Principal Component Analysis: Appendix

L. Graesser

March 14, 2016

Appendix 2: US County Median Household Income analysis: detailed results

```
disp('Analysis of US counties data: 15 features')
clear

load('counties_data.mat')
load('counties_result.mat')
X = counties_data;
% Y = median household income 2007 - 2013
y = counties_result(:,2);

% Variable selection
% Population measures:
% 1: pop 2014: 2
% 2: pop 2010 April: 3
% 3: pop 2010: 5
% 4: Veterans: 23
% 5: Housing Units: 25
% 6: Private non farm employment (abs): 33
% 7: Total number of firms: 36
% 8: Manfacturing shipments (value) 43
% 9: Retail sales (value) 45
% 10: Accommodation and food services sales (value) 47
% 11: Foreign born persons (%): 19
% 12: High school graduate or higher (%): 21
% 13: Bachelor's degree or higher (%): 22
% 14: Persons below the poverty level (%): 31
% 15: Women owned firms (%): 42

numsamples = (size(X,1));
finalX = zeros(numsamples,15);
index_data = [2, 3, 5, 23, 25, 33, 36, 43, 45, 47, 19, 21, 22, 31,
 42];
for i=1:15
    finalX(:,i) = X(:,index_data(i));
end

% Scaling and mean normalizing features
fig = figure
fig.Position = [100 100 950 950]
plotStateData2(finalX,y,'Counties 2: raw data')

finalX = meanNormalize(finalX);
finalX = scaleX(finalX);
disp('Feature covariance matrix')
cov_features = (finalX' * finalX) * (1 / (numsamples - 1))
[meanx varx] = meanVarsX(finalX);

fig = figure
fig.Position = [100 100 950 950]
plotStateData3(finalX,y,'Counties 2: norm,sc')

% Generating training and test datasets
```

```
[ Xtest, XTRAIN, Xtr1, Xtr2, Xtr3, Xtr4, Xtr5, Xtr6, Xtr7, Xtr8, Xtr9,
 Xtr10, ...
                  Xte1, Xte2, Xte3, Xte4, Xte5, Xte6, Xte7, Xte8, Xte9,
 Xte10, ...
  ytest, yTRAIN, ytr1, ytr2, ytr3, ytr4, ytr5, ytr6, ytr7, ytr8, ytr9,
 ytr10, ...
                  yte1, yte2, yte3, yte4, yte5, yte6, yte7, yte8, yte9,
 yte10 ...
           ] = createTestTrainSets_v2( finalX, y );

disp('Training models on training data...')
disp('Unregularised linear regression')
[b_lr res_norm_lr] = linearLS(XTRAIN,yTRAIN)
lm = linearRegression(XTRAIN,yTRAIN)
% Calculating R squared
RSS_lr = res_norm_lr^2;
totalRes = yTRAIN - mean(yTRAIN);
TSS_lr = norm(totalRes)^2;
disp('R Squared')
rsq_lr = 1 - RSS_lr / TSS_lr

disp('Ridge regression')
[ k, RR_rsq_train, RR_rsq_test, rr_opt_bvec, rr_opt_rsq, RR_all_bvec ]
 = ridgeFindK( Xtr1, Xtr2, Xtr3, Xtr4, Xtr5, ...
                                        Xtr6, Xtr7, Xtr8, Xtr9,
 Xtr10, ...
                                        ytr1, ytr2, ytr3, ytr4,
 ytr5, ...
                                        ytr6, ytr7, ytr8, ytr9,
 ytr10, ...
                                        Xte1, Xte2, Xte3, Xte4,
 Xte5, ...
                                        Xte6, Xte7, Xte8, Xte9,
 Xte10, ...
                                        yte1, yte2, yte3, yte4,
 yte5, ...
                                        yte6, yte7, yte8, yte9,
 yte10);

disp('Principle component analysis')
[ PCA_rsq, PCA_train_rsq, pcr_opt_rsq, pcr_opt_bvec, Xtest_PCA,
 XTRAIN_PCA...
 b_pcr1, b_pcr2, b_pcr3, b_pcr4, b_pcr5, b_pcr6, b_pcr7, b_pcr8,
 b_pcr9, b_pcr10] ...
    = pcrFindNumComponents(Xtest, XTRAIN, ytest, yTRAIN);

disp('b vector: linear regression')
disp(b_lr)
rsq_lr
disp('Ridge regression: test rsq')
RR_rsq_test
disp('Ridge regression: training rsq')
RR_rsq_train
disp('Ridge regression: b vector')
```

```
rr_opt_bvec
rr_opt_rsq
disp('Ridge regression: all bvectors')
RR_all_bvec
disp('Principal component regression: test rsq')
disp(PCA_rsq)
disp('Principal component regression: training rsq')
disp(PCA_train_rsq)
disp('Principal component regression: b vector')
disp(pcr_opt_bvec)
pcr_opt_rsq
disp('Principal component regression: all b vectors')
b_pcr1
b_pcr2
b_pcr3
b_pcr4
b_pcr5
b_pcr6
b_pcr7
b_pcr8
b_pcr9
b_pcr10

save USCountiesAnalysis2_Results.mat Xtest_PCA Xtest ytest b_lr pcr_opt_bvec rr
```

```
Analysis of US counties data: 15 features

fig =

  Figure (17) with properties:

      Number: 17
        Name: ''
       Color: [0.9400 0.9400 0.9400]
    Position: [440 378 560 420]
       Units: 'pixels'

  Use GET to show all properties


fig =

  Figure (17) with properties:

      Number: 17
        Name: ''
       Color: [0.9400 0.9400 0.9400]
    Position: [100 100 950 950]
       Units: 'pixels'

  Use GET to show all properties

Normalising X...
Scaling features....
```

