__',
          '__module__',
          '__summary',
          '_cache',
          'betas',
          'e_filtered',
          'e_pred',
          'h',
          'hth',
          'iter_stop',
          'iteration',
          'k',
          'mean_y',
```

```
         'n',
         'name_ds',
         'name_h',
         'name_q',
         'name_w',
         'name_x',
         'name_y',
         'name_yend',
         'name_z',
         'pr2',
         'pr2_e',
         'predy',
         'predy_e',
         'q',
         'rho',
         'sig2',
         'std_err',
         'std_y',
         'summary',
         'title',
         'u',
         'vm',
         'x',
         'y',
         'yend',
         'z',
         'z_stat']
```

```
In [24]: print combo3.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HOM)
-------------------------------------------------------------------
Data set            :         NAT
Weights matrix      :   nat_queen
Dependent Variable  :        HR60                Number of Observations:        3085
Mean dependent var  :      4.5041                Number of Variables   :           7
S.D. dependent var  :      5.6497                Degrees of Freedom    :        3078
Pseudo R-squared    :      0.3333
Spatial Pseudo R-squared:  0.2854
N. of iterations    :           1

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       0.3239850       0.2364887       1.3699807       0.1706929
               BLK60       0.0709563       0.0103216       6.8745561       0.0000000
                DV60       0.6158838       0.0969425       6.3530825       0.0000000
                PS60       0.1056035       0.0804090       1.3133301       0.1890717
                RD60       0.8086410       0.1364000       5.9284534       0.0000000
                UE60       0.0527400       0.0295573       1.7843270       0.0743705
              W_HR60       0.4495694       0.0652271       6.8923719       0.0000000
              lambda      -0.2729182       0.1021516      -2.6716980       0.0075469
------------------------------------------------------------------------------------
```

```
Instrumented: W_HR60
Instruments: W_BLK60, W_DV60, W_PS60, W_RD60, W_UE60
================================ END OF REPORT =====================================
```

Exogenous and Endogenous explanatory variables

```
In [25]: combo4 = pysal.spreg.GM_Combo_Hom(y,xe,yend=yend,
                       q=q,w=w,name_y=y_name,name_x=xe_names,
                       name_yend=yend_names,name_q=q_names,
                       name_w="nat_queen",name_ds="NAT")

In [26]: print combo4.summary

REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HOM)
-------------------------------------------------------------------
Data set            :         NAT
Weights matrix      :   nat_queen
Dependent Variable  :        HR60                Number of Observations:        3085
Mean dependent var  :      4.5041                Number of Variables   :           7
S.D. dependent var  :      5.6497                Degrees of Freedom    :        3078
Pseudo R-squared    :      0.3328
Spatial Pseudo R-squared:  0.2812
N. of iterations    :           1

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      -0.0617938       0.3188267      -0.1938164       0.8463197
               BLK60       0.0703134       0.0103875       6.7690271       0.0000000
                DV60       0.5412369       0.0977580       5.5365000       0.0000000
                PS60       0.0635952       0.0851447       0.7469065       0.4551200
                RD60       0.6902893       0.1403672       4.9177379       0.0000009
                UE60       0.1275722       0.0631550       2.0199859       0.0433848
               W_HR60       0.4819608       0.0602250       8.0026721       0.0000000
              lambda      -0.3104618       0.0989503      -3.1375537       0.0017036
------------------------------------------------------------------------------------
Instrumented: UE60, W_HR60
Instruments: FH60, FP59, GI59, W_BLK60, W_DV60, W_FH60, W_FP59, W_GI59,
             W_PS60, W_RD60
================================ END OF REPORT =====================================
```

0.4 GM Combo with Heteroskedastic Errors

Exogenous only

```
In [27]: combo5 = pysal.spreg.GM_Combo_Het(y,x,w=w,name_y=y_name,
                          name_x=x_names,name_w="nat_queen",
                          name_ds="NAT")

In [28]: print combo5.summary

REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET)
```

```
------------------------------------------------------------------
Data set            :         NAT
Weights matrix      :   nat_queen
Dependent Variable  :        HR60                Number of Observations:        3085
Mean dependent var  :      4.5041                Number of Variables   :           7
S.D. dependent var  :      5.6497                Degrees of Freedom    :        3078
Pseudo R-squared    :      0.3333
Spatial Pseudo R-squared:  0.2853
N. of iterations    :           1                Step1c computed       :          No

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       0.3264431       0.2256289       1.4468142       0.1479490
               BLK60       0.0711234       0.0122398       5.8108400       0.0000000
                DV60       0.6136830       0.1071702       5.7262456       0.0000000
                PS60       0.1064660       0.1016379       1.0475035       0.2948674
                RD60       0.8090534       0.1532549       5.2791348       0.0000001
                UE60       0.0537745       0.0287134       1.8728001       0.0610960
               W_HR60      0.4482870       0.0719587       6.2297823       0.0000000
              lambda      -0.4383045       0.0979689      -4.4739149       0.0000077
------------------------------------------------------------------------------------
Instrumented: W_HR60
Instruments: W_BLK60, W_DV60, W_PS60, W_RD60, W_UE60
================================ END OF REPORT =====================================
```

Exogenous and Endogenous explanatory variables

```
In [29]: combo6 = pysal.spreg.GM_Combo_Het(y,xe,yend=yend,q=q,w=w,
                                name_y=y_name,name_x=xe_names,
                                name_yend=yend_names,name_q=q_names,
                                name_w="nat_queen",name_ds="NAT")

In [30]: print combo6.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET)
------------------------------------------------------------------
Data set            :         NAT
Weights matrix      :   nat_queen
Dependent Variable  :        HR60                Number of Observations:        3085
Mean dependent var  :      4.5041                Number of Variables   :           7
S.D. dependent var  :      5.6497                Degrees of Freedom    :        3078
Pseudo R-squared    :      0.3328
Spatial Pseudo R-squared:  0.2810
N. of iterations    :           1                Step1c computed       :          No

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      -0.0622402       0.3447555      -0.1805344       0.8567330
               BLK60       0.0707237       0.0111402       6.3485286       0.0000000
                DV60       0.5402806       0.1044551       5.1723733       0.0000002
                PS60       0.0639946       0.0894662       0.7152939       0.4744275
```

```
               RD60       0.6916849       0.1395608       4.9561544       0.0000007
               UE60       0.1296600       0.0692818       1.8714876       0.0612775
              W_HR60      0.4790831       0.0645823       7.4181827       0.0000000
              lambda     -0.4644452       0.0926202      -5.0145116       0.0000005
------------------------------------------------------------------------------------
Instrumented: UE60, W_HR60
Instruments: FH60, FP59, GI59, W_BLK60, W_DV60, W_FH60, W_FP59, W_GI59,
             W_PS60, W_RD60
================================ END OF REPORT =====================================
```

In [ ]:

In [ ]:

Chapter 12 - Regimes (OLS only)

November 5, 2015

This notebook contains the PySAL/spreg code for Chapter 12 - Regimes, Non-Spatial (OLS only) in Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL.
by Luc Anselin and Sergio J. Rey


```
In [1]: __author__ = "Luc Anselin luc.anselin@asu.edu"
```

0.1 Regimes - Non-spatial - OLS

0.1.1 Baltimore Example

Basic Setup:

- import necessary modules (numpy and pysal)
- create a data object
- create variables as numpy arrays
- create regime variable (as list)
- create weights object(s) for diagnostics

```
In [1]: %pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

```
In [2]: import numpy as np
        import pysal
```

create data object

```
In [3]: db = pysal.open('data/baltim.dbf','r')
```

read in dependent variable and turn into numpy array y

```
In [4]: y_name = "PRICE"
        y = np.array([db.by_col(y_name)]).T
```

read in explanatory variables and turn into numpy array x

```
In [5]: x_names = ['NROOM','NBATH','PATIO','FIREPL','AC','GAR','AGE','LOTSZ','SQFT']
        x = np.array([db.by_col(var) for var in x_names]).T
```

create k = 4 nearest neighbor weights and row-standardize

```
In [6]: w = pysal.knnW_from_shapefile("data/baltim.shp",k=4,idVariable='STATION')
        w.transform = 'r'
```

use CITCOU as the regimes variable

```
In [7]: rvar = "CITCOU"
```

```
In [8]: regimes = db.by_col(rvar)    # note: regimes is a list
```

```
In [9]: regimes[:4]
```

```
Out[9]: [0.0, 1.0, 1.0, 1.0]
```

0.2 Regimes - Default Setting

With spatial diagnostics

```
In [10]: reg1 = pysal.spreg.OLS_Regimes(y,x,regimes,w=w,spat_diag=True,moran=True,
         name_y=y_name,name_x=x_names,name_regimes=rvar,name_w="baltim_k4",name_ds="baltim.dbf")
```

Various regime settings
Separate regressions by regime

```
In [11]: reg1.regime_err_sep
```

```
Out[11]: True
```

Different constant term in each regime

```
In [12]: reg1.constant_regi
```

```
Out[12]: 'many'
```

All coefficients are varying

```
In [13]: reg1.cols2regi
```

```
Out[13]: 'all'
```

Full output
Note the warning for islands in each of the regimes

```
In [14]: print reg1.summary
```

```
REGRESSION
----------

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION - REGIME 0
---------------------------------------------------------------
Data set            :  baltim.dbf
Weights matrix      :   baltim_k4
Dependent Variable  :     0_PRICE              Number of Observations:          83
Mean dependent var  :     31.5127              Number of Variables   :          10
S.D. dependent var  :     17.1598              Degrees of Freedom    :          73
R-squared           :      0.6129
Adjusted R-squared  :      0.5652
Sum squared residual:    9347.239              F-statistic           :     12.8414
Sigma-square        :     128.044              Prob(F-statistic)     :   5.381e-12
```

```
S.E. of regression  :      11.316              Log likelihood       :    -313.818
Sigma-square ML     :     112.617              Akaike info criterion :    647.635
S.E of regression ML:     10.6121              Schwarz criterion     :    671.824
```

Variable	Coefficient	Std.Error	t-Statistic	Probability
0_CONSTANT	8.1507931	6.1763252	1.3196833	0.1910646
0_AC	12.5725010	4.4326770	2.8363224	0.0059022
0_AGE	0.0480813	0.0638384	0.7531721	0.4537687
0_FIREPL	7.4664199	3.8221968	1.9534368	0.0546005
0_GAR	0.1065636	2.7762058	0.0383846	0.9694858
0_LOTSZ	0.1623115	0.0416145	3.9003550	0.0002116
0_NBATH	4.3791320	2.3582245	1.8569615	0.0673509
0_NROOM	1.2632458	1.4442233	0.8746887	0.3846120
0_PATIO	11.0843832	5.1221666	2.1640029	0.0337364
0_SQFT	-0.1616281	0.2387840	-0.6768797	0.5006216

Regimes variable: CITCOU
Warning: The regimes operation resulted in islands for regime 0.

