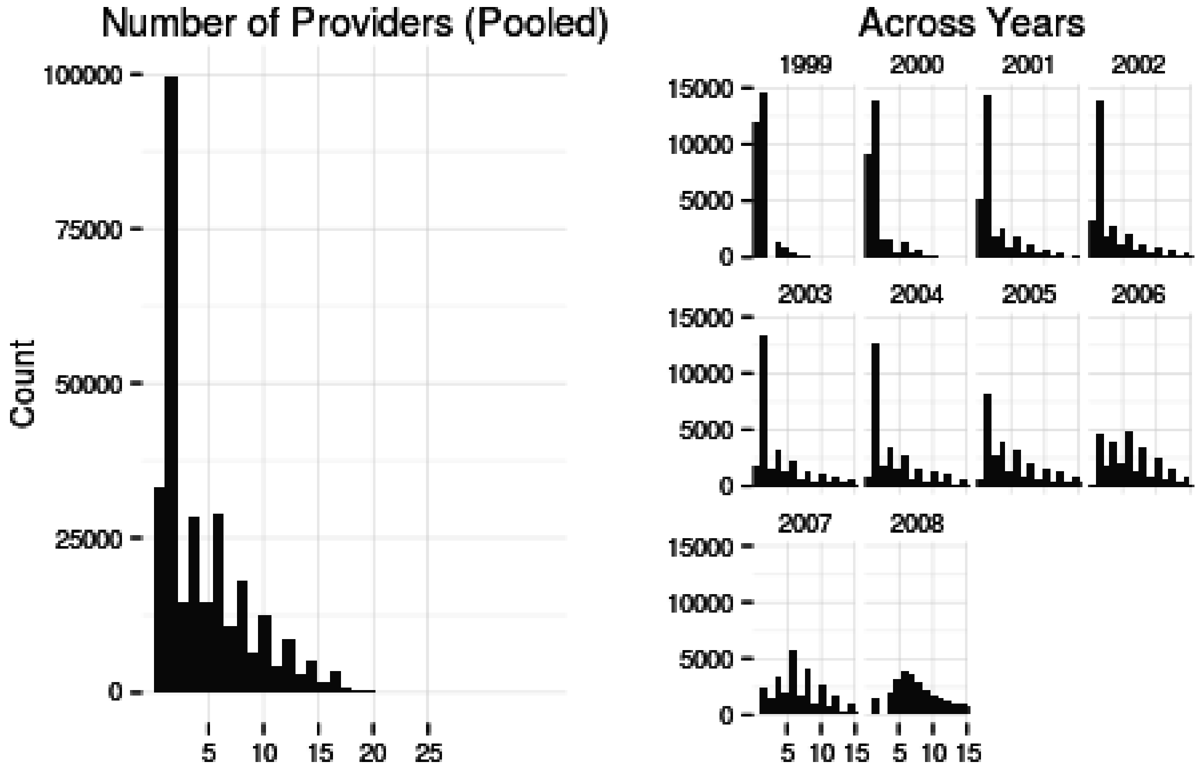
3-USDA Evaluation Poisson

Robert Dinterman

2015-10-08

Poisson Panel Regressions for Broadband Availability

The following regressions make use of the `glm` function for generalized linear models in `R` in order to identify the relationship between broadband availability and the USDA Broadband Loan Program. Loosely, we wish to model the number of broadband providers by zip code across the years 1999 to 2008 and determine whether or not the USDA Broadband Loan Program had an impact on broadband availability as this was one of the intended benefits of the subsidized loans. As a quick reference, here is what the dependent variable looks like across time:



These values have been rounded, the previous graph of these were not rounded which explains why these look different.

Data on broadband providers is measured twice a year (June 30 and December 31), therefore the two values are averaged for the yearly value. Further, the variable takes on a count value of 0, 1-3*, 4, 5, 6, . . . 31. The value 1-3* is a suppressed value of broadband providers for confidentiality purposes and has been coded as 2 to be consistent with the literature. In the sample, the mean across all years for providers is 4.7 and a sample variance of 16. For a Poisson distribution, the mean and variance are equal. This may be troublesome if the conditional mean and variance of the model differ by as much as the sample mean and variance. Therefore, a Negative Binomial is also considered (and likely to be the true process).