Feature covariance matrix

cov_features =

Columns 1 through 7

1.0000	0.9996	0.9996	0.9268	0.9933	0.9559	0.9815
0.9996	1.0000	1.0000	0.9253	0.9932	0.9557	0.9818
0.9996	1.0000	1.0000	0.9253	0.9932	0.9557	0.9818
0.9268	0.9253	0.9253	1.0000	0.9421	0.8816	0.8738
0.9933	0.9932	0.9932	0.9421	1.0000	0.9634	0.9744
0.9559	0.9557	0.9557	0.8816	0.9634	1.0000	0.9585
0.9815	0.9818	0.9818	0.8738	0.9744	0.9585	1.0000
0.8293	0.8291	0.8291	0.7525	0.8220	0.8283	0.7930
0.9759	0.9751	0.9751	0.9265	0.9771	0.9730	0.9715
0.8227	0.8214	0.8214	0.7869	0.8366	0.8683	0.8290
0.4730	0.4690	0.4690	0.4352	0.4698	0.4430	0.4682
0.0624	0.0621	0.0621	0.1253	0.0781	0.0788	0.0658
0.3228	0.3200	0.3200	0.3678	0.3425	0.3545	0.3239
-0.0628	-0.0618	-0.0618	-0.1009	-0.0646	-0.0673	-0.0696
0.2171	0.2175	0.2175	0.2681	0.2297	0.2001	0.1979

Columns 8 through 14

0.8293	0.9759	0.8227	0.4730	0.0624	0.3228	-0.0628
0.8291	0.9751	0.8214	0.4690	0.0621	0.3200	-0.0618
0.8291	0.9751	0.8214	0.4690	0.0621	0.3200	-0.0618
0.7525	0.9265	0.7869	0.4352	0.1253	0.3678	-0.1009
0.8220	0.9771	0.8366	0.4698	0.0781	0.3425	-0.0646
0.8283	0.9730	0.8683	0.4430	0.0788	0.3545	-0.0673
0.7930	0.9715	0.8290	0.4682	0.0658	0.3239	-0.0696
1.0000	0.8161	0.6317	0.3159	0.0569	0.2430	-0.0577
0.8161	1.0000	0.8457	0.4627	0.0893	0.3581	-0.0849
0.6317	0.8457	1.0000	0.4013	0.0594	0.2905	-0.0498
0.3159	0.4627	0.4013	1.0000	-0.1636	0.3579	-0.0508
0.0569	0.0893	0.0594	-0.1636	1.0000	0.6023	-0.6598
0.2430	0.3581	0.2905	0.3579	0.6023	1.0000	-0.4222
-0.0577	-0.0849	-0.0498	-0.0508	-0.6598	-0.4222	1.0000
0.1763	0.2152	0.1679	0.1731	0.0692	0.2736	0.0016

