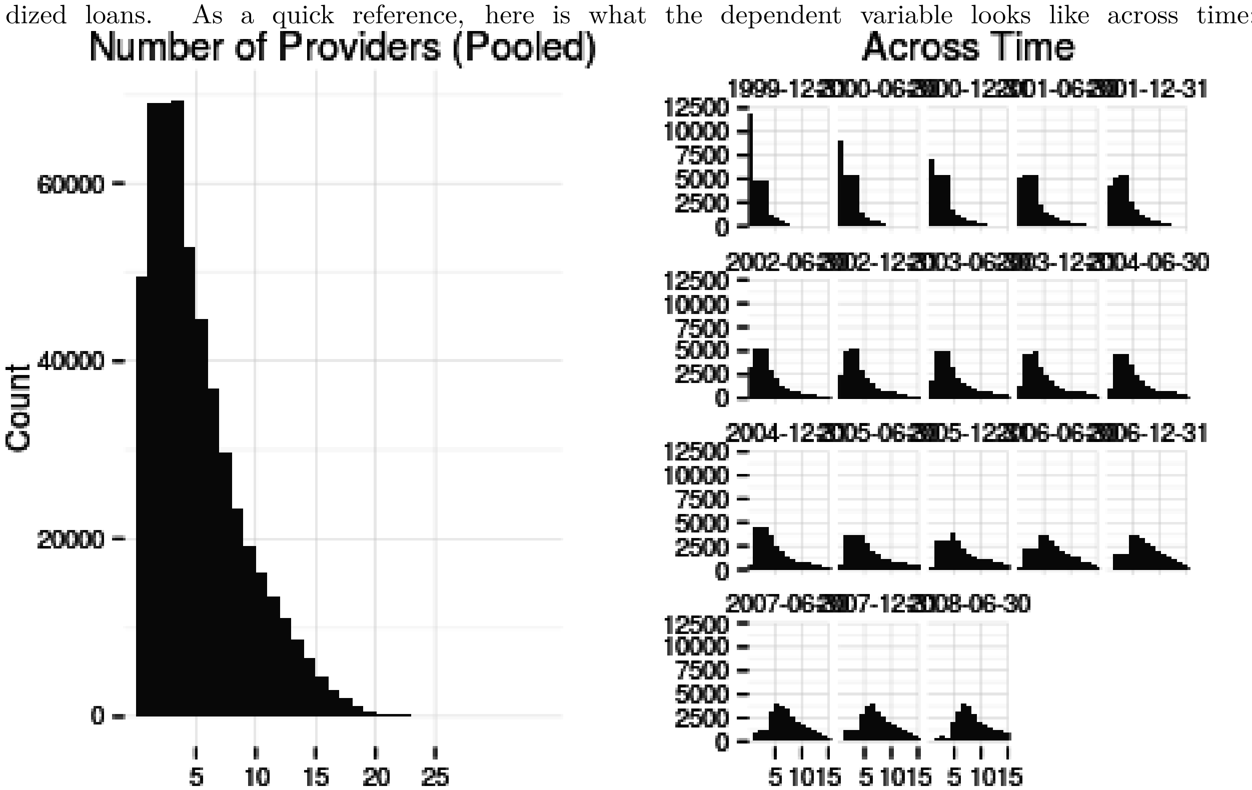
4-USDA Evaluation CAR Problem

Robert Dinterman

2015-11-29

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:



Suppressed values were drawn from a uniform distribution between 1 and 3 for visual purposes.

Data on broadband providers is measured twice a year (June 30 and December 31) and 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 15. For a Poisson distribution, the mean and variance are restricted to be equal which indicates here that the unconditional distribution is not likely to be Poisson. This may be troublesome if the conditional mean and variance of the model differ by as much as the sample mean and variance. Therefore, Quasi-Poisson and Negative Binomial models are 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 and the central focus. No loans were given before 2002, so this variable does vary by time. This can be broken down further to `ipilot` to indicate the smaller Pilot program which lasted from 2001 to 2003 and disbursed 28 loans totaling \$180 million as well as the Post-Pilot program, `ibip1234`, which was established by the 2002 Farm Bill. The Post-Pilot program began disbursing loans in 2003 and we have data until 2006 which totals \$1.22 billion across 70 loans.

- `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% 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).
- `logAPay_R2` - this also comes from the ZIP code business pattern data and is the total annual payroll for all establishments in the ZIP code divided by the number of employees. This proxies for wages and income to a degree. The suppressed values are replaced with the national average which would have the effect of dampening any effect. Suppressed values are not necessarily because of a low number of establishments in a ZIP code as it could also result from disclosure reasons as to not identify a dominant firm in a ZIP code. Since the focus of this analysis is on the effect of broadband loans on the number of providers, the only concern I would have with the suppression issues is if they systematically affected both the loans and the number of broadband providers.
- `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.

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:

$$Prov_{z,t}|X_{z,t}, \beta \sim Pois\left(\lambda_{z,t}\right)$$
$$log\left(\lambda_{z,t}\right) = \beta_0 + \beta_1 Loan_{z,t} + \beta_2 X_{z,t} + \tau_t + \varepsilon_{z,t}$$

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. The biannual values for the provider numbers gives a panel dataset where T = 18 and n = 29588.