Other variables used include:

- `iloans` - an indicator variable for whether or not a zip code received a loan. No loans were given before 2002, so this variable does vary by time.

- `log(est)` - this is from zip code business pattern data and is the number of establishments in a particular zip code. I take the log of this variable because the distribution is right-skewed. It is possible to substitute this variable with number of employees, annual payroll, or first quarter payroll from the ZBP but I choose not to because those variables are **suppressed for approximately 9.1% of the zip codes**. Establishments is highly correlated with the other variables anyway, so I would rather use a less precise proxy than potentially bias the sample.
- `log(Pop_IRS)` - IRS has data on number of tax returns filed by county from 1989 until 2013. I use the number of exemptions per county as a way to proxy for the population of a county. This variable is also right-skewed and therefore the log of population is taken instead of population. The alternative for population would be to use US Census data which produce yearly estimates at the county level. These estimates are based off of the 2000 Census and use the demographic age distribution of a county in order to project forward the birth rate and death rate to determine what the population in a county should be. Since this is simply a function of initial conditions in 2000, I choose to use IRS data because there is more variation in the data and it reflects changes in economic conditions across counties that would drive migration (population change).
- `logINC` - this is tabulated from the same IRS data above using Adjusted Gross Income (AGI) at the county level for each year. This is divided by the number of households for a county and is therefore a proxy for mean income, as reported to the IRS, per year. Again, this is a right-skewed variable which is the justification for taking the logarithm of the variable.
- `tri` - stands for Terrain Ruggedness Index which uses elevation data for a given polygon to calculate the feature changes in a given area relative to the entire domain. This is at the ZCTA level across the United States and is thought of as a proxy for increased costs of broadband deployment due to rough terrain. This does not vary across years and so zip code fixed effects will take away this variable.
- `ruc` - the rural-urban continuum code, but for this study I simply use 3 classifications of a county: Metro, Rural but adjacent to a metro county, and Rural but non-adjacent to a metro county. Counties do change across time, but only in years that end in 3 (1993, 2003, . . .). I choose to use the values for 2003 as this would be a little bit before the halfway point in the analysis.

(Quasi-)Poisson Regression Models

I start with by making use of the count nature of the broadband providers variable by assuming it follows a Poisson distribution:

$$Prov_{z,t}|X_{z,t}, \beta \sim Pois\left[exp\left(X'_{z,t}\beta\right)\right]$$

The variable $Prov_{z,t}$ is the number of providers in zip code z at time t. The $X_{z,t}$ are variables at the zip code or county level that determine the level of broadband providers. These are log of establishments, log of population, log of income, terrain ruggedness index, and rural: adjacent and non-adjacent. There is also a time fixed effect for each year included in these regressions.

Biannual Regression

In the basic panel setting, I averaged the two yearly values of broadband. As a naive start to use of the Poisson distribution, I will start with a Poisson regression with the dependent variable as the biannual values for the provider numbers, thus giving a regression where T = 18 and n = 29588.

There are two models: the first without year fixed effects and the second including these. An ANOVA test to determine whether the year fixed effects are jointly significant is performed at the bottom which provides evidence of year fixed effects as a significant predictor of broadband diffusion:

```
##
## Call:
```

