Technische Universität München

Zentrum Mathematik

Estimating the Trend in Bank-Branch Deposits in the New York State Using Multilevel Models

Project

by

Holger Schabenberger

Supervisor: Prof. Dr. Claudia Czado
Tutor: Vinzenz Erhardt

Deadline: 30 May 2008

List of Figures

List of Tables

Contents

Part I

Descriptive Statistics

Chapter 1

Hierarchical Model

The dataset considered in this project is the result of different prior updates upon which the following work is based. Hence, this paper reflects only our modifications. The dataset contains 3091 branch-year records of a major US bank in the state of New York with multiple branches. Each line in the dataset refers to one single branch of the bank. On a whle, 523 branches are included and observations cover the period from 1994 to 2002. During the project, the data with no ZIP Codes, no branch ID and the ones located outside of the state of New York have been removed, ending up with 2988 branch-year records with 506 branches. Also county assignment were corrected, too. The first six lines of the dataset are shown below as an example.

```
  b.ID year b.dep s.no.fail b.singleD s.m.share s.br.share s.dep.share s.dep.ave
1 4032 1994    26        18  180.6327    0.1364  0.9976636   1.0000000  112948.5
2 3190 1994    67        18  121.8393    0.1364  0.9976636   1.0000000  112948.5
3 4779 2001    92        11  273.2192    0.2208  0.6344262   0.7417029  264879.9
4 3604 1998   114        24  186.6597    0.2250  0.8272884   0.9462533  198819.9
5 4791 1997   255        70  250.4823    0.2218  0.8367347   0.9525438  179246.7
6 3217 1999   317        68  193.3312    0.2318  0.8333333   0.9551407  212462.7


   c.pop        c.inc c.unemp c.long c.lat c.ID   b.mmcD z.ZIP   c.name
 1308489  45525243000     5.0  73.24 40.77   59 8880.721 11021   Nassau
 1503909  83391843000     7.6  73.90 40.83   61 7240.861 10001 New York
 1549009 144033000000     6.4  73.80 40.58   61 7732.906 10001 New York
  282044   9991399000     3.4  73.98 41.21   87 7156.619 10923 Rockland
 1318211  52532377000     3.5  73.24 40.77   59 8587.163 11021   Nassau
 1535624 122577000000     5.9  73.90 40.83   61 7137.915 10001 New York
```

There are 18 different variables that can be divided into the following four groups: State Variables, County Variables, ZIP Code Variables and Branch Variables. For a better overview the variables have been renamed, i.e. all State Variables get the initial letter s., County Variables c., ZIP Variables z. and Branch Variables b.. See Figure 1.1 to understand hierarchical relations between state, counties, ZIP Codes and branches: The state can be divided into different counties (28) whereas each one is including a specific number of ZIP Codes. Finally, at least one branch in each ZIP Code is considered.

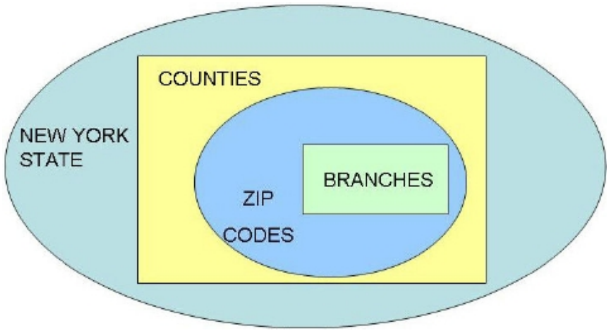

Figure 1.1: **Hierarchical structure of the four variable groups**

These four groups will be described in detail in the following paragraphs.

1.1 State Variables

The following variables are constant over ZIP Code and counties (NY) for each year and therefore only have different outcomes for each year.

- s.br.share
- s.dep.share
- s.dep.ave
- s.no.fail
- s.m.share

s.br.share
s.br.share is the fraction of the number of branches in New York State and the number of branches in the USA.

$$\left(\frac{number\ of\ branches\ in\ NY\ State}{number\ of\ branches\ in\ USA}\right)$$

s.dep.share
s.dep.share is the fraction of total deposits of the bank nationwide within New York State and the total deposits of bank in the USA.

$$\left(\frac{total\ deposits\ of\ the\ bank\ within\ the\ NY\ State}{total\ deposits\ of\ bank\ in\ USA}\right)$$

s.dep.ave
s.dep.ave is the average deposit per bank within New York State.

$$\left(\frac{total\ deposits\ of\ the\ bank\ within\ the\ NY\ State}{number\ of\ branches\ of\ the\ bank\ in\ the\ NY\ State}\right)$$

s.no.fail
The total number of the bank's branches that closed in NY during the year.
s.m.share
Market share in NY State.

The values of these five variables in the periode from 1994 until 2002 are described in Table 1.1 whereas *s.dep.ave* is given in 1000 US$.

	1994	1995	1996	1997	1998	1999	2000	2001	2002
s.br.share	0.9977	0.9976	0.9974	0.8367	0.8273	0.8333	0.8277	0.6344	0.6327
s.dep.share	1	1	1	0.9525	0.9463	0.9551	0.9549	0.7417	0.7658
s.dep.ave (/1000)	112.9	112.5	123.1	179.2	198.8	212.5	251.6	264.9	321.2
s.no.fail	18	30	131	70	24	68	11	11	7
s.m.share	0.1364	0.1323	0.1301	0.2218	0.225	0.2318	0.2232	0.2208	0.2359

Table 1.1: **Values of State Variables**

The time trend of these variables is displayed in Figure 1.2 which claryfies that the trend of *s.br.share* declines (from 0.9977 to 0.6327 during the period from 1994 to 2002) and of *s.dep.ave* rose (from 112900 to 321200). The values of *s.dep.share* fell until the year 2001 (from 1 to 0.7417) but increased again in 2002 (to 0.7658). The trends of the remaining variables are not as clear as the ones above.

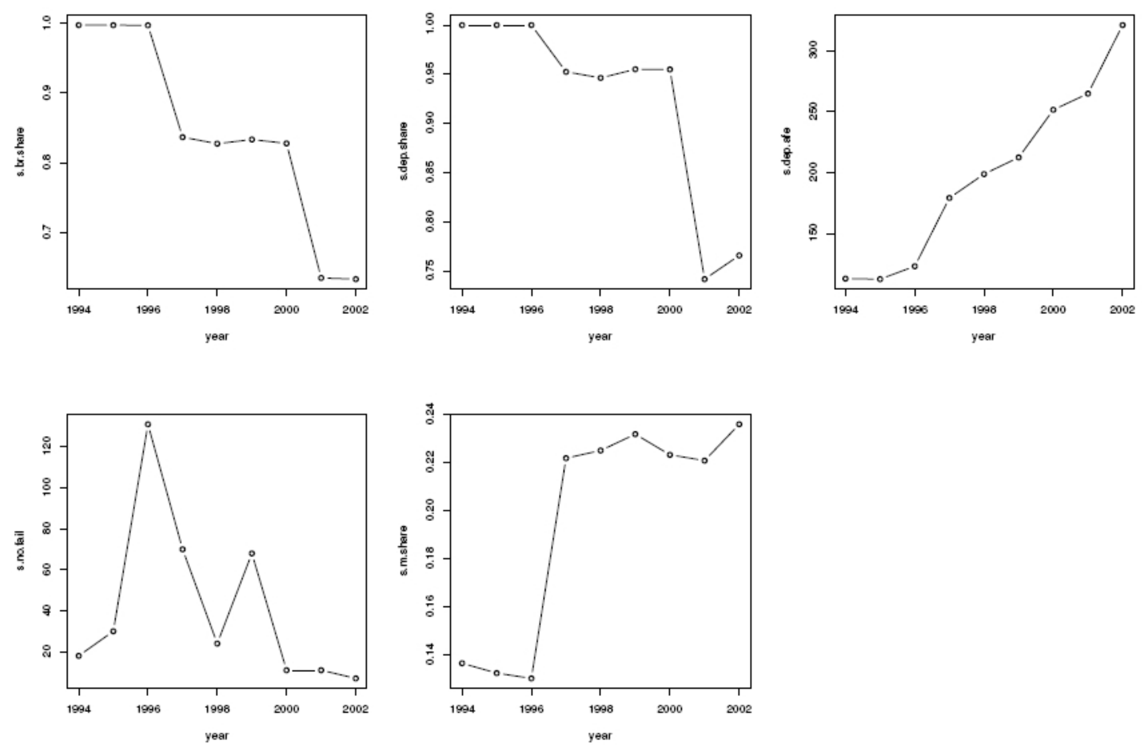

Figure 1.2: **Time plots of State Variables**

1.2 County Variables

The following variables are county and year specific.

- c.ID
- c.name
- c.pop
- c.inc
- c.inc.per.ca
- c.unemp

c.ID
c.ID is the number of every county of New York.

c.name
c.name is the countyname of NY.

c.pop
c.pop is he number of people living in the particular counties of NY (in the period from 1994 until 2002)

Table 1.2 depicts the population of the specific counties of the state of New York in 10000. Notice that the population of urban counties (e.g. Kings, Queens and New York) is higher than of the rural ones (e.g. Tioga, Genesee and Livingston) independent from time. During the period from 1994 until 2002 the population in the urban counties increased (e.g. New York from 1503900 to 1555400, Queens from 2047900 to 2227200 and Kings from 2361700 to 2475600) whereas the rural population seemed to be constant or insignificantly decreasing (e.g. Tioga from 53300 to 51800, Genesee from 61300 to 59900 and Livingston from 64000 to 64700). Italic values in the tablee were added for years in which there was no branch in the referring county completing the original and modified datasets on the basis of other references. See Figure 1.2 for a graphical presentation of the results. Altogether there are nine plots (one for every year) in which the size of the populations is presented in different colors; light colors referring to small sizes with darker colors to increasingly larger populations. The second parts of the maps shows the magnificated area of New York where most of the branches are located(see Figure 1.2).

	1994	1995	1996	1997	1998	1999	2000	2001	2002
Albany	30.03	*29.97*	*29.77*	29.62	29.51	29.47	*29.46*	*29.49*	*29.64*
Bronx	125.08	126.21	127.27	128.57	130.08	131.79	133.41	134.37	135.84
Broome	21.08	20.75	20.44	20.21	20.11	20.09	20.03	*20.02*	*20.02*
Chautauq	*14.30*	*14.25*	*14.25*	14.15	14.08	14.01	*13.96*	*13.87*	*13.83*
Chemung	9.41	9.37	9.30	9.22	9.18	9.13	9.10	9.07	9.08
Erie	97.71	97.39	97.04	96.30	95.66	95.25	94.94	94.66	94.28
Genesee	6.13	6.11	*6.13*	6.12	6.09	6.06	6.03	6.00	5.99
Herkimer	*6.70*	*6.67*	*6.65*	*6.58*	6.49	6.45	6.44	6.42	6.37
Kings	236.17	237.26	238.37	240.11	242.24	244.74	246.78	247.99	247.56
Livingst	*6.40*	*6.41*	*6.41*	6.42	6.39	6.42	6.44	6.47	6.47
Madison	*7.05*	*7.02*	*7.02*	6.96	6.95	6.95	6.94	6.98	6.98
Monroe	*73.44*	*73.36*	*73.43*	73.44	73.41	73.37	73.58	73.62	73.61
Nassau	130.85	131.29	131.58	131.82	132.29	133.08	133.66	133.93	133.93
New York	150.39	151.42	152.14	152.74	153.11	153.56	153.93	154.90	155.54
Niagara	*22.34*	*22.39*	*22.38*	22.31	22.19	22.06	*21.96*	*21.86*	*21.82*
Onondaga	47.27	47.02	46.66	46.24	45.99	45.86	45.85	45.93	45.95
Ontario	9.85	9.88	*9.95*	9.97	9.96	9.98	10.04	10.09	10.18
Orange	32.07	*32.35*	*32.60*	32.89	33.22	33.66	34.31	34.95	35.58
Oswego	*12.43*	*12.39*	*12.36*	12.32	12.22	12.22	12.25	12.26	12.27
Putnam	8.90	9.00	9.08	9.21	9.33	9.46	9.61	9.71	9.85
Queens	204.79	207.46	210.55	213.85	217.50	220.90	223.18	223.80	222.72
Rensselaer	15.55	15.50	*15.49*	*15.37*	*15.31*	*15.24*	*15.26*	*15.28*	*15.32*
Richmond	40.62	40.96	41.45	42.07	42.90	43.78	44.55	45.14	45.54
Rockland	27.49	27.66	27.77	27.91	28.20	28.53	28.75	28.94	29.12
Suffolk	135.40	136.14	136.84	137.72	139.07	140.62	142.43	144.33	145.57
Tioga	*5.33*	*5.26*	*5.23*	5.22	5.22	5.19	*5.18*	*5.15*	*5.18*
Wayne	*9.22*	*9.27*	*9.31*	9.33	9.36	9.38	9.38	9.39	9.37
Westchester	89.35	89.79	90.22	90.56	91.16	91.85	92.58	93.27	93.79

Table 1.2: **Table of Population Size per year and county** ***(in 10000)***

Figure 1.3: **Map of Population Size (as indicated by c.pop)** ***(in 10000)***

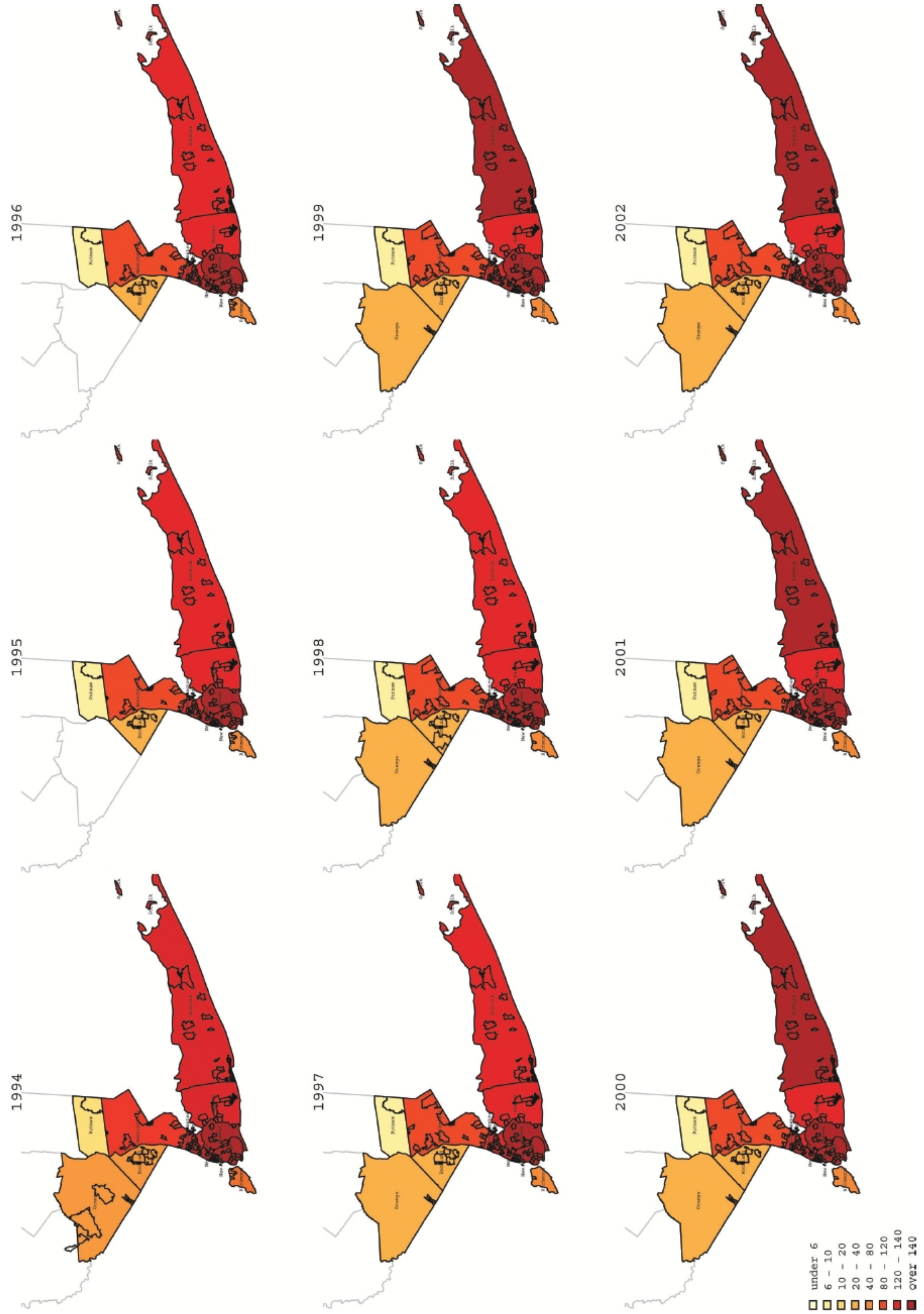

Figure 1.4: **Map of Population Size (as indicated by c.pop)** ***(in 10000) (enlarged)***

c.inc
c.inc is the final amount of money that a county earns (wealth of county).

Table 1.3 shows *c.inc* in million $ which increased over the years 1994 until 2002 in each county regardless whether urban or rural. However, it should be noticed that generally the values of the civic counties are higher than the rual ones (e.g. in the year 2002 New York with 131576.00, Queens with 64313.62 and Tioga with 1280.62 (all values in million)). In Table 1.3, again italic values have been added on the basis of other references, for empty cells in the dataset. These results are also demonstrated in plots (see Figure 1.2). Altogether there are nine plots (one for every year) in which the size of *c.inc* are highlighted in different colors. The second page again shows the enlarged area of New York where most of the branches are located (see Figure 1.2).

	1994	1995	1996	1997	1998	1999	2000	2001	2002
Albany	7729.12	*7935.51*	*8215.06*	7729.12	9166.37	9538.14	*10208.79*	*10802.00*	*10598.39*
Bronx	20842.01	21546.47	22220.01	22711.88	23656.21	24624.93	26046.40	26734.61	28457.34
Broome	4224.18	4301.23	4419.72	4566.58	4688.03	4835.02	5126.74	*5214.95*	*5222.62*
Chautauq	*2509.25*	*2549.60*	*2618.45*	2692.26	2820.75	2855.37	*2989.71*	*3037.55*	*3078.90*
Chemung	1780.29	1850.68	1926.66	1996.02	2084.71	2149.66	2298.45	2325.45	2230.09
Erie	21014.26	21942.53	22606.74	23398.49	24439.24	25204.33	26462.24	26964.09	27537.48
Genesee	1176.34	1217.73	*1245.83*	1300.86	1352.87	1395.42	1445.65	1469.18	1499.66
Herkimer	*1148.57*	*1181.51*	*1194.27*	*1228.45*	1286.53	1341.34	1411.71	1457.13	1452.81
Kings	46849.75	49013.31	51122.36	51983.63	54504.31	57746.76	61606.95	61432.10	62232.00
Livingst	*1204.92*	*1239.49*	*1280.87*	1329.51	1409.31	1456.69	1535.92	1569.23	1521.73
Madison	*1336.12*	*1391.49*	*1429.80*	1495.73	1593.48	1649.10	1742.90	1779.72	1839.21
Monroe	*18089.66*	*18884.74*	*19662.31*	20373.71	21283.51	21714.79	22677.09	23638.26	23926.87
Nassau	45525.24	47541.41	50173.93	52532.38	56116.26	57361.76	62006.99	63524.33	66351.28
New York	83391.84	90807.35	100284.00	105325.00	116793.00	122577.00	137797.00	144033.00	131576.00
Niagara	*4426.04*	*4613.91*	*4752.52*	4913.08	5073.53	5181.83	*5432.27*	*5490.24*	*5538.79*
Onondaga	10505.99	10818.50	11102.81	11427.24	12036.33	12409.02	13059.26	13378.09	13838.95
Ontario	2140.89	2258.51	*2321.59*	2430.62	2601.27	2725.65	2874.97	2954.00	2930.43
Orange	6929.86	*7182.84*	*7477.75*	7749.24	8313.77	8764.13	9384.07	9813.91	10322.74
Oswego	*2238.82*	*2299.79*	*2320.12*	2416.91	2492.95	2585.12	2746.56	2776.56	2744.67
Putnam	2426.87	2578.94	2743.72	2951.19	3136.97	3339.46	3627.76	3721.36	3885.42
Queens	47267.54	49660.17	51541.92	53295.48	56046.20	58675.41	62585.37	64789.25	64313.62
Rensselaer	3362.20	3371.44	*3437.03*	*3575.34*	*3765.70*	*3906.82*	*4184.05*	*4314.07*	*4440.63*
Richmond	10256.72	10872.01	11450.24	12001.49	12634.32	13380.44	14379.82	14548.02	15929.25
Rockland	8246.29	8467.63	8982.74	9410.89	9991.40	10441.26	11146.56	11335.49	12030.90
Suffolk	34540.47	37795.02	39683.16	42599.46	44911.73	47364.60	51132.71	52116.44	54807.83
Tioga	*946.11*	*970.57*	*1009.78*	1060.36	1085.18	1152.08	*1217.05*	*1247.92*	*1280.62*
Wayne	*1858.19*	*1933.10*	*2003.94*	2099.41	2149.99	2192.45	2297.23	2319.98	2394.83
Westchester	33417.98	35753.87	38554.47	40682.12	44441.42	46963.96	51272.74	52878.21	52073.72

Table 1.3: **Table of Income (indicated by c.inc)** ***(in million)***

Figure 1.5: **Map of Income (c.inc)** ***(in million)***

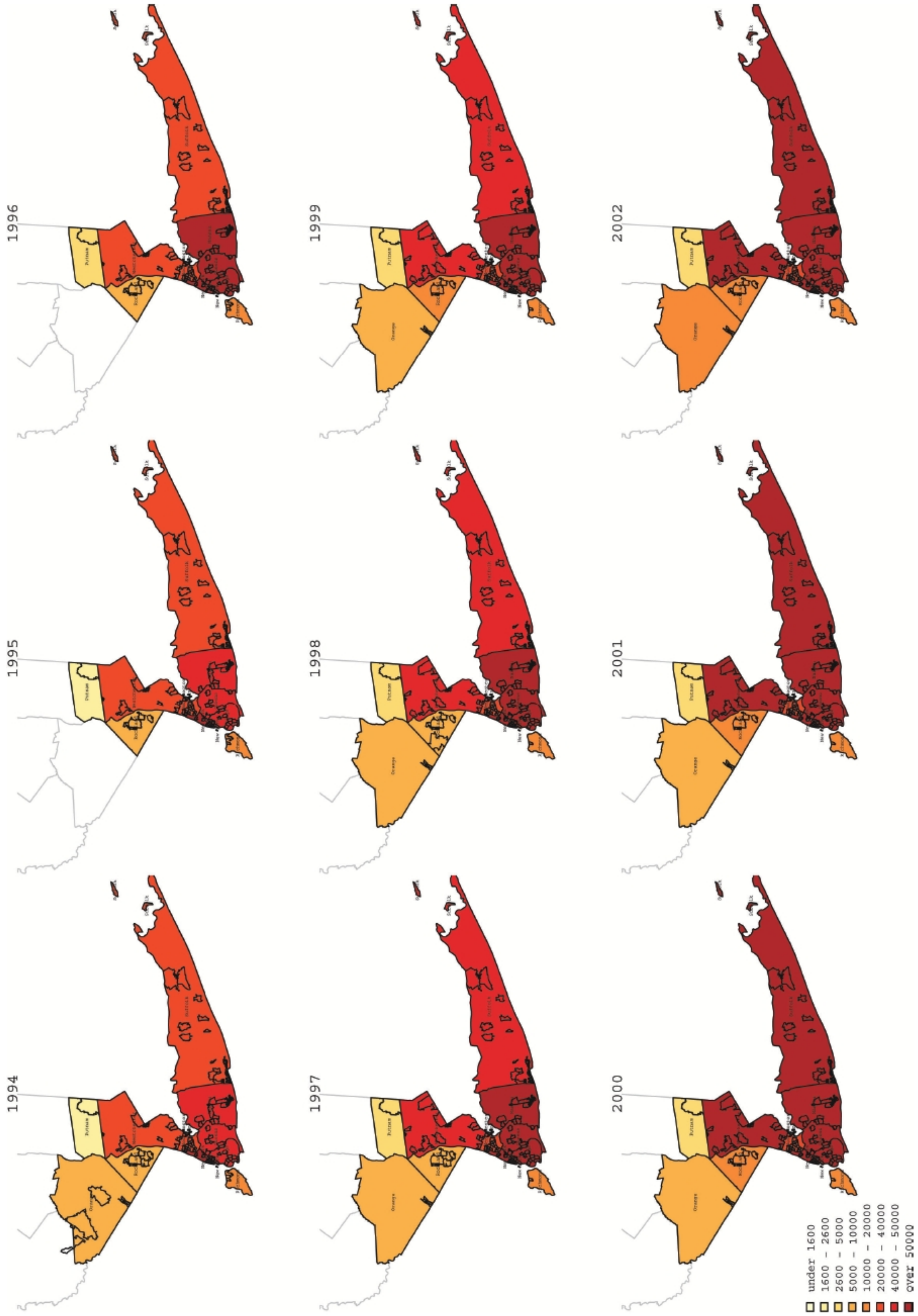

Figure 1.6: **Map of Income (c.inc)** *(in million)* **(enlarged)**

c.inc.per.ca

c.inc.per.ca is the fraction of the *c.inc* and the *c.pop*.

$$\left(\frac{c.inc}{c.pop}\right)$$

Table 1.4 shows the increasing trend of *c.inc.per.ca* from 1994 until 2002 in each county. In contrast to the *c.inc* and *c.pop* the values of the urban and rural regions converge. Indeed, New York City is still on top with 84591.18 in 2002 but the urban counties are no longer dominant (e.g. the urban county Bronx with the smallest value 20949.63 in 2002, Tioga with 24731.94, Kings with 25137.64) Again, italic values are given in addition to the orignal dataset. For a better overview see Figure 1.7. Altogether there are nine plots (one for every year) in which *c.inc.per.ca* is highlighted with different colors according to size. The second page shows the enlarged area of New York where most of the branches are located (see Figure 1.8). The purpose of these maps is to visualize income differences.