As a naive start to use of the Poisson distribution, I will start with two models: the first without time fixed effects and the second including these. An ANOVA test to determine whether the time fixed effects are jointly significant is performed at the bottom which provides evidence of time fixed effects as a significant predictor of broadband diffusion:

```
##
## Call:
## glm(formula = Prov_num ~ iloans + log(est) + log(Pop_IRS) + logAPay_R2 +
##     tri + ruc + poly(AREA_zcta, 2) + I(Pop_IRS/AREA_cty) + I(est/AREA_zcta),
##     family = poisson, data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.6120  -0.9492  -0.1417   0.6980   5.0418
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -1.736e+00  2.042e-02 -85.001  < 2e-16 ***
## iloans               3.296e-01  3.548e-03  92.901  < 2e-16 ***
## log(est)             2.482e-01  5.036e-04 492.777  < 2e-16 ***
## log(Pop_IRS)         8.273e-02  5.891e-04 140.439  < 2e-16 ***
## logAPay_R2           1.170e-01  2.096e-03  55.813  < 2e-16 ***
## tri                 -8.821e-04  3.658e-05 -24.110  < 2e-16 ***
## rucadj              -2.052e-02  2.110e-03  -9.725  < 2e-16 ***
## rucnonadj           -3.179e-02  2.619e-03 -12.141  < 2e-16 ***
## poly(AREA_zcta, 2)1 -3.215e+00  5.785e-01  -5.558 2.74e-08 ***
## poly(AREA_zcta, 2)2  1.564e+00  5.452e-01   2.868  0.00413 **
## I(Pop_IRS/AREA_cty)  4.732e-06  2.133e-07  22.185  < 2e-16 ***
## I(est/AREA_zcta)    -5.816e-06  9.377e-07  -6.203 5.55e-10 ***
## ---
## 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:  822893  on 532572  degrees of freedom
## AIC: 2411499
##
## Number of Fisher Scoring iterations: 5
```

```
##
## Call:
## glm(formula = Prov_num ~ iloans + log(est) + log(Pop_IRS) + logAPay_R2 +
##     tri + ruc + poly(AREA_zcta, 2) + I(Pop_IRS/AREA_cty) + I(est/AREA_zcta) +
##     factor(time), family = poisson, data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.3984  -0.6506  -0.0564   0.4914   3.9213
##
## Coefficients:
##                          Estimate Std. Error  z value Pr(>|z|)
## (Intercept)            -2.510e+00  2.092e-02 -120.028  < 2e-16 ***
## iloans                  9.819e-03  3.578e-03    2.744 0.006068 **
## log(est)                2.465e-01  5.038e-04  489.233  < 2e-16 ***
## log(Pop_IRS)            7.686e-02  5.895e-04  130.390  < 2e-16 ***
```

```
## logAPay_R2             9.281e-02  2.092e-03   44.369  < 2e-16 ***
## tri                   -1.141e-03  3.669e-05  -31.109  < 2e-16 ***
## rucadj                -3.492e-02  2.108e-03  -16.563  < 2e-16 ***
## rucnonadj             -5.036e-02  2.618e-03  -19.234  < 2e-16 ***
## poly(AREA_zcta, 2)1    -3.770e+00  5.784e-01   -6.518 7.11e-11 ***
## poly(AREA_zcta, 2)2     1.863e+00  5.426e-01    3.433 0.000597 ***
## I(Pop_IRS/AREA_cty)     4.020e-06  2.116e-07   18.997  < 2e-16 ***
## I(est/AREA_zcta)      -1.517e-06  9.241e-07   -1.642 0.100590
## factor(time)2000-06-30  2.207e-01  6.354e-03   34.743  < 2e-16 ***
## factor(time)2000-12-31  4.785e-01  6.025e-03   79.422  < 2e-16 ***
## factor(time)2001-06-30  6.478e-01  5.840e-03  110.916  < 2e-16 ***
## factor(time)2001-12-31  7.128e-01  5.777e-03  123.387  < 2e-16 ***
## factor(time)2002-06-30  8.306e-01  5.664e-03  146.659  < 2e-16 ***
## factor(time)2002-12-31  9.111e-01  5.597e-03  162.773  < 2e-16 ***
## factor(time)2003-06-30  1.006e+00  5.524e-03  182.031  < 2e-16 ***
## factor(time)2003-12-31  1.050e+00  5.492e-03  191.097  < 2e-16 ***
## factor(time)2004-06-30  1.079e+00  5.467e-03  197.400  < 2e-16 ***
## factor(time)2004-12-31  1.123e+00  5.438e-03  206.557  < 2e-16 ***
## factor(time)2005-06-30  1.270e+00  5.346e-03  237.631  < 2e-16 ***
## factor(time)2005-12-31  1.362e+00  5.296e-03  257.232  < 2e-16 ***
## factor(time)2006-06-30  1.389e+00  5.282e-03  263.073  < 2e-16 ***
## factor(time)2006-12-31  1.452e+00  5.250e-03  276.619  < 2e-16 ***
## factor(time)2007-06-30  1.515e+00  5.216e-03  290.383  < 2e-16 ***
## factor(time)2007-12-31  1.526e+00  5.211e-03  292.881  < 2e-16 ***
## factor(time)2008-06-30  1.717e+00  5.136e-03  334.331  < 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:  408991  on 532555  degrees of freedom
## AIC: 1997631
##
## Number of Fisher Scoring iterations: 5
```

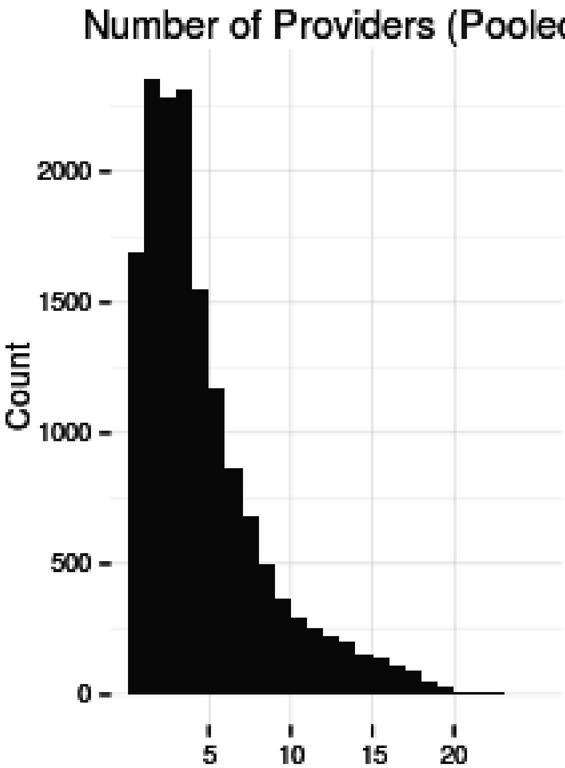
```
## Analysis of Deviance Table
##
## Model 1: Prov_num ~ iloans + log(est) + log(Pop_IRS) + logAPay_R2 + tri +
##     ruc + poly(AREA_zcta, 2) + I(Pop_IRS/AREA_cty) + I(est/AREA_zcta)
## Model 2: Prov_num ~ iloans + log(est) + log(Pop_IRS) + logAPay_R2 + tri +
##     ruc + poly(AREA_zcta, 2) + I(Pop_IRS/AREA_cty) + I(est/AREA_zcta) +
##     factor(time)
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1    532572     822893
## 2    532555     408991 17   413902 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

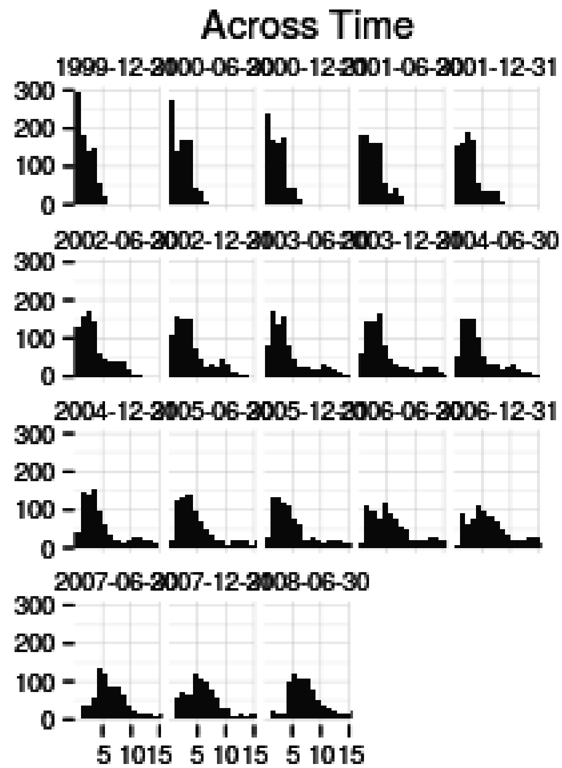
The associated coefficients appear to jive with expectations. There is a positive association with establishments, population, income, and metro setting. Further, tri and rural areas (non-adjacent even more) are associated with lower levels of broadband access.

Spatial Effects

I cannot get the package `CARBayesST` to run when I use all 30,000 ZIP codes across 18 time periods, but I can do this for Minnesota. Minnesota is a good candidate for examining spatial effects of the broadband loan programs because it is a state that received both Pilot and Farm Bill loans to both Metro and Rural areas. I begin with a recap of the previous results across MLE estimates of the Poisson models and relevant summary statistics:

Minnesota





The first step is to run the aspatial Poisson models with Minnesota to verify similar results to the rest of the United States:

