New Mexico Health Insurance Coverage, 2009-2013
Exploratory, Ordinary Least Squares, and Geographically Weighted Regression Using GeoDa-GWR, R, and QGIS

Larry Spear 4/13/2016 (Draft)

A dataset consisting of selected average statistics was derived from the U.S. Census Bureau's American Community Survey (ACS, 2009 – 2013) for New Mexico's census tracts (n=499). These data were originally processed and made available at the New Mexico Community Data Collaborative (NMCDC) ArcGIS Online web site as either feature layers or feature services. ArcGIS Desktop and SAS University Edition were used for further data preparation.

The results from ArcGIS exploratory regression (see Appendix) suggested that a fifth explanatory variable, percent white (P_WHITE) might produce a better fit. I did not use this model for ArcGIS GWR as the VIF was somewhat large (17.46), indicating more global multicollinearity. As a comparative example, I used this slightly different model and examined the results that were obtained from R using the GWmodel library and also GeoDa-GWR . QGIS was used to produce map output of results.

GeoDa-GWR Results:

The GeoDa-GWR results indicate that this model provides a better fit (Adj R^2 =0.7179 and AICc = 3127.20, see below). An Excel CSV file containing local estimates and diagnostics (regression coefficients, standard residuals, local R2, etc.) was produced. This file (after being joined to a census tract shapefile in QGIS) was used in GeoDa to derive a Global Moran's Index for the standardized residuals that confirmed clustering (Moran's I =0.0992, p=0.002, z=3.5201 – see Appendix). Also, QGIS maps of standardized residuals and local R2 (see Appendix) clearly depict the clustering of the standardized residuals and also the clustered pattern or the local R2 values. The areas where the model performed well (red and orange) and poorly (blues) is similar to the results from the other model (ArcGIS GWR with only four explanatory variables). However, there does seem to be a slight improvement of the strength of predictions in the southeastern part of the state although the strength looks slightly less in the northwest. (Note: Jenks Natural Breaks was used for the R2 maps).

GeoDa-GWR Output (portion only):

Program began at 4/8/2016 4:40:38 PM

****
Session: NMACS13_T1
Session control file: C:\gis\NMDOH\output\NMACS13_T1.ctl
****
Data filename: C:\gis\NMDOH\shapefiles\NMACS13P.dbf
Number of areas/points: 499

Model settings----------

Model type: Gaussian
Geographic kernel: adaptive bi-square
Method for optimal bandwidth search: Golden section search
Criterion for optimal bandwidth: AICc
Number of varying coefficients: 6
Number of fixed coefficients: 0
****
GWR (Geographically weighted regression) result

Bandwidth and geographic ranges
Bandwidth size: 179.328705

Coordinate	Min	Max	Range
X-coord	145161.532700	677569.209900	532407.677200
Y-coord	3518578.325200	4091505.497200	572927.172000

Diagnostic information
Residual sum of squares: 12615.482107
Effective number of parameters (model: trace(S)): 44.064843
Effective number of parameters (variance: trace(S'S)): 32.295010
Degree of freedom (model: n - trace(S)): 454.935157
Degree of freedom (residual: n - 2trace(S) + trace(S'S)): 443.165323
ML based sigma estimate: 5.028074
Unbiased sigma estimate: 5.335425
-2 log-likelihood: 3027.907572
Classic AIC: 3118.037259
AICc: 3127.203715
BIC/MDL: 3307.877693
CV: 35.960659
R square: 0.749545
Adjusted R square: 0.717919

R Results:

The R results (Adj R^2 =0.680791 and AICc = 3171.614), see below) using the GWmodel library are slightly different than those obtained from the GeoDa GWR. Both used a Gaussian kernel function but different bandwidths. I set the bandwidth in R to match what was used in ArcGIS GWR (bw = 97385) and GeoDa-GWR used a search method to obtain an optimal bandwidth (bw = 179). These results are still very useful and provide a valuable lesson about how the choice of kernel and bandwidth can influence results. However, the residual maps (see Appendix) are noticeably very similar. This supports the findings that GWR can produce a model that better fits the data than OLS. The choice of how to specify the model parameters for GWR are more complicated than OLS and do need some theoretical justification within the context of a given research question. These results proved a useful example, more research is necessary to develop a more justifiable and perhaps better model. Note: using a fixed kernel failed due to local multicollinearity between P_HistLat and the additional P_White variable.

```
*********************************************************************
   *                        Package   GWmodel                        *
   *********************************************************************
   Program starts at: 2016-04-13 13:01:23
   Call:
   gwr.basic(formula = Per_WO_Ins ~ Per_Capita + Per_Povert + P_AmIndian +
    P_HispLat + P_White, data = nmacs13.point.spdf, bw = 97385,
    kernel = "gaussian")

   Dependent (y) variable:  Per_WO_Ins
   Independent variables:  Per_Capita Per_Povert P_AmIndian P_HispLat P_White
   Number of data points: 499
   *********************************************************************
   *                    Results of Global Regression                   *
   *********************************************************************

Call:
lm(formula = formula, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-17.0703  -3.8766  -0.3501   3.4946  28.2314

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.188e+00  4.460e+00  -0.939    0.348
Per_Capita  -2.339e-04  3.732e-05  -6.269 7.95e-10 ***
Per_Povert   1.746e-01  3.270e-02   5.340 1.42e-07 ***
P_AmIndian   4.343e+01  4.714e+00   9.212  < 2e-16 ***
P_HispLat    2.907e+01  4.625e+00   6.285 7.23e-10 ***
P_White      2.051e+01  5.009e+00   4.094 4.95e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.966 on 493 degrees of freedom
Multiple R-squared:  0.6517,      Adjusted R-squared:  0.6481
F-statistic: 184.5 on 5 and 493 DF,  p-value: < 2.2e-16

   ***Extra Diagnostic information
   Residual sum of squares: 17544.89
   Sigma(hat): 5.941514
   AIC:  3206.497
   AICc:  3206.725
   *********************************************************************
   *          Results of Geographically Weighted Regression            *
   *********************************************************************

   ********************Model calibration information*********************
   Kernel function: gaussian
   Fixed bandwidth: 97385
   Regression points: the same locations as observations are used.
   Distance metric: Euclidean distance metric is used.

   ****************Summary of GWR coefficient estimates:******************
                     Min.    1st Qu.     Median    3rd Qu.     Max.
   Intercept  -6.528e+01 -9.458e+00  8.933e-01  1.661e+00   2.7250
   Per_Capita -4.523e-04 -2.503e-04 -1.724e-04 -1.493e-04  -0.0001
   Per_Povert -1.016e-01  1.348e-01  2.603e-01  2.894e-01   0.2976
   P_AmIndian  2.854e+01  2.965e+01  3.116e+01  5.380e+01 100.7000
   P_HispLat   1.675e+01  1.922e+01  2.170e+01  4.196e+01  84.0800
   P_White     6.769e+00  8.336e+00  1.114e+01  2.774e+01  81.9100
   ************************Diagnostic information*************************
   Number of data points: 499
   Effective number of parameters (2trace(S) - trace(S'S)): 35.75434
   Effective degrees of freedom (n-2trace(S) + trace(S'S)): 463.2457
   AICc (GWR book, Fotheringham, et al. 2002, p. 61, eq 2.33): 3171.614
   AIC (GWR book, Fotheringham, et al. 2002,GWR p. 96, eq. 4.22): 3138.944
   Residual sum of squares: 14924.28
   R-square value:  0.7037089
```

```
Adjusted R-square value:  0.680791

********************************************************************************
Program stops at: 2016-04-13 13:01:32
```

Appendix :

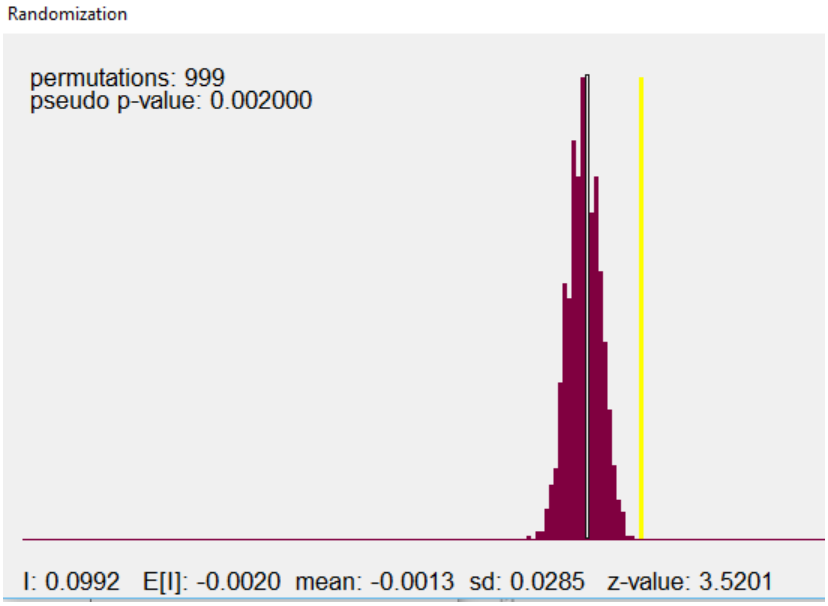
ArcGIS Exploratory Regression Output (portion only):

```
********************************************************************************
Choose 5 of 7 Summary
                                    Highest Adjusted R-Squared Results
AdjR2    AICc   JB K(BP)   VIF   SA   Model
 0.65 3206.72 0.00  0.00 17.46 0.00  -PER_CAPITA_INC***  +PER_POVERTY***  +P_HISPLAT***  +P_WHITE***  +P_AMINDIAN***
 0.64 3219.44 0.00  0.00 16.73 0.00  -MEDIAN_HOUSE_INC***  +PER_POVERTY***  +P_HISPLAT***  +P_WHITE***  +P_AMINDIAN***
 0.64 3220.12 0.00  0.00 17.44 0.00  -MEDIAN_HOUSE_INC***  -PER_CAPITA_INC***  +P_HISPLAT***  +P_WHITE***  +P_AMINDIAN***
       Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
********************************************************************************
```