	1994	1995	1996	1997	1998	1999	2000	2001	2002
Albany	25741.26	*26475.97*	*27596.94*	26095.41	31062.22	32366.47	*34653.98*	*36633.72*	*35763.09*
Bronx	16663.09	17072.41	17459.05	17665.23	18185.57	18685.04	19523.59	19896.30	20949.63
Broome	20036.61	20731.04	21621.31	22600.91	23315.21	24069.09	25593.13	*26043.13*	*26087.90*
Chautauq	*17542.18*	*17886.67*	*18378.15*	19022.39	20036.02	20382.80	*21418.74*	*21897.28*	*22263.08*
Chemung	18912.70	19756.39	20723.06	21657.68	22711.46	23549.15	25245.80	25637.74	24557.71
Erie	21507.23	22529.95	23296.92	24296.49	25548.61	26459.96	27872.50	28484.45	29208.37
Genesee	19195.44	19928.57	*20327.42*	21250.38	22222.97	23044.24	23964.29	24499.84	25023.98
Herkimer	*17133.10*	*17710.90*	*17971.33*	*18664.36*	19820.52	20808.20	21924.03	22707.40	22806.75
Kings	19837.57	20657.96	21446.61	21649.49	22499.81	23595.61	24964.54	24771.78	25137.64
Livingst	*18822.51*	*19322.90*	*19968.10*	20703.38	22038.42	22674.69	23855.97	24250.27	23512.14
Madison	*18955.28*	*19826.91*	*20381.79*	21504.59	22939.61	23720.24	25097.95	25499.28	26356.57
Monroe	*24632.49*	*25741.05*	*26778.51*	27742.20	28993.61	29597.13	30821.19	32107.83	32505.66
Nassau	34792.22	36209.61	38131.91	39851.27	42417.52	43102.73	46389.88	47430.96	49543.05
New York	55450.06	59971.14	65914.95	68957.77	76282.10	79822.27	89517.52	92983.97	84591.18
Niagara	*19812.17*	*20601.68*	*21236.33*	22018.41	22865.70	23494.41	*24737.43*	*25111.43*	*25380.53*
Onondaga	22225.87	23006.58	23794.42	24710.86	26172.65	27059.96	28482.64	29127.89	30118.91
Ontario	21732.97	22855.69	*23330.45*	24381.07	26116.34	27310.54	28634.59	29277.05	28788.12
Orange	21605.57	*22206.88*	*22937.87*	23563.94	25026.47	26034.91	27354.11	28081.46	29012.58
Oswego	*18008.66*	*18550.91*	*18768.33*	19611.89	20408.41	21155.03	22413.91	22640.09	22365.10
Putnam	27279.27	28664.74	30227.78	32029.80	33622.00	35283.34	37740.08	38315.19	39433.92
Queens	23081.16	23937.44	24479.43	24921.86	25768.57	26562.28	28042.02	28949.31	28876.81
Rensselaer	21619.07	21755.57	*22184.67*	*23268.92*	*24598.73*	*25636.63*	*27415.75*	*28229.76*	*28978.07*
Richmond	25249.92	26542.98	27625.49	28525.74	29453.92	30561.92	32278.16	32230.59	34979.75
Rockland	30002.34	30607.84	32345.27	33720.86	35424.97	36601.98	38767.28	39164.89	41310.65
Suffolk	25509.74	27762.06	29000.40	30931.87	32293.66	33683.62	35899.15	36109.25	37649.80
Tioga	*17764.40*	*18451.59*	*19312.31*	20329.03	20803.62	22201.57	*23516.98*	*24214.90*	*24731.94*
Wayne	*20153.23*	*20844.30*	*21518.17*	22506.07	22964.08	23376.17	24499.31	24706.43	25550.88
Westchester	37399.91	39817.44	42735.07	44921.24	48749.51	51129.37	55380.23	56690.77	55521.97

Table 1.4: **Table of Income per Capita (as indicated by c.inc.per.ca)**

Figure 1.7: **Map of Income per Capita (c.inc.per.ca)**

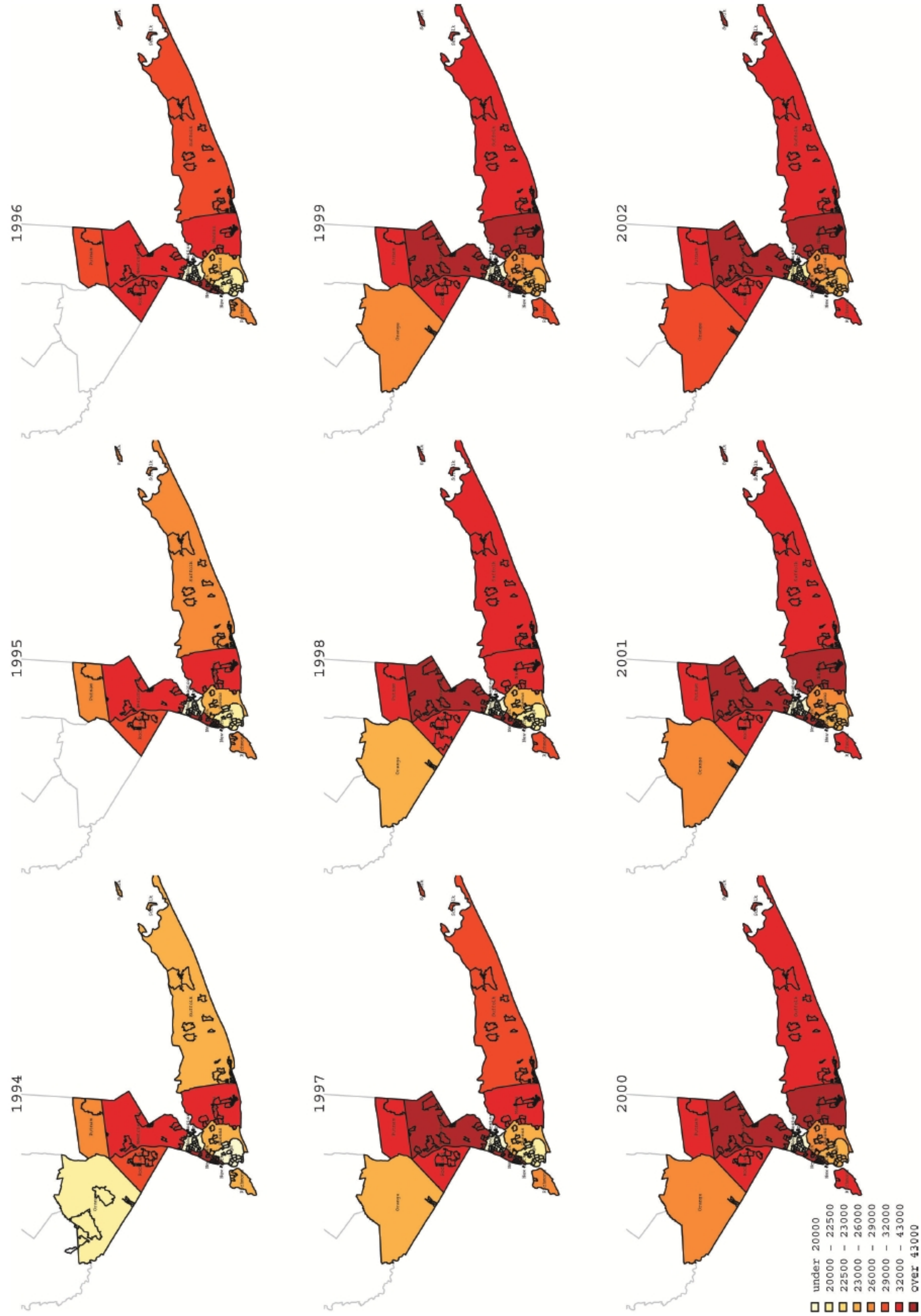

Figure 1.8: **Map of Income per Capita (c.inc.per.ca) (enlarged)**

c.unemp
c.unemp is the percentage of the number of unemployed people in a county of NY.

Table 1.5 shows *c.unemp* in percent for the different counties. There is no identifiable trend over the years 1994 until 2002, with a slight decline for 2000/2001 in each county and a subsequent rise in the following year. Furthermore, there is no difference between urban and rural regions as in the graphs presented so far. The Bronx shows the highest rate in 2002 (9.40) followed by Kings (8.80); the smallest one being Albany (3.30). Again, the figures are presented in a graph (see Figure 1.9). Altogether there are nine plots (one for every year) in which the heights of *c.unemp* is displayed in different colors. The second page shows the enlarged area of New York (see Figure 1.10). Furthermore, italic numbers were added to the original and modified datasets.

	1994	1995	1996	1997	1998	1999	2000	2001	2002
Albany	4.10	*4.20*	*3.80*	3.40	3.00	2.90	*2.80*	*2.70*	*3.30*
Bronx	10.20	9.70	10.60	11.70	9.90	8.20	7.00	7.10	9.40
Broome	6.70	5.30	4.40	4.30	4.10	4.10	3.20	*4.20*	*6.20*
Chautauq	*6.60*	*5.90*	*5.20*	5.70	5.30	5.10	*4.50*	*5.40*	*6.30*
Chemung	5.30	4.80	4.30	4.40	4.20	4.80	4.80	5.40	7.10
Erie	5.90	5.20	4.90	5.10	5.20	5.20	4.60	5.10	5.80
Genesee	6.40	5.80	*5.10*	5.60	5.30	5.20	4.80	5.10	6.50
Herkimer	*6.70*	*7.00*	*6.70*	*6.70*	5.90	5.20	4.80	4.80	5.80
Kings	9.90	9.30	10.00	10.70	9.30	8.00	6.70	6.70	8.80
Livingst	*5.80*	*5.40*	*5.20*	5.20	5.10	4.90	4.40	4.60	6.20
Madison	*6.00*	*5.80*	*5.30*	5.10	4.60	4.60	4.20	5.10	5.70
Monroe	*4.90*	*3.90*	*3.50*	3.60	3.60	3.90	3.50	4.30	5.70
Nassau	5.00	4.50	3.80	3.50	3.00	3.00	2.70	3.10	4.10
New York	7.60	7.00	7.40	7.80	6.70	5.90	5.40	6.40	8.50
Niagara	*6.90*	*6.00*	*6.10*	6.60	6.60	6.10	*5.70*	*6.70*	*7.30*
Onondaga	5.10	4.60	4.10	3.90	3.50	3.50	3.50	4.10	5.00
Ontario	5.30	5.00	*4.50*	4.20	4.00	4.20	3.90	4.40	5.80
Orange	5.40	*4.90*	*4.30*	4.20	3.50	3.70	3.30	3.70	4.40
Oswego	*8.70*	*8.50*	*7.00*	7.00	5.80	6.30	6.20	6.60	7.60
Putnam	4.70	3.90	3.60	3.10	2.70	2.90	2.40	2.70	3.50
Queens	8.30	7.70	8.10	8.50	6.90	6.10	4.90	4.90	6.70
Rensselaer	5.30	5.40	*4.90*	*4.60*	*4.10*	*3.90*	*3.70*	*3.70*	*4.50*
Richmond	7.80	7.50	7.80	8.40	6.90	5.80	4.90	4.90	6.80
Rockland	5.30	4.90	4.30	4.00	3.40	3.40	2.90	3.10	3.90
Suffolk	6.20	5.40	4.60	4.30	3.60	3.60	3.10	3.50	4.50
Tioga	*6.80*	*5.60*	*4.70*	4.00	3.80	3.80	*3.30*	*4.30*	*6.10*
Wayne	*6.00*	*5.60*	*5.10*	4.80	4.80	5.00	4.30	5.40	7.50
Westchester	4.90	4.40	4.10	3.70	3.40	3.40	2.90	3.50	4.30

Table 1.5: **Table of Unemployment Rate (c.unemp)**

Figure 1.9: **Map of Unemployment Rate (c.unemp)**

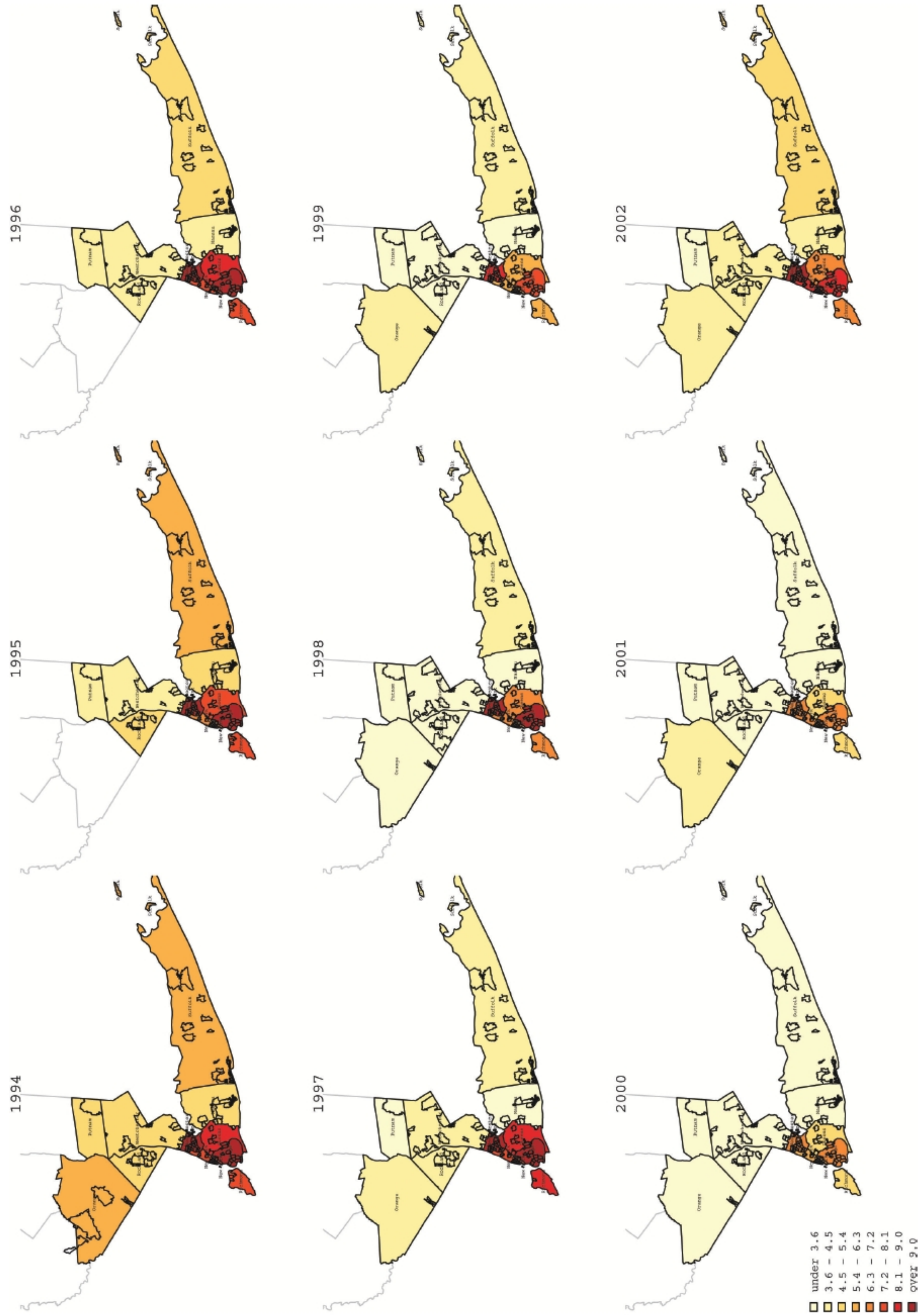

Figure 1.10: **Map of Unemployment Rate (c.unemp) (enlarged)**

1.3 ZIP Code Variables

z.ZIP is the only ZIP Code specific variable. It is a five-digits number identifying the exact area where a branch is located.

The following table shows the number of branches in a specific ZIP Code are and their county. There is a high concentration of branches (123) in ZIP Code 10001 (New York) followed by only 26 branches in 11001/11021 (Nassau). All others are distributed very sparsely.

# of branches	z.ZIP	c.name	# of branches	z.ZIP	c.name	# of branches	z.ZIP	c.name
123	10001	New York	1	10918	Orange	2	12084	Albany
4	10002	New York	1	10922	Orange	1	12180	Rensselaer
4	10003	New York	4	10923	Rockland	1	12729	Orange
2	10007	New York	3	10925	Orange	2	13027	Onondaga
17	10010	New York	1	10940	Orange	1	13029	Onondaga
1	10011	New York	2	10954	Rockland	1	13031	Onondaga
2	10012	New York	1	10956	Rockland	2	13032	Madison
2	10014	New York	1	10989	Rockland	2	13035	Madison
1	10015	New York	26	11001	Nassau	1	13036	Oswego
5	10021	New York	4	11003	Nassau	3	13057	Onondaga
1	10026	New York	2	11010	Nassau	3	13060	Onondaga
2	10027	New York	26	11021	Nassau	1	13214	Onondaga
2	10028	New York	2	11030	Nassau	1	13324	Herkimer
3	10034	New York	1	11096	Nassau	4	13732	Tioga
1	10041	New York	10	11101	Queens	1	13760	Broome
2	10104	New York	1	11104	Queens	4	13790	Broome
16	10167	New York	1	11203	Kings	4	13901	Broome
2	10301	Richmond	2	11204	Kings	2	14006	Erie
2	10302	Richmond	1	11207	Kings	2	14020	Genesee
3	10451	Bronx	1	11208	Kings	1	14043	Erie
3	10453	Bronx	3	11214	Kings	2	14052	Erie
1	10454	Bronx	3	11216	Kings	1	14072	Erie
1	10455	Bronx	1	11219	Kings	1	14094	Niagara
6	10458	Bronx	2	11220	Kings	3	14202	Erie
1	10459	Bronx	5	11223	Kings	1	14413	Wayne
2	10460	Bronx	1	11224	Kings	4	14420	Monroe
7	10461	Bronx	1	11225	Kings	1	14423	Livingst
9	10462	Bronx	3	11230	Kings	1	14428	Monroe
4	10463	Bronx	1	11364	Queens	2	14432	Ontario
1	10464	Bronx	4	11367	Queens	2	14435	Livingst
1	10472	Bronx	5	11368	Queens	1	14445	Monroe
6	10502	Westchester	1	11374	Queens	2	14450	Monroe
6	10504	Westchester	2	11416	Queens	1	14464	Monroe
7	10509	Putnam	4	11520	Nassau	1	14467	Monroe
2	10510	Westchester	1	11530	Nassau	1	14471	Ontario
1	10520	Westchester	4	11554	Nassau	1	14514	Monroe
2	10527	Westchester	3	11566	Nassau	1	14605	Monroe
1	10535	Westchester	2	11701	Suffolk	1	14606	Monroe
3	10538	Westchester	1	11702	Suffolk	2	14609	Monroe
1	10543	Westchester	4	11704	Suffolk	3	14611	Monroe
2	10550	Westchester	1	11713	Suffolk	2	14617	Monroe
2	10552	Westchester	1	11715	Suffolk	6	14701	Chautauq
2	10566	Westchester	2	11718	Suffolk	1	14722	Chautauq
1	10576	Westchester	1	11720	Suffolk	1	14733	Chautauq
1	10705	Westchester	3	11727	Suffolk	2	14814	Chemung
1	10901	Rockland	1	11901	Suffolk	1	14825	Chemung
1	10913	Rockland	1	11934	Suffolk			

Table 1.6: **Number of Branches in a specific ZIP Code and countyname**

1.4 Branch Variables

The following variables are constant within each branch over the same time period (1994-2002).