```
##
## Call:
## glm(formula = Prov_num ~ iloans + log(est) + log(Pop_IRS) + logAPay_R2 +
##     tri + ruc + poly(AREA_zcta, 2) + I(Pop_IRS/AREA_cty) + I(est/AREA_zcta),
##     family = "poisson", data = STdata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.7274  -0.8919  -0.1402   0.6509   3.6736
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -2.010e+00  1.432e-01 -14.033  < 2e-16 ***
## iloans               3.866e-01  1.016e-02  38.043  < 2e-16 ***
## log(est)             2.847e-01  3.922e-03  72.576  < 2e-16 ***
```

```
## log(Pop_IRS)         6.861e-02  6.043e-03  11.354  < 2e-16 ***
## logAPay_R2           1.276e-01  1.393e-02   9.162  < 2e-16 ***
## tri                  5.250e-03  7.921e-04   6.627 3.42e-11 ***
## rucadj               5.071e-02  1.429e-02   3.549 0.000387 ***
## rucnonadj           -2.544e-02  1.546e-02  -1.646 0.099816 .
## poly(AREA_zcta, 2)1 -2.577e+00  6.341e-01  -4.063 4.84e-05 ***
## poly(AREA_zcta, 2)2  2.182e+00  6.772e-01   3.222 0.001271 **
## I(Pop_IRS/AREA_cty)  9.789e-05  9.279e-06  10.550  < 2e-16 ***
## I(est/AREA_zcta)    -2.407e-05  6.874e-06  -3.502 0.000462 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 48417  on 15353  degrees of freedom
## Residual deviance: 21552  on 15342  degrees of freedom
## AIC: 65011
##
## Number of Fisher Scoring iterations: 5
```

```
##
## Call:
## glm(formula = Prov_num ~ iloans + log(est) + log(Pop_IRS) + logAPay_R2 +
##     tri + ruc + poly(AREA_zcta, 2) + I(Pop_IRS/AREA_cty) + I(est/AREA_zcta) +
##     factor(time), family = "poisson", data = STdata)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.75824  -0.57885  -0.00613   0.47851   2.47935
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -2.405e+00  1.438e-01 -16.723  < 2e-16 ***
## iloans                 6.135e-02  1.049e-02   5.846 5.04e-09 ***
## log(est)               2.986e-01  3.936e-03  75.861  < 2e-16 ***
## log(Pop_IRS)           6.456e-02  6.031e-03  10.704  < 2e-16 ***
## logAPay_R2             7.847e-02  1.375e-02   5.706 1.16e-08 ***
## tri                    2.939e-03  7.945e-04   3.700 0.000216 ***
## rucadj                 1.535e-02  1.423e-02   1.079 0.280795
## rucnonadj             -4.752e-02  1.542e-02  -3.081 0.002061 **
## poly(AREA_zcta, 2)1    -3.249e+00  6.334e-01  -5.129 2.91e-07 ***
## poly(AREA_zcta, 2)2     2.362e+00  6.737e-01   3.506 0.000455 ***
## I(Pop_IRS/AREA_cty)     6.865e-05  9.262e-06   7.412 1.24e-13 ***
## I(est/AREA_zcta)       -2.592e-05  6.981e-06  -3.712 0.000205 ***
## factor(time)2000-06-30  5.964e-02  3.860e-02   1.545 0.122342
## factor(time)2000-12-31  1.506e-01  3.778e-02   3.985 6.76e-05 ***
## factor(time)2001-06-30  3.417e-01  3.620e-02   9.440  < 2e-16 ***
## factor(time)2001-12-31  4.162e-01  3.565e-02  11.672  < 2e-16 ***
## factor(time)2002-06-30  6.212e-01  3.426e-02  18.133  < 2e-16 ***
## factor(time)2002-12-31  7.556e-01  3.351e-02  22.548  < 2e-16 ***
## factor(time)2003-06-30  8.410e-01  3.306e-02  25.441  < 2e-16 ***
## factor(time)2003-12-31  9.624e-01  3.250e-02  29.614  < 2e-16 ***
## factor(time)2004-06-30  8.833e-01  3.283e-02  26.901  < 2e-16 ***
## factor(time)2004-12-31  9.693e-01  3.245e-02  29.872  < 2e-16 ***
```