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 19.164

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	3.619	0.1637

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	9	51.012	0.0000
Koenker-Bassett test	9	36.345	0.0000

DIAGNOSTICS FOR SPATIAL DEPENDENCE

TEST	MI/DF	VALUE	PROB
Moran's I (error)	-0.0118	0.174	0.8619
Lagrange Multiplier (lag)	1	0.003	0.9551
Robust LM (lag)	1	0.003	0.9533
Lagrange Multiplier (error)	1	0.019	0.8900
Robust LM (error)	1	0.019	0.8893
Lagrange Multiplier (SARMA)	2	0.023	0.9888

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION - REGIME 1

```
Data set            :  baltim.dbf
Weights matrix      :   baltim_k4
Dependent Variable  :     1_PRICE              Number of Observations:         128
Mean dependent var  :     52.6036              Number of Variables   :          10
S.D. dependent var  :     23.5574              Degrees of Freedom    :         118
R-squared           :      0.6994
Adjusted R-squared  :      0.6765
```

Sum squared residual:	21182.550		F-statistic	:	30.5123
Sigma-square :	179.513		Prob(F-statistic)	:	7.144e-27
S.E. of regression :	13.398		Log likelihood	:	-508.594
Sigma-square ML :	165.489		Akaike info criterion	:	1037.188
S.E of regression ML:	12.8642		Schwarz criterion	:	1065.708

Variable	Coefficient	Std.Error	t-Statistic	Probability
1_CONSTANT	12.6503298	8.2390546	1.5354104	0.1273601
1_AC	1.9482904	3.2564121	0.5982935	0.5507907
1_AGE	-0.2262105	0.1319972	-1.7137522	0.0892000
1_FIREPL	11.4398872	3.2634124	3.5054985	0.0006452
1_GAR	9.0685246	2.3854458	3.8016058	0.0002291
1_LOTSZ	0.0473774	0.0183697	2.5791121	0.0111337
1_NBATH	13.6724100	2.8027934	4.8781370	0.0000034
1_NROOM	1.8803799	1.7763784	1.0585469	0.2919683
1_PATIO	6.8221115	3.4400819	1.9831247	0.0496762
1_SQFT	0.1398581	0.2406887	0.5810746	0.5622984

Regimes variable: CITCOU
Warning: The regimes operation resulted in islands for regime 1.

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 25.769

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	124.888	0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	9	87.423	0.0000
Koenker-Bassett test	9	27.036	0.0014

DIAGNOSTICS FOR SPATIAL DEPENDENCE

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.1795	3.116	0.0018
Lagrange Multiplier (lag)	1	19.126	0.0000
Robust LM (lag)	1	11.439	0.0007
Lagrange Multiplier (error)	1	7.728	0.0054
Robust LM (error)	1	0.041	0.8398
Lagrange Multiplier (SARMA)	2	19.167	0.0001

REGIMES DIAGNOSTICS - CHOW TEST

VARIABLE	DF	VALUE	PROB
CONSTANT	1	0.191	0.6621
AC	1	3.731	0.0534
AGE	1	3.500	0.0614
FIREPL	1	0.625	0.4292
GAR	1	5.995	0.0143
LOTSZ	1	6.384	0.0115

```
                NBATH          1            6.437           0.0112
                NROOM          1            0.073           0.7875
                PATIO          1            0.477           0.4897
                 SQFT          1            0.791           0.3739
          Global test         10           68.382           0.0000

DIAGNOSTICS FOR GLOBAL SPATIAL DEPENDENCE
Residuals are treated as homoskedastic for the purpose of these tests
TEST                           MI/DF       VALUE           PROB
Moran's I (error)              0.0895        2.520           0.0118
Lagrange Multiplier (lag)         1         15.683           0.0001
Robust LM (lag)                   1         12.271           0.0005
Lagrange Multiplier (error)       1          3.783           0.0518
Robust LM (error)                 1          0.371           0.5426
Lagrange Multiplier (SARMA)       2         16.054           0.0003

================================ END OF REPORT =====================================
```

0.2.1 Regime Options

regime_err_sep = False - forced homoskedasticity

using k nearest neighbor weights

```
In [15]: reg2 = pysal.spreg.OLS_Regimes(y,x,regimes,w=w,spat_diag=True,moran=True,
         regime_err_sep=False,
         name_y=y_name,name_x=x_names,name_regimes=rvar,name_w="baltim_k4",name_ds="baltim.dbf")

In [16]: print reg2.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES - REGIMES
---------------------------------------------------
Data set            :  baltim.dbf
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE               Number of Observations:         211
Mean dependent var  :     44.3072               Number of Variables   :          20
S.D. dependent var  :     23.6061               Degrees of Freedom    :         191
R-squared           :      0.7391
Adjusted R-squared  :      0.7132
Sum squared residual:   30529.788               F-statistic           :     28.4795
Sigma-square        :     159.842               Prob(F-statistic)     :   1.261e-45
S.E. of regression  :      12.643               Log likelihood        :    -824.216
Sigma-square ML     :     144.691               Akaike info criterion :    1688.433
S.E of regression ML:     12.0288               Schwarz criterion     :    1755.470

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
          0_CONSTANT       8.1507931       6.9007315       1.1811491       0.2390121
                0_AC      12.5725010       4.9525749       2.5385786       0.0119281
               0_AGE       0.0480813       0.0713259       0.6741077       0.5010579
            0_FIREPL       7.4664199       4.2704930       1.7483742       0.0820059
               0_GAR       0.1065636       3.1018203       0.0343552       0.9726298
             0_LOTSZ       0.1623115       0.0464954       3.4909141       0.0005978
```

```
        0_NBATH       4.3791320       2.6348149       1.6620264       0.0981478
        0_NROOM       1.2632458       1.6136128       0.7828680       0.4346742
        0_PATIO      11.0843832       5.7229332       1.9368360       0.0542408
         0_SQFT      -0.1616281       0.2667904      -0.6058241       0.5453508
     1_CONSTANT      12.6503298       7.7745361       1.6271492       0.1053540
           1_AC       1.9482904       3.0728153       0.6340408       0.5268128
          1_AGE      -0.2262105       0.1245552      -1.8161467       0.0709158
       1_FIREPL      11.4398872       3.0794210       3.7149475       0.0002668
          1_GAR       9.0685246       2.2509542       4.0287468       0.0000809
        1_LOTSZ       0.0473774       0.0173340       2.7332108       0.0068612
        1_NBATH      13.6724100       2.6447717       5.1695994       0.0000006
        1_NROOM       1.8803799       1.6762261       1.1217937       0.2633585
        1_PATIO       6.8221115       3.2461298       2.1016139       0.0368984
         1_SQFT       0.1398581       0.2271186       0.6157931       0.5387642
------------------------------------------------------------------------------------
Regimes variable: CITCOU

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           25.769

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2       150.287           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test               19       148.626           0.0000
Koenker-Bassett test             19        50.961           0.0001

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Moran's I (error)              0.0895       2.415           0.0157
Lagrange Multiplier (lag)         1        15.683           0.0001
Robust LM (lag)                   1        12.271           0.0005
Lagrange Multiplier (error)       1         3.783           0.0518
Robust LM (error)                 1         0.371           0.5426
Lagrange Multiplier (SARMA)       2        16.054           0.0003


REGIMES DIAGNOSTICS - CHOW TEST
                 VARIABLE        DF        VALUE           PROB
                 CONSTANT         1         0.187           0.6651
                       AC         1         3.323           0.0683
                      AGE         1         3.652           0.0560
                   FIREPL         1         0.570           0.4504
                      GAR         1         5.468           0.0194
                    LOTSZ         1         5.365           0.0205
                    NBATH         1         6.197           0.0128
                    NROOM         1         0.070           0.7908
                    PATIO         1         0.420           0.5171
                     SQFT         1         0.740           0.3895
              Global test        10        65.256           0.0000
================================ END OF REPORT =====================================
```