Column 15

0.2171
0.2175
0.2175
0.2681
0.2297
0.2001
0.1979
0.1763
0.2152
0.1679
0.1731
0.0692

```
    0.2736
    0.0016
    1.0000


fig =

  Figure (18) with properties:

      Number: 18
        Name: ''
       Color: [0.9400 0.9400 0.9400]
    Position: [440 378 560 420]
       Units: 'pixels'

  Use GET to show all properties


fig =

  Figure (18) with properties:

      Number: 18
        Name: ''
       Color: [0.9400 0.9400 0.9400]
    Position: [100 100 950 950]
       Units: 'pixels'

  Use GET to show all properties

Training models on training data...
Unregularised linear regression

b_lr =

   1.0e+05 *

    0.4586
    0.0090
    2.0790
   -1.9885
    0.0183
   -0.1330
    0.0154
   -0.0128
   -0.0002
    0.0132
   -0.0010
    0.0213
   -0.0021
    0.0451
   -0.0702
    0.0100
```

```
res_norm_lr =

   2.7947e+05


lm =


Linear regression model:
    y ~ [Linear formula with 16 terms in 15 predictors]

Estimated Coefficients:
                   Estimate          SE         tStat        pValue
                  ____________    ___________    ________
 ____________

    (Intercept)          45861        111.56       411.1
 0
    x1                   901.8        4234.2     0.21298
 0.83136
    x2                2.079e+05     1.8455e+05      1.1265
 0.26006
    x3               -1.9885e+05     1.843e+05      -1.079
 0.2807
    x4                   1830.8        512.32      3.5736
 0.00035874
    x5                   -13298        1300.4     -10.226
 4.4842e-24
    x6                   1543.7        640.75      2.4093
 0.016057
    x7                  -1282.6        872.04     -1.4708
 0.14148
    x8                   -24.74        224.66    -0.11012
 0.91232
    x9                   1316.1        776.62      1.6947
 0.090262
    x10                 -97.687         223.4    -0.43727
 0.66196
    x11                  2130.9         157.4      13.538
 2.3832e-40
    x12                 -208.26        204.16     -1.0201
 0.30779
    x13                    4514        182.36      24.753
 3.7e-121
    x14                 -7024.7        156.61     -44.854
 8.8932e-323
    x15                  1001.9        120.17      8.3375
 1.2366e-16


Number of observations: 2514, Error degrees of freedom: 2498
Root Mean Squared Error: 5.59e+03
```

```
R-squared: 0.783,  Adjusted R-Squared 0.782
F-statistic vs. constant model: 600, p-value = 0
R Squared

rsq_lr =

    0.7828

Ridge regression
Ridge regression

k_choices =

  Columns 1 through 7

    0.0010    0.0030    0.0100    0.0300    0.1000    0.3000    1.0000

  Columns 8 through 10

    3.0000   10.0000   30.0000


bnorms =

   1.0e+04 *

  Columns 1 through 7

    8.9069    5.9995    5.0447    4.8979    5.1718    4.9221    4.8470

  Columns 8 through 10

    4.8225    4.7219    4.6878

Principle component analysis
Coefficient matrix
  Columns 1 through 7

    0.3190   -0.0474   -0.0497   -0.0199   -0.0291    0.0359   -0.1789
    0.3188   -0.0484   -0.0515   -0.0245   -0.0305    0.0365   -0.1820
    0.3188   -0.0484   -0.0515   -0.0245   -0.0305    0.0365   -0.1820
    0.3022    0.0093    0.0020   -0.0678   -0.0025   -0.0614   -0.3415
    0.3197   -0.0364   -0.0388   -0.0298    0.0008   -0.0041   -0.1594
    0.3139   -0.0314   -0.0665   -0.0230    0.0551   -0.0478    0.2009
    0.3144   -0.0433   -0.0612    0.0025   -0.0104   -0.0424   -0.1189
    0.2704   -0.0476   -0.1214   -0.1423   -0.1173    0.6996    0.5371
    0.3178   -0.0228   -0.0518   -0.0174    0.0139   -0.0350   -0.0511
    0.2769   -0.0391   -0.0584    0.0133    0.1173   -0.6726    0.5765
    0.1646   -0.0341    0.4715    0.7463   -0.1611    0.0626   -0.0428
    0.0357    0.6320   -0.1850   -0.1830    0.2080   -0.0152   -0.2031
    0.1298    0.4976    0.2940    0.1844    0.5773    0.1637    0.1573
   -0.0344   -0.5720    0.1910   -0.1620    0.7046    0.0949   -0.1128
    0.0832    0.0959    0.7631   -0.5665   -0.2595   -0.0721    0.0355
```

```
  Columns 8 through 14

   -0.0294   -0.1049   -0.2575   -0.0752   -0.2679   -0.1683    0.8216
   -0.0304   -0.1173   -0.2644   -0.0963   -0.2678   -0.1713   -0.4023
   -0.0304   -0.1173   -0.2644   -0.0964   -0.2682   -0.1717   -0.4033
   -0.1065    0.7964    0.1209    0.0245    0.2738   -0.2184   -0.0075
   -0.0326    0.0072   -0.0876   -0.3272    0.1565    0.8534    0.0044
   -0.0385   -0.1978    0.6641   -0.5590    0.0195   -0.2351    0.0081
   -0.0401   -0.4692   -0.0631    0.3301    0.7218   -0.1545    0.0002
    0.1957    0.1843   -0.1046    0.0934    0.0778    0.0067   -0.0032
   -0.0747   -0.0482    0.4860    0.6574   -0.3968    0.2397   -0.0139
    0.1631    0.1564   -0.2513    0.0722   -0.0138    0.0061   -0.0035
    0.3967    0.0289    0.0495   -0.0053    0.0041   -0.0059   -0.0040
    0.6651   -0.0413    0.0515   -0.0039   -0.0094   -0.0082    0.0004
   -0.4683   -0.0048   -0.1037    0.0130   -0.0029   -0.0009   -0.0017
    0.3002    0.0004    0.0141    0.0179    0.0023   -0.0104   -0.0003
    0.0460   -0.0673    0.0110    0.0009   -0.0067   -0.0006    0.0016

  Column 15

   -0.0005
    0.7075
   -0.7067
   -0.0000
   -0.0002
    0.0001
   -0.0002
   -0.0000
    0.0000
    0.0000
   -0.0000
    0.0000
    0.0000
    0.0000
    0.0000

Singular values
  Columns 1 through 7

  173.8101         0         0         0         0         0         0
         0   80.3406         0         0         0         0         0
         0         0   57.5112         0         0         0         0
         0         0         0   51.0844         0         0         0
         0         0         0         0   38.5882         0         0
         0         0         0         0         0   33.7811         0
         0         0         0         0         0         0   26.4359
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
```

```
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0

  Columns 8 through 14

         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
   22.8197         0         0         0         0         0         0
         0   20.7514         0         0         0         0         0
         0         0   12.4585         0         0         0         0
         0         0         0    8.0815         0         0         0
         0         0         0         0    5.4425         0         0
         0         0         0         0         0    4.0326         0
         0         0         0         0         0         0    1.1972
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0

  Column 15

         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
    0.0226
         0
         0
         0
         0
         0

  Principal component regression: 12 components
```

```
b_pcr1 =

   1.0e+04 *

    4.5874
    0.1212
    0.6171
   -0.1701
   -0.3004
    0.3264
   -0.0374
   -0.1393
   -0.3840
   -0.0574
    0.0302
    0.3126
    0.4888


res_norm_pcr =

   2.6932e+05


lm =


Linear regression model:
    y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11 +
 x12

Estimated Coefficients:
                   Estimate      SE       tStat       pValue
                   ________    _______    _______    ___________

    (Intercept)      45874     119.41     384.18             0
    x1               1212.5    35.609      34.05    2.5066e-205
    x2               6170.9    83.158     74.206             0
    x3              -1701.3    114.88    -14.809     2.0397e-47
    x4              -3004.1    128.51    -23.376    2.2448e-108
    x5               3263.9     172.9     18.877     6.7069e-74
    x6              -374.49    179.24    -2.0894       0.036786
    x7              -1393.4    234.61     -5.939     3.3135e-09
    x8              -3840.1    292.66    -13.121     5.9185e-38
    x9                 -574    295.58    -1.9419       0.052268
    x10             302.47     496.47    0.60925        0.54242
    x11             3126.1      796.8     3.9233     8.9972e-05
    x12             4887.7     1204.5      4.058      5.119e-05


Number of observations: 2263, Error degrees of freedom: 2250
Root Mean Squared Error: 5.68e+03
R-squared: 0.777,  Adjusted R-Squared 0.776
F-statistic vs. constant model: 654, p-value = 0
```

```
R squared

rsq =

    0.7606

Principal component regression: 11 components

b_pcr2 =

   1.0e+04 *

    4.5867
    0.1205
    0.6182
   -0.1743
   -0.3054
    0.3233
   -0.0422
   -0.1497
   -0.3698
   -0.0441
    0.0135
    0.3107


res_norm_pcr =

   2.7033e+05


lm =


Linear regression model:
    y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11

Estimated Coefficients:
                   Estimate      SE       tStat       pValue
                   ________    _______    _______    ___________

    (Intercept)      45867     119.85     382.68              0
    x1               1204.8     35.298     34.132    3.8174e-206
    x2               6182.2     82.721     74.735              0
    x3              -1743.4     115.17    -15.138     2.2492e-49
    x4              -3054.1     128.77    -23.717     3.504e-111
    x5               3233.5     174.72     18.507     2.7134e-71
    x6              -421.65     177.13    -2.3804       0.017375
    x7              -1497.3     239.36    -6.2554     4.7321e-10
    x8              -3698.5     291.41    -12.692     1.0678e-35
    x9              -441.15     300.59    -1.4676        0.14235
    x10              135.15     510.56     0.2647        0.79126
    x11              3106.8     767.93     4.0457     5.3922e-05
```

```
Number of observations: 2263, Error degrees of freedom: 2251
Root Mean Squared Error: 5.7e+03
R-squared: 0.78,  Adjusted R-Squared 0.779
F-statistic vs. constant model: 725, p-value = 0
R squared

rsq =

    0.6958

Principal component regression: 10 components

b_pcr3 =

   1.0e+04 *

    4.5899
    0.1197
    0.6168
   -0.1606
   -0.2990
    0.3227
   -0.0365
   -0.1524
   -0.3637