```
## factor(time)2005-06-30  1.131e+00  3.180e-02  35.563  < 2e-16 ***
## factor(time)2005-12-31  1.112e+00  3.187e-02  34.907  < 2e-16 ***
## factor(time)2006-06-30  1.260e+00  3.139e-02  40.150  < 2e-16 ***
## factor(time)2006-12-31  1.374e+00  3.103e-02  44.270  < 2e-16 ***
## factor(time)2007-06-30  1.467e+00  3.076e-02  47.689  < 2e-16 ***
## factor(time)2007-12-31  1.376e+00  3.101e-02  44.355  < 2e-16 ***
## factor(time)2008-06-30  1.607e+00  3.042e-02  52.827  < 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: 48417.4  on 15353  degrees of freedom
## Residual deviance:  9986.4  on 15325  degrees of freedom
## AIC: 53480
##
## Number of Fisher Scoring iterations: 5
```

```
## Analysis of Deviance Table
##
## Model 1: Prov_num ~ iloans + log(est) + log(Pop_IRS) + logAPay_R2 + tri +
##     ruc + poly(AREA_zcta, 2) + I(Pop_IRS/AREA_cty) + I(est/AREA_zcta)
## Model 2: Prov_num ~ iloans + log(est) + log(Pop_IRS) + logAPay_R2 + tri +
##     ruc + poly(AREA_zcta, 2) + I(Pop_IRS/AREA_cty) + I(est/AREA_zcta) +
##     factor(time)
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1     15342    21551.9
## 2     15325     9986.4 17    11566 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The time fixed effects are still significant as seen with the F-test at the bottom of the output above. The major concern is that the coefficients across the Minnesota model and the United States model are similar and for the model without time fixed effects this is generally the case with exception to `tri` and `rucadj` having different signs and `logAPay_R2` having a stronger effect for Minnesota. The standard errors are noticeably larger for the Minnesota regression which is to be expected when reducing the number of ZIP codes. For the time effects, a the coefficient on `iloans` is larger for Minnesota while the same issues with `tri`, `rucadj`, and `logAPay_R2` arise. There may be better candidate states (Texas and Missouri come to mind while looking at the map of the broadband loan program disbursement), but Minnesota still seems plausible to analyze for spatial effects.