constant_regi='one' – one global constant
with regime_err_sep=True (default), i.e. groupwise heteroskedasticity

```
In [17]: reg3 = pysal.spreg.OLS_Regimes(y,x,regimes,w=w,spat_diag=True,moran=True,
         constant_regi='one',name_y=y_name,name_x=x_names,
         name_regimes=rvar,name_w="baltim_k4",name_ds="baltim.dbf")

In [18]: print reg3.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES - REGIMES (Group-wise heteroskedasticity)
----------------------------------------------------------------------------------
Data set            :  baltim.dbf
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE                Number of Observations:         211
Mean dependent var  :     44.3072                Number of Variables   :          19
S.D. dependent var  :     23.6061                Degrees of Freedom    :         192
R-squared           :      0.7389
Adjusted R-squared  :      0.7144

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
                0_AC      12.3629617       4.4713687       2.7649167       0.0062490
               0_AGE       0.0415673       0.0630325       0.6594583       0.5103917
            0_FIREPL       7.6921977       3.8431499       2.0015347       0.0467427
               0_GAR       0.3583408       2.7571385       0.1299684       0.8967275
             0_LOTSZ       0.1631251       0.0421765       3.8676805       0.0001504
             0_NBATH       4.2585971       2.3765053       1.7919577       0.0747144
             0_NROOM       0.9732288       1.3069987       0.7446287       0.4574066
             0_PATIO      11.1805981       5.1917191       2.1535445       0.0325226
              0_SQFT      -0.1448928       0.2392109      -0.6057117       0.5454216
                1_AC       1.9770250       3.2167442       0.6146044       0.5395437
               1_AGE      -0.2085209       0.1240057      -1.6815427       0.0942834
            1_FIREPL      11.3783766       3.2212699       3.5322642       0.0005160
               1_GAR       8.9112047       2.3293499       3.8256187       0.0001763
             1_LOTSZ       0.0472602       0.0181477       2.6041969       0.0099293
             1_NBATH      13.9019236       2.7192012       5.1125027       0.0000008
             1_NROOM       2.3460009       1.3965542       1.6798496       0.0946130
             1_PATIO       6.6782440       3.3829757       1.9740739       0.0498074
              1_SQFT       0.1226495       0.2345384       0.5229399       0.6016190
    _Global_CONSTANT       9.8248121       4.9652501       1.9787144       0.0492779
------------------------------------------------------------------------------------
Regimes variable: CITCOU
Warning: Residuals treated as homoskedastic for the purpose of diagnostics.

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Moran's I (error)              0.0854         2.319           0.0204
Lagrange Multiplier (lag)         1          15.213           0.0001
Robust LM (lag)                   1          12.214           0.0005
Lagrange Multiplier (error)       1           3.437           0.0638
Robust LM (error)                 1           0.438           0.5081
Lagrange Multiplier (SARMA)       2          15.651           0.0004
```

```
REGIMES DIAGNOSTICS - CHOW TEST
                 VARIABLE        DF        VALUE           PROB
                       AC         1           3.563           0.0591
                      AGE         1           3.446           0.0634
                   FIREPL         1           0.543           0.4610
                      GAR         1           5.799           0.0160
                    LOTSZ         1           6.371           0.0116
                    NBATH         1           7.292           0.0069
                    NROOM         1           0.837           0.3602
                    PATIO         1           0.530           0.4666
                     SQFT         1           0.655           0.4182
              Global test         9          68.189           0.0000
================================ END OF REPORT =====================================
```

with regime_err_sep=False, i.e. homoskedasticity

```
In [19]: reg4 = pysal.spreg.OLS_Regimes(y,x,regimes,w=w,spat_diag=True,moran=True,
         constant_regi='one',regime_err_sep=False,name_y=y_name,name_x=x_names,
         name_regimes=rvar,name_w="baltim_k4",name_ds="baltim.dbf")
```

```
In [20]: print reg4.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES - REGIMES
---------------------------------------------------
Data set            :  baltim.dbf
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE                Number of Observations:         211
Mean dependent var  :     44.3072                Number of Variables   :          19
S.D. dependent var  :     23.6061                Degrees of Freedom    :         192
R-squared           :      0.7389
Adjusted R-squared  :      0.7144
Sum squared residual:   30559.735                F-statistic           :     30.1790
Sigma-square        :     159.165                Prob(F-statistic)     :   2.471e-46
S.E. of regression  :      12.616                Log likelihood        :    -824.320
Sigma-square ML     :     144.833                Akaike info criterion :    1686.640
S.E of regression ML:     12.0347                Schwarz criterion     :    1750.325

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
                0_AC      12.3243111       4.9088481       2.5106320       0.0128770
               0_AGE       0.0403658       0.0689162       0.5857223       0.5587500
            0_FIREPL       7.7338436       4.2166122       1.8341368       0.0681815
               0_GAR       0.4047824       3.0179268       0.1341260       0.8934435
             0_LOTSZ       0.1632751       0.0463437       3.5231374       0.0005330
             0_NBATH       4.2363638       2.6085499       1.6240302       0.1060099
             0_NROOM       0.9197336       1.4019826       0.6560236       0.5125943
             0_PATIO      11.1983454       5.7047622       1.9629820       0.0510926
              0_SQFT      -0.1418060       0.2622738      -0.5406791       0.5893554
                1_AC       1.9738848       3.0657377       0.6438531       0.5204390
               1_AGE      -0.2104541       0.1188647      -1.7705350       0.0782249
```

```
          1_FIREPL      11.3850987       3.0703001       3.7081388       0.0002732
             1_GAR       8.9283971       2.2228329       4.0166749       0.0000846
           1_LOTSZ       0.0472730       0.0172956       2.7332401       0.0068575
           1_NBATH      13.8768417       2.5967453       5.3439364       0.0000003
           1_NROOM       2.2951164       1.3724545       1.6722714       0.0960996
           1_PATIO       6.6939663       3.2257527       2.0751641       0.0393043
            1_SQFT       0.1245301       0.2238656       0.5562716       0.5786729
  _Global_CONSTANT      10.1335935       5.1500216       1.9676798       0.0505449
------------------------------------------------------------------------------------
Regimes variable: CITCOU

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           22.573

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2         144.878           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test               18         147.991           0.0000
Koenker-Bassett test             18          51.428           0.0000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Moran's I (error)              0.0854          2.319           0.0204
Lagrange Multiplier (lag)         1           15.009           0.0001
Robust LM (lag)                   1           11.987           0.0005
Lagrange Multiplier (error)       1            3.437           0.0638
Robust LM (error)                 1            0.415           0.5195
Lagrange Multiplier (SARMA)       2           15.424           0.0004


REGIMES DIAGNOSTICS - CHOW TEST
                 VARIABLE        DF        VALUE           PROB
                       AC         1            3.205           0.0734
                      AGE         1            3.577           0.0586
                   FIREPL         1            0.493           0.4827
                      GAR         1            5.340           0.0208
                    LOTSZ         1            5.502           0.0190
                    NBATH         1            7.021           0.0081
                    NROOM         1            0.810           0.3681
                    PATIO         1            0.474           0.4911
                     SQFT         1            0.613           0.4336
              Global test         9           65.345           0.0000
================================ END OF REPORT =====================================
```

cols2regi – specifying variable specific regimes
set up the list with True for regimes, False for constant across regimes
follow the order in which the x array has been created
NROOM, NBATH, PATIO, FIREPL, AC, GAR, AGE, LOTSZ, SQFT
only NBATH, GAR and LOTSZ vary

```
In [21]: colsvari = [False,True,False,False,False,True,False,True,False]
```

must set constant_regi='one' to keep constant from varying across regimes - not included in cols2regi

with default regime_err_sep = True, k nearest neighbor weights

```
In [22]: reg5 = pysal.spreg.OLS_Regimes(y,x,regimes,w=w,spat_diag=True,moran=True,
         constant_regi='one',cols2regi=colsvari,
         name_y=y_name,name_x=x_names,
         name_regimes=rvar,name_w="baltim_k4",name_ds="baltim.dbf")
```

```
In [23]: print reg5.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES - REGIMES (Group-wise heteroskedasticity)
----------------------------------------------------------------------------------
Data set            :  baltim.dbf
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE            Number of Observations:         211
Mean dependent var  :     44.3072            Number of Variables   :          13
S.D. dependent var  :     23.6061            Degrees of Freedom    :         198
R-squared           :      0.7288
Adjusted R-squared  :      0.7123

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
               0_GAR      -0.5338910       2.7346384      -0.1952328       0.8454109
             0_LOTSZ       0.1442208       0.0388247       3.7146643       0.0002646
             0_NBATH       3.1935950       1.8607269       1.7163158       0.0876682
               1_GAR       8.8076244       2.1915833       4.0188408       0.0000831
             1_LOTSZ       0.0425182       0.0173812       2.4462143       0.0153096
             1_NBATH      14.3810934       2.1949150       6.5520047       0.0000000
    _Global_CONSTANT      10.9182891       4.8772065       2.2386358       0.0262928
          _Global_AC       6.1789275       2.4375839       2.5348573       0.0120226
         _Global_AGE      -0.0365910       0.0534059      -0.6851498       0.4940507
      _Global_FIREPL      10.1857847       2.4360365       4.1812940       0.0000435
       _Global_NROOM       1.3115815       1.1169156       1.1742888       0.2416898
       _Global_PATIO       8.6281564       2.8072460       3.0735305       0.0024136
        _Global_SQFT       0.0523376       0.1658904       0.3154953       0.7527179
------------------------------------------------------------------------------------
Regimes variable: CITCOU
Warning: Residuals treated as homoskedastic for the purpose of diagnostics.

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Moran's I (error)              0.0842        2.225           0.0261
Lagrange Multiplier (lag)         1         15.959           0.0001
Robust LM (lag)                   1         13.216           0.0003
Lagrange Multiplier (error)       1          3.343           0.0675
Robust LM (error)                 1          0.600           0.4386
Lagrange Multiplier (SARMA)       2         16.559           0.0003


REGIMES DIAGNOSTICS - CHOW TEST
                 VARIABLE        DF        VALUE           PROB
```

```
              GAR          1           7.364          0.0067
            LOTSZ          1           5.946          0.0147
            NBATH          1          37.809          0.0000
      Global test          3          59.755          0.0000
================================ END OF REPORT =====================================
```