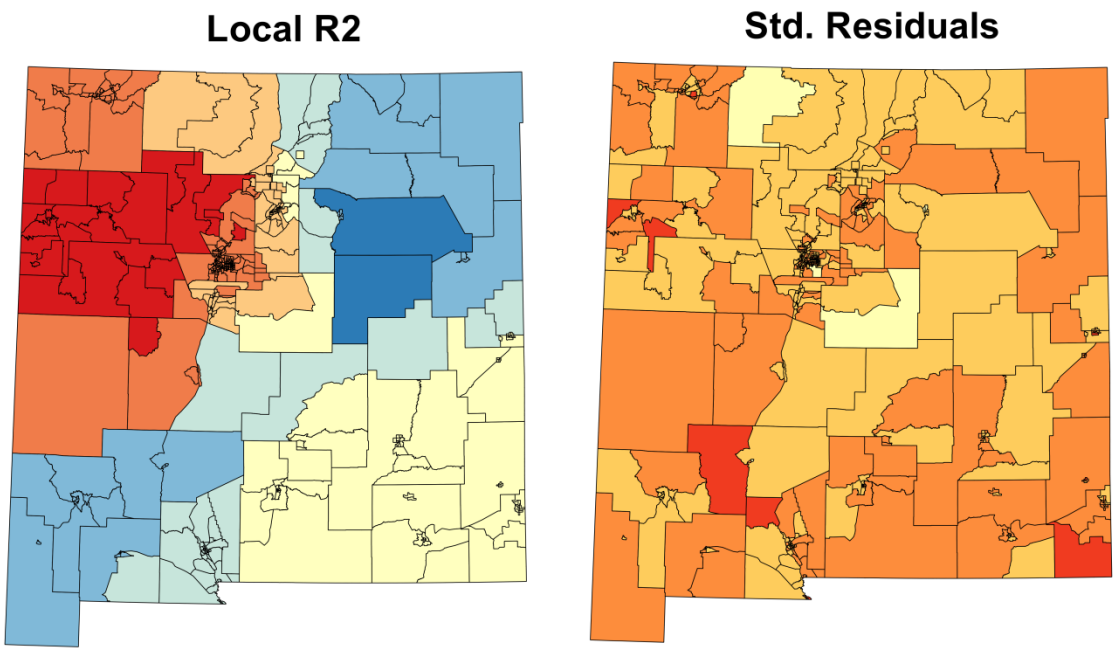
Additional GeoDa-GWR Results (portion only)

```
********************************************************************************
*          Semiparametric Geographically Weighted Regression          *
*                  Release 1.0.90 (GWR 4.0.90)                        *
*                      12 May 2015                                    *
*            (Originally coded by T. Nakaya: 1 Nov 2009)              *
*                                                                       *
*          Tomoki Nakaya(1), Martin Charlton(2), Chris Brunsdon (2)     *
*          Paul Lewis (2), Jing Yao (3), A Stewart Fotheringham (4)     *
*                  (c) GWR4 development team                          *
* (1) Ritsumeikan University, (2) National University of Ireland, Maynooth, *
*       (3) University of Glasgow, (4) Arizona State University         *
********************************************************************************
Variable settings----------------------------------
Area key: field4: GEOID
Easting (x-coord): field51 : UTMX
Northing (y-coord): field52: UTMY
Cartesian coordinates: Euclidean distance
Dependent variable: field20: Per_WO_Ins
Offset variable is not specified
Intercept: varying (Local) intercept
Independent variable with varying (Local) coefficient: field24: Per_Capita
Independent variable with varying (Local) coefficient: field30: Per_Povert
Independent variable with varying (Local) coefficient: field42: P_HispLat
Independent variable with varying (Local) coefficient: field44: P_White
Independent variable with varying (Local) coefficient: field46: P_AmIndian
```

GeoDa Moran's I Results:



GeoDa-GWR (QGIS Maps):



R-GWmodel (GISTools Maps):