- b.ID
- b.long
- b.lat
- b.dep
- b.log.dep
- b.singleD
- b.mmcD

b.ID
b.ID is the branch identity number.

b.long
b.long is the branches' geographic position given in degrees east or west of the meridian.

b.lat
b.lat is the distance north or south of the equator measured in degrees.

b.dep
b.dep is total deposits (in US$) in the branch.

b.log.dep
b.log.dep is log(*b.dep*) (Distribution of *b.log.dep* is relatively symmetric).

b.singleD
b.singleD is the sum of all distances between the focal branch (F) and all the branches of other banks which only have one single branch.

b.mmcD
b.mmcD is the sum of all distances between the focal branch (F) and all the branches of other banks which also have multiple branches.

Here, each data line is a record of a branch and therefore referred as the focal branch (F). *b.mmcD* and *b.singleD* are intended to be an auxiliary measure for the competitive situation of a branch but may have been calculated based on the longitude and latitude readings (accurate only to two decimal places). For a better understanding of *b.mmcD* and *b.singleD* see Figure 1.11.

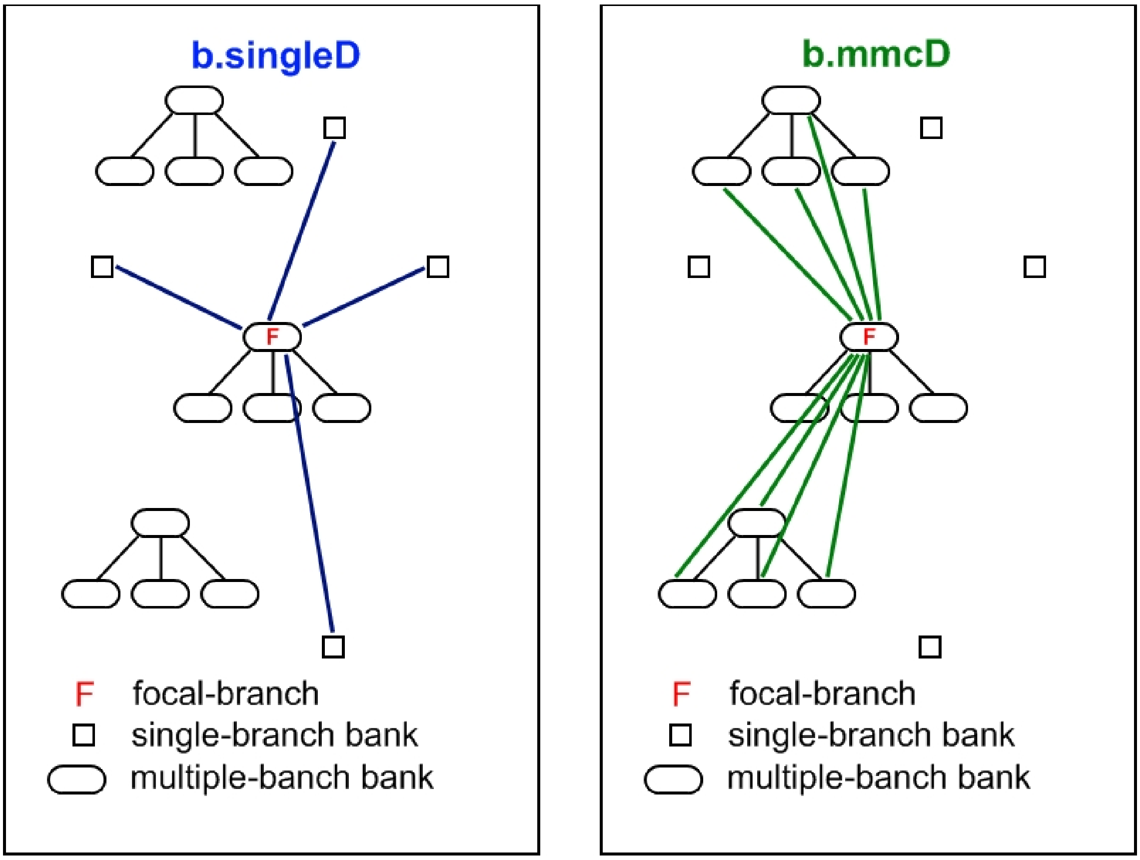

Figure 1.11: **Graphics of construction of variables b.singleD and b.mmcD**

The values of *b.singleD* range between 121.8 and 1092.0 and *b.mmcD* is between 6870 and 21140.
See Figure 1.12 for a better overview. Altogether there are nine plots (one for every year) in which the height (divided in quantiles) of the *b.singleD* are marked with different colors.
The color around NY city (see Figures 1.13 is very light in all figures indicating the distances from the focal bank to the rival branches are shorter than in the rural counties. Due to their high concentration within this relatively small, urban area, competition is high among them.

Figure 1.14) and 1.15 shows *b.mmcD* in quantile steps.
Similar to Figure 1.12, the light color again indicates that also the multiple-branch banks have many branches around the urban area of New York City.

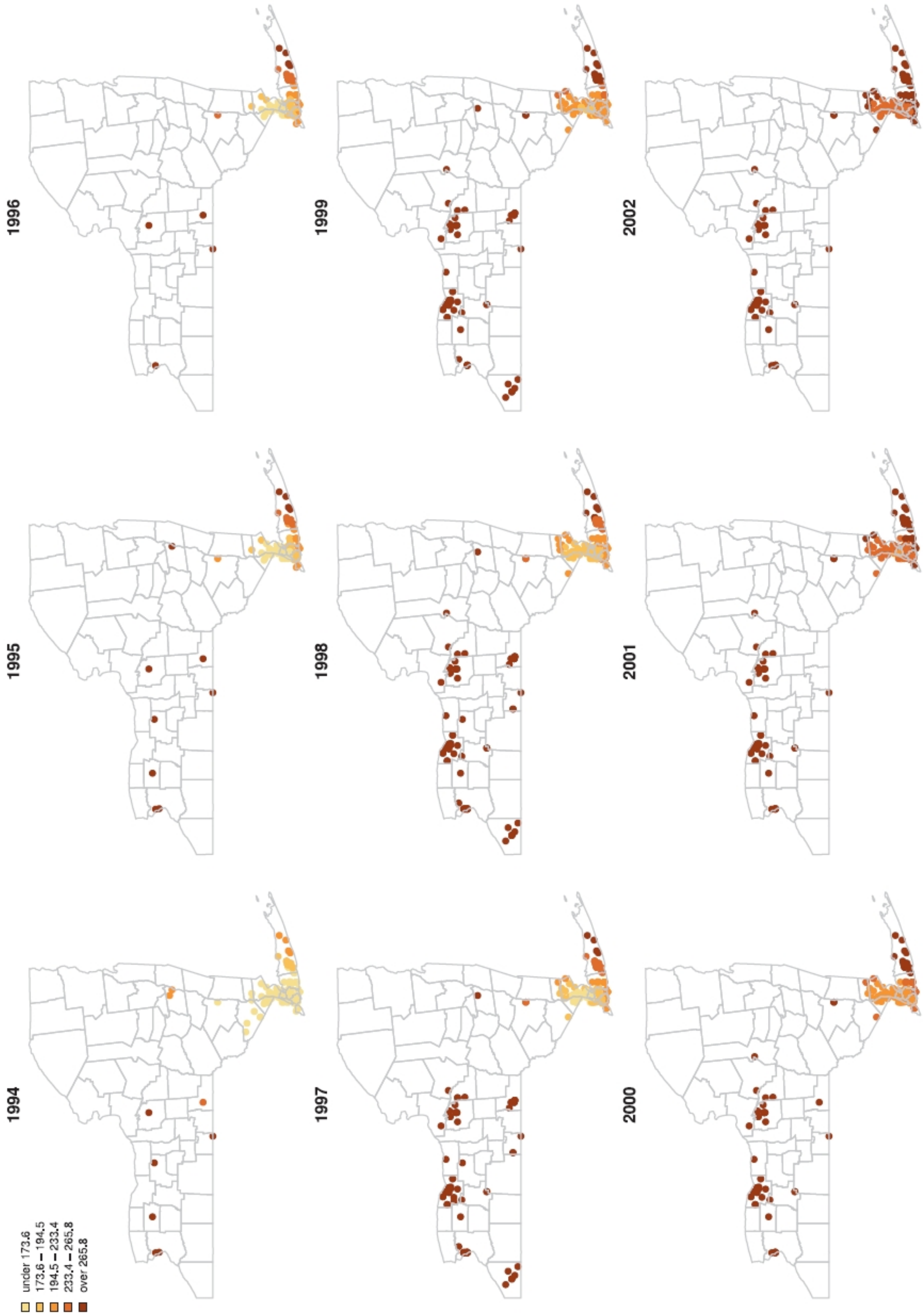

Figure 1.12: **Map of the sum distances between focal-branch and single-branch banks (as indicated by b.singleD)**

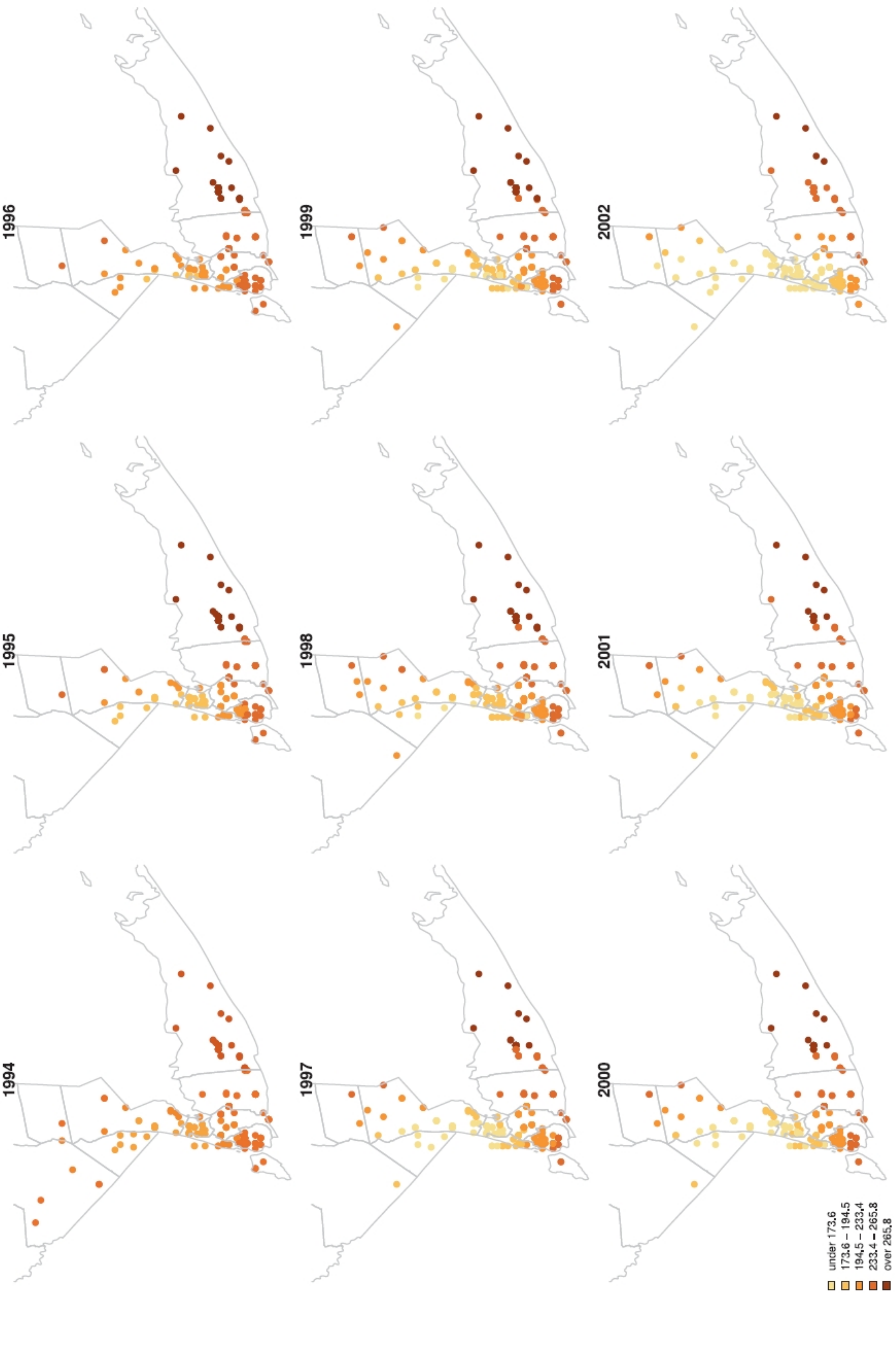

Figure 1.13: **Map of the sum distances between focal-branch and single-branch banks (indicated by b.singleD) (enlarged)**

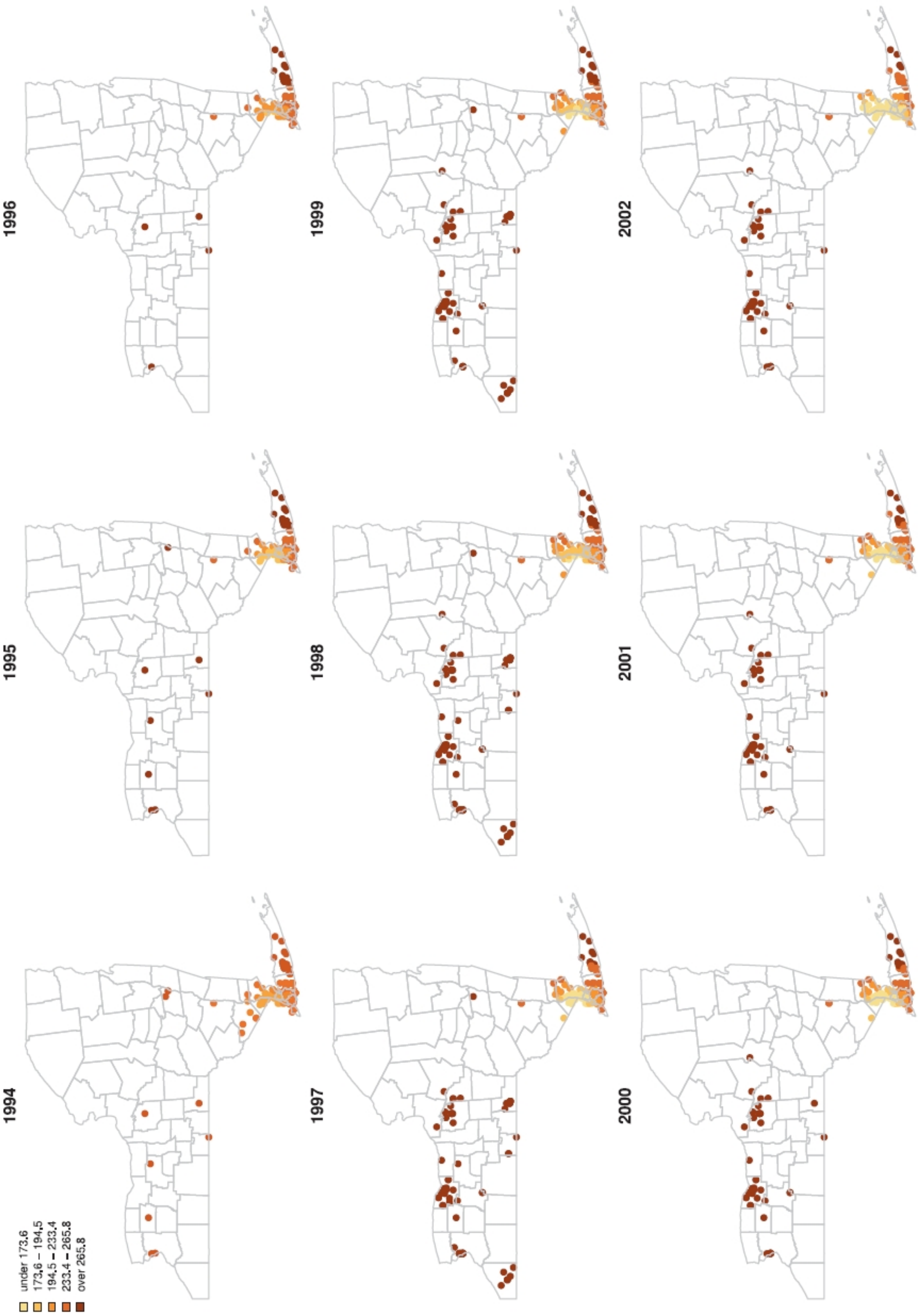

Figure 1.14: Map of the sum distances between focal-branch and multiple-branch banks (as indicated by b.mmcD)

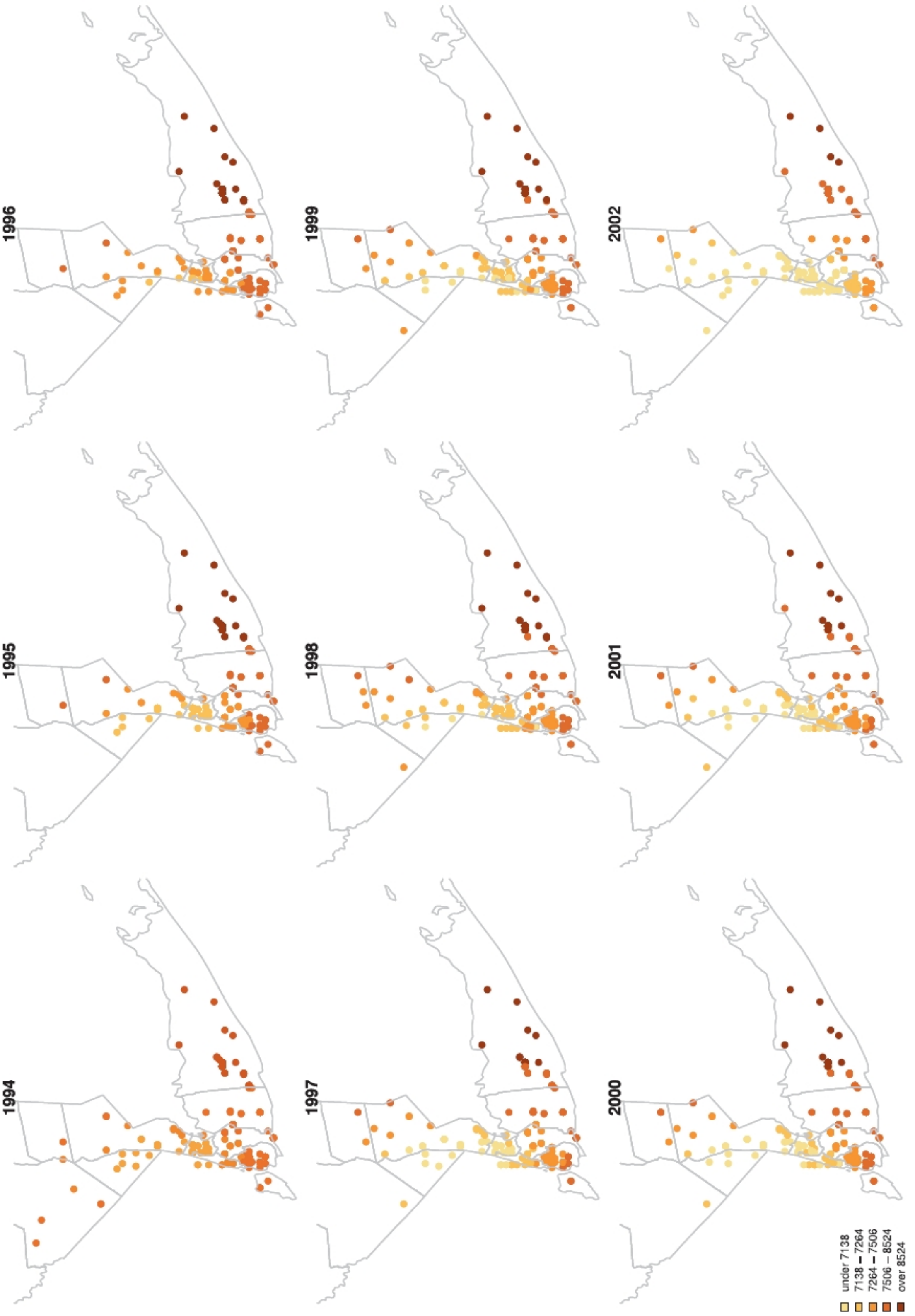

Figure 1.15: **Map of the sum distances between focal-branch and multiple-branch banks (as indicated by b.mmcD) (enlarged)**

Part II

Exploratory Data Analysis

The following chapters examine the correlation between response *b.dep* and the other variables. The dependent variable is taken in log form (*b.log.dep*), since the values of the *b.dep* are higher than zero and since there are some outliers which may otherwise have distorted our results. In addition *b.log.dep* is more normally distributed, so it is better suited for calculations than *b.dep*. Additionally, high monetary values are often shown in their log form in econometric analysis to allow for concepts, such as diminishing returns.

For this reason, the following examination is based solely on the response variable *b.log.dep*.

The following paragraphs are to present three models within the following. In the first model, the level of analysis are the branches of the bank. The response variable is *b.log.dep*. In the second and third model, we aggregate all Branch Variables to the county level and use them as response variables *c.log.dep.per.pop* (sum of county bank deposits divided by the population) and *c.log.dep* (sum of county bank deposits).

Chapter 2

Branch Model

2.1 State Variables

The lowess smoothing lines of the first five variables (*s.br.share*, *s.dep.share*, *s.dep.ave*, *s.no.fail*, *s.m.share*) shown in Figure 2.1 are already approximately linear to the response, hence, there is no need for transformation.

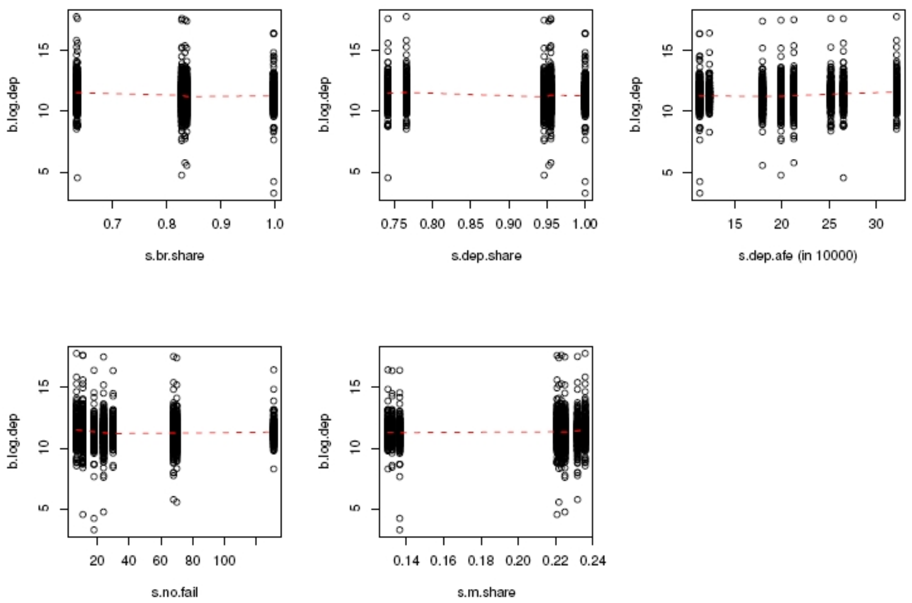

Figure 2.1: **Scatterplots of State Variables including lowess smoothing line (dashed red line)**

The response variable looks as follows:
For $i = 1, ..., 506$ branches and $t = 1, ..., 9$ years is defined:

$$\begin{aligned} y_{it} &= \text{response of } i\text{th branch in year } t \\ &= \beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{x}_t^{s.} + \boldsymbol{\varepsilon}_{it} \end{aligned}$$

where ε_{it} is assumed to be normally distributed with mean $\mathbf{0}$; $\boldsymbol{x}_t^{s.}$ are State Variables with $\boldsymbol{\beta}_1 := (\beta_1, \dots \beta_5)$.