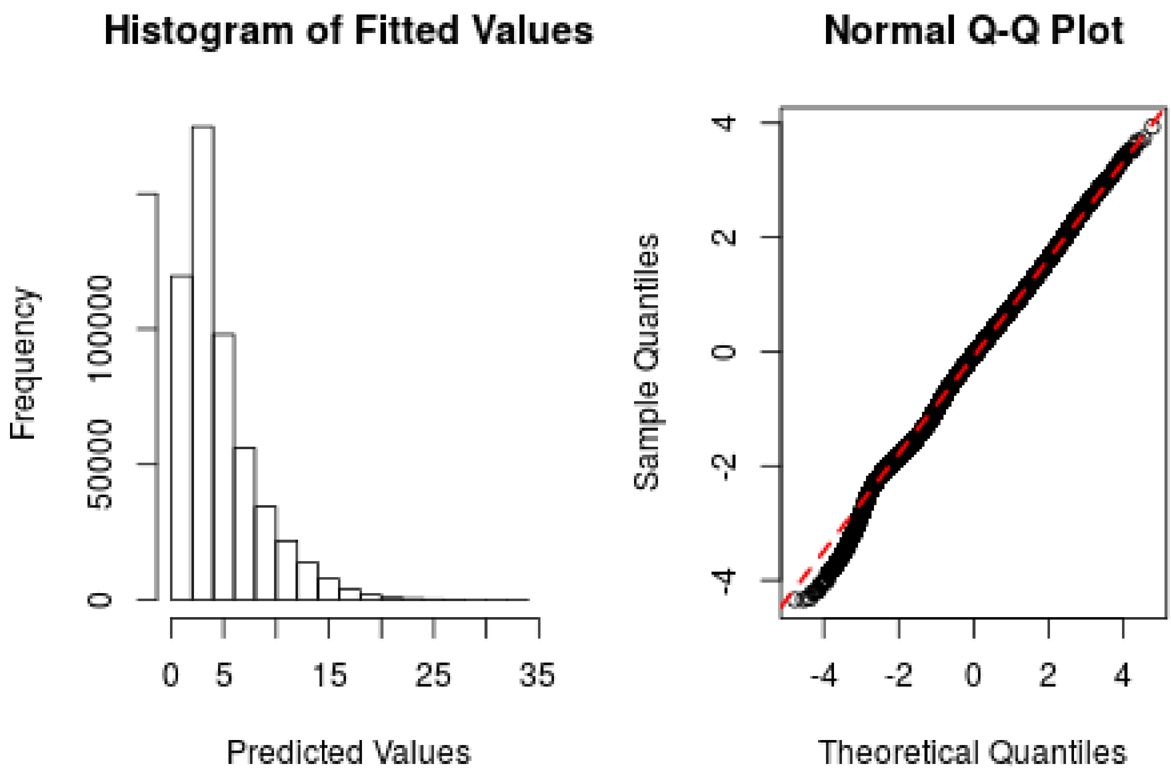
```
## glm(formula = round(Prov_num) ~ iloans + log(est) + log(Pop_IRS) +
##     logINC + tri + ruc, family = poisson, data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.6053  -0.9512  -0.1403   0.7031   4.9672
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.2839285  0.0302704   -9.38   <2e-16 ***
## iloans        0.3239243  0.0035446   91.39   <2e-16 ***
## log(est)      0.2568858  0.0004761  539.52   <2e-16 ***
## log(Pop_IRS)  0.0946551  0.0005661  167.21   <2e-16 ***
## logINC       -0.0420595  0.0029601  -14.21   <2e-16 ***
## tri          -0.0009407  0.0000359  -26.20   <2e-16 ***
## rucadj       -0.0296904  0.0021471  -13.83   <2e-16 ***
## rucnonadj    -0.0414214  0.0026223  -15.80   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1668357  on 532583  degrees of freedom
## Residual deviance:  826533  on 532576  degrees of freedom
## AIC: 2415132
##
## Number of Fisher Scoring iterations: 5
```

```
##
## Call:
## glm(formula = round(Prov_num) ~ iloans + log(est) + log(Pop_IRS) +
##     logINC + tri + factor(year) + ruc, family = poisson, data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.3407  -0.6572  -0.0577   0.4933   3.9359
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -3.014e+00  3.082e-02 -97.795  < 2e-16 ***
## iloans            4.862e-03  3.574e-03   1.360    0.174
## log(est)          2.506e-01  4.761e-04 526.367  < 2e-16 ***
## log(Pop_IRS)      7.857e-02  5.692e-04 138.028  < 2e-16 ***
## logINC            1.310e-01  2.965e-03  44.176  < 2e-16 ***
## tri              -1.330e-03  3.609e-05 -36.851  < 2e-16 ***
## factor(year)2000  3.556e-01  5.501e-03  64.653  < 2e-16 ***
## factor(year)2001  6.853e-01  5.300e-03 129.304  < 2e-16 ***
## factor(year)2002  8.809e-01  5.205e-03 169.234  < 2e-16 ***
## factor(year)2003  1.037e+00  5.140e-03 201.664  < 2e-16 ***
## factor(year)2004  1.107e+00  5.108e-03 216.762  < 2e-16 ***
## factor(year)2005  1.323e+00  5.038e-03 262.538  < 2e-16 ***
## factor(year)2006  1.426e+00  5.011e-03 284.574  < 2e-16 ***
## factor(year)2007  1.535e+00  4.992e-03 307.597  < 2e-16 ***
## factor(year)2008  1.727e+00  5.140e-03 335.974  < 2e-16 ***
```