   -0.0393
    0.0201


res_norm_pcr =

   2.7304e+05


lm =


Linear regression model:
    y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10

Estimated Coefficients:
                   Estimate      SE       tStat       pValue
                   ________    ______    _______    ____________

    (Intercept)      45899     120.97     379.41              0
    x1              1197.2     35.495     33.728    3.2161e-202
    x2              6168.2     84.552     72.952              0
    x3             -1605.9     117.46    -13.673     6.0848e-41
    x4             -2990.3     131.81    -22.686    8.7556e-103
    x5                3227     176.89     18.243     1.8636e-69
    x6             -364.66     178.92    -2.0381       0.041653
    x7             -1524.4     233.69    -6.5229     8.4886e-11
    x8               -3637     300.45    -12.105     1.0033e-32
```

```
    x9             -392.57     292.52    -1.3421         0.17971
    x10             200.91     495.27    0.40566         0.68503


Number of observations: 2263, Error degrees of freedom: 2252
Root Mean Squared Error: 5.75e+03
R-squared: 0.767,  Adjusted R-Squared 0.766
F-statistic vs. constant model: 743, p-value = 0
R squared

rsq =

    0.8104

Principal component regression: 9 components

b_pcr4 =

   1.0e+04 *

    4.5854
    0.1209
    0.6178
   -0.1768
   -0.3062
    0.3121
   -0.0400
   -0.1404
   -0.3807
   -0.0472


res_norm_pcr =

   2.7281e+05


lm =


Linear regression model:
    y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9

Estimated Coefficients:
                   Estimate      SE       tStat       pValue
                   ________    ______    _______    ___________

    (Intercept)      45854     120.84     379.45             0
    x1               1209.3    35.886     33.699    6.0378e-202
    x2               6178.2    83.327     74.144             0
    x3              -1767.8    116.27    -15.205       8.88e-50
    x4              -3061.5    130.24    -23.507    1.8366e-109
    x5               3121.4    175.52     17.784     2.5959e-66
    x6              -400.07    179.16     -2.233       0.025648
```

```
    x7             -1403.6     238.53    -5.8842     4.5986e-09
    x8             -3806.9     293.02    -12.992     2.8724e-37
    x9             -472.22     295.85    -1.5962        0.11059


Number of observations: 2263, Error degrees of freedom: 2253
Root Mean Squared Error: 5.75e+03
R-squared: 0.774,  Adjusted R-Squared 0.773
F-statistic vs. constant model: 859, p-value = 0
R squared

rsq =

    0.7481

Principal component regression: 8 components

b_pcr5 =

   1.0e+04 *

    4.5858
    0.1192
    0.6169
   -0.1636
   -0.2992
    0.3245
   -0.0307
   -0.1473
   -0.3849


res_norm_pcr =

   2.7086e+05


lm =


Linear regression model:
    y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8

Estimated Coefficients:
                   Estimate      SE       tStat       pValue
                   ________    ______    _______    ___________

    (Intercept)      45858     119.98     382.2             0
    x1               1191.7    37.116     32.108    1.3444e-186
    x2               6169.2    83.604     73.79             0
    x3              -1635.8    115.26    -14.192     7.5255e-44
    x4                -2992    132.48    -22.585    5.4975e-102
    x5               3245.5    175.88     18.452     6.4351e-71
    x6               -306.9    195.52    -1.5696        0.11664
```

```
    x7             -1472.9      257.8    -5.7132     1.2552e-08
    x8             -3848.9     299.41    -12.855     1.4955e-36


Number of observations: 2263, Error degrees of freedom: 2254
Root Mean Squared Error: 5.71e+03
R-squared: 0.771,  Adjusted R-Squared 0.77
F-statistic vs. constant model: 950, p-value = 0
R squared

rsq =

    0.7770

Principal component regression: 7 components

b_pcr6 =

   1.0e+04 *

    4.5847
    0.1203
    0.6202
   -0.1739
   -0.2940
    0.3123
   -0.0363
   -0.1300


res_norm_pcr =

   2.7977e+05


lm =


Linear regression model:
    y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7

Estimated Coefficients:
                   Estimate      SE       tStat       pValue
                   ________    ______    _______    __________

    (Intercept)       45847     123.88     370.08           0
    x1                1202.9     36.786     32.699    2.721e-192
    x2                6201.5     85.981     72.126           0
    x3               -1739.3     118.63    -14.662    1.4614e-46
    x4               -2940.2     133.39    -22.041    1.1586e-97
    x5                3122.6     180.06     17.342     2.423e-63
    x6               -363.39     183.08    -1.9849      0.047276
    x7               -1300.2      236.9    -5.4883    4.5122e-08
```

```
Number of observations: 2263, Error degrees of freedom: 2255
Root Mean Squared Error: 5.89e+03
R-squared: 0.76,  Adjusted R-Squared 0.759
F-statistic vs. constant model: 1.02e+03, p-value = 0
R squared

rsq =

    0.7188

Principal component regression: 6 components

b_pcr7 =

   1.0e+04 *

    4.5757
    0.1186
    0.6160
   -0.1637
   -0.2914
    0.3299
   -0.0492


res_norm_pcr =

   2.7882e+05


lm =


Linear regression model:
    y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6

Estimated Coefficients:
                   Estimate      SE       tStat       pValue
                   ________    ______    _______    ___________

    (Intercept)      45757     123.42     370.75             0
    x1               1185.9    37.072      31.99    1.7217e-185
    x2               6160.4    86.117     71.535             0
    x3              -1636.9    118.82    -13.776     1.6289e-41
    x4              -2913.6    132.28    -22.026     1.5086e-97
    x5               3298.6    181.09     18.215     2.8219e-69
    x6              -492.45    200.53    -2.4557       0.014134


Number of observations: 2263, Error degrees of freedom: 2256
Root Mean Squared Error: 5.87e+03
R-squared: 0.753,  Adjusted R-Squared 0.752
F-statistic vs. constant model: 1.14e+03, p-value = 0
```

```
R squared

rsq =

    0.7481

Principal component regression: 5 components

b_pcr8 =

   1.0e+04 *

    4.5860
    0.1190
    0.6143
   -0.1710
   -0.2960
    0.3027


res_norm_pcr =

   2.8438e+05


lm =


Linear regression model:
    y ~ 1 + x1 + x2 + x3 + x4 + x5

Estimated Coefficients:
                   Estimate      SE       tStat       pValue
                   ________    ______    _______    ___________

    (Intercept)      45860     125.86     364.37             0
    x1               1190.1    36.642     32.479    3.3541e-190
    x2               6142.8    86.752     70.808             0
    x3              -1710.2    120.57    -14.184     8.2896e-44
    x4                -2960    134.48    -22.011     1.9853e-97
    x5               3027.2    181.28     16.699     4.0753e-59


Number of observations: 2263, Error degrees of freedom: 2257
Root Mean Squared Error: 5.99e+03
R-squared: 0.751,  Adjusted R-Squared 0.751
F-statistic vs. constant model: 1.36e+03, p-value = 0
R squared

rsq =

    0.7577

Principal component regression: 4 components
```

```
b_pcr9 =

   1.0e+04 *

    4.5817
    0.1221
    0.6152
   -0.1682
   -0.2948


res_norm_pcr =

   3.0249e+05


lm =


Linear regression model:
    y ~ 1 + x1 + x2 + x3 + x4

Estimated Coefficients:
                   Estimate      SE       tStat       pValue
                   ________    ______    _______    ___________

    (Intercept)      45817     133.86     342.29             0
    x1                1221     41.467     29.444    1.5124e-161
    x2              6152.1     93.133     66.057             0
    x3             -1681.9     129.81    -12.956     4.3989e-37
    x4             -2948.3      143.3    -20.575     2.3611e-86


Number of observations: 2263, Error degrees of freedom: 2258
Root Mean Squared Error: 6.37e+03
R-squared: 0.719,  Adjusted R-Squared 0.719
F-statistic vs. constant model: 1.45e+03, p-value = 0
R squared

rsq =

    0.7164

Principal component regression: 3 components

b_pcr10 =

   1.0e+04 *

    4.5812
    0.1239
    0.6186
   -0.1738
```

```
res_norm_pcr =

   3.3220e+05


lm =


Linear regression model:
    y ~ 1 + x1 + x2 + x3

Estimated Coefficients:
                   Estimate      SE       tStat       pValue
                   ________    ______    _______    ___________

    (Intercept)      45812     146.96     311.74             0
    x1              1239.4      52.38     23.662    9.1615e-111
    x2              6185.6        103     60.056             0
    x3             -1738.4     144.98    -11.991      3.688e-32


Number of observations: 2263, Error degrees of freedom: 2259
Root Mean Squared Error: 6.99e+03
R-squared: 0.667,  Adjusted R-Squared 0.667
F-statistic vs. constant model: 1.51e+03, p-value = 0
R squared

rsq =

    0.6435


numcomps =

    12    11    10     9     8     7     6     5     4     3


rsqAll =

    0.7606    0.7772
    0.6958    0.7800
    0.8104    0.7673
    0.7481    0.7742
    0.7770    0.7712
    0.7188    0.7597
    0.7481    0.7528
    0.7577    0.7514
    0.7164    0.7192
    0.6435    0.6672

b vector: linear regression
   1.0e+05 *
```

```
    0.4586
    0.0090
    2.0790
   -1.9885
    0.0183
   -0.1330
    0.0154
   -0.0128
   -0.0002
    0.0132
   -0.0010
    0.0213
   -0.0021
    0.0451
   -0.0702
    0.0100


rsq_lr =

    0.7828

Ridge regression: test rsq

RR_rsq_test =

  Columns 1 through 7

    0.7924    0.7936    0.8109    0.7468    0.6489    0.7623    0.7335

  Columns 8 through 10

    0.7950    0.7811    0.7544

Ridge regression: training rsq

RR_rsq_train =

  Columns 1 through 7

    0.7817    0.7803    0.7776    0.7861    0.7892    0.7843    0.7872

  Columns 8 through 10

    0.7798    0.7798    0.7797

Ridge regression: b vector

rr_opt_bvec =

   1.0e+04 *

    4.5826
```