with regime_err_sep = False (homoskedasticity), k nearest neighbors

```
In [24]: reg6 = pysal.spreg.OLS_Regimes(y,x,regimes,w=w,spat_diag=True,moran=True,
         constant_regi='one',cols2regi=colsvari,regime_err_sep=False,
         name_y=y_name,name_x=x_names,
         name_regimes=rvar,name_w="baltim_k4",name_ds="baltim.dbf")

In [25]: print reg6.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES - REGIMES
---------------------------------------------------
Data set            :  baltim.dbf
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE                Number of Observations:         211
Mean dependent var  :     44.3072                Number of Variables   :          13
S.D. dependent var  :     23.6061                Degrees of Freedom    :         198
R-squared           :      0.7288
Adjusted R-squared  :      0.7123
Sum squared residual:   31739.937                F-statistic           :     44.3338
Sigma-square        :     160.303                Prob(F-statistic)     :   1.532e-49
S.E. of regression  :      12.661                Log likelihood        :    -828.317
Sigma-square ML     :     150.426                Akaike info criterion :    1682.635
S.E of regression ML:     12.2648                Schwarz criterion     :    1726.209

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
               0_GAR      -0.6329806       2.9769198      -0.2126294       0.8318348
             0_LOTSZ       0.1416988       0.0421173       3.3643857       0.0009211
             0_NBATH       3.1414902       1.9529611       1.6085780       0.1093020
               1_GAR       8.7399064       2.1002731       4.1613191       0.0000472
             1_LOTSZ       0.0425588       0.0166214       2.5604851       0.0111978
             1_NBATH      14.2191980       2.1820910       6.5163178       0.0000000
    _Global_CONSTANT      11.0645165       5.0453798       2.1929997       0.0294723
          _Global_AC       5.6662011       2.4240823       2.3374623       0.0204140
         _Global_AGE      -0.0488391       0.0563876      -0.8661322       0.3874660
      _Global_FIREPL      10.5115191       2.4440989       4.3007749       0.0000267
       _Global_NROOM       1.3482450       1.1497597       1.1726319       0.2423523
       _Global_PATIO       8.3790398       2.7820285       3.0118454       0.0029353
        _Global_SQFT       0.0734034       0.1685032       0.4356203       0.6635868
------------------------------------------------------------------------------------
Regimes variable: CITCOU

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           22.081

TEST ON NORMALITY OF ERRORS
```

```
TEST                              DF        VALUE           PROB
Jarque-Bera                        2         179.995           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                              DF        VALUE           PROB
Breusch-Pagan test                12         162.258           0.0000
Koenker-Bassett test              12          51.857           0.0000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                            MI/DF       VALUE           PROB
Moran's I (error)               0.0840         2.220           0.0264
Lagrange Multiplier (lag)          1          15.791           0.0001
Robust LM (lag)                    1          13.047           0.0003
Lagrange Multiplier (error)        1           3.325           0.0682
Robust LM (error)                  1           0.581           0.4459
Lagrange Multiplier (SARMA)        2          16.372           0.0003


REGIMES DIAGNOSTICS - CHOW TEST
                 VARIABLE         DF        VALUE           PROB
                      GAR          1           6.859           0.0088
                    LOTSZ          1           4.962           0.0259
                    NBATH          1          35.148           0.0000
              Global test          3          57.519           0.0000
================================ END OF REPORT =====================================
```

default is constant varies across regimes

```
In [26]: reg7 = pysal.spreg.OLS_Regimes(y,x,regimes,w=w,spat_diag=True,moran=True,
         cols2regi=colsvari,
         name_y=y_name,name_x=x_names,
         name_regimes=rvar,name_w="baltim_k4",name_ds="baltim.dbf")

In [27]: print reg7.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES - REGIMES (Group-wise heteroskedasticity)
-----------------------------------------------------------------------------------
Data set            :  baltim.dbf
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE               Number of Observations:         211
Mean dependent var  :     44.3072               Number of Variables   :          14
S.D. dependent var  :     23.6061               Degrees of Freedom    :         197
R-squared           :      0.7289
Adjusted R-squared  :      0.7110

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
          0_CONSTANT      11.4001349       5.3053644       2.1487939       0.0328709
               0_GAR      -0.5037883       2.7523302      -0.1830406       0.8549544
             0_LOTSZ       0.1441176       0.0390178       3.6936367       0.0002863
             0_NBATH       2.9233414       2.1945768       1.3320753       0.1843744
```

```
         1_CONSTANT      10.2791300       5.6240112       1.8277222       0.0691040
              1_GAR       8.7912487       2.1942467       4.0064997       0.0000873
            1_LOTSZ       0.0428939       0.0174692       2.4553981       0.0149398
            1_NBATH      14.6915880       2.5853633       5.6826009       0.0000000
         _Global_AC       6.0650203       2.4776377       2.4479044       0.0152446
        _Global_AGE      -0.0400820       0.0552525      -0.7254330       0.4690473
     _Global_FIREPL      10.1648793       2.4460652       4.1556043       0.0000483
      _Global_NROOM       1.3433861       1.1284255       1.1904960       0.2352841
      _Global_PATIO       8.5584205       2.8253938       3.0291071       0.0027814
       _Global_SQFT       0.0543615       0.1664689       0.3265568       0.7443498
------------------------------------------------------------------------------------
Regimes variable: CITCOU
Warning: Residuals treated as homoskedastic for the purpose of diagnostics.

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Moran's I (error)              0.0829         2.223           0.0262
Lagrange Multiplier (lag)         1          16.023           0.0001
Robust LM (lag)                   1          13.440           0.0002
Lagrange Multiplier (error)       1           3.241           0.0718
Robust LM (error)                 1           0.658           0.4172
Lagrange Multiplier (SARMA)       2          16.681           0.0002


REGIMES DIAGNOSTICS - CHOW TEST
                 VARIABLE        DF        VALUE           PROB
                 CONSTANT         1           0.054           0.8160
                      GAR         1           7.216           0.0072
                    LOTSZ         1           5.830           0.0158
                    NBATH         1          14.378           0.0001
              Global test         4          59.478           0.0000
================================ END OF REPORT =====================================
```

0.3 Practice

Use the Boston example (see Chapter 5 notebook) with CHAS as the regime variable. Experiment with the different options. For example, using the results of the Chow test in the default setup for OLS regimes, let only those coefficient vary that are significant in the individual Chow tests.

In [ ]:

Chapter 12 - Regime OLS practice solutions

November 5, 2015

0.1 Solutions to Practice Example Chapter 12 (OLS only)

```
In [1]: %pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

```
In [2]: import numpy as np
        import pysal
```

```
In [3]: db = pysal.open('data/boston.dbf','r')
```

```
In [4]: len(db)
```

```
Out[4]: 506
```

```
In [5]: db.header
```

```
Out[5]: ['POLY_ID',
         'ID',
         'TOWN',
         'TOWNNO',
         'TRACT',
         'LON',
         'LAT',
         'x',
         'y',
         'MEDV',
         'CMEDV',
         'CRIM',
         'ZN',
         'INDUS',
         'CHAS',
         'NOX',
         'RM',
         'AGE',
         'DIS',
         'RAD',
         'TAX',
         'PTRATIO',
         'B',
         'LSTAT']
```

```
In [6]: y_name = 'MEDV'
        y = np.array([db.by_col(y_name)]).T
```

```
In [7]: x_names = ['CRIM','NOX','RM','AGE','DIS','LSTAT']
        x = np.array([db.by_col(var) for var in x_names]).T

In [8]: rvar = "CHAS"
        regimes = db.by_col(rvar)

In [9]: w = pysal.knnW_from_shapefile('data/boston.shp',
                                        k=4,idVariable='POLY_ID')
        w.transform = 'r'
```

Default regime regression

```
In [10]: reg1 = pysal.spreg.OLS_Regimes(y,x,regimes,w=w,spat_diag=True,moran=True,
         name_y=y_name,name_x=x_names,name_regimes=rvar,name_w="boston_k4",name_ds="boston.dbf")

In [11]: print reg1.summary
```

```
REGRESSION
----------

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION - REGIME 0
---------------------------------------------------------------
Data set            :  boston.dbf
Weights matrix      :   boston_k4
Dependent Variable  :      0_MEDV              Number of Observations:         471
Mean dependent var  :     22.0938              Number of Variables   :           7
S.D. dependent var  :      8.8314              Degrees of Freedom    :         464
R-squared           :      0.6815
Adjusted R-squared  :      0.6774
Sum squared residual:   11673.935              F-statistic           :    165.4969
Sigma-square        :      25.159              Prob(F-statistic)     :  6.533e-112
S.E. of regression  :       5.016              Log likelihood        :   -1424.335
Sigma-square ML     :      24.785              Akaike info criterion :    2862.671
S.E of regression ML:      4.9785              Schwarz criterion     :    2891.755

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
          0_CONSTANT       8.1040354       3.8968336       2.0796462       0.0381065
               0_AGE      -0.0257461       0.0140762      -1.8290484       0.0680340
              0_CRIM      -0.1237989       0.0303482      -4.0792867       0.0000532
               0_DIS      -0.9839729       0.1877386      -5.2411850       0.0000002
             0_LSTAT      -0.5310903       0.0547369      -9.7026084       0.0000000
               0_NOX     -10.3946003       3.7540335      -2.7689151       0.0058496
                0_RM       5.1863159       0.4323591      11.9953890       0.0000000
------------------------------------------------------------------------------------
Regimes variable: CHAS
Warning: The regimes operation resulted in islands for regime 0.