```
## rucadj           -1.732e-02  2.151e-03  -8.054 8.03e-16 ***
## rucnonadj        -3.344e-02  2.626e-03 -12.733  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1668357  on 532583  degrees of freedom
## Residual deviance:  413700  on 532567  degrees of freedom
## AIC: 2002317
##
## Number of Fisher Scoring iterations: 5
```

```
## Analysis of Deviance Table
##
## Model 1: round(Prov_num) ~ iloans + log(est) + log(Pop_IRS) + logINC +
##     tri + ruc
## Model 2: round(Prov_num) ~ iloans + log(est) + log(Pop_IRS) + logINC +
##     tri + factor(year) + ruc
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1    532576     826533
## 2    532567     413700  9   412833 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The associated coefficients appear to jive with expectations. There is a positive association with establishments, population, income, and metro setting. Further, tri and rural areas (non-adjacent even moreso) are associated with lower levels of broadband access. However, because the functional form of the model is non-linear I hold off on interpretation of the effects. The main point to take-away here is that the loans are not significant. To check if this is a suitable fit of a model, I turn to a histogram of fitted values and Q-Q plot for the model with year fixed effects:



Visual inspection of the fit indicates use of a count model is superior to the previous panel regression. The Fitted Values range versus actual provider counts:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.211   2.134   3.603   4.693   6.203  33.560
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   2.000   4.000   4.693   7.000  31.000
```