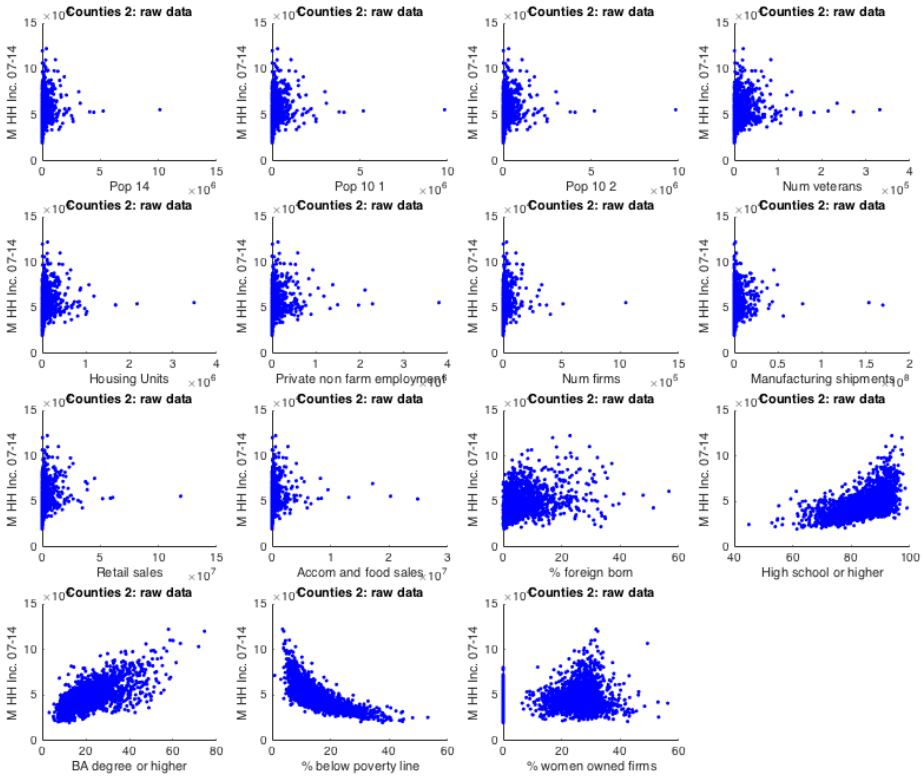
```
    0.2622
    0.3188
    0.3143
    0.1315
   -1.0756
   -0.0044
   -0.1610
    0.0172
    0.2015
   -0.0210
    0.2073
   -0.0202
    0.4502
   -0.6987
    0.1070


rr_opt_rsq =

    0.7950

Ridge regression: all bvectors

RR_all_bvec =

   1.0e+04 *

  Columns 1 through 7

    4.5911    4.5891    4.5852    4.5830    4.5894    4.5888    4.5841
    0.0075    0.2818    0.1538    0.0369   -0.3314   -0.0780    0.0832
    5.7169    2.7447    1.2900    0.6970    0.9632    0.5243    0.3861
   -4.7995   -2.0013   -0.2785    0.1972    0.7696    0.4727    0.3728
    0.1959    0.2425    0.1677    0.1768    0.1842    0.2006    0.1808
   -1.2990   -1.5563   -1.3345   -1.2724   -1.7932   -1.3436   -1.1597
    0.1776    0.1702   -0.0010    0.1460    0.1689    0.1670    0.1257
   -0.0976   -0.1052   -0.2279   -0.1275    0.0237   -0.0959   -0.0979
   -0.0003    0.0139    0.0072    0.0029    0.0248    0.0144    0.0123
    0.1227    0.1926    0.2313    0.1467    0.0464    0.1387    0.1000
   -0.0312   -0.0008   -0.0244   -0.0124    0.0065   -0.0053   -0.0055
    0.2142    0.2106    0.2080    0.2235    0.1900    0.2091    0.2061
   -0.0216   -0.0214   -0.0199   -0.0226   -0.0425   -0.0208   -0.0299
    0.4464    0.4564    0.4504    0.4544    0.4675    0.4478    0.4610
   -0.7056   -0.7021   -0.7000   -0.6953   -0.7128   -0.7136   -0.7112
    0.0988    0.1078    0.0913    0.0988    0.0965    0.1011    0.0962

  Columns 8 through 10

    4.5826    4.5794    4.5886
    0.2622    0.1498    0.0830
    0.3188    0.1428    0.0829
    0.3143    0.1412    0.0826
    0.1315    0.0939    0.0660
   -1.0756   -0.6999   -0.3881
```