Next, I run the previous models as Bayesian models for comparability of MLE and Bayesian results (and because I have only been able to figure out how to implement spatial models with Bayesian methods):

```
##
## Iterations = 1001:11000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##                             Mean        SD  Naive SE Time-series SE
```

```
## (Intercept)         -2.006e+00 1.395e-01 1.395e-03      1.070e-02
## iloans               3.865e-01 9.545e-03 9.545e-05      6.693e-04
## log(est)             2.852e-01 3.616e-03 3.616e-05      2.625e-04
## log(Pop_IRS)         6.858e-02 5.989e-03 5.989e-05      4.610e-04
## logAPay_R2           1.270e-01 1.312e-02 1.312e-04      9.739e-04
## tri                  5.268e-03 7.559e-04 7.559e-06      5.575e-05
## rucadj               5.119e-02 1.399e-02 1.399e-04      1.038e-03
## rucnonadj           -2.556e-02 1.636e-02 1.636e-04      1.177e-03
## poly(AREA_zcta, 2)1 -2.655e+00 5.645e-01 5.645e-03      3.996e-02
## poly(AREA_zcta, 2)2  2.191e+00 7.068e-01 7.068e-03      5.475e-02
## I(Pop_IRS/AREA_cty)  9.779e-05 8.872e-06 8.872e-08      6.474e-07
## I(est/AREA_zcta)    -2.454e-05 6.803e-06 6.803e-08      5.076e-07
##
## 2. Quantiles for each variable:
##
##                           2.5%        25%        50%        75%      97.5%
## (Intercept)         -2.258e+00 -2.113e+00 -2.005e+00 -1.898e+00 -1.755e+00
## iloans               3.686e-01  3.800e-01  3.861e-01  3.923e-01  4.052e-01
## log(est)             2.781e-01  2.827e-01  2.851e-01  2.876e-01  2.920e-01
## log(Pop_IRS)         5.738e-02  6.416e-02  6.826e-02  7.273e-02  8.100e-02
## logAPay_R2           1.014e-01  1.180e-01  1.270e-01  1.360e-01  1.508e-01
## tri                  3.815e-03  4.775e-03  5.249e-03  5.807e-03  6.756e-03
## rucadj               2.510e-02  4.103e-02  5.144e-02  6.047e-02  8.119e-02
## rucnonadj           -5.576e-02 -3.812e-02 -2.426e-02 -1.510e-02  7.093e-03
## poly(AREA_zcta, 2)1 -3.815e+00 -3.027e+00 -2.639e+00 -2.272e+00 -1.576e+00
## poly(AREA_zcta, 2)2  8.368e-01  1.734e+00  2.235e+00  2.645e+00  3.465e+00
## I(Pop_IRS/AREA_cty)  7.745e-05  9.283e-05  9.753e-05  1.036e-04  1.155e-04
## I(est/AREA_zcta)    -3.763e-05 -2.883e-05 -2.462e-05 -2.003e-05 -1.158e-05
```

```
##
## Iterations = 1001:11000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##                              Mean SD Naive SE Time-series SE
## (Intercept)            -2.405e+00  0        0              0
## iloans                  6.135e-02  0        0              0
## log(est)                2.986e-01  0        0              0
## log(Pop_IRS)            6.456e-02  0        0              0
## logAPay_R2              7.847e-02  0        0              0
## tri                     2.939e-03  0        0              0
## rucadj                  1.535e-02  0        0              0
## rucnonadj              -4.752e-02  0        0              0
## poly(AREA_zcta, 2)1    -3.249e+00  0        0              0
## poly(AREA_zcta, 2)2     2.362e+00  0        0              0
## I(Pop_IRS/AREA_cty)     6.866e-05  0        0              0
## I(est/AREA_zcta)       -2.592e-05  0        0              0
## factor(time)2000-06-30  5.964e-02  0        0              0
## factor(time)2000-12-31  1.506e-01  0        0              0
## factor(time)2001-06-30  3.417e-01  0        0              0
```

```
## factor(time)2001-12-31  4.162e-01  0        0              0
## factor(time)2002-06-30  6.212e-01  0        0              0
## factor(time)2002-12-31  7.556e-01  0        0              0
## factor(time)2003-06-30  8.410e-01  0        0              0
## factor(time)2003-12-31  9.624e-01  0        0              0
## factor(time)2004-06-30  8.833e-01  0        0              0
## factor(time)2004-12-31  9.693e-01  0        0              0
## factor(time)2005-06-30  1.131e+00  0        0              0
## factor(time)2005-12-31  1.112e+00  0        0              0
## factor(time)2006-06-30  1.260e+00  0        0              0
## factor(time)2006-12-31  1.374e+00  0        0              0
## factor(time)2007-06-30  1.467e+00  0        0              0
## factor(time)2007-12-31  1.376e+00  0        0              0
## factor(time)2008-06-30  1.607e+00  0        0              0
##
## 2. Quantiles for each variable:
##
##                              2.5%        25%        50%        75%
## (Intercept)            -2.405e+00 -2.405e+00 -2.405e+00 -2.405e+00
## iloans                  6.135e-02  6.135e-02  6.135e-02  6.135e-02
## log(est)                2.986e-01  2.986e-01  2.986e-01  2.986e-01
## log(Pop_IRS)            6.456e-02  6.456e-02  6.456e-02  6.456e-02
## logAPay_R2              7.847e-02  7.847e-02  7.847e-02  7.847e-02
## tri                     2.939e-03  2.939e-03  2.939e-03  2.939e-03
## rucadj                  1.535e-02  1.535e-02  1.535e-02  1.535e-02
## rucnonadj              -4.752e-02 -4.752e-02 -4.752e-02 -4.752e-02
## poly(AREA_zcta, 2)1    -3.249e+00 -3.249e+00 -3.249e+00 -3.249e+00
## poly(AREA_zcta, 2)2     2.362e+00  2.362e+00  2.362e+00  2.362e+00
## I(Pop_IRS/AREA_cty)     6.866e-05  6.866e-05  6.866e-05  6.866e-05
## I(est/AREA_zcta)       -2.592e-05 -2.592e-05 -2.592e-05 -2.592e-05
## factor(time)2000-06-30  5.964e-02  5.964e-02  5.964e-02  5.964e-02
## factor(time)2000-12-31  1.506e-01  1.506e-01  1.506e-01  1.506e-01
## factor(time)2001-06-30  3.417e-01  3.417e-01  3.417e-01  3.417e-01
## factor(time)2001-12-31  4.162e-01  4.162e-01  4.162e-01  4.162e-01
## factor(time)2002-06-30  6.212e-01  6.212e-01  6.212e-01  6.212e-01
## factor(time)2002-12-31  7.556e-01  7.556e-01  7.556e-01  7.556e-01
## factor(time)2003-06-30  8.410e-01  8.410e-01  8.410e-01  8.410e-01
## factor(time)2003-12-31  9.624e-01  9.624e-01  9.624e-01  9.624e-01
## factor(time)2004-06-30  8.833e-01  8.833e-01  8.833e-01  8.833e-01
## factor(time)2004-12-31  9.693e-01  9.693e-01  9.693e-01  9.693e-01
## factor(time)2005-06-30  1.131e+00  1.131e+00  1.131e+00  1.131e+00
## factor(time)2005-12-31  1.112e+00  1.112e+00  1.112e+00  1.112e+00
## factor(time)2006-06-30  1.260e+00  1.260e+00  1.260e+00  1.260e+00
## factor(time)2006-12-31  1.374e+00  1.374e+00  1.374e+00  1.374e+00
## factor(time)2007-06-30  1.467e+00  1.467e+00  1.467e+00  1.467e+00
## factor(time)2007-12-31  1.376e+00  1.376e+00  1.376e+00  1.376e+00
## factor(time)2008-06-30  1.607e+00  1.607e+00  1.607e+00  1.607e+00
##                             97.5%
## (Intercept)            -2.405e+00
## iloans                  6.135e-02
## log(est)                2.986e-01
## log(Pop_IRS)            6.456e-02
## logAPay_R2              7.847e-02
## tri                     2.939e-03
```

```
## rucadj                    1.535e-02
## rucnonadj                -4.752e-02
## poly(AREA_zcta, 2)1      -3.249e+00
## poly(AREA_zcta, 2)2       2.362e+00
## I(Pop_IRS/AREA_cty)       6.866e-05
## I(est/AREA_zcta)         -2.592e-05
## factor(time)2000-06-30  5.964e-02
## factor(time)2000-12-31  1.506e-01
## factor(time)2001-06-30  3.417e-01
## factor(time)2001-12-31  4.162e-01
## factor(time)2002-06-30  6.212e-01
## factor(time)2002-12-31  7.556e-01
## factor(time)2003-06-30  8.410e-01
## factor(time)2003-12-31  9.624e-01
## factor(time)2004-06-30  8.833e-01
## factor(time)2004-12-31  9.693e-01
## factor(time)2005-06-30  1.131e+00
## factor(time)2005-12-31  1.112e+00
## factor(time)2006-06-30  1.260e+00
## factor(time)2006-12-31  1.374e+00
## factor(time)2007-06-30  1.467e+00
## factor(time)2007-12-31  1.376e+00
## factor(time)2008-06-30  1.607e+00
```

The time fixed effects model is running into problems with estimation, likely due to the number of parameters to estimate. Either way, the spatial methods used below account for time fixed effects as a random effect and that is good enough for me as I do not care about the estimates on the time fixed effects but only wish to control for the diffusion process of broadband.

I now turn to the `CARBayesST` package in `R` to estimate three separate spatial models. The package models spatio-temporal autocorrelation by random effects, which are assigned conditional autoregressive (CAR) prior distributions. The weight matrix used for the estimation procedure is through a contiguity based, where I present the Moran's I test results for residual spatial correlation for each time period of the preferred Poisson model with time fixed effects:

time	Statistic	Pval
1999-12-31	0.2721126	0
2000-06-30	0.2272970	0
2000-12-31	0.2042153	0
2001-06-30	0.1468049	0
2001-12-31	0.1556255	0
2002-06-30	0.1947414	0
2002-12-31	0.1770871	0
2003-06-30	0.1927477	0
2003-12-31	0.1922888	0
2004-06-30	0.1247874	0
2004-12-31	0.1739186	0
2005-06-30	0.1963677	0
2005-12-31	0.1577641	0
2006-06-30	0.1993229	0
2006-12-31	0.2346561	0
2007-06-30	0.2454667	0
2007-12-31	0.2318501	0
2008-06-30	0.3159622	0