Table 2.1 illustrates the chosen proceeding. Starting with the full model including all State Variables, insignificant effects according to the F-test are taken out sequentially (backward selection). The first model includes the 5 State Variables mentioned. As indicated in the ANOVA table, the variable *s.br.share* is insignificant with a p-value of 0.8165 so it is removed from the first model. Afterwards, the model without the *s.br.share* still has a variable with an insignificant p-value of 0.9281 namely *s.dep.share*; hence, it was also removed along with *s.no.fail* that has a p-value of 0.2066.

model	remove (p-value)
$b.log.dep = \beta_0 + \beta_1 s.br.share + \beta_2 s.dep.share + \beta_3 s.dep.ave + +\beta_4 s.no.fail + \beta_5 s.m.share$	*s.br.share* (0.8165)
$b.log.dep = \beta_0 + \beta_1 s.dep.share + \beta_2 s.dep.ave + \beta_3 s.no.fail + +\beta_4 s.m.share$	*s.dep.share* (0.9281)
$b.log.dep = \beta_0 + \beta_1 s.dep.ave + \beta_2 s.no.fail + \beta_3 s.m.share$	*s.no.fail* (0.2066)
$b.log.dep = \beta_0 + \beta_1 s.dep.ave + \beta_2 s.m.share$	

Table 2.1: **Backward selection for the State Variables**

Table 2.2 (See Appendix A.1.0.1 for the R commands) shows the details of the final model (s.B) with the significant variables *s.dep.ave* and *s.m.share* their p-values and the adjusted R-Squared of 0.01752.

	model
	s.B
s.br.share	
s.dep.share	
s.dep.ave	3.44e-13 ***
s.no.fail	
s.m.share	1.54e-07 ***
Adjusted R^2	0.01752

Table 2.2: **p-values (according to the F-test) of the State Variables in the Branch Model**[2]

Notice that R produces a statistic called adjusted R-Squared. The adjusted R-Squared is a modification of R-Squared that adjusts for the degrees of freedom in the model, and penalizes an unnecessarily complex linear model. The adjusted R-Squared is defined as

$$_{adj}r^2_{\boldsymbol{x}y} = 1 - \left(1 - r^2_{\boldsymbol{x}y}\right)\frac{n-1}{n-p-1} = 1 - \frac{s^2_{y|\boldsymbol{x}}/df_{y|\boldsymbol{x}}}{s^2_y/df_y}$$

[2] All regressions include a constant. ***,**,* indicate significance at the 1, 5, and 10% level, respectively.

where p is the total number of regressors in the linear model (though not counting the constant term), n is the sample size and $df_{y|\boldsymbol{x}}/df_y$ are the degrees of freedom for $s^2_{y|\boldsymbol{x}}/s^2_y$ and $s^2_{y|\boldsymbol{x}}$ is the square of the error of a linear regression of $\boldsymbol{x}_i$ on y_i by the equation $y = a + \boldsymbol{b}^t\boldsymbol{x}$:

$$s^2_{y|\boldsymbol{x}} = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - a - \boldsymbol{b}^t\boldsymbol{x}_i\right)^2$$

s^2_y is just the variance of y:

$$s^2_y = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

The adjusted R-Squared can be negative, and will always be less than or equal to R-Squared then unlike the R-Squared, the adjusted R-Squared increases only if the new term improves the model.

2.2 Adding County Variables

Now County Variables will be added to the above model s.B.
The plot in Figure 2.2 plots the response *b.log.dep* against *c.pop* (in 10000) / *c.inc* (in million) / *c.unemp* as well as the response against *c.inc.per.ca*.
On the second plot (Figure 2.3) there are 9 colored "LOWESS"-lines referring to the respective nine years under examination (1994 until 2002).

A very popular technique for curve fitting complicated data sets is called "LOWESS" (locally weighted smoothing scatter plots, sometimes called *loess*). In "LOWESS", the data is modeled locally by a polynomial weighted least squares regression, the weights giving more importance to the local data points. This method of approximating data sets is called *locally weighted polynomial regression*.

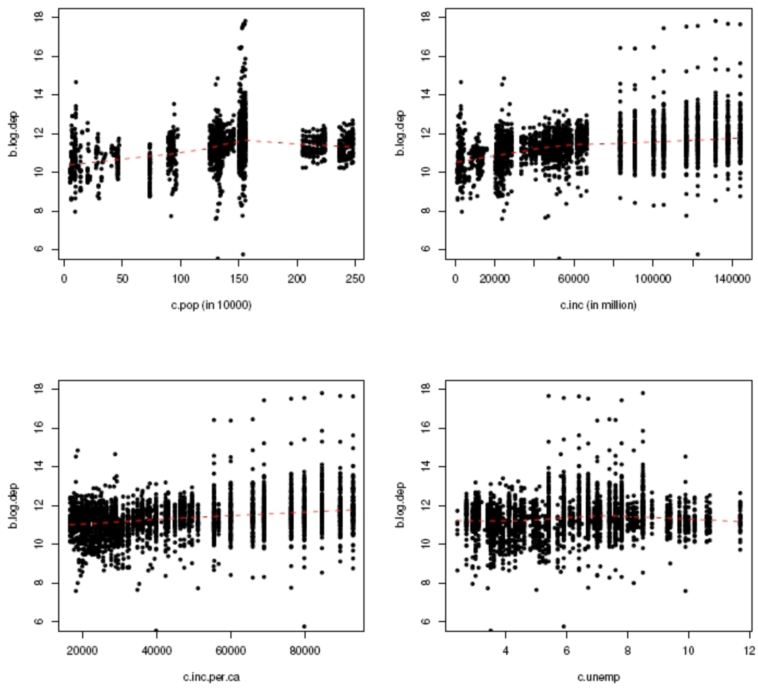

Figure 2.2: **Scatterplots of County Variables including lowess smoothing line (dashed red line)**

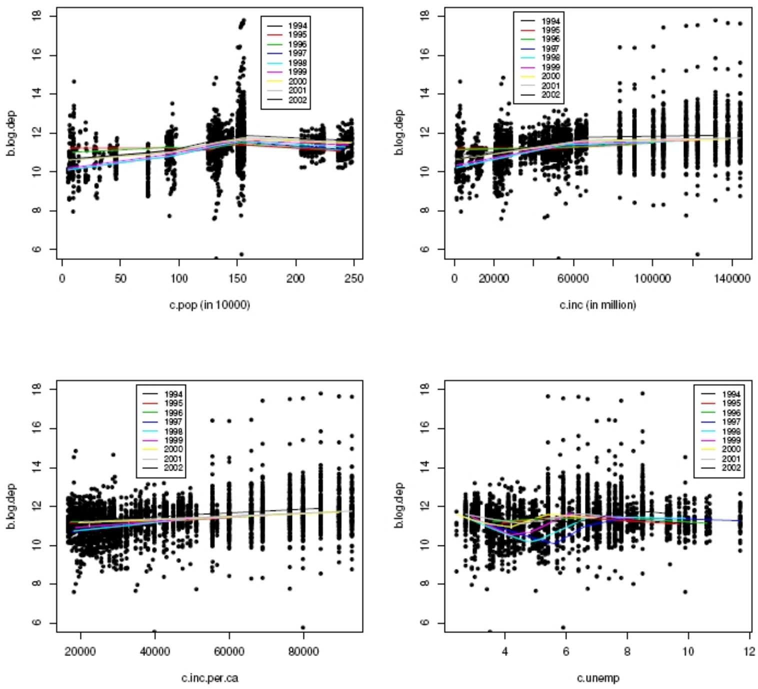

Figure 2.3: **Scatterplots of County Variables including the 9 different lowess smoothing lines for each year**

Transformation of County Variables

As the red dashed line of the scatterplot *b.log.dep* against *c.inc* (in million) approximately has the functional form of a square root function this variable is transformed to *c.sqrt.inc* as shown in Figure 2.4.

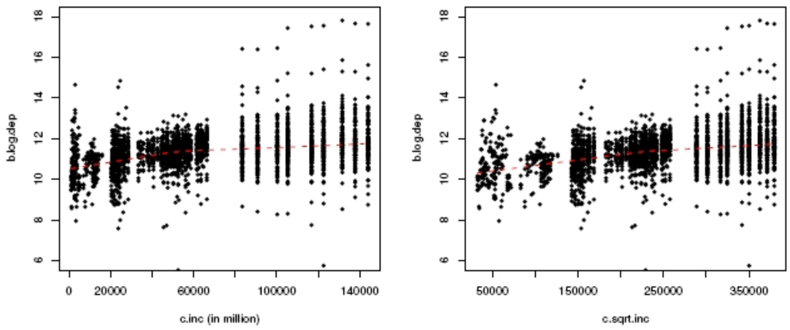

Figure 2.4: **Scatterplots of the transformation of County Variable** *c.inc* **including lowess smoothing lines**

As the variables *c.inc* and *c.inc.per.ca* are highly correlated (due to *c.inc.per.ca* being a standardized variable of *c.inc*), they will not be used in the same regression, so we devise two different models where the first one (model c.Ba) includes only the variables *c.pop*, *c.sqrt.inc* and *c.unemp*, the second one (model c.Bb) includes *c.pop*, *c.inc.per.ca* and *c.unemp*. After comparing both models, model c.Ba is chosen for the further analysis. (see Table 2.4)

In addition to the branches, the model now contains also the county observations: For $i = 1, ..., 506$ branches, $j = 1, ..., 28$ counties and $t = 1, ..., 9$ years is defined:

$$\begin{aligned} y_{ijt} &= \text{response of } i\text{th branch in county } j \text{ in year } t \\ &= \beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{x}_t^{s\cdot} + \boldsymbol{\beta}_2^T \boldsymbol{x}_{jt}^{c\cdot} + \boldsymbol{\varepsilon}_{ijt} \end{aligned}$$

where ε_{ijt} is assumed to be normally distributed with mean $\mathbf{0}$, $\boldsymbol{x}_t^{s\cdot}$ are State Variables and $\boldsymbol{x}_{jt}^{c\cdot}$ are County Variables with $\boldsymbol{\beta}_1 := (\beta_1, \beta_2)$ and $\boldsymbol{\beta}_2 := (\beta_3, \ldots, \beta_5)$.

2.2.1 Model c.Ba (with *c.pop* + *c.sqrt.inc* + *c.unemp*)

Model c.Ba includes the two State Variables mentioned above and three County Variables (*c.pop*, *c.sqrt.inc*, *c.unemp*). To assess the statistical significance of these three County Variables, we proceed via backward selection again.

Because all state and County Variables are statistically significant and improve the R-Squared as well as the adjusted R-Squared, the backward elimination does not result in the removal of any variable.

The final model c.Ba with the variables *s.dep.ave*, *s.m.share*, *c.pop*, *c.sqrt.inc*, *c.unemp*, their p-values and the adjusted R-Squared with 0.1534 is shown in Table 2.4. (See Appendix A.1.0.2 for the R-Code)

2.2.2 Model c.Bb (with *c.pop* + *c.inc.per.ca* + *c.unemp*)

The following Table 2.3 presents the backward elimination including the two State Variables *s.dep.ave*, *s.m.share* and the three County Variables *c.pop*, *c.inc.per.ca* and *c.unemp*. According to the ANOVA table the variable *c.unemp* is insignificant with a p-value of 0.1419 so it is removed from the first model. The result is a statistically significant model c.Bb.

model	remove (p-value)
$b.log.dep = \beta_0 + \beta_1 s.dep.ave + \beta_2 s.m.share + \beta_3 c.pop +$ $+\beta_4 c.inc.per.ca + \beta_5 c.unemp$	*c.unemp* (0.1419)
$b.log.dep = \beta_0 + \beta_1 s.dep.ave + \beta_2 s.m.share + \beta_3 c.pop +$ $+\beta_4 c.inc.per.ca$	

Table 2.3: **Backward selection for the County Variables (model c.Bb)**

The final model c.Bb with the variables *s.dep.ave*, *s.m.share*, *c.pop*, *c.inc.per.ca*, their p-values and the adjusted R-Squared with 0.1515 is shown in Table 2.4 below. (See Appendix A.1.0.3 for the R commands)

	model	
	c.Ba	c.Bb
s.br.share		
s.dep.share		
s.dep.ave	7.04e-07 ***	3.68e-06 ***
s.no.fail		
s.m.share	0.00101 **	0.000264 ***
c.pop	0.00523 **	$<$ 2e-16 ***
c.sqrt.inc	$<$ 2e-16 ***	
c.inc.per.ca		$<$ 2e-16 ***
c.unemp	0.03402 *	
Adjusted R^2	0.1534	0.1515

Table 2.4: **p-values of the County Variables in the Branch Model**

The model c.Ba with *c.sqrt.inc* has an adjusted R-squared of 0.1534 instead of *c.inc.per.ca* with adjusted R-squared of 0.1515 (see Table 2.4). So the model c.Ba seems to be the better one and model c.Bb was rejected.

2.2.3 Interaction effects with time (c.B.ct)

We want to examine all County Variables for an interaction with time.

Therefore, we analyze the interaction plots in Figure 2.5.

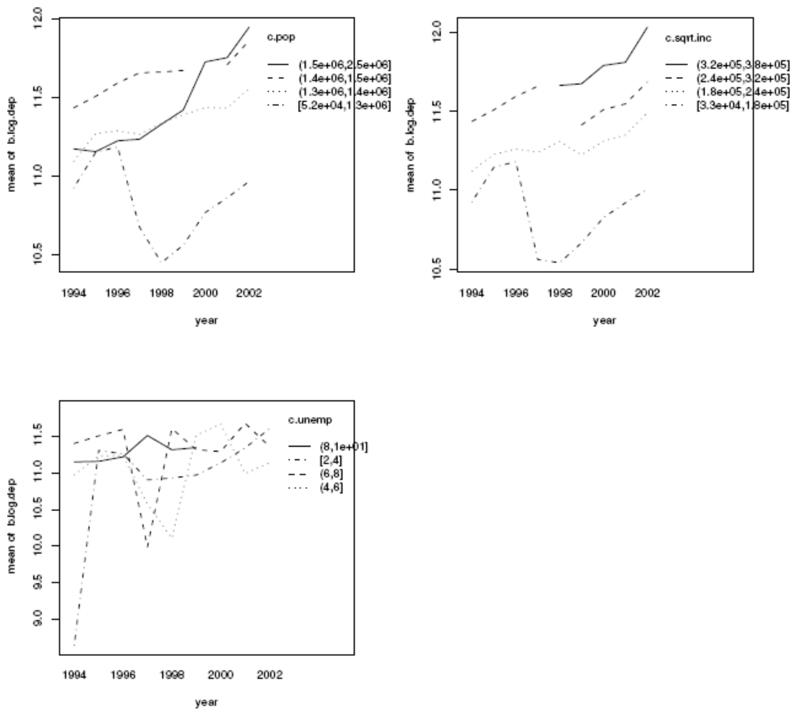

Figure 2.5: **Interactions between time and the County Variables**

Generally, an interaction plot displays the levels of one variable on the X-axis and separate lines for the means of each level of the other variable. In this case, the graphs below depict the interaction of time (variable *year* plotted on the X-axis) and the County Variables (here the four quantiles of *c.pop*, *c.sqrt.inc* and *c.unemp*). The Y-axis shows the dependent variable (here *b.log.dep*).

- If the lines are nearly parallel to each other there is no evidence for an interaction effect
- A nonparallel shape is an indication of an interaction between the factors

The plots in Figure 2.5 suggest possible interaction effects between the variable *year* and the other County Variables. Consequently they were added to the model c.Ba linearly. By adding the variable *year* singularities are produced because the State Variables display a timetrend so *year* was omitted.

This leads to the following model specification:
For $i = 1, ..., 506$ branches, $j = 1, ..., 28$ counties and $t = 1, ..., 9$ years is defined:

$$\begin{aligned} y_{ijt} &= \text{response of } i\text{th branch in county } j \text{ in year } t \\ &= \beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{x}_t^{s.} + \boldsymbol{\beta}_2^T \boldsymbol{x}_{jt}^{c.} + \sum_{z=2}^{9} \beta_{6+z-2} D(z) x_{jt}^{c.} + \boldsymbol{\varepsilon}_{ijt} \end{aligned}$$

where $D(z) := \begin{cases} 0 & \text{if } z \neq t \\ 1 & \text{if } z = t \end{cases}$

Table 2.5 below shows the backward elimination. The first model includes the five variables from the final model above and in addition the three intaractions with the time variable. But with the ANOVA table we see, that the interaction *c.sqrt.inc* : *year* is insignificant with a p-value of 0.9586 so it is removed from the first model. The model without the *c.sqrt.inc* : *year* still had a variable with an insignificant p-value of 0.7779 namely *c.unemp* : *year* and is also removed.

model	remove (p-value)
$b.log.dep = \beta_0 + \beta_1 s.dep.ave + \beta_2 s.m.share + \beta_3 c.pop +$ $+\beta_4 c.sqrt.inc + \beta_5 c.pop : year + \beta_6 c.sqrt.inc : year +$ $+\beta 7 c.unemp : year$	*c.sqrt.inc* : *year* (0.9586)
$b.log.dep = \beta_0 + \beta_1 s.dep.ave + \beta_2 s.m.share + \beta_3 c.pop +$ $+\beta_4 c.sqrt.inc + \beta_5 c.pop : year + \beta_6 c.unemp : year$	*c.unemp* : *year* (0.7779)
$b.log.dep = \beta_0 + \beta_1 s.dep.ave + \beta_2 s.m.share + \beta_3 c.pop +$ $+\beta_4 c.sqrt.inc + \beta_5 c.pop : year$	

Table 2.5: **Backward selection for the interactions county with *year* variables**

See Table 2.6 of the final model c.B.ct with the variables *s.dep.ave*, *s.m.share* and their p-values the adjusted R-Squared with 0.1563. (The R-Code is shown in the Appendix A.1.0.4).

	model
	c.B.ct
s.br.share	
s.dep.share	
s.dep.ave	0.000430 ***
s.no.fail	
s.m.share	2.83e-06 ***
c.pop	0.172197
c.sqrt.inc	< 2e-16 ***
c.unemp	0.072345 .
c.pop:year	
Adjusted R^2	0.1563

Table 2.6: **p-values of the interactions county with *year* in the Branch Model**

2.2.4 Interactions among County Variables in the Branch Model (c.B.cc)

We are also interested in the interaction effects among the County Variables.

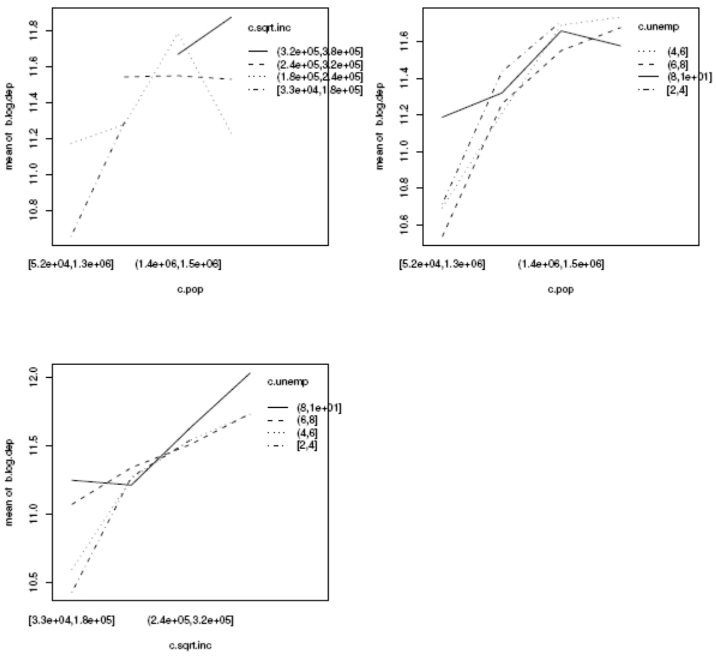

Figure 2.6: **Interactions among the County Variables in the Branch Model**

Figure 2.6 displays the effects. Since the lines are all nonparallel, all variables should be included to the model.

With the backward elimination we see that from the full model the variables *c.unem* : *c.sqrt.inc* and *c.pop* : *c.sqrt.inc* have to be removed because the p-values in the ANOVA tabel was 0.1858 for *c.unem* : *c.sqrt.inc* and in the next model (without *c.unem* : *c.sqrt.inc*) the p-value for *c.pop* : *c.sqrt.inc* was 0.1340. So we only keep the interaction *c.pop* : *c.unemp*.
The Table 2.7 of the final model c.B.cc with their p-values the adjusted R-Squared with 0.1592 is shown below (see also the R-Code in the Appendix A.1.0.5).

	model
	c.B.cc
s.dep.ave	0.000430 ***
s.m.share	2.83e-06 ***
c.pop	0.172197
c.sqrt.inc	$<$ 2e-16 ***
c.unemp	0.072345 .
c.pop:year	0.01951 *
c.pop:c.unemp	0.000799 ***
Adjusted R^2	0.1592

Table 2.7: **p-values of the interactions among County Variables in the Branch Model**

2.2.5 Adding Branch Variables

After the County Variables we want to add the Branch Variables *b.singleD* and *b.mmcD*.

The plots in Figure 2.7 and 2.8 (with 9 colored "LOWESS"-lines) again show the response against *b.singleD* and *b.mmcD*.