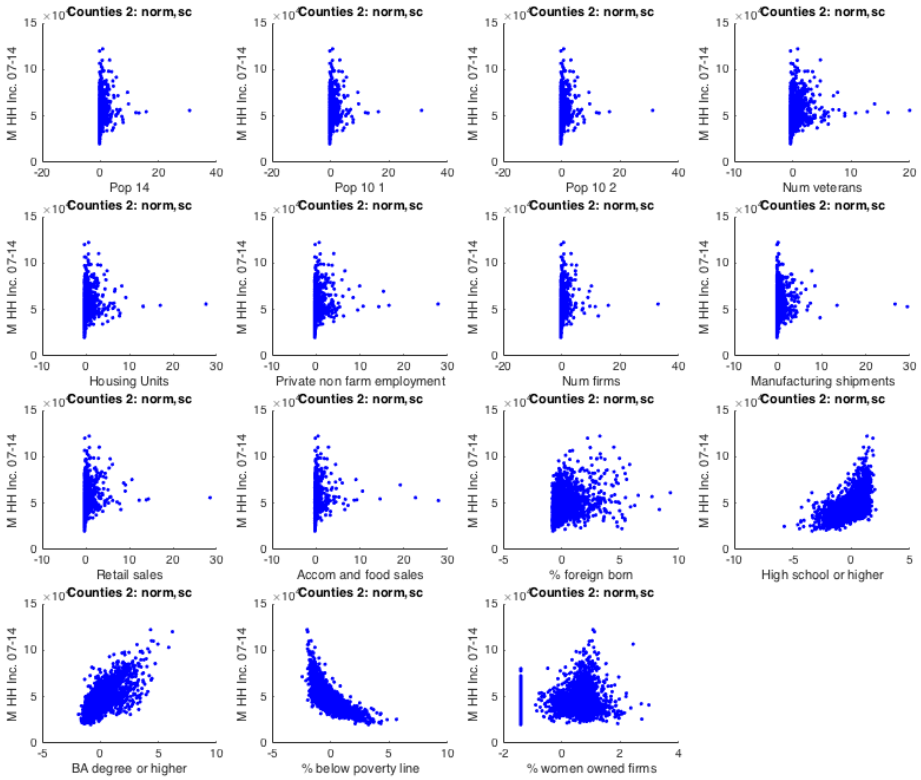
```
   -0.0044    0.0074   -0.0020
   -0.1610   -0.0212   -0.0629
    0.0172    0.0155    0.0273
    0.2015    0.1760    0.1129
   -0.0210   -0.0104   -0.0101
    0.2073    0.2253    0.2122
   -0.0202   -0.0125   -0.0138
    0.4502    0.4391    0.4399
   -0.6987   -0.6937   -0.6916
    0.1070    0.0881    0.0994

Principal component regression: test rsq
  Columns 1 through 7

    0.7606    0.6958    0.8104    0.7481    0.7770    0.7188    0.7481

  Columns 8 through 10

    0.7577    0.7164    0.6435

Principal component regression: training rsq
  Columns 1 through 7

    0.7772    0.7800    0.7673    0.7742    0.7712    0.7597    0.7528

  Columns 8 through 10

    0.7514    0.7192    0.6672

Principal component regression: b vector
   1.0e+04 *

    4.5860
    0.1190
    0.6143
   -0.1710
   -0.2960
    0.3027


pcr_opt_rsq =

    0.7577

Principal component regression: all b vectors

b_pcr1 =

   1.0e+04 *

    4.5874
    0.1212
    0.6171
   -0.1701
```