Those are extremely small p-values, so I am a little concerned about having screwed something up. But across the board there appears to be positive spatial autocorrelation in each time period, although it appears to trend downward from 1999 until 2004 then tick back up. I believe there was a bill in 2005 which deregulated an aspect of the broadband market and this may be the source of spatial autocorrelation. I have made a note to double check the regulation, but the important aspect here is that there appears to be spatial autocorrelation that I should attempt to account for to see if the main result related to the broadband loan program still holds. I next inspect a few Bayesian models which are explained in some more depth through the CARBayesST Vignette. All spatial models begin in the form of:

$$Prov_{z,t}|X_{z,t},\beta \sim Pois\left(\lambda_{z,t}\right)$$
$$log\left(\lambda_{z,t}\right) = \beta_0 + \beta_1 Loan_{z,t} + \beta_2 X_{z,t} + M_{z,t}$$

where the $M_{z,t}$ term is a latent component for ZIP code z and time period t that captures spatio-temporal autocorrelation. This term takes on different functional forms in the `CARBayesST` package, checking to see how much the term associated with the broadband loan programs changes is the main interest here. If there are large changes, this likely indicates that spatial dependence is an issue in evaluating the broadband loan programs. The spatial model which best fits the data would be most appropriate for evaluation. However, if the coefficients of interest do not substantially change, this would indicate that the broadband loan programs are not affected by the spatial nature of broadband diffusion. Spatial methods, while potentially a better fit, may not be appropriate and could hinder efficiency of estimates.

The first spatial model is ST.CARanova, which models the $M_{z,t}$ term as:

$$M_{z,t} = \phi_z + \delta_t + \gamma_{z,t},$$
$$\phi_z|\phi_{-z},\mathbf{W} \sim N\left(\frac{\rho_\phi \mathbf{W}\phi_j}{\rho_\phi \mathbf{W} + 1 - \rho_\phi}, \frac{\tau^2_\phi}{\rho_\phi \mathbf{W} + 1 - \rho_\phi}\right)$$
$$\delta_t|\delta_{-t},\mathbf{D} \sim N\left(\frac{\rho_\delta \mathbf{D}\delta_j}{\rho_\delta \mathbf{D} + 1 - \rho_\delta}, \frac{\tau^2_\delta}{\rho_\delta \mathbf{D} + 1 - \rho_\delta}\right)$$
$$\gamma_{z,t} \sim N(0,\tau^2_\gamma)$$
$$\tau^2_\phi,\tau^2_\delta,\tau^2_\gamma \sim \text{Inverse-Gamma}(a,b)$$
$$\rho_\phi,\rho_\delta \sim \text{Uniform}(0,1)$$

where $\mathbf{D}$ is a $T \times T$ temporal neighborhood matrix where the elements equal 1 for time periods before and after that of interest.

```
##
## #################
## #### Model fitted
## #################
## Likelihood model - Poisson (log link function)
## Latent structure model - spatial and temporal main effects and an interaction
## Regression equation - Prov_num ~ iloans + log(est) + log(Pop_IRS) + logAPay_R2 + tri +
##     ruc + poly(AREA_zcta, 2) + I(Pop_IRS/AREA_cty) + I(est/AREA_zcta)
##
## ############
## #### Results
## ############
## Posterior quantities for selected parameters and DIC
##
##                        Median    2.5%   97.5% n.sample % accept n.effective
```

```
## (Intercept)         -0.5849 -1.0767 -0.0062     9000     33.4         39.0
## iloans               0.0672  0.0365  0.0968     9000     33.4        132.6
## log(est)             0.2505  0.2377  0.2663     9000     33.4         68.6
## log(Pop_IRS)         0.0375  0.0054  0.0749     9000     33.4         14.2
## logAPay_R2           0.0266 -0.0117  0.0663     9000     33.4         92.6
## tri                  0.0039 -0.0009  0.0093     9000     33.4         32.2
## rucadj              -0.0177 -0.1228  0.0474     9000     33.4         21.0
## rucnonadj           -0.0144 -0.1482  0.1140     9000     33.4         12.5
## poly(AREA_zcta, 2)1  4.7314  2.4236  6.9671     9000     33.4         71.0
## poly(AREA_zcta, 2)2 -2.0593 -3.9833 -0.2500     9000     33.4        128.6
## I(Pop_IRS/AREA_cty)  0.0000 -0.0001  0.0001     9000     33.4         15.2
## I(est/AREA_zcta)     0.0000 -0.0001  0.0000     9000     33.4         71.4
## tau2.phi             0.1323  0.1114  0.1563     9000    100.0        703.4
## tau2.delta           0.0212  0.0106  0.0506     9000    100.0       1621.9
## tau2.gamma           0.0004  0.0002  0.0009     9000    100.0          4.6
## rho                  0.9715  0.9129  0.9958     9000     44.9        314.8
## lambda               0.9834  0.9195  0.9986     9000     46.4        704.9
##                     Geweke.diag
## (Intercept)                -0.9
## iloans                     -0.8
## log(est)                    0.1
## log(Pop_IRS)                1.7
## logAPay_R2                 -0.8
## tri                         0.8
## rucadj                      1.3
## rucnonadj                   1.0
## poly(AREA_zcta, 2)1        -0.1
## poly(AREA_zcta, 2)2         0.3
## I(Pop_IRS/AREA_cty)         0.5
## I(est/AREA_zcta)            2.1
## tau2.phi                   -1.3
## tau2.delta                 -0.4
## tau2.gamma                  7.1
## rho                        -1.2
## lambda                     -0.5
##
## DIC =  51881.93       p.d =  576.5914       LMPL =  -25391.73
```

I am a little perplexed with the documentation on the model and the `lambda` term as it does not show up anywhere. I *think* that the `lambda` term is really the ρ_δ term and the `rho` term is the ρ_ϕ term. But, we can verify how the residual dependence declines through Moran's I tests:

time	Statistic	Pval
1999-12-31	0.2387921	0
2000-06-30	0.1759464	0
2000-12-31	0.1456818	0
2001-06-30	0.1434816	0
2001-12-31	0.1451806	0
2002-06-30	0.1900692	0
2002-12-31	0.1717411	0
2003-06-30	0.3056672	0
2003-12-31	0.3557057	0
2004-06-30	0.1603597	0

time	Statistic	Pval
2004-12-31	0.2130481	0
2005-06-30	0.2478610	0
2005-12-31	0.1786113	0
2006-06-30	0.1923907	0
2006-12-31	0.2701748	0
2007-06-30	0.3238801	0
2007-12-31	0.2775035	0
2008-06-30	0.5299309	0