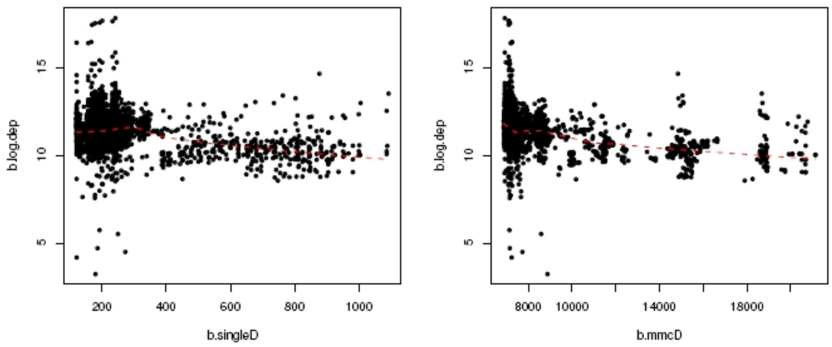

Figure 2.7: **Scatterplots of Branch Variables including lowess smoothing line (dashed red line)**

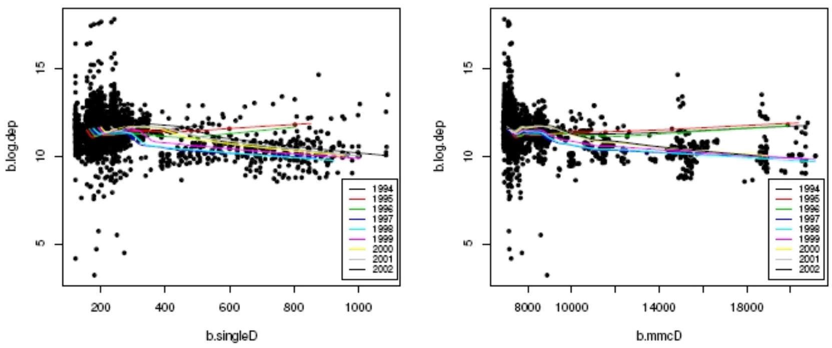

Figure 2.8: **Scatterplots of Branch Variables including 9 lowess smoothing lines for the years**

There seems to be no need to transform them so they are added to the model c.B.cc. Implementing the backward elimination as before, we see that the variable *b.mmcD* is insignificant with an p-value of 0.7241 and is therfore removed.
Table 2.8 shows the final model b.B with their p-values the adjusted R-Squared with 0.1829. (R-Code see Appendix A.1.0.6)

Interactions between time and the Branch Variable

In addition, we examine the interaction between the time variable and the Branch Variable.

Even though Figure 2.9 indicates that it might be worthwhile to add them because the lines are not parallel, the ANOVA table shows that the interaction is insignificant with a p-value of 0.107.
The R-Code of the ANOVA table is shown in the Appendix A.1.0.7.

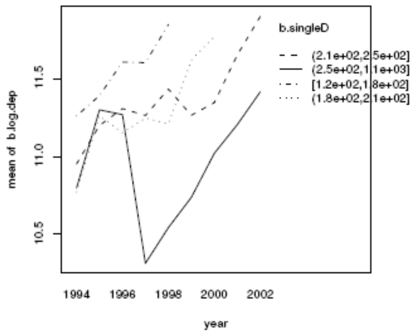

Figure 2.9: **Interactions between time and the Branch Variable**

Finally, the best Branch Model we found is the one before model b.B with an adjusted R-Squared of 0.1829 (see Table 2.8).

	model
	b.B
s.br.share	
s.dep.share	
s.dep.ave	2.02e-05 ***
s.no.fail	
s.m.share	0.01611 *
c.pop	0.00602 **
c.sqrt.inc	3.00e-16 ***
c.unemp	7.52e-05 ***
b.singleD	$<$ 2e-16 ***
b.mmcD	
c.pop:year	0.3019
c.pop:c.unemp	8.18e-05 ***
Adjusted R^2	0.1829

Table 2.8: **p-values of the Branch Variables in the Branch Model**

2.3 Interaction effects between all variables

So far, only the county level was analyzed and possible interactions at every stage were examined, i.e. only among County Variables. With the State Variables this did not make sense because there was only one observation per year for each of them and since we had only one Branch Variable left, interaction effects were obsolete anyway. Here we want to examine all interactions, so we firstly focus on the whole model with all State, County and Branch Variables as well as their interactions between all of them, which means moving along several levels of analysis. Using the backward selection we eliminate all insignificant variables step by step. Again we examined the two models all.Ba and all.Bb with *c.inc* respectively *c.inc.per.ca* and end with the following models (see Table 2.9 and the R-Code for model all.Ba and model all.Bb in the Appendix A.1.0.8 and A.1.0.9)

Model all.Ba have an adjusted R-Squared of 0.2089 which is better than model b.B before with 0.1829 and better than model all.Bb, so this is our best Branch Model. All three models have very low explanatory power with an adjusted R-Squared of 0.21 or lower.

	model		
	b.B	all.Ba	all.Bb
s.br.share		0.37675	0.005632 **
s.dep.share		0.78138	0.009223 **
s.dep.ave	2.02e-05 ***	0.03181 *	0.001037 **
s.no.fail			
s.m.share	0.01611 *	0.27215	0.020027 *
c.pop	0.00602 **	1.84e-12 ***	0.012856 *
c.sqrt.inc	3.00e-16 ***	2.64e-16 ***	
c.inc.per.ca			0.001577 **
c.unemp	7.52e-05 ***	< 2e-16 ***	2.65e-10 ***
b.singleD	< 2e-16 ***	0.01047 *	0.000305 ***
b.mmcD		0.03734 *	0.000912 ***
c.pop:year	0.3019		
c.pop:c.sqrt.inc		0.00280 **	
c.pop:c.inc.per.ca			0.024600 *
c.pop:c.unemp	8.18e-05 ***	4.13e-08 ***	0.012764 *
c.pop:b.singleD			0.000425 ***
c.pop:b.mmcD		7.60e-15 ***	0.000141 ***
c.sqrt.inc:b.mmcD		< 2e-16 ***	
c.inc.per.ca:b.mmcD			3.56e-07 ***
c.unemp:b.mmcD		3.23e-14 ***	1.70e-09 ***
s.br.share:b.singleD		1.30e-06 ***	0.000222 ***
s.dep.share:b.singleD		7.62e-06 ***	0.000451 ***
s.dep.ave:b.singleD		5.49e-05 ***	2.43e-07 ***
s.m.share:b.singleD		0.01038 *	0.004412 **
s.br.share:b.mmcD		3.21e-05 ***	0.002071 **
s.dep.share:b.mmcD		0.00782 **	0.003930 **
s.dep.ave:b.mmcD		0.00777 **	0.000108 ***
s.m.share:b.mmcD			0.012176 *
s.dep.ave:c.pop			0.032713 *
s.dep.ave:c.unemp		0.07978 .	0.046054 *
Adjusted R^2	0.1829	0.2089	0.2015

Table 2.9: **p-values of the final Branch Models**

Chapter 3

County Model 1

We now turn to the question which model has higher explanatory power and improve the R-Squared. Our first idea was to aggregate the response *b.dep* and divide by the population, so we get a new response variable, namely

$$c.log.dep.per.pop := log\left(\frac{\sum_{county} b.dep}{c.pop}\right)$$

where $\sum_{county}$ is the sum of all branches in a certain county. Like discussed extensively in Chapter 2 for the Branch Model, in this section, we first examine the state, then the County Variables. The Branch Variables are dropped since we use a new response variable.

3.1 State Variables

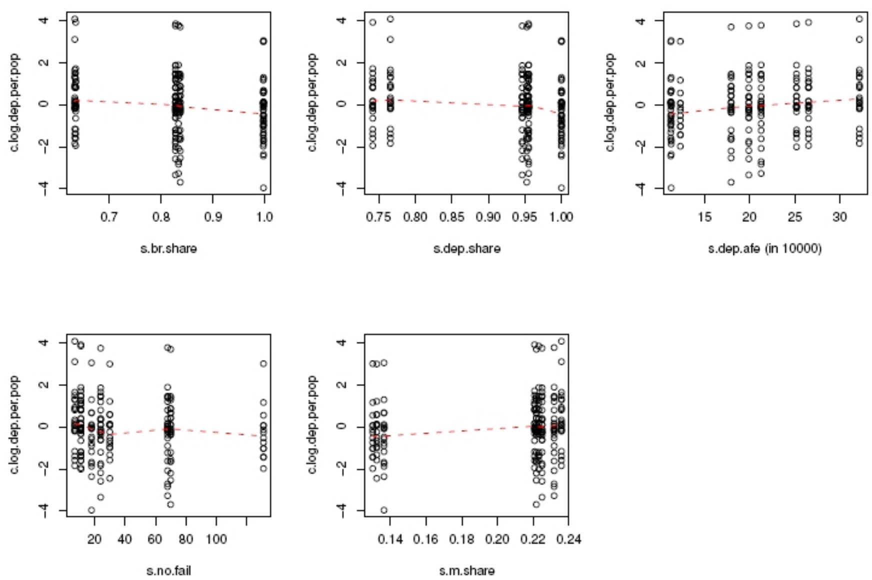

Figure 3.1: **Scatterplots of State Variables including lowess smoothing line (dashed red line)**

The Figure 3.1 is similar to the plot before in the branch model (see Figure 2.1) so there is no need to transform the State Variables.

The model is specified as follows.
For $j = 1, ..., 28$ counties and $t = 1, ..., 9$ years is defined:

$$\begin{aligned} y_{jt} &= \text{response of } j\text{th county in year } t \\ &= \beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{x}_t^{s.} + \boldsymbol{\varepsilon}_{jt} \end{aligned}$$

The backward selection with this new response *c.log.per.pop* eliminates all State Variables except *s.br.share*. For further details, see the Tables 3.1 and 3.2 below.

model	remove (p-value)
$c.log.dep.per.pop = \beta_0 + \beta_1 s.br.share + \beta_2 s.dep.share + \beta_3 s.dep.ave + \beta_4 s.no.fail + \beta_5 s.m.share$	$s.no.fail$ (0.964)
$c.log.dep.per.pop = \beta_0 + \beta_1 s.br.share + \beta_2 s.dep.share + \beta_3 s.dep.ave + \beta_4 s.m.share$	$s.dep.ave$ (0.711)
$c.log.dep.per.pop = \beta_0 + \beta_1 s.br.share + \beta_2 s.dep.share + \beta_3 s.m.share$	$s.dep.share$ (0.313)
$c.log.dep.per.pop = \beta_0 + \beta_1 s.br.share + \beta_2 s.m.share$	$s.m.share$ (0.838)
$c.log.dep.per.pop = \beta_0 + \beta_1 s.br.share$	

Table 3.1: **Backward selection for the State Variables**

	model
	s.C1
s.br.share	0.0153 *
s.dep.share	
s.dep.ave	
s.no.fail	
s.m.share	
Adjusted R^2	0.02479

Table 3.2: **p-values of the State Variables in the County Model 1**

3.2 Adding County Variables

Furthermore the scatterplots of the new response against the County Variables *c.inc* (in million), *c.inc.per.ca* and *c.unemp* is shown in Figure 3.3 and 3.3 with 9 colored "LOWESS"-lines referring to the nine years. The variable *c.pop* is indirect in the new response so it is omitted. As the dashed lines display a u-shape, we transform the variables and add quadratics.

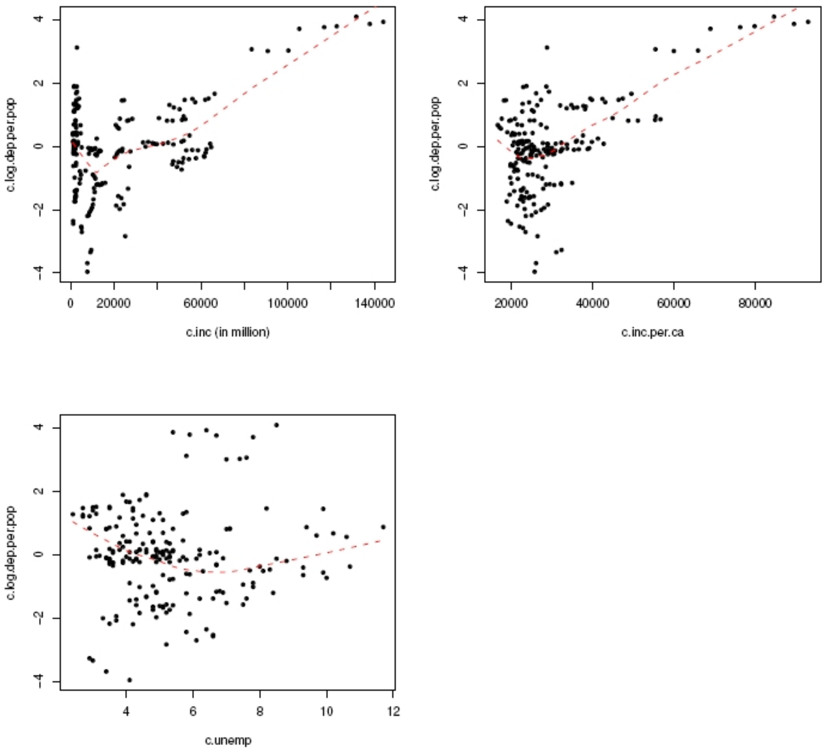

Figure 3.2: **Scatterplots of County Variables including lowess smoothing line (dashed red line)**

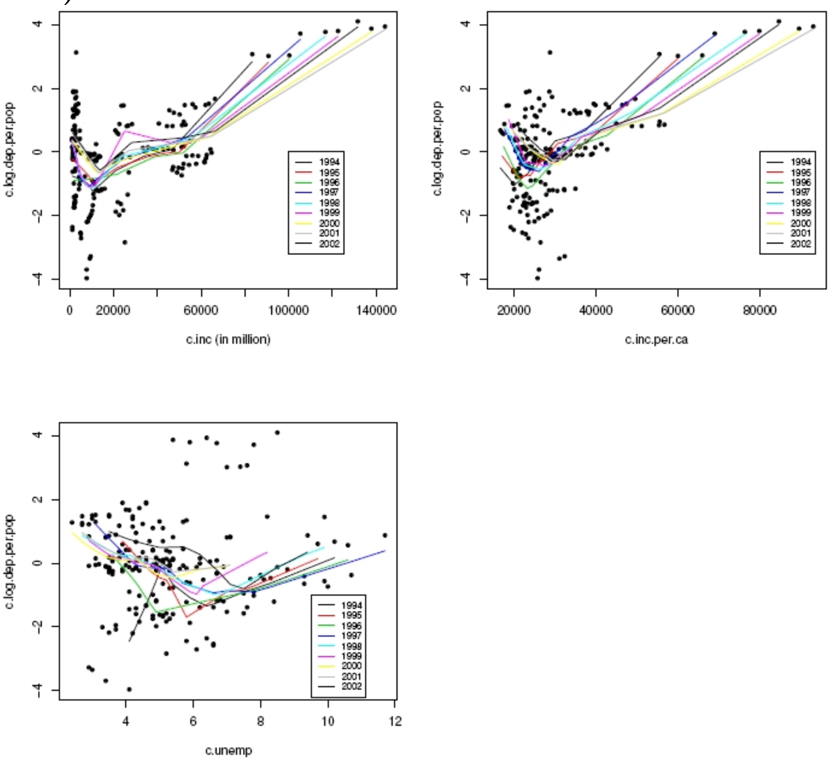

Figure 3.3: **Scatterplots of County Variables including lowess smoothing lines**

This leads to the following model specification:
For $j = 1, ..., 28$ counties and $t = 1, ..., 9$ years is defined:

$$\begin{aligned} y_{jt} &= \text{response of } j\text{th county in year } t \\ &= \beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{x}_t^{s.} + \boldsymbol{\beta}_2^T \boldsymbol{x}_{jt}^{c.} + \boldsymbol{\varepsilon}_{jt} \end{aligned}$$

With the same motivation like above we examine two models, one (model c.C1a) with *c.inc* and the other (model c.C1b) with *c.inc.per.ca* to decide which one is the better one afterwards.

3.2.1 Model c.C1a (with *c.inc*)

The backward elimination removed no variable and results an R-Squared of 0.3578. (see R-Code in Appendix A.2.0.11)

3.2.2 Model c.C1b (with *c.inc.per.ca*)

The backward selection of the second model removed the quadratic terms as well as *s.br.share* and left only *c.inc.per.ca* and *c.unemp* (see Table 3.3), with an R-Squared of 0.3706 (see R-Code in Appendix A.2.0.12)

model	remove (p-value)
$c.log.dep.per.pop = \beta_0 + \beta_1 s.br.share + \beta_2 poly(c.inc.per.ca, 2)[, 1]+$ $+\beta_3 poly(c.inc.per.ca, 2)[, 2] + \beta_4 poly(c.unemp, 2)[, 1]+$ $+\beta_5 poly(c.unemp, 2)[, 2]$	$poly(c.inc.per.ca, 2)[, 2]$ (0.3158)
$c.log.dep.per.pop = \beta_0 + \beta_1 s.br.share + \beta_2 poly(c.inc.per.ca, 2)[, 1]+$ $+\beta_3 poly(c.unemp, 2)[, 1] + \beta_4 poly(c.unemp, 2)[, 2]$	$poly(c.unemp, 2)[, 2]$ (0.1957)
$c.log.dep.per.pop = \beta_0 + \beta_1 s.br.share + \beta_2 poly(c.inc.per.ca, 2)[, 1]+$ $+\beta_3 poly(c.unemp, 2)[, 1]+$	$s.br.share$ (0.1237)
$c.log.dep.per.pop = \beta_0 + \beta_1 poly(c.inc.per.ca, 2)[, 1]+$ $+\beta_2 poly(c.unemp, 2)[, 1]+$	

Table 3.3: **Backward selection for the County Model 1 with the variable c.inc.per.ca**

Comparing the two County Models on the basis of the R-Squared (see Table 3.4), we see that the one with *c.inc.per.ca* is a bit better, so we continue with the final model c.C1b.

	model	
	c.C1a	c.C1b
s.br.share	0.0337 *	
s.dep.share		
s.dep.ave		
s.no.fail		
s.m.share		
c.inc	2.84e-15 ***	
c.inc^2	1.65e-06 ***	
c.inc.per.ca		<2e-16 ***
c.unemp	0.3450	0.0139 *
c.unemp^2	0.0120 *	
Adjusted R^2	0.3578	0.3706

Table 3.4: **p-values of the County Variables in the County Model 1**

3.3 Interaction effects with time (c.C1.ct)

Figure 3.4 shows an interaction plot between the variables *c.inc.per.ca*, *c.unemp* and the variable *year*. As the lines do not display a clear parallel format we add the dummy-variable *year* and the interaction with *c.inc.per.ca* and *c.unemp* to the model.

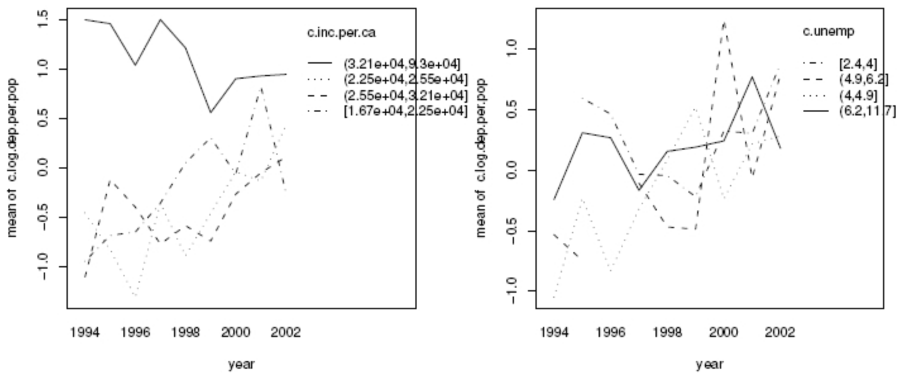

Figure 3.4: **Interactions between time and the County Variables**

Again, the model specification is:
For $j = 1, ..., 28$ counties and $t = 1, ..., 9$ years is defined:

$$\begin{aligned} y_{jt} &= \text{response of } j\text{th county in year } t \\ &= \beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{x}_{jt}^{c.} + \sum_{z=2}^{9} \beta_{3+z-2} D(z) + \sum_{z=2}^{9} \beta_{10+z-2} D(z) x_{jt}^{c.} + \boldsymbol{\varepsilon}_{jt} \end{aligned}$$

But the R-Code (shown in Appendix A.2.0.13) confirms that these variables are not significant, so we remove them.

There are no other variables left the first model c.C1b had an adjusted R-Squared of 0.3706. (see Table 3.4)

3.4 Interaction effects among County Variables in the County Model 1 (c.C1.cc)

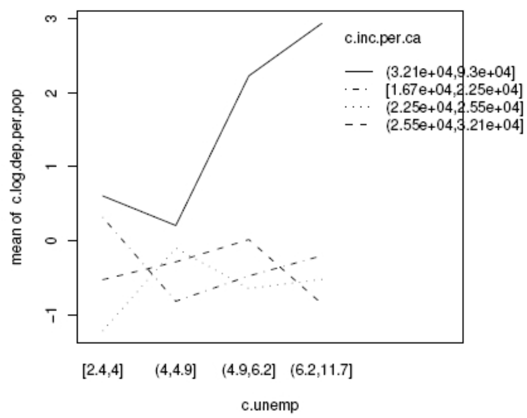

Figure 3.5: **Interactions among the County Variables in the County Model 1**

Figure 3.5 displays the effects. Since the lines are all nonparallel, all interactions should be included to the model.

With the backward elimination we see that *c.unemp* : *c.inc.per.ca* has to be removed because the p-values in the ANOVA tabel is 0.778. So we only keep the model c.C1b with the variables *c.inc.per.ca* and *c.unemp*.

Table 2.7 of the final model c.B.cc with their p-values the adjusted R-Squared with 0.1592 is shown below (see also the R-Code in the Appendix A.2.0.14).

3.5 Interaction effects between all variables

As in the analysis of the Branch Model (see Section 2), we now examine interaction effects among all levels. All variables and their interactions are put into the model and with the backward selection eleminated step by step. In the end we achieve the following Table 3.5 of model all.C1a (with *c.inc*) and for model all.C1b (with *c.inc.per.ca*). (R command see Appendix A.2.0.15 and A.2.0.16)

	model	
	all.C1a	all.C1b
s.br.share		
s.dep.share		
s.dep.ave	0.082932 .	0.083720 .
s.no.fail		
s.m.share		
c.pop		
c.inc	2.28e-09 ***	
$c.inc^2$	5.19e-07 ***	
c.inc.per.ca		2.76e-09 ***
$c.inc.per.ca^2$		0.013445 *
c.unemp	0.914189	0.831877
$c.unemp^2$	0.000313 ***	
c.inc:c.unemp	0.039120 *	
$c.inc{:}c.unemp^2$	0.054449 .	
$c.inc12{:}c.unemp^2$	0.019547 *	
$c.inc.per.ca^2{:}c.unemp$		0.000943 ***
s.dep.ave:c.inc	1.61e-05 ***	
s.dep.ave:c.inc.per.ca		9.89e-05 ***
Adjusted R^2	0.4094	0.4233

Table 3.5: **p-values of the final County Model 1**

Model all.C1a has an adjusted R-Squared of 0.4094 while model all.C1b achieves an adjusted R-Squared of 0.4233. Hence, on the basis of adjusted R-squared the latter model is the best County Model 1 because the best one before (c.C1b) has only an adjusted R-Squared of 0.3706. Again, these models have low explanatory power with a adjusted R-Squared of 0.4233 or lower.