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           47.618

TEST ON NORMALITY OF ERRORS
TEST                             DF         VALUE           PROB
Jarque-Bera                       2          546.616           0.0000
```

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	6	64.772	0.0000
Koenker-Bassett test	6	20.345	0.0024

DIAGNOSTICS FOR SPATIAL DEPENDENCE

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.4883	16.237	0.0000
Lagrange Multiplier (lag)	1	105.840	0.0000
Robust LM (lag)	1	0.028	0.8669
Lagrange Multiplier (error)	1	244.718	0.0000
Robust LM (error)	1	138.906	0.0000
Lagrange Multiplier (SARMA)	2	244.746	0.0000

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION - REGIME 1

Data set	boston.dbf		
Weights matrix	boston_k4		
Dependent Variable	1_MEDV	Number of Observations	35
Mean dependent var	28.4400	Number of Variables	7
S.D. dependent var	11.8166	Degrees of Freedom	28
R-squared	0.7414		
Adjusted R-squared	0.6860		
Sum squared residual	1227.556	F-statistic	13.3815
Sigma-square	43.841	Prob(F-statistic)	4.126e-07
S.E. of regression	6.621	Log likelihood	-111.918
Sigma-square ML	35.073	Akaike info criterion	237.836
S.E of regression ML	5.9222	Schwarz criterion	248.723

Variable	Coefficient	Std.Error	t-Statistic	Probability
1_CONSTANT	57.0766870	21.4502496	2.6608868	0.0127561
1_AGE	0.0492828	0.1002633	0.4915339	0.6268801
1_CRIM	1.6986344	0.6470628	2.6251462	0.0138749
1_DIS	-1.5463540	2.4097200	-0.6417152	0.5262771
1_LSTAT	-1.0254282	0.3165284	-3.2396094	0.0030797
1_NOX	-48.4520409	14.6412959	-3.3092727	0.0025786
1_RM	1.4362819	2.1043742	0.6825221	0.5005170

Regimes variable: CHAS
Warning: The regimes operation resulted in islands for regime 1.

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 52.476

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	0.056	0.9726

```
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                6           17.560           0.0074
Koenker-Bassett test              6           19.414           0.0035

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Moran's I (error)              0.2619         2.304           0.0212
Lagrange Multiplier (lag)         1          12.889           0.0003
Robust LM (lag)                   1          11.139           0.0008
Lagrange Multiplier (error)       1           2.024           0.1548
Robust LM (error)                 1           0.274           0.6008
Lagrange Multiplier (SARMA)       2          13.163           0.0014


REGIMES DIAGNOSTICS - CHOW TEST
                 VARIABLE        DF        VALUE           PROB
                 CONSTANT         1           5.046           0.0247
                      AGE         1           0.549           0.4587
                     CRIM         1           7.915           0.0049
                      DIS         1           0.054           0.8160
                    LSTAT         1           2.368           0.1238
                      NOX         1           6.340           0.0118
                       RM         1           3.047           0.0809
              Global test         7          30.223           0.0001

DIAGNOSTICS FOR GLOBAL SPATIAL DEPENDENCE
Residuals are treated as homoskedastic for the purpose of these tests
TEST                           MI/DF       VALUE           PROB
Moran's I (error)              0.4478        16.208           0.0000
Lagrange Multiplier (lag)         1         168.223           0.0000
Robust LM (lag)                   1          10.157           0.0014
Lagrange Multiplier (error)       1         233.376           0.0000
Robust LM (error)                 1          75.311           0.0000
Lagrange Multiplier (SARMA)       2         243.533           0.0000

================================ END OF REPORT =====================================
```

Keeping AGE, DIS, LSTAT and RM constant and letting constant, CRIM and NOX vary order of variables in x is CRIM, NOX, RM, AGE, DIS, LSTAT

```
In [12]: colsvari = [True,True,False,False,False,False]
```

```
In [13]: reg2 = pysal.spreg.OLS_Regimes(y,x,regimes,w=w,spat_diag=True,moran=True,
         cols2regi=colsvari,
         name_y=y_name,name_x=x_names,
         name_regimes=rvar,name_w="boston_k4",name_ds="boston.dbf")
```

```
In [14]: print reg2.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES - REGIMES (Group-wise heteroskedasticity)
-----------------------------------------------------------------------------------
```

```
Data set            :  boston.dbf
Weights matrix      :   boston_k4
Dependent Variable  :        MEDV                Number of Observations:         506
Mean dependent var  :     22.5328                Number of Variables   :          10
S.D. dependent var  :      9.1971                Degrees of Freedom    :         496
R-squared           :      0.6932
Adjusted R-squared  :      0.6876

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
          0_CONSTANT       9.1431225       3.8154187       2.3963615       0.0169281
              0_CRIM      -0.1227281       0.0301084      -4.0762043       0.0000533
               0_NOX     -10.7282290       3.7046976      -2.8958447       0.0039483
          1_CONSTANT      27.0896561       7.1209694       3.8042091       0.0001600
              1_CRIM       2.0303936       0.6580158       3.0856305       0.0021447
               1_NOX     -41.1760368      11.3959164      -3.6132273       0.0003333
         _Global_AGE      -0.0234270       0.0138239      -1.6946805       0.0907637
         _Global_DIS      -0.9944373       0.1860103      -5.3461400       0.0000001
       _Global_LSTAT      -0.5445706       0.0536013     -10.1596519       0.0000000
          _Global_RM       5.0579970       0.4212336      12.0075822       0.0000000
------------------------------------------------------------------------------------
Regimes variable: CHAS
Warning: Residuals treated as homoskedastic for the purpose of diagnostics.

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Moran's I (error)              0.4475      15.935           0.0000
Lagrange Multiplier (lag)         1       162.290           0.0000
Robust LM (lag)                   1         8.339           0.0039
Lagrange Multiplier (error)       1       233.065           0.0000
Robust LM (error)                 1        79.114           0.0000
Lagrange Multiplier (SARMA)       2       241.404           0.0000


REGIMES DIAGNOSTICS - CHOW TEST
                 VARIABLE        DF        VALUE           PROB
                 CONSTANT         1         7.938           0.0048
                     CRIM         1        10.672           0.0011
                      NOX         1         6.806           0.0091
              Global test         3        20.205           0.0002
================================ END OF REPORT =====================================
```

In [ ]:

In [ ]:

Chapter 13 - Regimes, Spatial (Lag only)

November 5, 2015

This notebook contains the PySAL/spreg code for Chapter 13 - Regimes, Spatial (Lag only) in Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL. by Luc Anselin and Sergio J. Rey


```
In [1]: __author__ = "Luc Anselin luc.anselin@asu.edu"
```

0.1 Regimes, Spatial - Spatial Lag

0.1.1 Baltimore Example

Basic Setup:

- import necessary modules (numpy and pysal)
- create a data object
- create variables as numpy arrays
- create regime variable (as list)
- create weights object(s) for diagnostics

```
In [1]: %pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

```
In [2]: import numpy as np
        import pysal
```

create data object

```
In [3]: db = pysal.open('data/baltim.dbf','r')
```

read in dependent variable and turn into numpy array y

```
In [4]: y_name = "PRICE"
        y = np.array([db.by_col(y_name)]).T
```

read in explanatory variables and turn into numpy array x

```
In [5]: x_names = ['NROOM','NBATH','PATIO','FIREPL','AC','GAR','AGE','LOTSZ','SQFT']
        x = np.array([db.by_col(var) for var in x_names]).T
```

create k = 4 nearest neighbor weights and row-standardize

```
In [6]: w = pysal.knnW_from_shapefile("data/baltim.shp",k=4,idVariable='STATION')
        w.transform = 'r'
```

creating a regimes variable

```
In [7]: rvar = "CITCOU"
        regimes = db.by_col(rvar)    # note: regimes is a list
```