So there is still a bit of a right-skew, but the fit appears to match well. I take this as strong evidence that a count model should be used here as opposed to a least squares panel methods.

Annual Regression

To further inspect this relationship, the $Prov_{z,t}$ has been averaged annually and rounded to be an integer value. This is to be consistent with the basic panel data I performed earlier and for ease of interpretation. There are two models presented: Poisson and Quasi-Poisson. The advantage of the Quasi-Poisson model is that the variance is not restricted to be equal to the mean of the distribution as the Poisson regression does:

```
##
## Call:
## glm(formula = round(Prov_num) ~ iloans + log(est) + log(Pop_IRS) +
##     logINC + tri + ruc + factor(year), family = poisson, data = pdata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.2993  -0.6297  -0.0684   0.4722   3.5452
##
```

```
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -2.954e+00  4.108e-02 -71.905  < 2e-16 ***
## iloans            3.828e-03  4.659e-03   0.822    0.411
## log(est)          2.477e-01  6.344e-04 390.353  < 2e-16 ***
## log(Pop_IRS)      7.433e-02  7.605e-04  97.747  < 2e-16 ***
## logINC            1.315e-01  3.970e-03  33.118  < 2e-16 ***
## tri              -1.358e-03  4.818e-05 -28.181  < 2e-16 ***
## rucadj           -1.650e-02  2.865e-03  -5.758 8.51e-09 ***
## rucnonadj        -3.308e-02  3.494e-03  -9.468  < 2e-16 ***
## factor(year)2000  3.573e-01  6.169e-03  57.924  < 2e-16 ***
## factor(year)2001  6.846e-01  5.810e-03 117.826  < 2e-16 ***
## factor(year)2002  8.803e-01  5.636e-03 156.200  < 2e-16 ***
## factor(year)2003  1.036e+00  5.515e-03 187.781  < 2e-16 ***
## factor(year)2004  1.106e+00  5.456e-03 202.786  < 2e-16 ***
## factor(year)2005  1.323e+00  5.322e-03 248.673  < 2e-16 ***
## factor(year)2006  1.426e+00  5.269e-03 270.625  < 2e-16 ***
## factor(year)2007  1.536e+00  5.232e-03 293.530  < 2e-16 ***
## factor(year)2008  1.728e+00  5.146e-03 335.726  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 955360  on 295879  degrees of freedom
## Residual deviance: 212297  on 295863  degrees of freedom
## AIC: 1093451
##
## Number of Fisher Scoring iterations: 5
```

```
##
## Call:
## glm(formula = round(Prov_num) ~ iloans + log(est) + log(Pop_IRS) +
##     logINC + tri + ruc + factor(year), family = quasipoisson,
##     data = pdata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.2993  -0.6297  -0.0684   0.4722   3.5452
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.9536948  0.0328235 -89.987  < 2e-16 ***
## iloans            0.0038283  0.0037232   1.028    0.304
## log(est)          0.2476575  0.0005070 488.513  < 2e-16 ***
## log(Pop_IRS)      0.0743327  0.0006077 122.327  < 2e-16 ***
## logINC            0.1314734  0.0031722  41.445  < 2e-16 ***
## tri              -0.0013578  0.0000385 -35.268  < 2e-16 ***
## rucadj           -0.0164957  0.0022892  -7.206 5.78e-13 ***
## rucnonadj        -0.0330789  0.0027918 -11.849  < 2e-16 ***
## factor(year)2000  0.3573115  0.0049291  72.490  < 2e-16 ***
## factor(year)2001  0.6845882  0.0046427 147.455  < 2e-16 ***
## factor(year)2002  0.8802704  0.0045032 195.479  < 2e-16 ***
## factor(year)2003  1.0355403  0.0044065 235.001  < 2e-16 ***
```

```
## factor(year)2004  1.1063063  0.0043593 253.780  < 2e-16 ***
## factor(year)2005  1.3233855  0.0042524 311.206  < 2e-16 ***
## factor(year)2006  1.4259932  0.0042105 338.678  < 2e-16 ***
## factor(year)2007  1.5356479  0.0041804 367.342  < 2e-16 ***
## factor(year)2008  1.7277668  0.0041123 420.149  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 0.6385025)
##
##     Null deviance: 955360  on 295879  degrees of freedom
## Residual deviance: 212297  on 295863  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

WOW, the residual fit for these models is vastly improved. The Quasi-Poisson model appears to be a better fit as the dispersion parameter is significant and fits the data much better as the implied variance in the predicted values is closer to the sample variance. To note, the point estimates are identical and only the standard errors differ between the two models.

From a policy perspective, it appears that the loan program was not effective as we cannot find a significant effect in the regression. Keep in mind that there is likely attenuation bias due to imprecise measurement of the dependent variable of interest and this is only the beginning of modeling. I have yet to disaggregate the loan variables or interact them with their intended beneficiaries: rural areas. We know that the loans were disbursed to both metro and rural regions, which may muddy the waters of finding this as a program that can increase access for under-served and remote areas. Further, these loans were not exclusive to broadband deployment either.

A few more checks of the model is necessary before I can start fidgeting around with interaction terms of the loans. For one, serial correlation appears to still be a problem. We can see this through a correlation matrix of residuals across years for both the Poisson and Quasi-Poisson Regressions:

This is not good, although the typical applied remedy is to simply use robust standard errors in significance testing. The `vcovHAC` function in `R` allows for heteroskedastic and autocorrelation consistent covariance matrix estimation. Applying this to our previous models would yield different standard errors, although this does not change our qualitative understanding of the loan program so far. Because this is a computationally intensive process (and has yet to converge for any models so far), I will suppress the output for the time being. However, I will present the `vcovHC` results though which are the White standard errors:

```
##
## z test of coefficients:
##
##                     Estimate  Std. Error  z value  Pr(>|z|)
## (Intercept)      -2.9537e+00  3.5103e-02 -84.1439 < 2.2e-16 ***
## iloans            3.8283e-03  3.5050e-03   1.0923    0.2747
## log(est)          2.4766e-01  5.0133e-04 493.9958 < 2.2e-16 ***
## log(Pop_IRS)      7.4333e-02  6.6249e-04 112.2015 < 2.2e-16 ***
## logINC            1.3147e-01  3.4335e-03  38.2915 < 2.2e-16 ***
## tri              -1.3578e-03  3.6783e-05 -36.9128 < 2.2e-16 ***
## rucadj           -1.6496e-02  2.1063e-03  -7.8315  4.82e-15 ***
## rucnonadj        -3.3079e-02  2.6699e-03 -12.3894 < 2.2e-16 ***
## factor(year)2000  3.5731e-01  5.4344e-03  65.7495 < 2.2e-16 ***
## factor(year)2001  6.8459e-01  5.1034e-03 134.1436 < 2.2e-16 ***
## factor(year)2002  8.8027e-01  4.9762e-03 176.8973 < 2.2e-16 ***
```