Chapter 4

County Model 2

Alternatively, we investigated another model class based on an unstandardized response

$$c.log.dep := log\left(\sum_{county} b.dep\right)$$

which will not be divided by the population (as in County Model 1). Therefore, we may consider the population as an additional covariate.

4.1 State Variables

Figure 4.1 looks again very similar to the ones above so that there is no need to transform the State Variables.

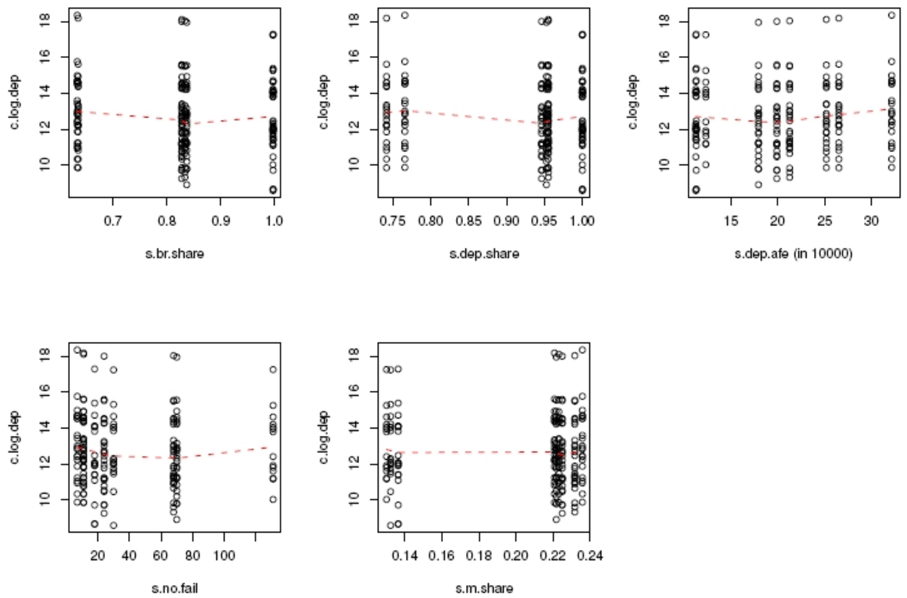

Figure 4.1: **Scatterplots of State Variables including lowess smoothing line (dashed red line)**

The backward selection found no significant State Variables, so none were added to the model and we continue by adding County Variables.

4.2 Adding County Variables

The scatterplots of *c.log.dep* against the variables *c.pop* (in 10000), *c.inc* (in million), *c.inc.per.ca* and *c.unemp* is shown in Figure 4.2 and 4.3 with 9 colored "LOWESS"-lines.

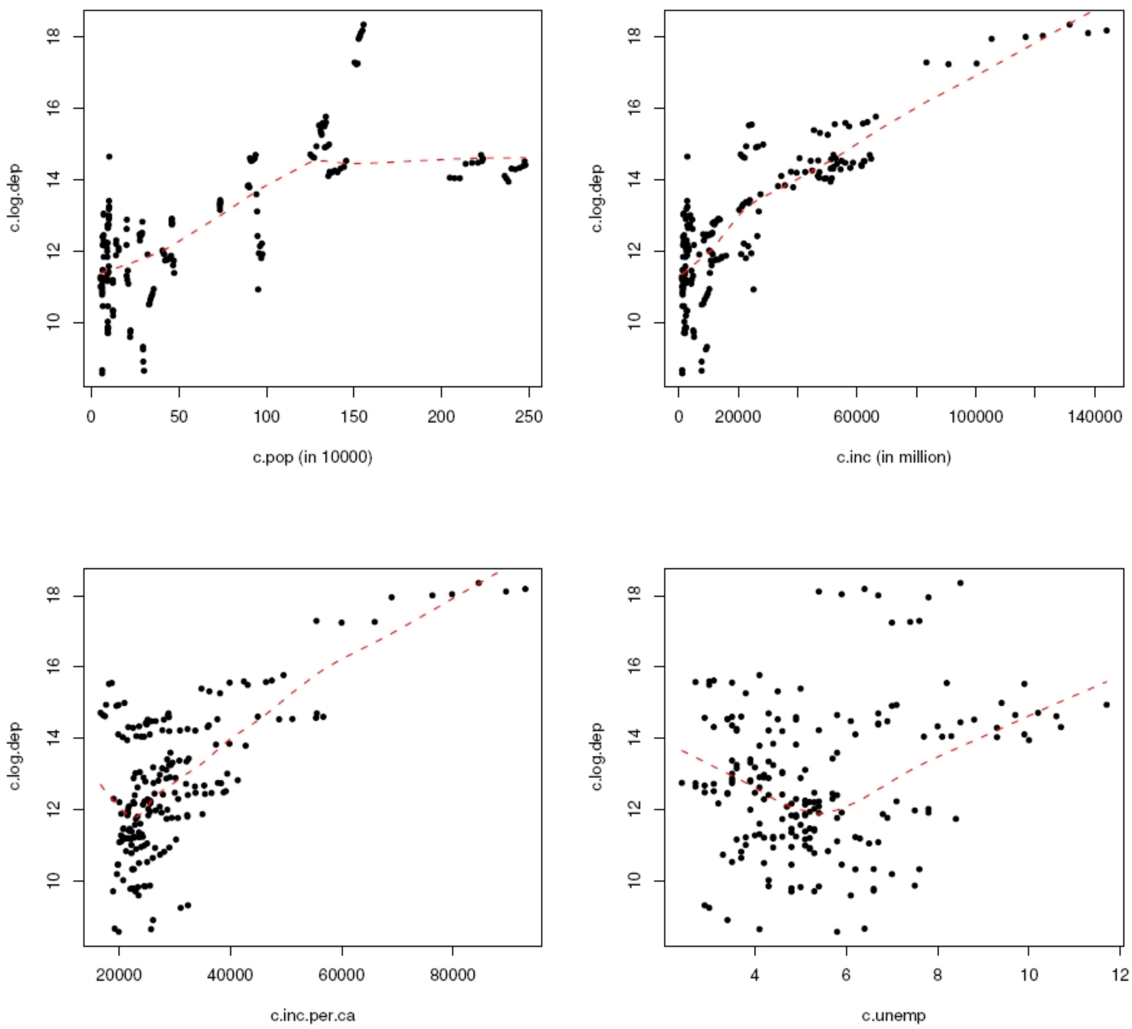

Figure 4.2: **Scatterplots of County Variables including lowess smoothing line (dashed red line)**

All of them exept *c.inc* (looks already linear) seem to be quadratic, consequently we transform them accordingly.

Then we apply the same procedure as before and create two separate versions. The first one (model c.C2a) with *c.inc* the second (model c.C2b) with *c.inc.per.ca* instead of *c.inc*.

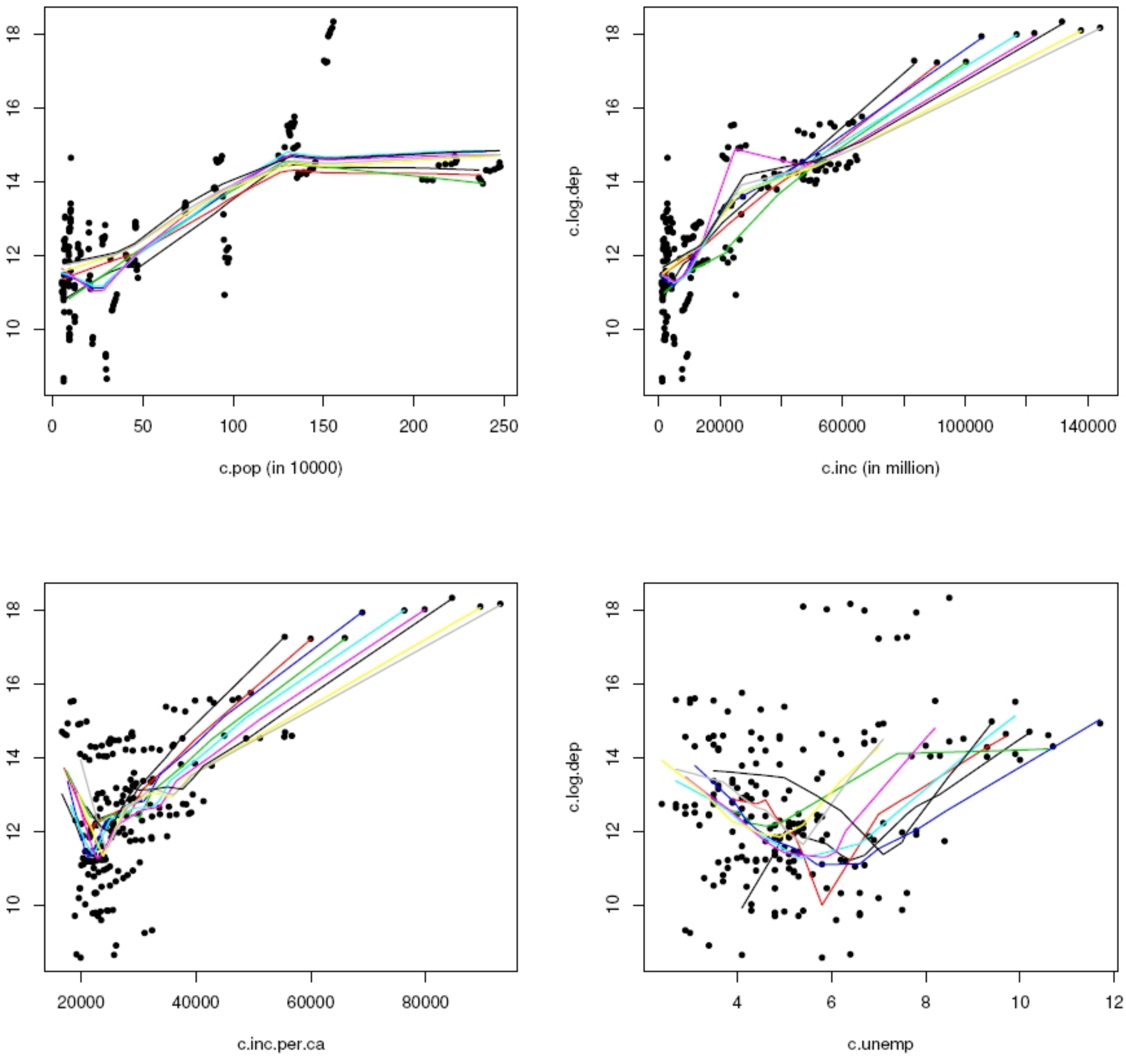

Figure 4.3: **Scatterplots of County Variables including lowess smoothing lines**

4.2.1 Model c.C2a (with *c.inc*)

In the first model the backward selection eliminates no variables which means that all of them are significant and end with an adjusted R-Squared of 0.7478 (see R-Code Appendix A.3.0.17)

4.2.2 Model c.C2b (with *c.inc.per.ca*)

The elimination procedure of the second model remove only the quadratic *c.inc.per.ca* term and ends with an adjusted R-Squared of 0.7564.

model	remove (p-value)
$c.log.dep = \beta_0 + \beta_1 poly(c.pop,2)[,1] + \beta_2 poly(c.pop,2)[,2]+$ $+\beta_3 poly(c.inc.per.ca,2)[,1] + \beta_4 poly(c.inc.per.ca,2)[,2]+$ $+\beta_5 poly(c.unemp,2)[,1] + \beta 6 poly(c.unemp,2)[,2]$	$poly(c.inc.per.ca,2)[,2]$ (0.8557)
$c.log.dep = \beta_0 + \beta_1 poly(c.pop,2)[,1] + \beta_2 poly(c.pop,2)[,2]+$ $+\beta_3 poly(c.inc.per.ca,2)[,1] + \beta_4 poly(c.unemp,2)[,1]+$ $+\beta_6 poly(c.unemp,2)[,2]$	

Table 4.1: **Backward selection for the County Model 2 with the variable c.inc.per.ca**

	model	
	c.C2a	c.C2b
c.pop	0.126788	< 2e-16 ***
c.pop^2	0.000326 ***	0.000728 ***
c.inc	< 2e-16 ***	
c.inc.per.ca		< 2e-16 ***
c.unemp	0.340369	0.025595 *
c.unemp^2	0.000551 ***	0.038429 *
Adjusted R^2	0.7478	0.7564

Table 4.2: **p-values of the County Variables in the County Model 1**

Model c.C2b to be the better fit (adjusted R-Squared of 0.7564 vs. 0.7478), so we continue with this one only. (see Table 4.2)

4.2.3 Interaction effects with time (c.C2.ct)

Figure 4.4 shows that every variable could have an interaction effect with the variable *year*, so we add *year* as well as the interaction between *year* and the County Variables.

However, since the backward selection removes all interactions between the variable year and the County Variables and since the variable *year* also is insignificant, we again arrived at the same model (c.C2b) as before with an adjusted R-Squared of 0.7564.

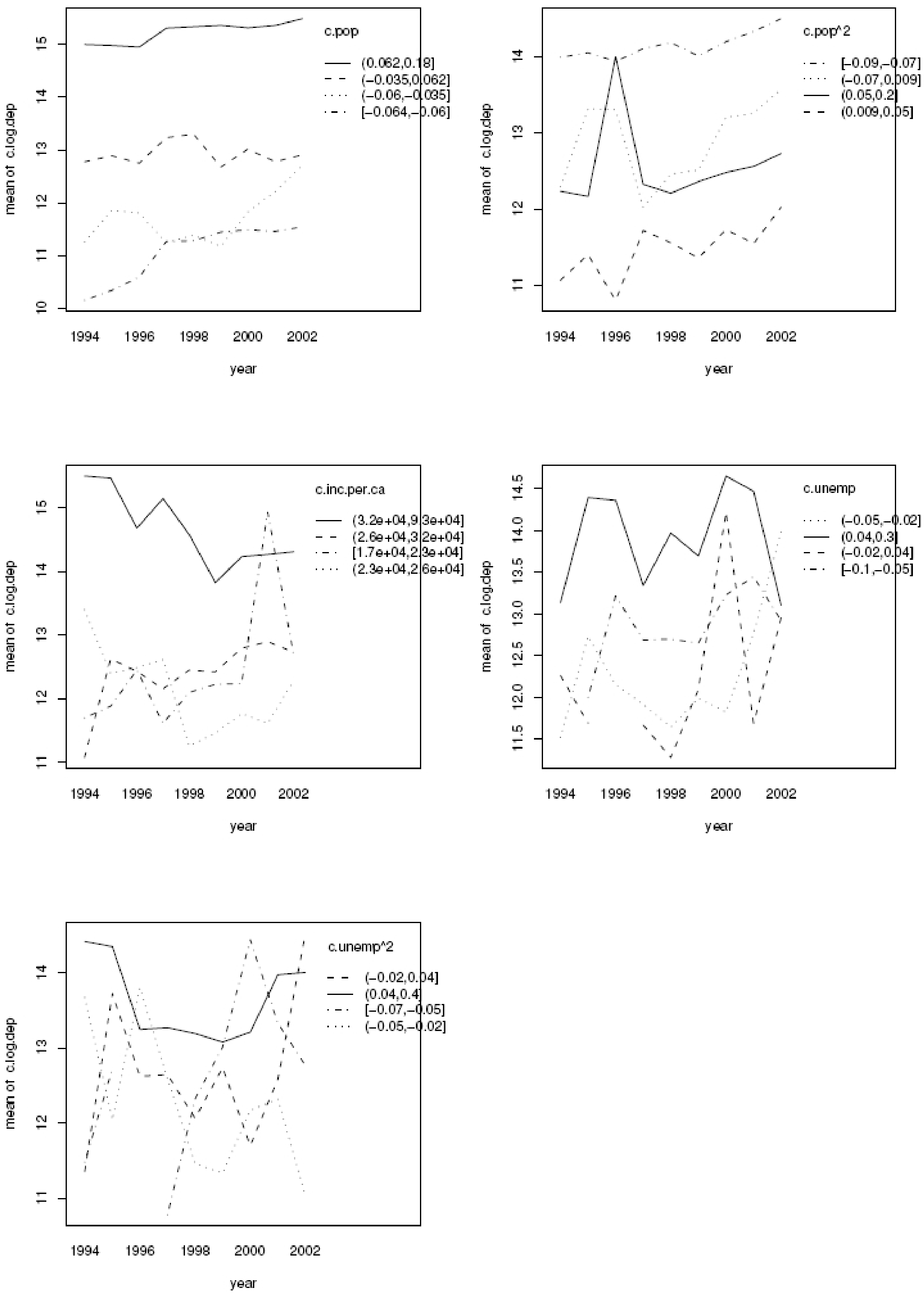

Figure 4.4: **Interactions between time and the County Variables**

4.3 Interaction effects among County Variables in the County Model 2 (c.C2.cc)

Furthermore the interaction among the County Variables could be added to the model (see Figure 4.5).

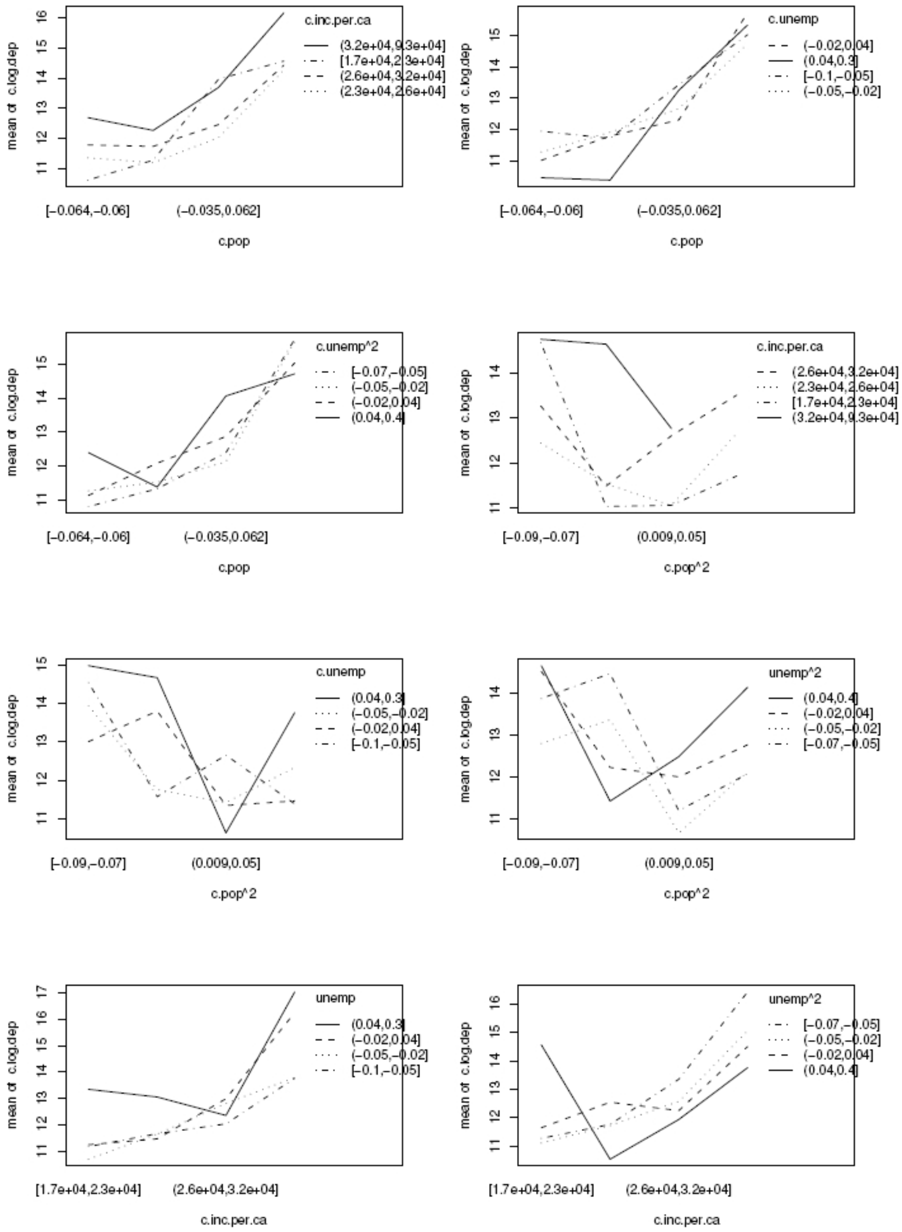

Figure 4.5: **Interactions among the County Variables in the County Model 2**

The final model c.C2b is improved, because the variables *c.pop* and its quadratic term $c.pop^2$ are with the interactions *c.inc* and *c.unemp* significant as well as $c.pop : c.unemp^2$ and $s.dep.ave : c.pop^2$. So we get the model c.C2.cc (see Table 4.3) which had now an R-Squared of 0.7898 (the R-Code is shown in the Appendix A.3.0.19).

4.4 Interaction effects between all variables

Model all.C2a with an adjusted R-Squared of 0.8007 is a bit better than all.C2b (with 0.7998) and the one before c.C2.cc (with 0.7898), meaning that model all.C2a with the interactions between all variables improves the model compare to the model c.C2.cc with only interactions at every stage (see Table 4.3).

	model		
	c.C2.cc	all.C2a	all.C2b
s.dep.ave	0.002135 **	0.07491 .	
c.pop	< 2e-16 ***	1.06e-08 ***	< 2e-16 ***
$c.pop^2$	0.768515	0.032162 *	0.55809
c.inc	0.000111 ***		
c.inc.per.ca	< 2e-16 ***		1.58e-09 ***
$c.inc.per.ca^2$			
c.unemp	0.011768 *	0.069014 .	0.00525 **
$c.unemp^2$	0.181607		0.10938
c.pop:c.inc	0.000817 ***		
c.pop:c.unemp	0.077541 .	0.000260 ***	0.02510 *
$c.inc.per.ca^2$:c.unemp			
s.dep.ave:c.inc.per.ca		0.00429 **	
$c.pop^2$:c.inc	2.27e-05 ***		
$c.pop^2$:c.unemp	4.44e-08 ***	1.10e-05 ***	5.50e-09 ***
c.pop:$c.unemp^2$	0.000554 ***		7.81e-05 ***
s.dep.ave:$c.pop^2$	0.001642 **		
Adjusted R^2	0.7898	0.8007	0.7998

Table 4.3: **p-values of the final County Model 2**

Chapter 5

Results and Interpretation

Table 5.1 shows the p-values of the best Branch Model, County Model 1 and County Model 2 and their adjusted R-Squared. Generally, one can see that the aggregated responses *c.log.dep.per.pop* in County Model 1 (with an adjusted R-squared of 0.4233) and in County Model 2 with the dependent variable *c.log.dep* (and an adjusted R-squared of 0.8007) are essentially better than the best Branch Model (0.2089), meaning that the explanatory power of the given variables for the variability of the data is best at the county level.