Next, we have ST.CARsepspatial:

$$
\begin{gathered}
M_{z,t} = \phi_{z,t} + \delta_t, \\
\phi_{z,t} | \phi_{-z,t}, \mathbf{W} \sim N\left(\frac{\rho \mathbf{W} \phi_t}{\rho \mathbf{W} + 1 - \rho}, \frac{\tau_\phi^2}{\rho \mathbf{W} + 1 - \rho}\right) \\
\delta_t | \delta_{t-1} \sim N\left(\delta_{t-1}, \tau_\delta^2\right), \\
\tau_\phi^2, \tau_\delta^2 \sim \text{Inverse-Gamma}(a, b) \\
\rho \sim \text{Uniform}(0, 1)
\end{gathered}
$$

```
##
## #################
## #### Model fitted
## #################
## Likelihood model - Poisson (log link function)
## Latent structure model - A random walk time trend with separate spatial effects
## Regression equation - Prov_num ~ iloans + log(est) + log(Pop_IRS) + logAPay_R2 + tri +
##     ruc + poly(AREA_zcta, 2) + I(Pop_IRS/AREA_cty) + I(est/AREA_zcta)
##
## ############
## #### Results
## ############
## Posterior quantities for selected parameters and DIC
##
##                      Median    2.5%   97.5% n.sample % accept n.effective
## (Intercept)         -1.4156 -1.6918 -1.1432     9000     33.6       191.1
## iloans               0.0518  0.0302  0.0738     9000     33.6       264.0
## log(est)             0.2910  0.2830  0.2984     9000     33.6       228.4
## log(Pop_IRS)         0.0641  0.0508  0.0770     9000     33.6       121.5
## logAPay_R2           0.0677  0.0431  0.0949     9000     33.6       234.1
## tri                  0.0023  0.0005  0.0042     9000     33.6       263.0
## rucadj               0.0089 -0.0236  0.0399     9000     33.6       154.6
## rucnonadj           -0.0573 -0.0940 -0.0245     9000     33.6       132.1
## poly(AREA_zcta, 2)1 -1.4747 -2.7844 -0.2299     9000     33.6       283.9
## poly(AREA_zcta, 2)2  1.2849 -0.0456  2.6705     9000     33.6       243.2
## I(Pop_IRS/AREA_cty)  0.0001  0.0000  0.0001     9000     33.6       185.3
## I(est/AREA_zcta)     0.0000  0.0000  0.0000     9000     33.6       763.9
## tau2.1               0.0845  0.0322  0.1440     9000    100.0        43.8
## tau2.2               0.0130  0.0033  0.0402     9000    100.0         8.5
## tau2.3               0.0030  0.0008  0.0335     9000    100.0         6.3
```

```
## tau2.4               0.0037  0.0006  0.0140     9000    100.0         8.5
## tau2.5               0.0032  0.0013  0.0107     9000    100.0         9.9
## tau2.6               0.0112  0.0036  0.0305     9000    100.0        10.1
## tau2.7               0.0125  0.0052  0.0334     9000    100.0        11.6
## tau2.8               0.0173  0.0064  0.0345     9000    100.0        21.5
## tau2.9               0.0229  0.0095  0.0409     9000    100.0        28.6
## tau2.10              0.0092  0.0035  0.0224     9000    100.0        17.3
## tau2.11              0.0129  0.0052  0.0270     9000    100.0        21.1
## tau2.12              0.0153  0.0067  0.0312     9000    100.0        18.9
## tau2.13              0.0069  0.0031  0.0148     9000    100.0        27.1
## tau2.14              0.0083  0.0037  0.0166     9000    100.0        22.5
## tau2.15              0.0083  0.0030  0.0214     9000    100.0        14.1
## tau2.16              0.0088  0.0021  0.0181     9000    100.0        13.1
## tau2.17              0.0051  0.0025  0.0128     9000    100.0        17.7
## tau2.18              0.0100  0.0048  0.0179     9000    100.0        32.8
## sig2                 0.0174  0.0094  0.0376     9000    100.0      7541.3
## rho                  0.9898  0.9821  0.9945     9000     45.3        60.3
## delta.1             -0.8596 -0.9156 -0.8028    9000     44.1       478.2
## delta.2             -0.7894 -0.8394 -0.7406    9000     44.1       783.3
## delta.3             -0.6894 -0.7419 -0.6409    9000     44.1       483.5
## delta.4             -0.5021 -0.5474 -0.4586    9000     44.1       813.4
## delta.5             -0.4216 -0.4649 -0.3799    9000     44.1      1670.5
## delta.6             -0.2315 -0.2717 -0.1930    9000     44.1       771.3
## delta.7             -0.1134 -0.1561 -0.0732    9000     44.1       148.9
## delta.8             -0.0380 -0.0774  0.0009    9000     44.1       337.0
## delta.9              0.0671  0.0268  0.1016    9000     44.1       208.5
## delta.10             0.0296 -0.0062  0.0648    9000     44.1       543.3
## delta.11             0.1029  0.0665  0.1366    9000     44.1       254.3
## delta.12             0.2527  0.2179  0.2862    9000     44.1       221.0
## delta.13             0.2613  0.2291  0.2941    9000     44.1       695.6
## delta.14             0.4182  0.3899  0.4490    9000     44.1       939.5
## delta.15             0.5366  0.5095  0.5638    9000     44.1       721.4
## delta.16             0.6376  0.6104  0.6648    9000     44.1       692.1
## delta.17             0.5458  0.5193  0.5738    9000     44.1       941.2
## delta.18             0.7955  0.7709  0.8217    9000     44.1       699.7
##                     Geweke.diag
## (Intercept)                 4.5
## iloans                     -1.0
## log(est)                    0.5
## log(Pop_IRS)               -5.8
## logAPay_R2                 -2.2
## tri                        -2.0
## rucadj                     -0.9
## rucnonadj                   0.4
## poly(AREA_zcta, 2)1        -1.7
## poly(AREA_zcta, 2)2         0.1
## I(Pop_IRS/AREA_cty)         3.3
## I(est/AREA_zcta)           -1.1
## tau2.1                     -3.8
## tau2.2                      7.1
## tau2.3                      9.2
## tau2.4                      0.1
## tau2.5                      5.8
## tau2.6                      1.5
```

```
## tau2.7                      0.7
## tau2.8                     -0.9
## tau2.9                      2.1
## tau2.10                    -4.7
## tau2.11                     1.4
## tau2.12                    -2.9
## tau2.13                     1.2
## tau2.14                    -1.0
## tau2.15                    -1.4
## tau2.16                    -0.2
## tau2.17                    -1.4
## tau2.18                    -0.2
## sig2                        1.8
## rho                         1.1
## delta.1                     4.6
## delta.2                    -2.8
## delta.3                   -10.0
## delta.4                     2.9
## delta.5                    -5.8
## delta.6                    -1.2
## delta.7                    -0.4
## delta.8                     3.0
## delta.9                    -1.9
## delta.10                    7.6
## delta.11                   -4.1
## delta.12                    2.0
## delta.13                   -1.3
## delta.14                    1.2
## delta.15                    1.2
## delta.16                    0.9
## delta.17                    4.1
## delta.18                    3.6
##
## DIC =  53194.35       p.d =  303.3422       LMPL =  -26305.14
```