```
## factor(year)2003  1.0355e+00  4.8491e-03 213.5526 < 2.2e-16 ***
## factor(year)2004  1.1063e+00  4.6958e-03 235.5926 < 2.2e-16 ***
## factor(year)2005  1.3234e+00  4.5643e-03 289.9458 < 2.2e-16 ***
## factor(year)2006  1.4260e+00  4.5202e-03 315.4694 < 2.2e-16 ***
## factor(year)2007  1.5356e+00  4.5454e-03 337.8460 < 2.2e-16 ***
## factor(year)2008  1.7278e+00  4.4988e-03 384.0534 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This is slightly puzzling that the standard errors are generally smaller in this setting. This gives me pause as to how to approach the calculation of standard errors in this setting…

Extension

How about let's do the natural extension to this model to see if the impact of the loans varied across the **rural** counties. Since the Quasi-Poisson model is less restrictive, that will be the estimation technique I will use. The specific regression formula can be seen at the top of the output:

```
##
## Call:
## glm(formula = round(Prov_num) ~ iloans + iloans:ruc + log(est) +
##     log(Pop_IRS) + logINC + tri + ruc + factor(year), family = quasipoisson,
##     data = pdata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.3026  -0.6287  -0.0684   0.4720   3.5507
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.945e+00  3.281e-02 -89.750   <2e-16 ***
## iloans           -4.754e-02  4.942e-03  -9.620   <2e-16 ***
## log(est)          2.474e-01  5.070e-04 488.017   <2e-16 ***
## log(Pop_IRS)      7.450e-02  6.078e-04 122.576   <2e-16 ***
## logINC            1.307e-01  3.171e-03  41.221   <2e-16 ***
## tri              -1.351e-03  3.849e-05 -35.097   <2e-16 ***
## rucadj           -2.176e-02  2.324e-03  -9.367   <2e-16 ***
## rucnonadj        -3.830e-02  2.834e-03 -13.516   <2e-16 ***
## factor(year)2000  3.573e-01  4.927e-03  72.519   <2e-16 ***
## factor(year)2001  6.846e-01  4.641e-03 147.502   <2e-16 ***
## factor(year)2002  8.801e-01  4.502e-03 195.520   <2e-16 ***
## factor(year)2003  1.035e+00  4.405e-03 235.056   <2e-16 ***
## factor(year)2004  1.106e+00  4.358e-03 253.752   <2e-16 ***
## factor(year)2005  1.323e+00  4.251e-03 311.292   <2e-16 ***
## factor(year)2006  1.426e+00  4.209e-03 338.819   <2e-16 ***
## factor(year)2007  1.536e+00  4.179e-03 367.480   <2e-16 ***
## factor(year)2008  1.728e+00  4.111e-03 420.315   <2e-16 ***
## iloans:rucadj     1.197e-01  8.562e-03  13.985   <2e-16 ***
## iloans:rucnonadj  1.236e-01  1.038e-02  11.910   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 0.6380285)
```

```
##
##     Null deviance: 955360  on 295879  degrees of freedom
## Residual deviance: 212128  on 295861  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

Interesting, although not necessarily unexpected. One thing we did here is split up a heterogeneous group into three fairly homogeneous groups in order to evaluate the effects. In this setting, the loan program applied to metro counties actually suppressed broadband providers across this time at -0.0475, but loans applied to rural adjacent counties would be a combination of three parameters. The sum of these parameters is 0.0504323 and for the rural nonadjacent counties it is 0.0378034. In order to test the significance of these effects, we can construct a Chisq test to test whether the summation of the respective variables are different from zero:

```
## Linear hypothesis test
##
## Hypothesis:
## iloans  + rucadj  + iloans:rucadj = 0
##
## Model 1: restricted model
## Model 2: round(Prov_num) ~ iloans + iloans:ruc + log(est) + log(Pop_IRS) +
##     logINC + tri + ruc + factor(year)
##
##   Res.Df Df  Chisq Pr(>Chisq)
## 1 295862
## 2 295861  1 51.297  7.939e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Linear hypothesis test
##
## Hypothesis:
## iloans  + rucnonadj  + iloans:rucnonadj = 0
##
## Model 1: restricted model
## Model 2: round(Prov_num) ~ iloans + iloans:ruc + log(est) + log(Pop_IRS) +
##     logINC + tri + ruc + factor(year)
##
##   Res.Df Df  Chisq Pr(>Chisq)
## 1 295862
## 2 295861  1 17.149  3.456e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can see that the effects are both statistically significant, which gives a bit of a puzzling result on the loans effectiveness. Rural areas see a significant increase in availability of broadband via both the Pilot and Farm Bill loan programs.