It is interesting to see that the interaction of $b.singleD : b.mmcD$ in all of the models are insignificant. The coefficient on population (*c.pop*) is positiv as well as the total income (*c.inc*) (respectively *c.sqrt.inc* and *c.inc.per.ca*) and highly significant, that means they have positive influence on the deposits, the higher the population and the total income in one county the higher the deposits in the branches of this county. But the coefficient on unemployment rate (*c.unemp*) is negative and in the County Models only the interactions are really significant, so it has a negative effect, the higher the unemployment rate the smaller the deposits (see also Appendix A.3.0.20).

Finally, we want to examine the best model all.C2a with regards to the interaction effects of the variables *c.pop* : *c.unemp* and *c.inc* : *c.pop* therefore see the perspective plots in Figure 5.1 and 5.2. Here, the Z-axis is now the response *c.dep* (without the logarithm). In both Figures you can see 9 plots referring to the respective nine years under examination. Clearly, the *year* has in both Figures an effect on the deposits because of the different shapes in the 9 plots.

In the first Figure the larger *c.pop* and *c.unemp* the higher the leverage of the interaction *c.pop* : *c.unemp* on the response *c.dep*. It makes intuitive sense that if there are many people unemployed there will be less money around. So this is an evidence of low deposits in the branches.

The second one shows that the influence of the interaction *c.inc* : *c.pop* on the *c.dep* is quite considerable if *c.inc* is large and at the same time *c.pop* is small. The interaction shows that high *c.inc* : *c.pop* effects high deposits. So a high income in a county partitioned on a few people obviously can redress the deposits in a branch.

Due to the hierarchical structure in the data (state, county and branch level) the next step should be to examine Mixed Models. As one of the advantages of Mixed Models is that they can provide improved estimates of the within-subject coeffcients (the random effects) by pooling information across subjects. Pooled estimates of the random effects provide so-called *best-linear-unbiased predictors* (or BLUPs).

	Branch Model	County Model 1	County Model 2
	all.Ba	all.C1b	all.C2a
s.br.share	0.37675		
s.dep.share	0.78138		
s.dep.ave	0.03181 *	0.083720 .	0.002135 **
s.m.share	0.27215		
c.pop	1.84e-12 ***		1.06e-08 ***
c.pop^2			0.032162 *
c.inc			0.000111 ***
c.sqrt.inc	2.64e-16 ***		
c.inc.per.ca		2.76e-09 ***	
c.inc.per.ca^2		0.013445 *	
c.unemp	$<$ 2e-16 ***	0.831877	0.069014 .
b.singleD	0.01047 *		
b.mmcD	0.03734 *		
c.pop:c.sqrt.inc	0.00280 **		
c.pop:c.inc			0.000817 ***
c.pop:c.unemp	4.13e-08 ***		0.000260 ***
c.pop:b.mmcD	7.60e-15 ***		
c.sqrt.inc:b.mmcD	$<$ 2e-16 ***		
c.unemp:b.mmcD	3.23e-14 ***		
s.br.share:b.singleD	1.30e-06 ***		
s.dep.share:b.singleD	7.62e-06 ***		
s.dep.ave:b.singleD	5.49e-05 ***		
s.m.share:b.singleD	0.01038 *		
s.br.share:b.mmcD	3.21e-05 ***		
s.dep.share:b.mmcD	0.00782 **		
s.dep.ave:b.mmcD	0.00777 **		
s.dep.ave:c.unemp	0.07978 .		
c.inc.per.ca^2:c.unemp		0.000943 ***	
s.dep.ave:c.inc.per.ca		9.89e-05 ***	
c.pop^2:c.inc			2.27e-05 ***
c.pop^2:c.unemp			1.10e-05 ***
s.dep.ave:c.pop^2			0.001642 **
Adjusted R^2	0.2089	0.4233	0.8007

Table 5.1: **p-values of the final Models**

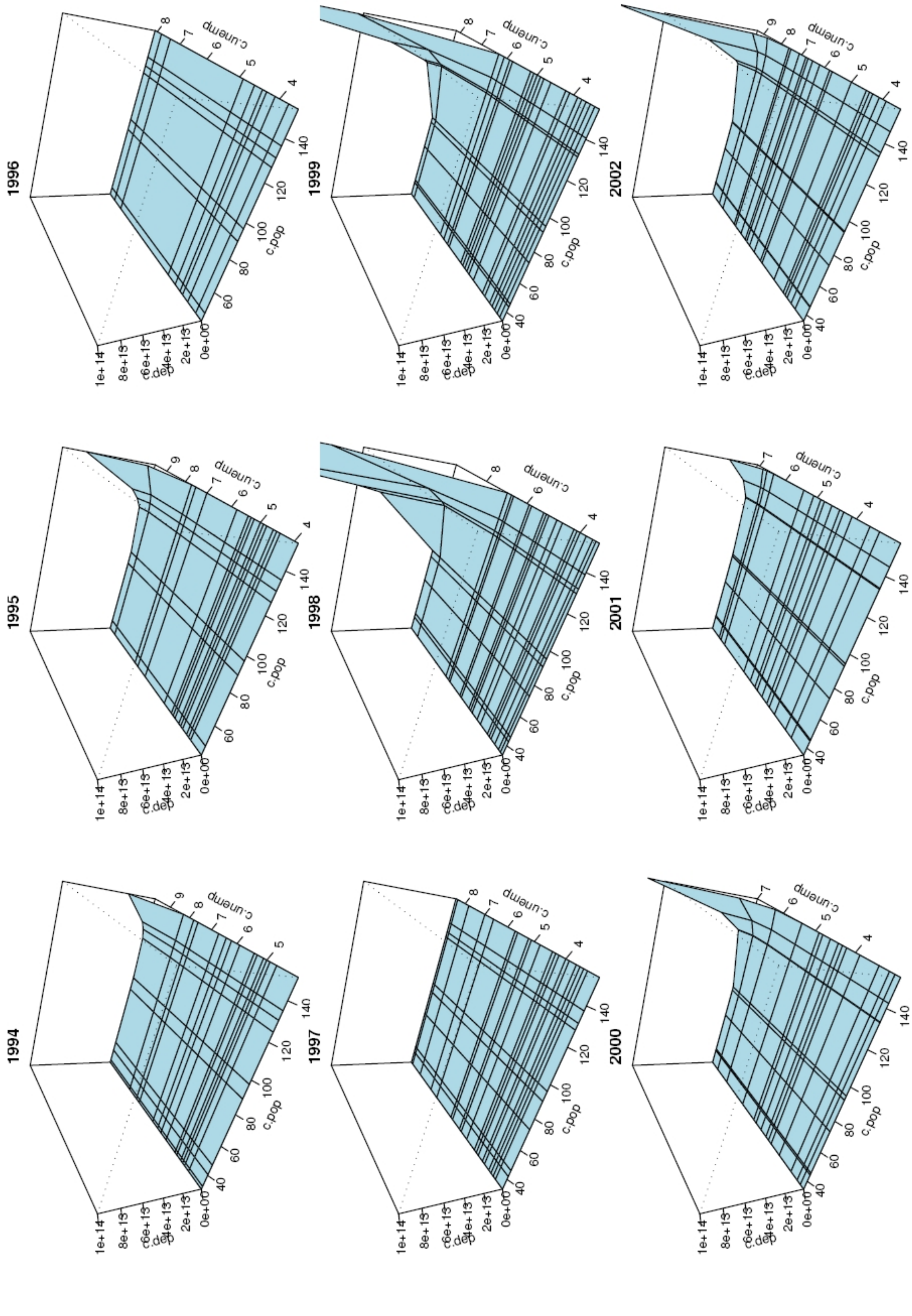

Figure 5.1: **3D interaction plot of c.pop × c.unemp (c.pop *(in 10000)*)**

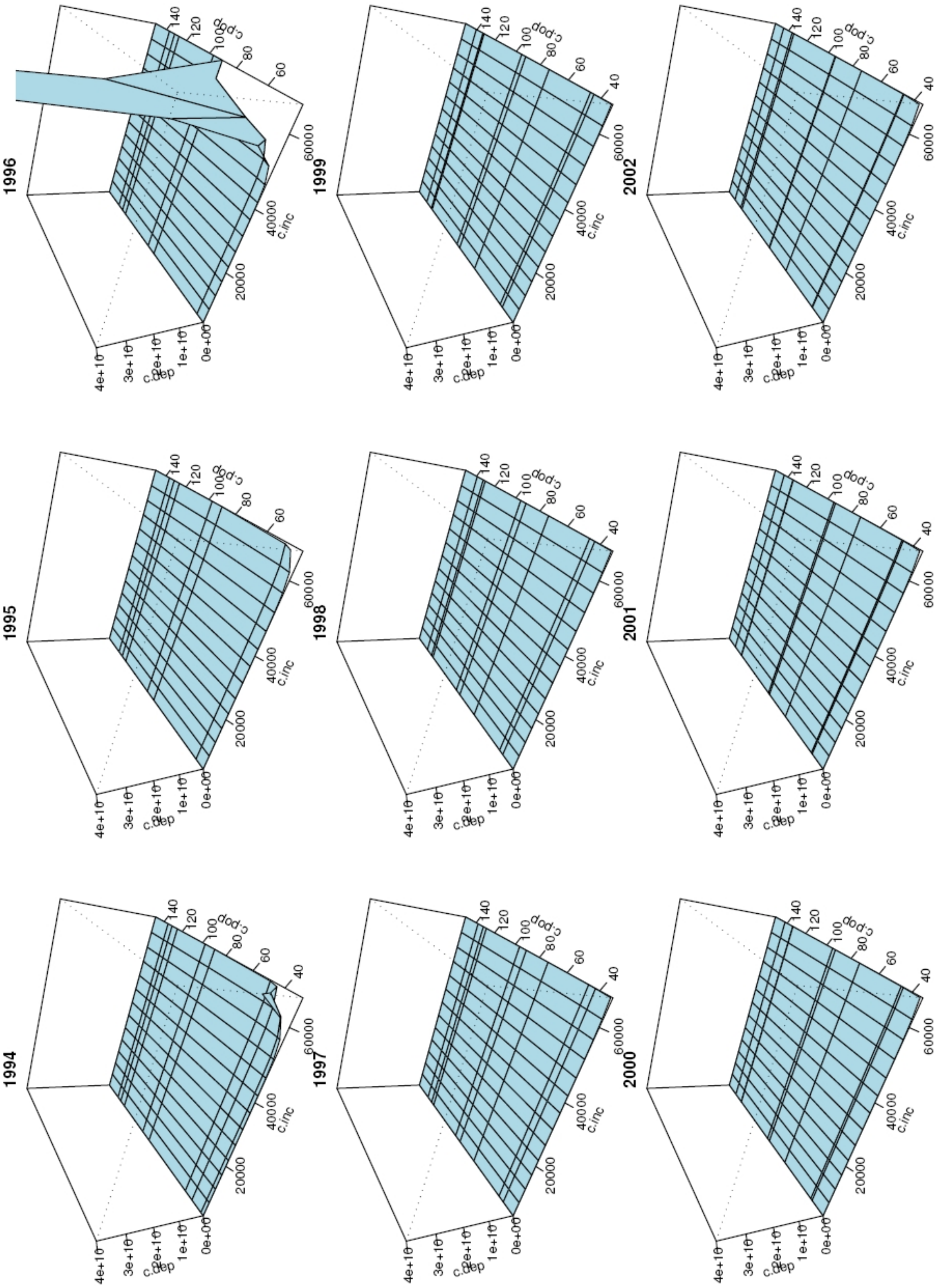

Figure 5.2: **3D interaction plot of c.inc × c.pop**

Appendix A

R program codes

A.1 Branch Model

A.1.0.1 Summary of model s.B

```
> fit0 <- lm(b.log.dep ~ s.dep.ave + s.m.share)
> summary(fit0)

Call:
lm(formula = b.log.dep ~ s.dep.ave + s.m.share)

Residuals:
     Min       1Q   Median       3Q      Max
-8.02488 -0.46533  0.01064  0.46576  6.30987

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.144e+01  9.285e-02 123.152  < 2e-16 ***
s.dep.ave    3.473e-06  4.751e-07   7.309 3.44e-13 ***
s.m.share   -3.991e+00  7.588e-01  -5.260 1.54e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9802 on 2985 degrees of freedom
Multiple R-Squared: 0.01818,    Adjusted R-squared: 0.01752
F-statistic: 27.63 on 2 and 2985 DF,  p-value: 1.290e-12
```

A.1.0.2 Summary of model c.Ba

```
> fit0 <- lm(b.log.dep ~ s.dep.ave + s.m.share + c.pop + c.sqrt.inc + c.unemp)
> summary(fit0)

Call:
lm(formula = b.log.dep ~ s.dep.ave + s.m.share + c.pop + c.sqrt.inc +
    c.unemp)

Residuals:
      Min        1Q    Median        3Q       Max
-7.905970 -0.414074 -0.008766  0.450766  5.888570

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.025e+01  1.141e-01  89.815  < 2e-16 ***
s.dep.ave    2.214e-06  4.454e-07   4.971 7.04e-07 ***
s.m.share   -2.350e+00  7.140e-01  -3.292  0.00101 **
c.pop        1.311e-07  4.692e-08   2.795  0.00523 **
c.sqrt.inc   3.376e-06  2.455e-07  13.751  < 2e-16 ***
```

```
c.unemp      1.938e-02  9.137e-03   2.121  0.03402 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9099 on 2982 degrees of freedom
Multiple R-Squared: 0.1548,     Adjusted R-squared: 0.1534
F-statistic: 109.2 on 5 and 2982 DF,  p-value: < 2.2e-16
```

A.1.0.3 Summary of model c.Bb

```
> fit0 <- lm(b.log.dep ~ s.dep.ave + s.m.share + c.pop + c.inc.per.ca)
> summary(fit0)

Call:
lm(formula = b.log.dep ~ s.dep.ave + s.m.share + c.pop + c.inc.per.ca)

Residuals:
      Min        1Q    Median        3Q       Max
-7.911170 -0.418458 -0.003897  0.450543  5.915909

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.036e+01  1.019e-01 101.671  < 2e-16 ***
s.dep.ave     2.075e-06  4.474e-07   4.637 3.68e-06 ***
s.m.share    -2.591e+00  7.095e-01  -3.652 0.000264 ***
c.pop         4.459e-07  3.308e-08  13.479  < 2e-16 ***
c.inc.per.ca  9.909e-06  7.361e-07  13.462  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9109 on 2983 degrees of freedom
Multiple R-Squared: 0.1526,     Adjusted R-squared: 0.1515
F-statistic: 134.3 on 4 and 2983 DF,  p-value: < 2.2e-16
```

A.1.0.4 Summary of model c.B.ct

```
> fit0 <- lm(b.log.dep ~ s.dep.ave + s.m.share + c.pop + c.sqrt.inc + c.unemp + c.pop:year)
> summary(fit0)

Call:
lm(formula = b.log.dep ~ s.dep.ave + s.m.share + c.pop + c.sqrt.inc +
    c.unemp + c.pop:year)

Residuals:
     Min       1Q   Median       3Q      Max
-7.90359 -0.41293 -0.00738  0.44086  5.86866

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.134e+01  2.939e-01  38.587  < 2e-16 ***
s.dep.ave       4.085e-06  1.159e-06   3.525 0.000430 ***
s.m.share      -9.366e+00  1.996e+00  -4.692 2.83e-06 ***
c.pop          -1.163e-07  8.517e-08  -1.366 0.172197
c.sqrt.inc      3.280e-06  2.467e-07  13.293  < 2e-16 ***
c.unemp         1.787e-02  9.942e-03   1.798 0.072345 .
c.pop:year1995  2.070e-08  5.185e-08   0.399 0.689710
c.pop:year1996 -3.646e-09  5.484e-08  -0.066 0.947002
c.pop:year1997  3.769e-07  9.433e-08   3.995 6.62e-05 ***
c.pop:year1998  3.560e-07  9.314e-08   3.822 0.000135 ***
c.pop:year1999  3.846e-07  9.759e-08   3.941 8.29e-05 ***
c.pop:year2000  2.649e-07  9.259e-08   2.861 0.004256 **
c.pop:year2001  2.210e-07  9.368e-08   2.359 0.018410 *
c.pop:year2002  2.585e-07  1.140e-07   2.267 0.023451 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.9083 on 2974 degrees of freedom
Multiple R-Squared:  0.16,      Adjusted R-squared: 0.1563
F-statistic: 43.57 on 13 and 2974 DF,  p-value: < 2.2e-16
```

A.1.0.5 Summary of model c.B.cc

```
> fit0 <- lm(b.log.dep ~ s.dep.ave + s.m.share + c.sqrt.inc + c.pop + c.unemp +
c.pop:year + c.pop:c.unemp)
> summary(fit0)

Call:
lm(formula = b.log.dep ~ s.dep.ave + s.m.share + c.sqrt.inc +
    c.pop + c.unemp + c.pop:year + c.pop:c.unemp)

Residuals:
     Min       1Q   Median       3Q      Max
-7.93448 -0.42126 -0.01005  0.43501  5.86723

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.058e+01  3.716e-01  28.457  < 2e-16 ***
s.dep.ave       4.093e-06  1.157e-06   3.538 0.000410 ***
s.m.share      -7.823e+00  2.045e+00  -3.825 0.000134 ***
c.sqrt.inc      2.949e-06  2.653e-07  11.116  < 2e-16 ***
c.pop           3.868e-07  1.723e-07   2.245 0.024858 *
c.unemp         1.037e-01  2.742e-02   3.781 0.000159 ***
c.pop:year1995  2.003e-08  5.176e-08   0.387 0.698810
c.pop:year1996  5.456e-09  5.481e-08   0.100 0.920725
c.pop:year1997  3.041e-07  9.663e-08   3.147 0.001665 **
c.pop:year1998  2.664e-07  9.673e-08   2.754 0.005921 **
c.pop:year1999  2.783e-07  1.024e-07   2.716 0.006642 **
c.pop:year2000  1.583e-07  9.773e-08   1.620 0.105377
c.pop:year2001  1.208e-07  9.817e-08   1.231 0.218458
c.pop:year2002  1.658e-07  1.171e-07   1.416 0.156949
c.pop:c.unemp  -6.598e-08  1.966e-08  -3.357 0.000799 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9067 on 2973 degrees of freedom
Multiple R-Squared: 0.1631,     Adjusted R-squared: 0.1592
F-statistic:  41.4 on 14 and 2973 DF,  p-value: < 2.2e-16
```

A.1.0.6 Summary of model b.B

```
> fit0 <- lm(b.log.dep ~ s.dep.ave + s.m.share + c.sqrt.inc + c.pop + c.unemp +
c.pop:year + c.pop:c.unemp + b.singleD)
> summary(fit0)

Call:
lm(formula = b.log.dep ~ s.dep.ave + s.m.share + c.sqrt.inc +
    c.pop + c.unemp + c.pop:year + c.pop:c.unemp + b.singleD)

Residuals:
     Min       1Q   Median       3Q      Max
-7.93402 -0.41611 -0.01795  0.38604  5.86119

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.040e+01  3.668e-01  28.342  < 2e-16 ***
s.dep.ave       4.884e-06  1.144e-06   4.270 2.02e-05 ***
s.m.share      -4.912e+00  2.040e+00  -2.408  0.01611 *
c.sqrt.inc      2.239e-06  2.724e-07   8.220 3.00e-16 ***
c.pop           4.675e-07  1.701e-07   2.748  0.00602 **
c.unemp         1.072e-01  2.703e-02   3.965 7.52e-05 ***
b.singleD      -1.166e-03  1.247e-04  -9.347  < 2e-16 ***
c.pop:year1995  6.152e-08  5.122e-08   1.201  0.22981
```

```
c.pop:year1996  5.607e-08  5.430e-08   1.033  0.30188
c.pop:year1997  1.765e-07  9.623e-08   1.834  0.06675 .
c.pop:year1998  1.262e-07  9.653e-08   1.308  0.19110
c.pop:year1999  1.262e-07  1.023e-07   1.234  0.21722
c.pop:year2000  1.051e-08  9.763e-08   0.108  0.91424
c.pop:year2001  2.654e-09  9.759e-08   0.027  0.97831
c.pop:year2002  1.405e-08  1.166e-07   0.121  0.90407
c.pop:c.unemp  -7.656e-08  1.941e-08  -3.945 8.18e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8938 on 2972 degrees of freedom
Multiple R-Squared: 0.187,      Adjusted R-squared: 0.1829
F-statistic: 45.59 on 15 and 2972 DF,  p-value: < 2.2e-16
```

A.1.0.7 Summary of model b.B.bt

```
> ## Backward Selection (ANOVA)
> fit0 <- lm(b.log.dep ~ s.dep.ave + s.m.share + c.sqrt.inc + c.pop + c.unemp +
c.pop:year + c.pop:c.unemp + b.singleD + b.singleD:year)
> fit1 <- update(fit0, . ~ . - b.singleD:year)
>
> anova(fit0, fit1, test="F")
Analysis of Variance Table

Model 1: b.log.dep ~ s.dep.ave + s.m.share + c.sqrt.inc + c.pop + c.unemp +
     c.pop:year + c.pop:c.unemp + b.singleD + b.singleD:year
Model 2: b.log.dep ~ s.dep.ave + s.m.share + c.sqrt.inc + c.pop + c.unemp +
    b.singleD + c.pop:year + c.pop:c.unemp
  Res.Df    RSS   Df Sum of Sq     F Pr(>F)
1   2964 2364.0
2   2972 2374.5   -8     -10.5 1.645  0.107
```

A.1.0.8 Summary of model all.Ba

```
> summary(fit0)

Call:
lm(formula = b.log.dep ~ s.br.share + s.dep.share + s.dep.ave +
    s.m.share + c.pop + c.sqrt.inc + c.unemp + b.singleD + b.mmcD +
    c.pop:c.sqrt.inc + c.pop:c.unemp + c.pop:b.mmcD + c.sqrt.inc:b.mmcD +
    c.unemp:b.mmcD + s.br.share:b.singleD + s.dep.share:b.singleD +
    s.dep.ave:b.singleD + s.m.share:b.singleD + s.br.share:b.mmcD +
    s.dep.share:b.mmcD + s.dep.ave:b.mmcD + s.dep.ave:c.unemp)

Residuals:
      Min        1Q    Median        3Q       Max
-7.858314 -0.405982 -0.005695  0.399555  5.761910

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)          -1.039e+00  6.179e+00  -0.168  0.86645
s.br.share            1.005e+01  1.137e+01   0.884  0.37675
s.dep.share          -3.079e+00  1.109e+01  -0.278  0.78138
s.dep.ave             1.646e-05  7.662e-06   2.148  0.03181 *
s.m.share            -1.201e+01  1.093e+01  -1.098  0.27215
c.pop                -4.787e-06  6.765e-07  -7.076 1.84e-12 ***
c.sqrt.inc            3.961e-05  4.809e-06   8.236 2.64e-16 ***
c.unemp               7.814e-01  9.327e-02   8.378  < 2e-16 ***
b.singleD            -5.929e-02  2.314e-02  -2.562  0.01047 *
b.mmcD                2.996e-03  1.438e-03   2.083  0.03734 *
c.pop:c.sqrt.inc      1.805e-12  6.034e-13   2.991  0.00280 **
c.pop:c.unemp        -1.122e-07  2.040e-08  -5.499 4.13e-08 ***
c.pop:b.mmcD          7.142e-10  9.139e-11   7.815 7.60e-15 ***
c.sqrt.inc:b.mmcD    -5.457e-09  6.471e-10  -8.433  < 2e-16 ***
c.unemp:b.mmcD       -8.016e-05  1.051e-05  -7.626 3.23e-14 ***
```

```
s.br.share:b.singleD   2.093e-01  4.314e-02   4.850 1.30e-06 ***
s.dep.share:b.singleD -1.730e-01  3.857e-02  -4.484 7.62e-06 ***
s.dep.ave:b.singleD    1.147e-07  2.840e-08   4.040 5.49e-05 ***
s.m.share:b.singleD    8.775e-02  3.422e-02   2.564  0.01038 *
s.br.share:b.mmcD     -6.028e-03  1.448e-03  -4.164 3.21e-05 ***
s.dep.share:b.mmcD     4.065e-03  1.527e-03   2.662  0.00782 **
s.dep.ave:b.mmcD      -4.686e-09  1.759e-09  -2.664  0.00777 **
s.dep.ave:c.unemp     -2.391e-07  1.364e-07  -1.753  0.07978 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8795 on 2965 degrees of freedom
Multiple R-Squared: 0.2148,     Adjusted R-squared: 0.2089
F-statistic: 36.86 on 22 and 2965 DF,  p-value: < 2.2e-16
```