0.2 Spatial Lag Regimes (IV)

0.2.1 Default setup (with spatial diagnostics)

regime_lag_sep = False and regime_err_sep = True

one spatial lag coefficient and heteroskedasticity (White standard errors)

```
In [8]: reg1 = pysal.spreg.GM_Lag_Regimes(y,x,regimes,w=w,spat_diag=True,name_y=y_name,
        name_x=x_names,name_regimes=rvar,name_w="baltim_k4",name_ds="baltim.dbf")
```

```
In [9]: print reg1.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES - REGIMES
------------------------------------------------------------
Data set            :  baltim.dbf
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE              Number of Observations:         211
Mean dependent var  :     44.3072              Number of Variables   :          21
S.D. dependent var  :     23.6061              Degrees of Freedom    :         190
Pseudo R-squared    :      0.7609
Spatial Pseudo R-squared:  0.7511

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
          0_CONSTANT      -7.9106111      11.1708343      -0.7081486       0.4788530
                0_AC      11.4234750       2.8020744       4.0767921       0.0000457
               0_AGE       0.0816839       0.0894822       0.9128510       0.3613209
            0_FIREPL       4.2998763       4.3728174       0.9833194       0.3254502
               0_GAR       0.0476798       2.8106893       0.0169637       0.9864655
             0_LOTSZ       0.1097621       0.0494611       2.2191611       0.0264758
             0_NBATH       4.3881697       2.3185401       1.8926434       0.0584053
             0_NROOM       2.3011044       2.2304548       1.0316750       0.3022244
             0_PATIO      11.7423138       5.2046237       2.2561312       0.0240624
              0_SQFT      -0.1562391       0.3466908      -0.4506583       0.6522359
          1_CONSTANT       4.5831279       8.3669344       0.5477667       0.5838521
                1_AC       2.0822932       3.2396167       0.6427591       0.5203804
               1_AGE      -0.1704200       0.1395672      -1.2210611       0.2220629
            1_FIREPL       8.8750302       2.8130485       3.1549510       0.0016053
               1_GAR       6.9821813       3.0976615       2.2540169       0.0241951
             1_LOTSZ       0.0416225       0.0260264       1.5992443       0.1097663
             1_NBATH      11.6968197       3.1909777       3.6655912       0.0002468
             1_NROOM       0.9591934       1.5702331       0.6108605       0.5412919
             1_PATIO       4.4690966       3.4572318       1.2926806       0.1961215
              1_SQFT       0.1057631       0.2267528       0.4664246       0.6409116
     _Global_W_PRICE       0.3462294       0.0972955       3.5585333       0.0003729
------------------------------------------------------------------------------------
Instrumented: _Global_W_PRICE
Instruments: 0_W_AC, 0_W_AGE, 0_W_FIREPL, 0_W_GAR, 0_W_LOTSZ, 0_W_NBATH,
             0_W_NROOM, 0_W_PATIO, 0_W_SQFT, 1_W_AC, 1_W_AGE, 1_W_FIREPL,
             1_W_GAR, 1_W_LOTSZ, 1_W_NBATH, 1_W_NROOM, 1_W_PATIO, 1_W_SQFT
Regimes variable: CITCOU
```

```
DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test             1          0.990           0.3197

REGIMES DIAGNOSTICS - CHOW TEST
                 VARIABLE        DF        VALUE           PROB
                 CONSTANT         1          0.911            0.3398
                       AC         1          4.977            0.0257
                      AGE         1          2.324            0.1274
                   FIREPL         1          0.743            0.3886
                      GAR         1          2.741            0.0978
                    LOTSZ         1          1.684            0.1944
                    NBATH         1          3.560            0.0592
                    NROOM         1          0.249            0.6176
                    PATIO         1          1.375            0.2410
                     SQFT         1          0.397            0.5289
              Global test        10         38.732            0.0000
================================ END OF REPORT =====================================
```

0.2.2 Constant Lag Coefficient, Homoskedasticity

regime_lag_sep = False and regime_err_sep = False

```
In [10]: reg2 = pysal.spreg.GM_Lag_Regimes(y,x,regimes,w=w,spat_diag=True,
         regime_err_sep=False,name_y=y_name,
         name_x=x_names,name_regimes=rvar,name_w="baltim_k4",name_ds="baltim.dbf")
```

```
In [11]: print reg2.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES - REGIMES
------------------------------------------------------------
Data set            :  baltim.dbf
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE               Number of Observations:         211
Mean dependent var  :     44.3072               Number of Variables   :          21
S.D. dependent var  :     23.6061               Degrees of Freedom    :         190
Pseudo R-squared    :      0.7609
Spatial Pseudo R-squared:  0.7511

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
          0_CONSTANT      -7.9106111       7.1357915      -1.1085822       0.2676105
                0_AC      11.4234750       4.5180780       2.5283926       0.0114586
               0_AGE       0.0816839       0.0653582       1.2497887       0.2113768
            0_FIREPL       4.2998763       3.9468124       1.0894555       0.2759531
               0_GAR       0.0476798       2.8256726       0.0168738       0.9865373
             0_LOTSZ       0.1097621       0.0437727       2.5075473       0.0121572
             0_NBATH       4.3881697       2.4002213       1.8282354       0.0675142
             0_NROOM       2.3011044       1.4860482       1.5484722       0.1215086
             0_PATIO      11.7423138       5.2152186       2.2515478       0.0243509
              0_SQFT      -0.1562391       0.2430390      -0.6428561       0.5203175
          1_CONSTANT       4.5831279       7.2825586       0.6293293       0.5291335
```

```
             1_AC       2.0822932       2.7993648       0.7438449       0.4569703
            1_AGE      -0.1704200       0.1140699      -1.4939970       0.1351764
         1_FIREPL       8.8750302       2.8565945       3.1068569       0.0018909
            1_GAR       6.9821813       2.0969242       3.3297252       0.0008693
          1_LOTSZ       0.0416225       0.0158369       2.6281958       0.0085839
          1_NBATH      11.6968197       2.4448284       4.7843111       0.0000017
          1_NROOM       0.9591934       1.5392131       0.6231713       0.5331720
          1_PATIO       4.4690966       2.9981975       1.4905945       0.1360680
           1_SQFT       0.1057631       0.2070209       0.5108812       0.6094343
    _Global_W_PRICE       0.3462294       0.0727891       4.7566105       0.0000020
------------------------------------------------------------------------------------
Instrumented: _Global_W_PRICE
Instruments: 0_W_AC, 0_W_AGE, 0_W_FIREPL, 0_W_GAR, 0_W_LOTSZ, 0_W_NBATH,
             0_W_NROOM, 0_W_PATIO, 0_W_SQFT, 1_W_AC, 1_W_AGE, 1_W_FIREPL,
             1_W_GAR, 1_W_LOTSZ, 1_W_NBATH, 1_W_NROOM, 1_W_PATIO, 1_W_SQFT
Regimes variable: CITCOU

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test             1           0.990           0.3197

REGIMES DIAGNOSTICS - CHOW TEST
                 VARIABLE        DF        VALUE           PROB
                 CONSTANT         1           1.687            0.1939
                       AC         1           3.087            0.0789
                      AGE         1           3.713            0.0540
                   FIREPL         1           0.909            0.3403
                      GAR         1           3.887            0.0487
                    LOTSZ         1           2.170            0.1408
                    NBATH         1           4.550            0.0329
                    NROOM         1           0.386            0.5343
                    PATIO         1           1.456            0.2275
                     SQFT         1           0.673            0.4119
              Global test        10          42.108            0.0000
================================ END OF REPORT =====================================
```

0.2.3 Different Lag Coefficient - Heteroskedasticity

regime_lag_sep = True and regime_err_sep = True

```
In [12]: reg3 = pysal.spreg.GM_Lag_Regimes(y,x,regimes,w=w,spat_diag=True,cores=False,
         regime_lag_sep=True,regime_err_sep=True,name_y=y_name,
         name_x=x_names,name_regimes=rvar,name_w="baltim_k4",name_ds="baltim.dbf")

In [13]: print reg3.summary
```

```
REGRESSION
----------

SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES ESTIMATION - REGIME 0
------------------------------------------------------------------------
Data set            :  baltim.dbf
Weights matrix      :   baltim_k4
Dependent Variable  :     0_PRICE                Number of Observations:          83
Mean dependent var  :     31.5127                Number of Variables   :          11
```

```
S.D. dependent var  :     17.1598             Degrees of Freedom    :          72
Pseudo R-squared    :      0.6074
Spatial Pseudo R-squared:  0.6097

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
          0_CONSTANT       3.5115654       8.0774135       0.4347388       0.6637520
                0_AC      12.1021336       4.2247793       2.8645599       0.0041759
               0_AGE       0.0536825       0.0606711       0.8848118       0.3762581
            0_FIREPL       6.7001786       3.7261107       1.7981695       0.0721502
               0_GAR      -0.0778926       2.6314990      -0.0296001       0.9763860
             0_LOTSZ       0.1466405       0.0436008       3.3632565       0.0007703
             0_NBATH       4.5436972       2.2361256       2.0319508       0.0421586
             0_NROOM       1.5804082       1.4165268       1.1156924       0.2645538
             0_PATIO      11.2170868       4.8404878       2.3173464       0.0204849
              0_SQFT      -0.1815703       0.2268049      -0.8005573       0.4233880
           0_W_PRICE       0.1158306       0.1394951       0.8303560       0.4063376
------------------------------------------------------------------------------------
Instrumented: 0_W_PRICE
Instruments: 0_W_AC, 0_W_AGE, 0_W_FIREPL, 0_W_GAR, 0_W_LOTSZ, 0_W_NBATH,
             0_W_NROOM, 0_W_PATIO, 0_W_SQFT
Regimes variable: CITCOU

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test            1           0.473           0.4916
----------

SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES ESTIMATION - REGIME 1
----------------------------------------------------------------------
Data set            :  baltim.dbf
Weights matrix      :   baltim_k4
Dependent Variable  :     1_PRICE             Number of Observations:         128
Mean dependent var  :     52.6036             Number of Variables   :          11
S.D. dependent var  :     23.5574             Degrees of Freedom    :         117
Pseudo R-squared    :      0.7521
Spatial Pseudo R-squared:  0.7272

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
          1_CONSTANT       4.7259639       7.4124347       0.6375724       0.5237521
                1_AC       1.1979943       2.8447583       0.4211234       0.6736650
               1_AGE      -0.2141868       0.1151314      -1.8603687       0.0628334
            1_FIREPL       8.3104564       2.9354673       2.8310506       0.0046395
               1_GAR       7.3539573       2.1171974       3.4734396       0.0005138
             1_LOTSZ       0.0444901       0.0160317       2.7751414       0.0055178
             1_NBATH      11.0543992       2.5172425       4.3914716       0.0000113
             1_NROOM       1.2158803       1.5564972       0.7811645       0.4347057
             1_PATIO       3.8386833       3.0773523       1.2473981       0.2122516
              1_SQFT       0.1181056       0.2099339       0.5625846       0.5737178
           1_W_PRICE       0.3460901       0.0797089       4.3419243       0.0000141
------------------------------------------------------------------------------------
```

```
Instrumented: 1_W_PRICE
Instruments: 1_W_AC, 1_W_AGE, 1_W_FIREPL, 1_W_GAR, 1_W_LOTSZ, 1_W_NBATH,
             1_W_NROOM, 1_W_PATIO, 1_W_SQFT
Regimes variable: CITCOU
Warning: The regimes operation resulted in islands for regime 1.

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test             1           0.042          0.8375

REGIMES DIAGNOSTICS - CHOW TEST
                 VARIABLE        DF        VALUE           PROB
                 CONSTANT         1           0.012           0.9118
                       AC         1           4.583           0.0323
                      AGE         1           4.237           0.0396
                   FIREPL         1           0.115           0.7343
                      GAR         1           4.842           0.0278
                    LOTSZ         1           4.835           0.0279
                    NBATH         1           3.739           0.0532
                    NROOM         1           0.030           0.8625
                    PATIO         1           1.655           0.1983
                     SQFT         1           0.940           0.3322
                  W_PRICE         1           2.054           0.1518
              Global test        11          41.450           0.0000
================================ END OF REPORT =====================================
```