However, the urban areas that received the loans actually saw a decrease in availability on average. There is potential that some (potentially unobservable) factor is retarding the growth in broadband for these areas that I am not controlling for. If this is the case, there is little I can do to further examine this issue. But one might also consider asking the question of why these particular loans were given out in the first place. These areas were not the targetted recepients of the loans and one interpretation of this is that loans to these areas would not be expected to see increases. There is potential for fraud, but this is equally untestable as is an unobservable driving the results.

Here is a regression with the program broken into the Pilot and the Farm Bill program:

```
##
## Call:
## glm(formula = round(Prov_num) ~ ipilot + icur + log(est) + log(Pop_IRS) +
##     logINC + tri + ruc + factor(year), family = quasipoisson,
##     data = pdata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.2995  -0.6297  -0.0684   0.4722   3.5451
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.9505902  0.0328322 -89.869  < 2e-16 ***
## ipilot            0.0265110  0.0072756   3.644 0.000269 ***
## icur             -0.0049819  0.0042629  -1.169 0.242544
## log(est)          0.2476513  0.0005070 488.516  < 2e-16 ***
## log(Pop_IRS)      0.0744481  0.0006085 122.346  < 2e-16 ***
## logINC            0.1310553  0.0031740  41.290  < 2e-16 ***
## tri              -0.0013590  0.0000385 -35.299  < 2e-16 ***
## rucadj           -0.0162837  0.0022898  -7.111 1.15e-12 ***
## rucnonadj        -0.0330851  0.0027918 -11.851  < 2e-16 ***
## factor(year)2000  0.3573180  0.0049289  72.494  < 2e-16 ***
## factor(year)2001  0.6845723  0.0046425 147.457  < 2e-16 ***
## factor(year)2002  0.8800139  0.0045036 195.405  < 2e-16 ***
## factor(year)2003  1.0352821  0.0044069 234.920  < 2e-16 ***
## factor(year)2004  1.1061270  0.0043595 253.730  < 2e-16 ***
## factor(year)2005  1.3233709  0.0042523 311.213  < 2e-16 ***
## factor(year)2006  1.4261713  0.0042104 338.727  < 2e-16 ***
## factor(year)2007  1.5357878  0.0041803 367.389  < 2e-16 ***
## factor(year)2008  1.7279229  0.0041122 420.199  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 0.6384543)
##
##     Null deviance: 955360  on 295879  degrees of freedom
## Residual deviance: 212288  on 295862  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

In this setting, we see that the pilot loan was statistically significant with a coefficient estimate of 0.0265 while the Farm Bill program is not statistically different from zero with its coefficient of -0.00498. This appears to give evidence that the loans take time to be effective. The Pilot loans were disbursed in 2002 while the panel goes up to 2008, which appears to give adequate time for the programs to contribute towards an increase in availability. On the other hand, the Farm Bill loans were disbursed in 2005 and 2006.

And finally, a cumbersome interaction between the Pilot/Farm Bill with the Rural-Urban Continuum:

```
##
```

```
## Call:
## glm(formula = round(Prov_num) ~ ipilot + icur + ipilot:ruc +
##     icur:ruc + log(est) + log(Pop_IRS) + logINC + tri + ruc +
##     factor(year), family = quasipoisson, data = pdata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.3034  -0.6287  -0.0683   0.4720   3.5508
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.937e+00  3.282e-02 -89.490  < 2e-16 ***
## ipilot            2.762e-02  9.985e-03   2.766  0.00567 **
## icur             -7.086e-02  5.614e-03 -12.622  < 2e-16 ***
## log(est)          2.473e-01  5.070e-04 487.814  < 2e-16 ***
## log(Pop_IRS)      7.472e-02  6.088e-04 122.746  < 2e-16 ***
## logINC            1.298e-01  3.173e-03  40.896  < 2e-16 ***
## tri              -1.353e-03  3.848e-05 -35.147  < 2e-16 ***
## rucadj           -2.162e-02  2.324e-03  -9.305  < 2e-16 ***
## rucnonadj        -3.806e-02  2.834e-03 -13.428  < 2e-16 ***
## factor(year)2000  3.573e-01  4.927e-03  72.533  < 2e-16 ***
## factor(year)2001  6.845e-01  4.640e-03 147.518  < 2e-16 ***
## factor(year)2002  8.799e-01  4.501e-03 195.477  < 2e-16 ***
## factor(year)2003  1.035e+00  4.405e-03 235.011  < 2e-16 ***
## factor(year)2004  1.106e+00  4.358e-03 253.709  < 2e-16 ***
## factor(year)2005  1.323e+00  4.250e-03 311.331  < 2e-16 ***
## factor(year)2006  1.426e+00  4.208e-03 338.901  < 2e-16 ***
## factor(year)2007  1.536e+00  4.178e-03 367.552  < 2e-16 ***
## factor(year)2008  1.728e+00  4.110e-03 420.402  < 2e-16 ***
## ipilot:rucadj    -1.795e-02  1.833e-02  -0.979  0.32739
## ipilot:rucnonadj  1.981e-02  1.764e-02   1.124  0.26121
## icur:rucadj       1.585e-01  9.559e-03  16.584  < 2e-16 ***
## icur:rucnonadj    1.651e-01  1.280e-02  12.894  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 0.6378272)
##
##     Null deviance: 955360  on 295879  degrees of freedom
## Residual deviance: 212061  on 295858  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

Under this regression, we can see that the Pilot program appears to have benefited the metro areas. The rural adjacent areas receiving the Pilot loans appeared to have a decrease in availability with an associated coefficient of -0.0119498. The rural nonadjacent areas did appear to benefit with an associated coefficient of 0.0093809. However, these are not statistically different from zero:

```
## Linear hypothesis test
##
## Hypothesis:
## ipilot  + rucadj  + ipilot:rucadj = 0
##
```

```
## Model 1: restricted model
## Model 2: round(Prov_num) ~ ipilot + icur + ipilot:ruc + icur:ruc + log(est) +
##     log(Pop_IRS) + logINC + tri + ruc + factor(year)
##
##   Res.Df Df  Chisq Pr(>Chisq)
## 1 295859
## 2 295858  1 0.6011     0.4382
```

```
## Linear hypothesis test
##
## Hypothesis:
## ipilot  + rucnonadj  + ipilot:rucnonadj = 0
##
## Model 1: restricted model
## Model 2: round(Prov_num) ~ ipilot + icur + ipilot:ruc + icur:ruc + log(est) +
##     log(Pop_IRS) + logINC + tri + ruc + factor(year)
##
##   Res.Df Df Chisq Pr(>Chisq)
## 1 295859
## 2 295858  1 0.415     0.5194
```

If we extend this to the Farm Bill loans, we see that the metro areas had a statistically significant decrease in availability. The rural areas appear to have a statistically significant increase of 0.0660515 for adjacent areas and 0.0561825 for nonadjacent areas.

```
## Linear hypothesis test
##
## Hypothesis:
## icur  + rucadj  + icur:rucadj = 0
##
## Model 1: restricted model
## Model 2: round(Prov_num) ~ ipilot + icur + ipilot:ruc + icur:ruc + log(est) +
##     log(Pop_IRS) + logINC + tri + ruc + factor(year)
##
##   Res.Df Df  Chisq Pr(>Chisq)
## 1 295859
## 2 295858  1 71.948  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Linear hypothesis test
##
## Hypothesis:
## icur  + rucnonadj  + icur:rucnonadj = 0
##
## Model 1: restricted model
## Model 2: round(Prov_num) ~ ipilot + icur + ipilot:ruc + icur:ruc + log(est) +
##     log(Pop_IRS) + logINC + tri + ruc + factor(year)
##
##   Res.Df Df  Chisq Pr(>Chisq)
## 1 295859
## 2 295858  1 23.884  1.023e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Cool stuff. I imagine this is something that I could run with and have the first part of this chapter effectively done except for one other problem . . . .