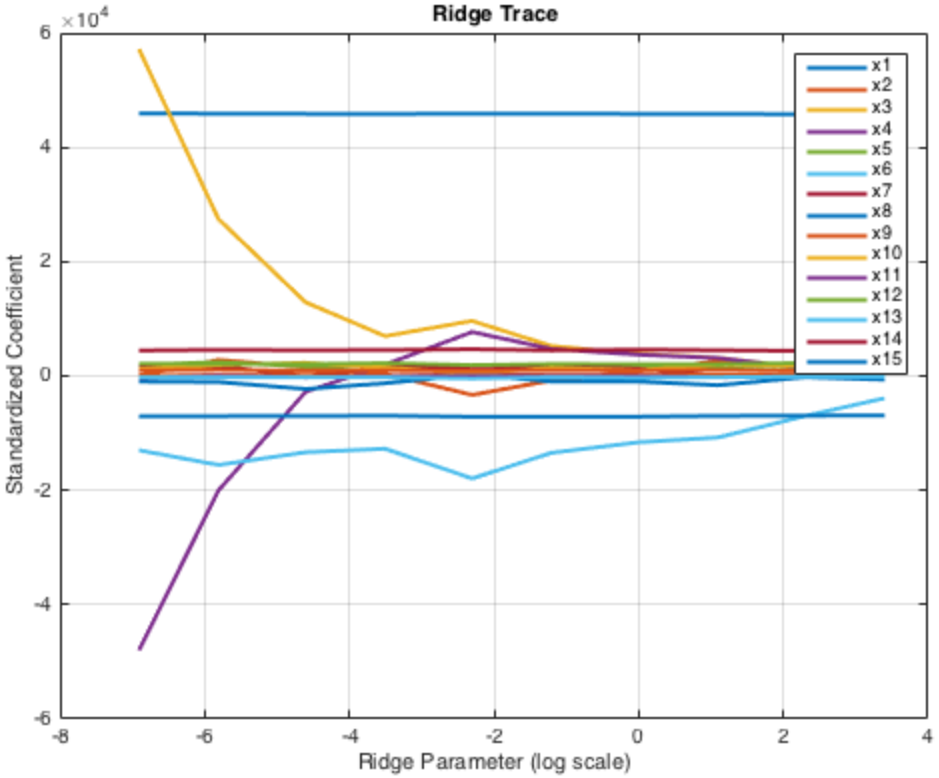
```
   -0.3004
    0.3264
   -0.0374
   -0.1393
   -0.3840
   -0.0574
    0.0302
    0.3126
    0.4888


b_pcr2 =

   1.0e+04 *

    4.5867
    0.1205
    0.6182
   -0.1743
   -0.3054
    0.3233
   -0.0422
   -0.1497
   -0.3698
   -0.0441
    0.0135
    0.3107


b_pcr3 =

   1.0e+04 *

    4.5899
    0.1197
    0.6168
   -0.1606
   -0.2990
    0.3227
   -0.0365
   -0.1524
   -0.3637
   -0.0393
    0.0201


b_pcr4 =

   1.0e+04 *

    4.5854
    0.1209
    0.6178
   -0.1768
```