The decline in dependence:

time	Statistic	Pval
1999-12-31	0.1556947	0
2000-06-30	0.1865953	0
2000-12-31	0.1755787	0
2001-06-30	0.1524100	0
2001-12-31	0.1717291	0
2002-06-30	0.1775744	0
2002-12-31	0.1456873	0
2003-06-30	0.1703799	0
2003-12-31	0.1320779	0
2004-06-30	0.1530479	0
2004-12-31	0.1600830	0
2005-06-30	0.1341154	0
2005-12-31	0.1820677	0
2006-06-30	0.1426663	0
2006-12-31	0.1639555	0

time	Statistic	Pval
2007-06-30	0.1758554	0
2007-12-31	0.1671453	0
2008-06-30	0.1416563	0

And finally, ST.CARar:

$$M_{z,t} = \phi_{z,t},$$
$$\phi_t|\phi_{t-1} \sim N\left(\gamma\phi_{t-1}, \tau^2\mathbf{Q}\left(\mathbf{W},\rho\right)^{-1}\right) t = 2,\ldots,T,$$
$$\phi_1 \sim N\left(0, \tau^2\mathbf{Q}\left(\mathbf{W},\rho\right)^{-1}\right)$$
$$\tau^2 \sim \text{Inverse-Gamma}(a,b)$$
$$\rho,\gamma \sim \text{Uniform}(0,1)$$

with:

$$\mathbf{Q}\left(\mathbf{W},\rho\right) = \rho\left(\text{diag}\left(\mathbf{W1}\right) - \mathbf{W}\right) + (1-\rho)\,\mathbf{I}$$

```
##
## #################
## #### Model fitted
## #################
## Likelihood model - Poisson (log link function)
## Latent structure model - Autoregressive CAR model
## Regression equation - Prov_num ~ iloans + log(est) + log(Pop_IRS) + logAPay_R2 + tri +
##     ruc + poly(AREA_zcta, 2) + I(Pop_IRS/AREA_cty) + I(est/AREA_zcta)
##
## ############
## #### Results
## ############
## Posterior quantities for selected parameters and DIC
##
##                       Median    2.5%   97.5% n.sample % accept n.effective
## (Intercept)          -0.6963 -1.4859 -0.2094     9000     33.6        14.3
## iloans                0.0547  0.0120  0.1426     9000     33.6        12.7
## log(est)              0.2610  0.2500  0.2735     9000     33.6        57.9
## log(Pop_IRS)          0.0330  0.0144  0.0808     9000     33.6        15.0
## logAPay_R2            0.0406  0.0038  0.0781     9000     33.6        52.2
## tri                   0.0010 -0.0017  0.0042     9000     33.6        26.6
## rucadj               -0.0456 -0.1074  0.0099     9000     33.6        21.2
## rucnonadj            -0.1206 -0.1942 -0.0610     9000     33.6        12.9
## poly(AREA_zcta, 2)1   3.8211  1.1163  5.4326     9000     33.6        64.8
## poly(AREA_zcta, 2)2 -1.6646 -3.5859  0.2080     9000     33.6        87.3
## I(Pop_IRS/AREA_cty)   0.0001  0.0000  0.0001     9000     33.6        13.8
## I(est/AREA_zcta)      0.0000  0.0000  0.0000     9000     33.6        21.9
## tau2                  0.0136  0.0118  0.0197     9000    100.0        12.7
## rho                   0.9994  0.9989  0.9996     9000     45.5        82.7
## gamma                 0.9613  0.9457  0.9767     9000    100.0        30.8
##                      Geweke.diag
```

```
## (Intercept)              -12.9
## iloans                    10.2
## log(est)                   8.0
## log(Pop_IRS)              13.3
## logAPay_R2                 6.2
## tri                       -1.0
## rucadj                     3.6
## rucnonadj                  3.0
## poly(AREA_zcta, 2)1       -3.7
## poly(AREA_zcta, 2)2        5.7
## I(Pop_IRS/AREA_cty)        0.8
## I(est/AREA_zcta)          -0.3
## tau2                       5.7
## rho                       -7.3
## gamma                     -0.9
##
## DIC =  51869.55       p.d =  873.9536       LMPL =  -25127.9
```

And lastly, the decline in dependence:

time	Statistic	Pval
1999-12-31	0.1984310	0.0000000
2000-06-30	0.1356184	0.0000000
2000-12-31	0.0921001	0.0000026
2001-06-30	0.0554628	0.0028360
2001-12-31	0.0441565	0.0134159
2002-06-30	0.0490504	0.0070840
2002-12-31	0.0123768	0.2540449
2003-06-30	0.0367920	0.0318504
2003-12-31	0.0187042	0.1658119
2004-06-30	-0.0055517	0.5846582
2004-12-31	0.0285712	0.0731453
2005-06-30	0.0342606	0.0417586
2005-12-31	0.0618543	0.0010408
2006-06-30	0.0049039	0.3832972
2006-12-31	0.0581942	0.0018684
2007-06-30	0.0331394	0.0468819
2007-12-31	0.0246846	0.1033064
2008-06-30	-0.0148646	0.7481477

It appears that across the spatial models, the effects on the loans are similar. Which is really all I want to show. We *can* account for Spatial effects, but the result for the broadband loan program appears to be robust across models. Further, it appears that the `ar` model is best at alleviating the spatial correlation as evidenced by the Moran's I tests. So I am contempt with this loose analysis, but I would prefer for this to be shown at a larger scale than 1 state.

But why can't I get this to work for the entire United States? Is it because 30,000 ZIP codes across 18 time periods (about half a million observations) too computationally intensive? Or is the `CARBayesST` package not the best option for computing a spatio-temporal model? I only have a loose understanding of the OpenBUGS software but that might be my best option.