0.3 Hybrid models

0.3.1 One Global Constant

constant_regi='one'

```
In [14]: reg5 = pysal.spreg.GM_Lag_Regimes(y,x,regimes,w=w,spat_diag=True,
         constant_regi='one',cores=False,name_y=y_name,
         name_x=x_names,name_regimes=rvar,name_w="baltim_k4",name_ds="baltim.dbf")

In [15]: print reg5.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES - REGIMES
------------------------------------------------------------
Data set            :  baltim.dbf
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE              Number of Observations:         211
Mean dependent var  :     44.3072              Number of Variables   :          20
S.D. dependent var  :     23.6061              Degrees of Freedom    :         191
Pseudo R-squared    :      0.7595
Spatial Pseudo R-squared:  0.7517

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
                0_AC      10.8305600       2.6406091       4.1015386       0.0000410
```

```
             0_AGE       0.0590564       0.0879669       0.6713475       0.5019992
          0_FIREPL       5.2053119       4.1505682       1.2541203       0.2097983
             0_GAR       0.8472330       2.4770299       0.3420358       0.7323239
           0_LOTSZ       0.1155147       0.0474209       2.4359445       0.0148530
           0_NBATH       4.0065534       2.1938432       1.8262715       0.0678094
           0_NROOM       1.3214078       1.6595792       0.7962306       0.4258980
           0_PATIO      12.0066771       5.1942737       2.3115218       0.0208040
            0_SQFT      -0.1036572       0.3555538      -0.2915372       0.7706405
              1_AC       2.1424985       3.2823072       0.6527416       0.5139229
             1_AGE      -0.1317404       0.1375002      -0.9581109       0.3380068
          1_FIREPL       8.8840193       2.8091612       3.1625168       0.0015641
             1_GAR       6.7344280       3.0925902       2.1776012       0.0294357
           1_LOTSZ       0.0416920       0.0259131       1.6089145       0.1076350
           1_NBATH      12.3620424       3.1066516       3.9792176       0.0000691
           1_NROOM       2.1219371       1.5752501       1.3470478       0.1779649
           1_PATIO       4.2694651       3.4803283       1.2267420       0.2199196
            1_SQFT       0.0669139       0.2241789       0.2984843       0.7653335
  _Global_CONSTANT      -1.6461745       7.3712409      -0.2233239       0.8232834
    _Global_W_PRICE       0.3252752       0.1002870       3.2434437       0.0011809
------------------------------------------------------------------------------------
Instrumented: _Global_W_PRICE
Instruments: 0_W_AC, 0_W_AGE, 0_W_FIREPL, 0_W_GAR, 0_W_LOTSZ, 0_W_NBATH,
             0_W_NROOM, 0_W_PATIO, 0_W_SQFT, 1_W_AC, 1_W_AGE, 1_W_FIREPL,
             1_W_GAR, 1_W_LOTSZ, 1_W_NBATH, 1_W_NROOM, 1_W_PATIO, 1_W_SQFT
Regimes variable: CITCOU

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test             1           1.031           0.3098

REGIMES DIAGNOSTICS - CHOW TEST
                 VARIABLE        DF        VALUE           PROB
                       AC         1           4.476           0.0344
                      AGE         1           1.560           0.2117
                   FIREPL         1           0.516           0.4726
                      GAR         1           2.265           0.1323
                    LOTSZ         1           2.155           0.1421
                    NBATH         1           5.676           0.0172
                    NROOM         1           0.248           0.6183
                    PATIO         1           1.553           0.2127
                     SQFT         1           0.167           0.6828
              Global test         9          33.636           0.0001
================================ END OF REPORT =====================================
```

0.3.2 Fixed and Varying Coefficients

Hybrid models with cols2regi

set up the list with True for regimes, False for constant across regimes
follow the order in which the x array has been created
NROOM, NBATH, PATIO, FIREPL, AC, GAR, AGE, LOTSZ, SQFT
only NBATH and PATIO vary

```
In [16]: colsvari = [False,True,True,False,False,False,False,False,False]
```

```
In [17]: reg6 = pysal.spreg.GM_Lag_Regimes(y,x,regimes,w=w,spat_diag=True,
```

```
        constant_regi='one',cols2regi=colsvari,name_y=y_name,
        name_x=x_names,name_regimes=rvar,name_w="baltim_k4",name_ds="baltim.dbf")
```

```
In [18]: print reg6.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES - REGIMES
------------------------------------------------------------
Data set            :  baltim.dbf
Weights matrix      :   baltim_k4
Dependent Variable  :       PRICE               Number of Observations:         211
Mean dependent var  :     44.3072               Number of Variables   :          13
S.D. dependent var  :     23.6061               Degrees of Freedom    :         198
Pseudo R-squared    :      0.7459
Spatial Pseudo R-squared:  0.7356

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
             0_NBATH       4.1276399       1.7139905       2.4082047       0.0160312
             0_PATIO      12.9459029       4.6526505       2.7824791       0.0053945
             1_NBATH      11.4479575       2.5909710       4.4184044       0.0000099
             1_PATIO       4.4631096       3.6356118       1.2276090       0.2195938
    _Global_CONSTANT      -2.6226230       6.6995390      -0.3914632       0.6954549
          _Global_AC       5.1196107       2.4862269       2.0591888       0.0394762
         _Global_AGE       0.0054693       0.0895005       0.0611088       0.9512725
      _Global_FIREPL       7.8224574       2.3564903       3.3195373       0.0009017
         _Global_GAR       4.1538938       2.2214006       1.8699436       0.0614917
       _Global_LOTSZ       0.0463523       0.0248417       1.8659085       0.0620542
       _Global_NROOM       1.2195542       1.3069629       0.9331208       0.3507576
        _Global_SQFT       0.0586837       0.2183021       0.2688189       0.7880690
     _Global_W_PRICE       0.4021719       0.1120659       3.5887093       0.0003323
------------------------------------------------------------------------------------
Instrumented: _Global_W_PRICE
Instruments: 0_W_AC, 0_W_AGE, 0_W_FIREPL, 0_W_GAR, 0_W_LOTSZ, 0_W_NBATH,
             0_W_NROOM, 0_W_PATIO, 0_W_SQFT, 1_W_AC, 1_W_AGE, 1_W_FIREPL,
             1_W_GAR, 1_W_LOTSZ, 1_W_NBATH, 1_W_NROOM, 1_W_PATIO, 1_W_SQFT
Regimes variable: CITCOU

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test             1           2.647          0.1038

REGIMES DIAGNOSTICS - CHOW TEST
                 VARIABLE        DF        VALUE           PROB
                    NBATH         1          22.672           0.0000
                    PATIO         1           2.192           0.1387
              Global test         2          23.284           0.0000
================================ END OF REPORT =====================================
```

0.4 Practice

As for Chapter 12, we will use the Boston example with CHAS as the regime variable. Experiment with the different options for fixed and varying spatial lag coefficient, and with hybrid specifications (some coefficients fixed, some varying).

```
In [ ]:
```

Chapter 13 - Regime Spatial Lag practice solutions

November 5, 2015

0.1 Solutions to Practice Example Chapter 13 (Spatial Lag only)

```
In [1]: %pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

```
In [2]: import numpy as np
        import pysal
```

```
In [3]: db = pysal.open('data/boston.dbf','r')
```

```
In [4]: len(db)
```

```
Out[4]: 506
```

```
In [5]: db.header
```

```
Out[5]: ['POLY_ID',
         'ID',
         'TOWN',
         'TOWNNO',
         'TRACT',
         'LON',
         'LAT',
         'x',
         'y',
         'MEDV',
         'CMEDV',
         'CRIM',
         'ZN',
         'INDUS',
         'CHAS',
         'NOX',
         'RM',
         'AGE',
         'DIS',
         'RAD',
         'TAX',
         'PTRATIO',
         'B',
         'LSTAT']
```

```
In [6]: y_name = 'MEDV'
        y = np.array([db.by_col(y_name)]).T
```

```
In [7]: x_names = ['CRIM','NOX','RM','AGE','DIS','LSTAT']
        x = np.array([db.by_col(var) for var in x_names]).T
```