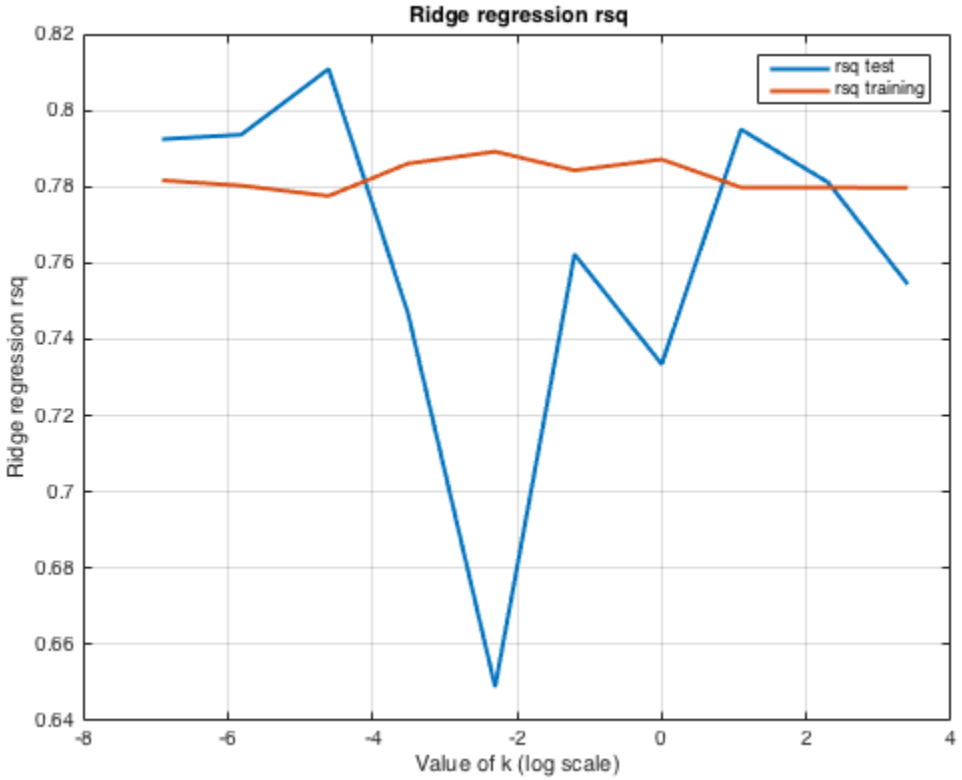
```
   -0.3062
    0.3121
   -0.0400
   -0.1404
   -0.3807
   -0.0472


b_pcr5 =

   1.0e+04 *

    4.5858
    0.1192
    0.6169
   -0.1636
   -0.2992
    0.3245
   -0.0307
   -0.1473
   -0.3849


b_pcr6 =

   1.0e+04 *

    4.5847
    0.1203
    0.6202
   -0.1739
   -0.2940
    0.3123
   -0.0363
   -0.1300


b_pcr7 =

   1.0e+04 *

    4.5757
    0.1186
    0.6160
   -0.1637
   -0.2914
    0.3299
   -0.0492


b_pcr8 =

   1.0e+04 *
```

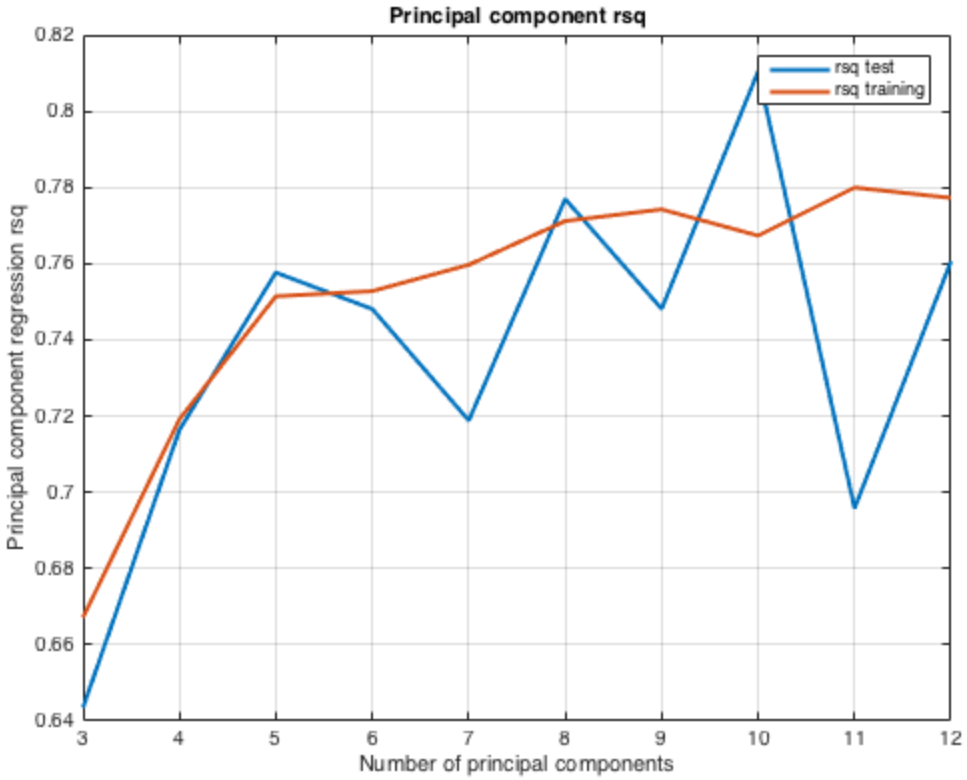
```
    4.5860
    0.1190
    0.6143
   -0.1710
   -0.2960
    0.3027


b_pcr9 =

   1.0e+04 *

    4.5817
    0.1221
    0.6152
   -0.1682
   -0.2948


b_pcr10 =

   1.0e+04 *

    4.5812
    0.1239
    0.6186
   -0.1738
```


Counties 2: raw data
M HH Inc. 07-14
Pop 14
Pop 10 1
Pop 10 2
Num veterans
Housing Units
Private non farm employment
Num firms
Manufacturing shipments
Retail sales
Accom and food sales
% foreign born
High school or higher
BA degree or higher
% below poverty line
% women owned firms


Counties 2: norm,sc
M HH Inc. 07-14
Pop 14
Pop 10 1
Pop 10 2
Num veterans
Housing Units
Private non farm employment
Num firms
Manufacturing shipments
Retail sales
Accom and food sales
% foreign born
High school or higher
BA degree or higher
% below poverty line
% women owned firms


Ridge Trace
×10^4
6
4
2
0
-2
-4
-6
-8
-6
-4
-2
0
2
4
Standardized Coefficient
Ridge Parameter (log scale)
x1
x2
x3
x4
x5
x6
x7
x8
x9
x10
x11
x12
x13
x14
x15


Ridge regression rsq
0.82
0.8
0.78
0.76
0.74
0.72
0.7
0.68
0.66
0.64
-8
-6
-4
-2
0
2
4
Ridge regression rsq
Value of k (log scale)
rsq test
rsq training



Published with MATLAB® R2015b