A.1.0.9 Summary of model all.Bb

```
> summary(fit0)

Call:
lm(formula = b.log.dep ~ s.br.share + s.dep.share + s.dep.ave +
    s.m.share + c.pop + c.inc.per.ca + c.unemp + b.singleD +
    b.mmcD + c.pop:c.inc.per.ca + c.pop:c.unemp + c.pop:b.singleD +
    c.pop:b.mmcD + c.inc.per.ca:b.mmcD + c.unemp:b.mmcD + s.br.share:b.singleD +
    s.dep.share:b.singleD + s.dep.ave:b.singleD + s.m.share:b.singleD +
    s.br.share:b.mmcD + s.dep.share:b.mmcD + s.dep.ave:b.mmcD +
    s.m.share:b.mmcD + s.dep.ave:c.pop + s.dep.ave:c.unemp)

Residuals:
     Min       1Q   Median       3Q      Max
-7.91596 -0.40818 -0.01131  0.39045  5.79449

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)           -3.398e+01  1.435e+01  -2.369 0.017899 *
s.br.share             1.539e+02  5.553e+01   2.771 0.005632 **
s.dep.share           -1.372e+02  5.265e+01  -2.605 0.009223 **
s.dep.ave              2.776e-05  8.455e-06   3.284 0.001037 **
s.m.share              1.664e+02  7.152e+01   2.327 0.020027 *
c.pop                 -1.421e-06  5.710e-07  -2.489 0.012856 *
c.inc.per.ca           4.809e-05  1.520e-05   3.163 0.001577 **
c.unemp                5.741e-01  9.056e-02   6.340 2.65e-10 ***
b.singleD             -1.901e-01  5.259e-02  -3.615 0.000305 ***
b.mmcD                 1.128e-02  3.398e-03   3.320 0.000912 ***
c.pop:c.inc.per.ca     1.390e-11  6.183e-12   2.249 0.024600 *
c.pop:c.unemp         -5.656e-08  2.270e-08  -2.492 0.012764 *
c.pop:b.singleD       -4.124e-09  1.169e-09  -3.528 0.000425 ***
c.pop:b.mmcD           2.699e-10  7.084e-11   3.811 0.000141 ***
c.inc.per.ca:b.mmcD   -8.596e-09  1.685e-09  -5.103 3.56e-07 ***
c.unemp:b.mmcD        -6.206e-05  1.027e-05  -6.043 1.70e-09 ***
s.br.share:b.singleD   7.318e-01  1.979e-01   3.698 0.000222 ***
s.dep.share:b.singleD -6.542e-01  1.863e-01  -3.512 0.000451 ***
s.dep.ave:b.singleD    1.675e-07  3.238e-08   5.175 2.43e-07 ***
s.m.share:b.singleD    7.275e-01  2.553e-01   2.849 0.004412 **
s.br.share:b.mmcD     -3.984e-02  1.292e-02  -3.083 0.002071 **
s.dep.share:b.mmcD     3.532e-02  1.224e-02   2.886 0.003930 **
s.dep.ave:b.mmcD      -7.600e-09  1.960e-09  -3.878 0.000108 ***
s.m.share:b.mmcD      -4.193e-02  1.671e-02  -2.509 0.012176 *
s.dep.ave:c.pop        2.360e-12  1.105e-12   2.137 0.032713 *
s.dep.ave:c.unemp     -3.129e-07  1.568e-07  -1.996 0.046054 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8836 on 2962 degrees of freedom
Multiple R-Squared: 0.2082,     Adjusted R-squared: 0.2015
F-statistic: 31.15 on 25 and 2962 DF,  p-value: < 2.2e-16
```

A.2 County Model 1

A.2.0.10 Summary of model s.C1

```
> lm.5 <- lm(c.log.dep.per.pop ~ s.br.share)
> summary(lm.5)

Call:
lm(formula = c.log.dep.per.pop ~ s.br.share)

Residuals:
     Min       1Q   Median       3Q      Max
-3.64703 -0.63106 -0.07575  0.75328  3.89770

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.5801     0.6653   2.375   0.0185 *
s.br.share   -1.9398     0.7931  -2.446   0.0153 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.397 on 195 degrees of freedom
Multiple R-Squared: 0.02976,    Adjusted R-squared: 0.02479
F-statistic: 5.982 on 1 and 195 DF,  p-value: 0.01534
\scriptsize
\begin{verbatim}
```

A.2.0.11 Summary of model c.C1a

```
> summary(fit0)

Call:
lm(formula = c.log.dep.per.pop ~ s.br.share + poly(c.inc, 2)[,
    1] + poly(c.inc, 2)[, 2] + poly(c.unemp, 2)[, 1] + poly(c.unemp,
    2)[, 2])

Residuals:
     Min       1Q   Median       3Q      Max
-3.62326 -0.65481  0.04734  0.73609  3.38255

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)             1.1293     0.5474   2.063   0.0404 *
s.br.share             -1.3963     0.6528  -2.139   0.0337 *
poly(c.inc, 2)[, 1]    10.1185     1.1763   8.602 2.84e-15 ***
poly(c.inc, 2)[, 2]     5.8809     1.1888   4.947 1.65e-06 ***
poly(c.unemp, 2)[, 1]  -1.1191     1.1821  -0.947   0.3450
poly(c.unemp, 2)[, 2]   3.0059     1.1850   2.537   0.0120 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.134 on 191 degrees of freedom
Multiple R-Squared: 0.3742,     Adjusted R-squared: 0.3578
F-statistic: 22.84 on 5 and 191 DF,  p-value: < 2.2e-16
```

A.2.0.12 Summary of model c.C1b

```
> summary(fit0)

Call:
lm(formula = c.log.dep.per.pop ~ poly(c.inc.per.ca, 2)[, 1] +
    poly(c.unemp, 2)[, 1])

Residuals:
```

```
     Min       1Q   Median       3Q      Max 
-3.52595 -0.65161  0.01292  0.79182  3.16235 

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)                -0.02872    0.07996  -0.359   0.7199    
poly(c.inc.per.ca, 2)[, 1] 12.29092    1.13705  10.810   <2e-16 ***
poly(c.unemp, 2)[, 1]       2.82390    1.13705   2.484   0.0139 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 1.122 on 194 degrees of freedom
Multiple R-Squared: 0.3771,     Adjusted R-squared: 0.3706 
F-statistic: 58.71 on 2 and 194 DF,  p-value: < 2.2e-16 
```

A.2.0.13 Summary of model c.C1.ct

```
> summary(fit0)

Call:
lm(formula = c.log.dep.per.pop ~ c.inc.per.ca + c.unemp)

Residuals:
     Min       1Q   Median       3Q      Max 
-3.52595 -0.65161  0.01292  0.79182  3.16235 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -2.662e+00  3.352e-01  -7.942 1.57e-13 ***
c.inc.per.ca  6.890e-05  6.374e-06  10.810  < 2e-16 ***
c.unemp       1.107e-01  4.457e-02   2.484   0.0139 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 1.122 on 194 degrees of freedom
Multiple R-Squared: 0.3771,     Adjusted R-squared: 0.3706 
F-statistic: 58.71 on 2 and 194 DF,  p-value: < 2.2e-16 
```

A.2.0.14 Summary of model c.C1.cc

```
> ## Backward Selection (ANOVA)
> fit0 <- lm(c.log.dep.per.pop ~ c.inc.per.ca + c.unemp + c.unemp:c.inc)
> fit1 <- update(fit0, . ~ . - c.unemp:c.inc)
> 
> anova(fit0, fit1, test="F")
Analysis of Variance Table

Model 1: c.log.dep.per.pop ~ c.inc.per.ca + c.unemp + c.unemp:c.inc
Model 2: c.log.dep.per.pop ~ c.inc.per.ca + c.unemp
  Res.Df     RSS  Df Sum of Sq      F Pr(>F)
1    193 244.279                          
2    194 244.380  -1    -0.101 0.0797  0.778
> 
> # c.unemp:c.inc removed
> fit0 <- lm(c.log.dep.per.pop ~ c.inc.per.ca + c.unemp)
> summary(fit0)

Call:
lm(formula = c.log.dep.per.pop ~ c.inc.per.ca + c.unemp)

Residuals:
     Min       1Q   Median       3Q      Max 
-3.52595 -0.65161  0.01292  0.79182  3.16235 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
```

```
(Intercept)  -2.662e+00  3.352e-01  -7.942 1.57e-13 ***
c.inc.per.ca  6.890e-05  6.374e-06  10.810  < 2e-16 ***
c.unemp       1.107e-01  4.457e-02   2.484   0.0139 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.122 on 194 degrees of freedom
Multiple R-Squared: 0.3771,     Adjusted R-squared: 0.3706
F-statistic: 58.71 on 2 and 194 DF,  p-value: < 2.2e-16
```

A.2.0.15 Summary of model all.C1a

```
> summary(fit0)

Call:
lm(formula = c.log.dep.per.pop ~ s.dep.ave + poly(c.inc, 2)[,
    1] + poly(c.inc, 2)[, 2] + poly(c.unemp, 2)[, 1] + poly(c.unemp,
    2)[, 2] + poly(c.inc, 2)[, 1]:poly(c.unemp, 2)[, 1] + poly(c.inc,
    2)[, 1]:poly(c.unemp, 2)[, 2] + poly(c.inc, 2)[, 2]:poly(c.unemp,
    2)[, 2] + s.dep.ave:poly(c.inc, 2)[, 1])

Residuals:
     Min       1Q   Median       3Q      Max
-3.45219 -0.53901  0.08295  0.67371  2.90118

Coefficients:
                                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                               -2.772e-01  2.859e-01  -0.970 0.333514
s.dep.ave                                  2.147e-06  1.232e-06   1.743 0.082932 .
poly(c.inc, 2)[, 1]                        3.333e+01  5.305e+00   6.283 2.28e-09 ***
poly(c.inc, 2)[, 2]                        1.252e+01  2.407e+00   5.200 5.19e-07 ***
poly(c.unemp, 2)[, 1]                      1.537e-01  1.425e+00   0.108 0.914189
poly(c.unemp, 2)[, 2]                      6.114e+00  1.665e+00   3.673 0.000313 ***
poly(c.inc, 2)[, 1]:poly(c.unemp, 2)[, 1] -5.468e+01  2.632e+01  -2.078 0.039120 *
poly(c.inc, 2)[, 1]:poly(c.unemp, 2)[, 2]  5.496e+01  2.840e+01   1.935 0.054449 .
poly(c.inc, 2)[, 2]:poly(c.unemp, 2)[, 2]  8.537e+01  3.625e+01   2.355 0.019547 *
s.dep.ave:poly(c.inc, 2)[, 1]             -8.586e-05  1.938e-05  -4.429 1.61e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.087 on 187 degrees of freedom
Multiple R-Squared: 0.4365,     Adjusted R-squared: 0.4094
F-statistic:  16.1 on 9 and 187 DF,  p-value: < 2.2e-16
```

A.2.0.16 Summary of model all.C1b

```
> summary(fit0)

Call:
lm(formula = c.log.dep.per.pop ~ s.dep.ave + poly(c.inc.per.ca,
    2)[, 1] + poly(c.inc.per.ca, 2)[, 2] + poly(c.unemp, 2)[,
    1] + poly(c.inc.per.ca, 2)[, 2]:poly(c.unemp, 2)[, 1] + s.dep.ave:poly(c.inc.per.ca,
    2)[, 1])

Residuals:
     Min       1Q   Median       3Q      Max
-3.46899 -0.48360  0.04477  0.65390  3.15467

Coefficients:
                                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                                      -5.420e-01  2.588e-01  -2.095 0.037536 *
s.dep.ave                                         2.089e-06  1.202e-06   1.739 0.083720 .
poly(c.inc.per.ca, 2)[, 1]                        2.936e+01  4.704e+00   6.242 2.76e-09 ***
poly(c.inc.per.ca, 2)[, 2]                        3.522e+00  1.412e+00   2.495 0.013445 *
poly(c.unemp, 2)[, 1]                             2.923e-01  1.375e+00   0.213 0.831877
poly(c.inc.per.ca, 2)[, 2]:poly(c.unemp, 2)[, 1]  6.504e+01  1.936e+01   3.360 0.000943 ***
```

```
s.dep.ave:poly(c.inc.per.ca, 2)[, 1]           -8.635e-05  2.171e-05  -3.978 9.89e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.074 on 190 degrees of freedom
Multiple R-Squared: 0.441,      Adjusted R-squared: 0.4233
F-statistic: 24.98 on 6 and 190 DF,  p-value: < 2.2e-16
```

A.3 County Model 2

A.3.0.17 Summary of model c.C2a

```
> summary(lm.1)

Call:
lm(formula = c.log.dep ~ poly(c.pop, 2)[, 1] + poly(c.pop, 2)[,
    2] + c.inc + poly(c.unemp, 2)[, 1] + poly(c.unemp, 2)[, 2])

Residuals:
     Min       1Q   Median       3Q      Max
-3.23076 -0.40516  0.07746  0.56909  3.48254

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            1.161e+01  1.303e-01  89.050  < 2e-16 ***
poly(c.pop, 2)[, 1]    3.116e+00  2.032e+00   1.534 0.126788
poly(c.pop, 2)[, 2]   -4.085e+00  1.116e+00  -3.660 0.000326 ***
c.inc                  4.957e-11  4.742e-12  10.452  < 2e-16 ***
poly(c.unemp, 2)[, 1]  1.140e+00  1.193e+00   0.956 0.340369
poly(c.unemp, 2)[, 2]  3.760e+00  1.070e+00   3.514 0.000551 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.008 on 191 degrees of freedom
Multiple R-Squared: 0.7542,     Adjusted R-squared: 0.7478
F-statistic: 117.2 on 5 and 191 DF,  p-value: < 2.2e-16
```

A.3.0.18 Summary of model c.C2b

```
> summary(lm.2)

Call:
lm(formula = c.log.dep ~ poly(c.pop, 2)[, 1] + poly(c.pop, 2)[,
    2] + poly(c.inc.per.ca, 2)[, 1] + poly(c.unemp, 2)[, 1] +
    poly(c.unemp, 2)[, 2])

Residuals:
    Min      1Q  Median      3Q     Max
-3.1499 -0.5320  0.1323  0.5162  3.1015

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                12.74372    0.07057 180.573  < 2e-16 ***
poly(c.pop, 2)[, 1]        14.85095    1.25909  11.795  < 2e-16 ***
poly(c.pop, 2)[, 2]        -3.78730    1.10270  -3.435 0.000728 ***
poly(c.inc.per.ca, 2)[, 1] 12.96548    1.18412  10.949  < 2e-16 ***
poly(c.unemp, 2)[, 1]       2.70650    1.20293   2.250 0.025595 *
poly(c.unemp, 2)[, 2]       2.13750    1.02534   2.085 0.038429 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9906 on 191 degrees of freedom
Multiple R-Squared: 0.7626,     Adjusted R-squared: 0.7564
F-statistic: 122.7 on 5 and 191 DF,  p-value: < 2.2e-16
```

A.3.0.19 Summary of model c.C2.cc

```
> summary(fit0)

Call:
lm(formula = c.log.dep ~ c.inc.per.ca + poly(c.pop, 2)[, 1] +
    poly(c.pop, 2)[, 2] + poly(c.unemp, 2)[, 1] + poly(c.unemp,
    2)[, 2] + poly(c.pop, 2)[, 1]:poly(c.unemp, 2)[, 1] + poly(c.pop,
    2)[, 1]:poly(c.unemp, 2)[, 2] + poly(c.pop, 2)[, 2]:poly(c.unemp,
    2)[, 1])

Residuals:
    Min      1Q  Median      3Q     Max
-3.1537 -0.3879  0.1001  0.5607  3.1197

Coefficients:
                                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                                1.052e+01  2.032e-01  51.786  < 2e-16 ***
c.inc.per.ca                               7.326e-05  6.261e-06  11.702  < 2e-16 ***
poly(c.pop, 2)[, 1]                        2.159e+01  1.786e+00  12.088  < 2e-16 ***
poly(c.pop, 2)[, 2]                       -4.339e-01  1.472e+00  -0.295 0.768515
poly(c.unemp, 2)[, 1]                     -5.381e+00  2.115e+00  -2.544 0.011768 *
poly(c.unemp, 2)[, 2]                     -2.142e+00  1.598e+00  -1.341 0.181607
poly(c.pop, 2)[, 1]:poly(c.unemp, 2)[, 1]  3.718e+01  2.095e+01   1.775 0.077541 .
poly(c.pop, 2)[, 1]:poly(c.unemp, 2)[, 2]  7.601e+01  2.163e+01   3.514 0.000554 ***
poly(c.pop, 2)[, 2]:poly(c.unemp, 2)[, 1] -1.043e+02  1.828e+01  -5.706 4.44e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9202 on 188 degrees of freedom
Multiple R-Squared: 0.7984,     Adjusted R-squared: 0.7898
F-statistic: 93.05 on 8 and 188 DF,  p-value: < 2.2e-16
```

A.3.0.20 Summary of model all.C2a

```
> summary(fit0)

Call:
lm(formula = c.log.dep ~ s.dep.ave + poly(c.pop, 2)[, 1] + poly(c.pop,
    2)[, 2] + c.inc + poly(c.unemp, 2)[, 1] + poly(c.pop, 2)[,
    1]:c.inc + poly(c.pop, 2)[, 1]:poly(c.unemp, 2)[, 1] + poly(c.pop,
    2)[, 2]:c.inc + poly(c.pop, 2)[, 2]:poly(c.unemp, 2)[, 1] +
    s.dep.ave:poly(c.pop, 2)[, 2])

Residuals:
     Min       1Q   Median       3Q      Max
-2.73872 -0.41772  0.04774  0.64213  2.92522

Coefficients:
                                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                                1.118e+01  2.515e-01  44.443  < 2e-16 ***
s.dep.ave                                  3.234e-06  1.038e-06   3.114 0.002135 **
poly(c.pop, 2)[, 1]                        3.199e+01  5.340e+00   5.990 1.06e-08 ***
poly(c.pop, 2)[, 2]                        1.298e+01  6.012e+00   2.159 0.032162 *
c.inc                                      9.373e-11  2.373e-11   3.950 0.000111 ***
poly(c.unemp, 2)[, 1]                     -2.285e+00  1.249e+00  -1.829 0.069014 .
poly(c.pop, 2)[, 1]:c.inc                 -9.025e-10  2.652e-10  -3.403 0.000817 ***
poly(c.pop, 2)[, 1]:poly(c.unemp, 2)[, 1]  6.203e+01  1.666e+01   3.724 0.000260 ***
poly(c.pop, 2)[, 2]:c.inc                 -3.242e-10  7.458e-11  -4.347 2.27e-05 ***
poly(c.pop, 2)[, 2]:poly(c.unemp, 2)[, 1] -7.171e+01  1.587e+01  -4.520 1.10e-05 ***
s.dep.ave:poly(c.pop, 2)[, 2]              4.800e-05  1.502e-05   3.195 0.001642 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.896 on 186 degrees of freedom
Multiple R-Squared: 0.8109,     Adjusted R-squared: 0.8007
F-statistic: 79.74 on 10 and 186 DF,  p-value: < 2.2e-16
```

A.3.0.21 Summary of model all.C2b

```
> summary(fit0)

Call:
lm(formula = c.log.dep ~ s.dep.ave + poly(c.pop, 2)[, 1] + poly(c.pop,
    2)[, 2] + poly(c.inc.per.ca, 2)[, 1] + poly(c.unemp, 2)[,
    1] + poly(c.unemp, 2)[, 2] + poly(c.pop, 2)[, 1]:poly(c.unemp,
    2)[, 1] + poly(c.pop, 2)[, 1]:poly(c.unemp, 2)[, 2] + poly(c.pop,
    2)[, 2]:poly(c.unemp, 2)[, 1] + s.dep.ave:poly(c.inc.per.ca,
    2)[, 1])

Residuals:
     Min       1Q   Median       3Q      Max
-2.92612 -0.36840  0.06388  0.57526  2.85501

Coefficients:
                                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                                1.232e+01  2.401e-01  51.323  < 2e-16 ***
s.dep.ave                                  1.878e-06  1.048e-06   1.791  0.07491 .
poly(c.pop, 2)[, 1]                        2.171e+01  1.744e+00  12.448  < 2e-16 ***
poly(c.pop, 2)[, 2]                       -8.673e-01  1.478e+00  -0.587  0.55809
poly(c.inc.per.ca, 2)[, 1]                 2.212e+01  3.482e+00   6.354 1.58e-09 ***
poly(c.unemp, 2)[, 1]                     -5.854e+00  2.072e+00  -2.825  0.00525 **
poly(c.unemp, 2)[, 2]                     -2.540e+00  1.579e+00  -1.609  0.10938
poly(c.pop, 2)[, 1]:poly(c.unemp, 2)[, 1]  4.747e+01  2.102e+01   2.258  0.02510 *
poly(c.pop, 2)[, 1]:poly(c.unemp, 2)[, 2]  8.614e+01  2.132e+01   4.040 7.81e-05 ***
poly(c.pop, 2)[, 2]:poly(c.unemp, 2)[, 1] -1.096e+02  1.791e+01  -6.117 5.50e-09 ***
s.dep.ave:poly(c.inc.per.ca, 2)[, 1]      -4.195e-05  1.451e-05  -2.891  0.00429 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8979 on 186 degrees of freedom
Multiple R-Squared: 0.8101,     Adjusted R-squared: 0.7998
F-statistic: 79.32 on 10 and 186 DF,  p-value: < 2.2e-16
```