```
In [8]: rvar = "CHAS"
        regimes = db.by_col(rvar)
```

```
In [9]: w = pysal.knnW_from_shapefile('data/boston.shp',
                                        k=4,idVariable='POLY_ID')
        w.transform = 'r'
```

Default regime regression - one spatial lag coefficient

```
In [10]: reg1 = pysal.spreg.GM_Lag_Regimes(y,x,regimes,w=w,spat_diag=True,name_y=y_name,
         name_x=x_names,name_regimes=rvar,name_w="boston_k4",name_ds="boston.dbf")
```

```
In [11]: print reg1.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES - REGIMES
------------------------------------------------------------
Data set            :  boston.dbf
Weights matrix      :   boston_k4
Dependent Variable  :        MEDV                Number of Observations:         506
Mean dependent var  :     22.5328                Number of Variables   :          15
S.D. dependent var  :      9.1971                Degrees of Freedom    :         491
Pseudo R-squared    :      0.7509
Spatial Pseudo R-squared:  0.7019

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
          0_CONSTANT       1.9971534       6.4730209       0.3085350       0.7576753
               0_AGE      -0.0257038       0.0148801      -1.7273909       0.0840975
              0_CRIM      -0.1049741       0.0255258      -4.1124706       0.0000391
               0_DIS      -0.8593821       0.1704692      -5.0412764       0.0000005
             0_LSTAT      -0.4388405       0.1159908      -3.7834062       0.0001547
               0_NOX      -7.3824231       3.5670211      -2.0696326       0.0384868
                0_RM       4.9734098       0.8047620       6.1799759       0.0000000
          1_CONSTANT      52.7381380      22.4111206       2.3532129       0.0186120
               1_AGE       0.0327541       0.0759571       0.4312184       0.6663095
              1_CRIM       1.4064698       0.7682942       1.8306396       0.0671544
               1_DIS      -1.6276184       2.2266950      -0.7309570       0.4648054
             1_LSTAT      -0.9440847       0.3188578      -2.9608333       0.0030681
               1_NOX     -43.2094420      10.6978217      -4.0390879       0.0000537
                1_RM       1.0185629       2.7944265       0.3644980       0.7154861
       _Global_W_MEDV      0.1878789       0.0893492       2.1027493       0.0354877
------------------------------------------------------------------------------------
Instrumented: _Global_W_MEDV
Instruments: 0_W_AGE, 0_W_CRIM, 0_W_DIS, 0_W_LSTAT, 0_W_NOX, 0_W_RM,
             1_W_AGE, 1_W_CRIM, 1_W_DIS, 1_W_LSTAT, 1_W_NOX, 1_W_RM
Regimes variable: CHAS

DIAGNOSTICS FOR SPATIAL DEPENDENCE
```

```
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test            1          29.464           0.0000

REGIMES DIAGNOSTICS - CHOW TEST
                 VARIABLE        DF        VALUE           PROB
                 CONSTANT         1          4.753            0.0292
                      AGE         1          0.575            0.4481
                     CRIM         1          3.860            0.0494
                      DIS         1          0.119            0.7299
                    LSTAT         1          2.231            0.1353
                      NOX         1         10.586            0.0011
                       RM         1          1.883            0.1700
              Global test         7         25.091            0.0007
================================ END OF REPORT =====================================
```

Different spatial lag coefficient by regime

```
In [12]: reg2 = pysal.spreg.GM_Lag_Regimes(y,x,regimes,w=w,spat_diag=True,cores=False,
         regime_lag_sep=True,regime_err_sep=True,name_y=y_name,
         name_x=x_names,name_regimes=rvar,name_w="boston_k4",name_ds="boston.dbf")

In [13]: print reg2.summary
```

```
REGRESSION
----------

SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES ESTIMATION - REGIME 0
------------------------------------------------------------------------
Data set            :  boston.dbf
Weights matrix      :   boston_k4
Dependent Variable  :      0_MEDV                Number of Observations:         471
Mean dependent var  :     22.0938                Number of Variables   :           8
S.D. dependent var  :      8.8314                Degrees of Freedom    :         463
Pseudo R-squared    :      0.6753
Spatial Pseudo R-squared:  0.6815

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
          0_CONSTANT       8.6803461       4.3732000       1.9848958       0.0471561
               0_AGE      -0.0258782       0.0141152      -1.8333562       0.0667496
              0_CRIM      -0.1259612       0.0312996      -4.0243655       0.0000571
               0_DIS      -0.9955440       0.1922645      -5.1779932       0.0000002
             0_LSTAT      -0.5404006       0.0634028      -8.5232917       0.0000000
               0_NOX     -10.6753594       3.8826769      -2.7494844       0.0059689
                0_RM       5.2162763       0.4452429      11.7155742       0.0000000
            0_W_MEDV      -0.0200673       0.0685075      -0.2929213       0.7695823
------------------------------------------------------------------------------------
Instrumented: 0_W_MEDV
Instruments: 0_W_AGE, 0_W_CRIM, 0_W_DIS, 0_W_LSTAT, 0_W_NOX, 0_W_RM
Regimes variable: CHAS

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test            1          46.984           0.0000
```

```
----------

SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES ESTIMATION - REGIME 1
------------------------------------------------------------------------
Data set            :  boston.dbf
Weights matrix      :   boston_k4
Dependent Variable  :      1_MEDV                Number of Observations:          35
Mean dependent var  :     28.4400                Number of Variables   :           8
S.D. dependent var  :     11.8166                Degrees of Freedom    :          27
Pseudo R-squared    :      0.8519
Spatial Pseudo R-squared:  0.8103

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
          1_CONSTANT      44.8287190      14.7772808       3.0336244       0.0024164
               1_AGE      -0.0074283       0.0690503      -0.1075780       0.9143305
              1_CRIM       0.2228650       0.5476273       0.4069647       0.6840339
               1_DIS      -2.4169243       1.6430422      -1.4710056       0.1412896
             1_LSTAT      -0.7484767       0.2230083      -3.3562731       0.0007900
               1_NOX     -29.1037026      10.8087813      -2.6925980       0.0070898
                1_RM       1.2008034       1.4257911       0.8422015       0.3996752
              1_W_MEDV       0.3584167       0.0797971       4.4915983       0.0000071
------------------------------------------------------------------------------------
Instrumented: 1_W_MEDV
Instruments: 1_W_AGE, 1_W_CRIM, 1_W_DIS, 1_W_LSTAT, 1_W_NOX, 1_W_RM
Regimes variable: CHAS
Warning: The regimes operation resulted in islands for regime 1.

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test             1           1.307           0.2530

REGIMES DIAGNOSTICS - CHOW TEST
                 VARIABLE        DF        VALUE           PROB
                 CONSTANT         1           5.502            0.0190
                      AGE         1           0.069            0.7935
                     CRIM         1           0.404            0.5248
                      DIS         1           0.738            0.3902
                    LSTAT         1           0.805            0.3695
                      NOX         1           2.575            0.1086
                       RM         1           7.227            0.0072
                   W_MEDV         1          12.951            0.0003
              Global test         8          51.489            0.0000
================================ END OF REPORT =====================================
```

Keeping AGE, DIS, LSTAT and RM constant and letting constant, CRIM and NOX vary order of variables in x is CRIM, NOX, RM, AGE, DIS, LSTAT

```
In [14]: colsvari = [True,True,False,False,False,False]
```

```
In [15]: reg3 = pysal.spreg.GM_Lag_Regimes(y,x,regimes,w=w,spat_diag=True,
         constant_regi='one',cols2regi=colsvari,name_y=y_name,
         name_x=x_names,name_regimes=rvar,name_w="boston_k4",name_ds="boston.dbf")
```

```
In [16]: print reg3.summary
```

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES - REGIMES
------------------------------------------------------------
Data set            :  boston.dbf
Weights matrix      :   boston_k4
Dependent Variable  :        MEDV                Number of Observations:         506
Mean dependent var  :     22.5328                Number of Variables   :          10
S.D. dependent var  :      9.1971                Degrees of Freedom    :         496
Pseudo R-squared    :      0.7476
Spatial Pseudo R-squared:  0.6835

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
              0_CRIM      -0.0912037       0.0268472      -3.3971428       0.0006809
               0_NOX     -11.8180423       3.6078832      -3.2756166       0.0010543
              1_CRIM       0.9410008       0.8584270       1.0961919       0.2729948
               1_NOX     -11.1960904       4.3850765      -2.5532258       0.0106730
    _Global_CONSTANT       4.8105023       6.1267270       0.7851667       0.4323558
         _Global_AGE      -0.0178455       0.0144821      -1.2322477       0.2178566
         _Global_DIS      -0.9561049       0.1762541      -5.4245828       0.0000001
       _Global_LSTAT      -0.4367124       0.1076675      -4.0561212       0.0000499
          _Global_RM       4.7390297       0.7951419       5.9599801       0.0000000
      _Global_W_MEDV       0.2299312       0.1006010       2.2855767       0.0222790
------------------------------------------------------------------------------------
Instrumented: _Global_W_MEDV
Instruments: 0_W_AGE, 0_W_CRIM, 0_W_DIS, 0_W_LSTAT, 0_W_NOX, 0_W_RM,
             1_W_AGE, 1_W_CRIM, 1_W_DIS, 1_W_LSTAT, 1_W_NOX, 1_W_RM
Regimes variable: CHAS

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test             1       28.773          0.0000

REGIMES DIAGNOSTICS - CHOW TEST
                 VARIABLE        DF        VALUE           PROB
                     CRIM         1        1.442           0.2298
                      NOX         1        0.076           0.7827
              Global test         2        3.458           0.1775
================================ END OF REPORT =====================================
```

In [ ]:

In [ ]: