Unpublished Supplement to

"Using Geographic Information Systems to Study Interstate Competition,"

***American Political Science Review* (November 2005), vol. 99, pp. 505-19.**

William D. Berry
Florida State University

Brady Baybeck
University of Missouri - St. Louis

Table of Contents

I. Measuring the Degree of Concern by State Officials Using GIS Software

How might we estimate the degree of concern about residents going to other states to play the lottery if we had a U.S. map showing state boundaries and the location of all adults in the United States? [Imagine a hypothetical map with hundreds of millions of points designating the location of each adult.] For specificity, consider the hypothetical state, S, depicted in Figure 1 (on p. 508 of the article). We could form a band internal to the border of state S of width *D* miles, and then exclude all areas of the band that are not within *D* miles of a state with a lottery in the previous year. This would yield the shaded region in the figure. The adults in this shaded region constitute what we have labeled the population of concern. For each adult in this region, we would measure his or her distance, d_i, from the nearest state with a lottery, and then transform d_i to a value for the propensity to play another state's lottery using the function $f(d_i)$. Finally, to obtain the degree of concern we would add up the propensity scores of all adults in the region and divide the sum by the adult population of state S.

The procedure for calculating degree of concern in our welfare application would be different. Here our starting map would need to show state boundaries, the location of all poor persons, and the center of each city with population greater than *p* (i.e., each city assumed "appealing" to the poor). Consider the hypothetical state, S, depicted in Figure 3 (on p. 511 of the article). A preliminary goal is to isolate the shaded region in the map (as the poor persons residing in this region constitute the population of concern). We could do this by drawing a circle of radius *M* around each designated city (with population greater than *p*) in the state, and taking all area inside the circle(s). From this area, we would exclude all sections that are within state S. We would also exclude all sections that are outside of state S but in a state with a welfare benefit higher, or less than ΔB dollars lower, than the benefit in state S. This would leave us with the shaded region in the figure. For each poor person in the shaded region, we would measure the distance, m_i, from the nearest designated city in state S, and then transform m_i to a propensity to migrate based on distance using the function $g(m_i)$. We would also compute the person's benefit difference, Δb_i, with state S, and transform Δb_i to a propensity to migrate based on benefit difference using the function $h(\Delta b_i)$. Then, for each poor person in the shaded region, we would multiply the propensity to migrate based on distance by the propensity to migrate based on benefit difference to compute the propensity to migrate for better welfare benefits. Next, if there were no poor persons in the shaded region (i.e., no persons in the population of concern), we would set degree of concern to zero. But if the population of concern were not empty, to obtain degree of concern, we would sum the propensity values of all poor persons in the shaded region and divide the total by the poor population of state S.

While such computational procedures on individual-level data are theoretically feasible using GIS software, they are impractical. First, the necessary individual-level residential data are not available. Second, even if these data were available, the computational requirements for the calculations would be enormous, and would likely overwhelm even the most powerful computer available to the typical political scientist. Thus, to keep the computations manageable, we make counties our units of analysis and assume that all individuals in a county are at its *geographic center*.[1] This then implies that scores for m_i and d_i for all individuals within a county are identical.

How do we measure degree of concern in the lottery model using counties as computational units? For specificity, take state S in Figure 1 of the manuscript. Recall that the shaded region denotes the location of the population of concern. We identify each county, c, having a geographic center inside the shaded region. For each such county, we measure the distance, d_c, from the center of the county to the nearest state with a lottery, and convert this distance to a score for propensity to play another state's lottery using the function f. Multiplying this *county* propensity score by the adult population of the county yields a value that is equal to the sum of the *individual* propensity scores across all adults living in the county. [In effect, because we assume that everyone in a county is at the same location, multiplying county adult population by the county propensity score is mathematically equivalent to summing up that same propensity score over all adults in the county.] Then we sum the propensity-population products

[1] GIS software typically identifies the geographic center of a county by forming a polygon around the boundary of the county, determining the coordinates of each of the polygon's vertices in two orthogonal dimensions (x and y), and then computing the average x- and y-coordinates across all vertices.

just calculated across all identified counties. This gives us the sum of *individual* propensity scores across all persons in the population of concern (assuming all persons in a county are at its center) – which is the numerator of degree of concern. The final step is to divide by the poor population of state S. Incorporating all the steps in this procedure into a single computational formula,

$$\text{Degree of concern}_S = \left\{ \sum_{\substack{\text{each county, } c, \\ \text{with a center in} \\ \text{the shaded} \\ \text{region of state S}}} [\text{adult population}_c \times f(d_c)] \right\} / \left(\text{adult population}_S\right)$$

Theoretically, GIS could be used to implement this entire procedure. However, because we have more experience with non-spatial statistical software, we chose to use GIS only for the first step in the procedure: measuring the distance of county centers from the nearest state with a lottery. These distance values are then used as input data for a program written in Stata to perform the remaining operations necessary to compute our measures of degree of concern.

We begin our GIS calculations for the lottery application with an electronic "map" of the continental United States showing boundaries for all states and the geographic center of each county.[2] Boundaries and point locations are projected on a two-dimensional map using an *equidistant conic projection,* which tends to preserve relatively accurate distance measures.[3] To illustrate our methodology, we consider one of our years of observation: 1985. Our task is to identify, for each state without a lottery in 1985, all counties within the state having a center located within *D* miles of the border of a state with a lottery in the previous year, 1984 (where *D* is the maximum distance that a person would travel to purchase a lottery ticket). Then for each of these counties, we must determine the distance of that

---

[2] Information about the location of state borders and the center of each county is from U.S. Bureau of the Census (2005) boundary files, which define locations by latitude and longitude coordinates. To make the boundaries consistent over time we make some minor adjustments to the boundary files (U.S. Centers for Disease Control and Prevention 1998).

[3] *Projection* refers to the process of flattening the earth for display on a two-dimensional map, and the projection one chooses is very important when calculating geographic measures like area or distance. Some projections are designed to look nice on a map, some are designed to preserve relative distances, and some are designed to preserve relative areas (Longley, *et al.* 2001, 89-94).

county's center from the nearest state with a lottery in 1984. For state S in Figure 1, this would involve measuring the distance between each county with a center in the shaded region and the nearest state with a lottery. For specificity, we describe the procedure for calculating these distances when D is set to 170 miles.

In GIS analysis, a *buffer* is a band of a user-specified width around a geographic observation of interest (e.g., a state).[4] We "draw" 18 concentric buffers – the first 5 miles wide, and each of the others 10 miles wide – external to each state in the U.S. *with a lottery in 1984*. The first buffer external to a state contains all locations less than 5 miles from the state; the second contains all points greater than 5 miles, but less than or equal to 15 miles, from the state; and so forth. The eighteenth buffer contains all locations greater than 165 miles, but less than or equal to 175 miles, from the state. We denote these buffers by enclosing two fundamental parameters – the state around which they are drawn, and the distance from the middle of the buffer to the state – in parentheses.[5] For example, the buffer around Missouri containing all points between 75 and 85 miles from Missouri is labeled (MO, 80). Figure S-2 shows the 18 buffers for Missouri – (MO, 0), (MO, 10), (MO, 20),…, (MO, 160) and (MO, 170) – and a blow-up of the areas in which these buffers extend into Missouri's neighbors to the northwest: Nebraska, Iowa, and small sections of Kansas and South Dakota. We now have a map showing the center of each county in states *without* a lottery in 1985, and overlaying these county centers are eighteen 5- or 10-mile wide buffers external to each state *with* a lottery in 1984.

We then join the county and buffer data by "attaching" to each county each buffer containing its center. Consider, for example, Stanton County, in northeast Nebraska – which is highlighted in both parts of Figure S-2. The geographic center of Stanton County – which is 116 miles from the closest point in Missouri – is denoted by the dot. Since we assume that all persons in Stanton county are at its center, all

[4] When the geographic observation is larger than a point, the buffers can be either external or internal. Most GIS programs allow the user to build buffers of various distances through a point-and-click interface or simple programming.

[5] The one exception is the first buffer external to a state, which is assigned a distance parameter of zero.

persons in Stanton County are assumed to be located in the buffer (MO, 120), and therefore the buffer (MO, 120) – but no other Missouri buffers – will get attached to Stanton County. Note that some counties will have buffers from multiple states attached (i.e., counties with a largest city that is less than 175 miles from more than one state with a lottery in 1984); other counties will have no buffers attached (i.e., counties the largest city of which is not within 175 miles of any state with a lottery in 1984). For each county with at least one buffer attached, we program the computer to choose the buffer with the smallest distance value (regardless of the state around which the buffer is drawn). The distance value of this buffer is the distance (rounded to the nearest multiple of 10 miles) between the geographic center of the county and the nearest state with a lottery in 1984.[6] This is the precisely the distance we seek to measure.

We now turn to the measurement of degree of concern in our welfare application. Consider state S in Figure 3 of the manuscript. We begin by identifying all counties having a center inside the shaded region. For each such county, we measure the distance, m_c, between the county's center and the center of the nearest city with population greater than p in state S.[7] Next, using the function g, we convert this distance to a *county* score for the propensity to migrate based on distance. We also compute a benefit difference score, Δb_c, for the county, and transform this value to a county score for the propensity to migrate based on benefit difference using the function h. Then, for each county, we take the product of these two propensity scores, which produces a county score for the propensity to migrate for better welfare benefits. We multiply this county propensity score by the poor population of the county, yielding the sum of individual propensity scores for all poor persons in the county (assuming all persons in a county are at its center). We sum the propensity-population products over all identified counties, giving us the numerator of degree of concern. Finally, we divide by the poor population of state S. Incorporating these steps into a single computational formula,

---

[6] We could measure distance with more precision by decreasing the width of the concentric buffers to less than 10 miles. For example, we could measure distance to the nearest mile if we drew 171 buffers around states, the first a half-mile wide and the rest one mile wide. The tradeoff would be a dramatic increase in the memory (and time) required to run the computer program.

[7] Information about the location of cities is from GeoLytics, Inc. (2001).

$$\text{Degree of concern}_S = \Big\{ \sum_{\substack{\text{each county, } c, \\ \text{with a center in} \\ \text{the shaded} \\ \text{region of state S}}} [\text{poor population}_c \times g(m_c) \times h(\Delta b_c)] \Big\} \,/\, \big(\text{adult population}_S\big)$$

Thus, with respect to the GIS procedures required for measuring degree of concern, the major difference between the lottery and welfare applications is that for the former we compute distances from a county center to a state with a lottery, while for the latter we calculate distances from a county center to the center of a city with population exceeding some threshold.[8] Consequently, for the welfare application, our electronic map must show not only state boundaries and the location of all county centers, but also the location of the center of each city in the U.S. with population exceeding p. Furthermore, our buffers are concentric "doughnuts" centered at a city's geographic center, rather than bands external to the border of a state. To be precise, the first buffer consists of all points inside a circle of radius 5 miles centered at the city's center, the second contains all locations between circles of radii 10 and 15 miles, the third contains all points between circles of radii 15 and 25 miles, and so forth. Such buffers allow us to calculate the distance between a county's center and the center of a city accurate to the nearest multiple of 10 miles.

[8] Actually, this is the procedure only for models setting $p > 0$. When $p = 0$, since the intent is to assume that any location that is close enough and in which the benefit level is large enough is appealing, we determine the population of concern of a state by drawing a band of width M external to the border of the state rather than circles of radius M around every community in the state (i.e., all cities with population greater than zero).

REFERENCES

GeoLytics, Inc. 2001. *CensusCD 1980*. East Brunswick, NJ: GeoLytics, Inc.

Longley, Paul A., Michael F. Goodchild, David J. Maguire, and David W. Rhind. 2001. *Geographic Information Systems and Science.* Hoboken, NJ: John Wiley & Sons.

United States Bureau of the Census. 2005. "Cartographic Boundary Files." http://www.census.gov/population/www/censusdata/cencounts.html (July 18, 2001; March 29, 2005).

United States Centers for Disease Control and Prevention. 1998. "Scientific Data Documentation, Compressed Mortality, 1968-1995." http://wonder.cdc.gov/wonder/sci_data/mort/cmf/type_txt/cmf6895.asp (February 1998).

FIGURE S-1. Specific Functions Mapping an Individual's Benefit Difference (Δb) into the Propensity to Migrate Based on Benefit Difference

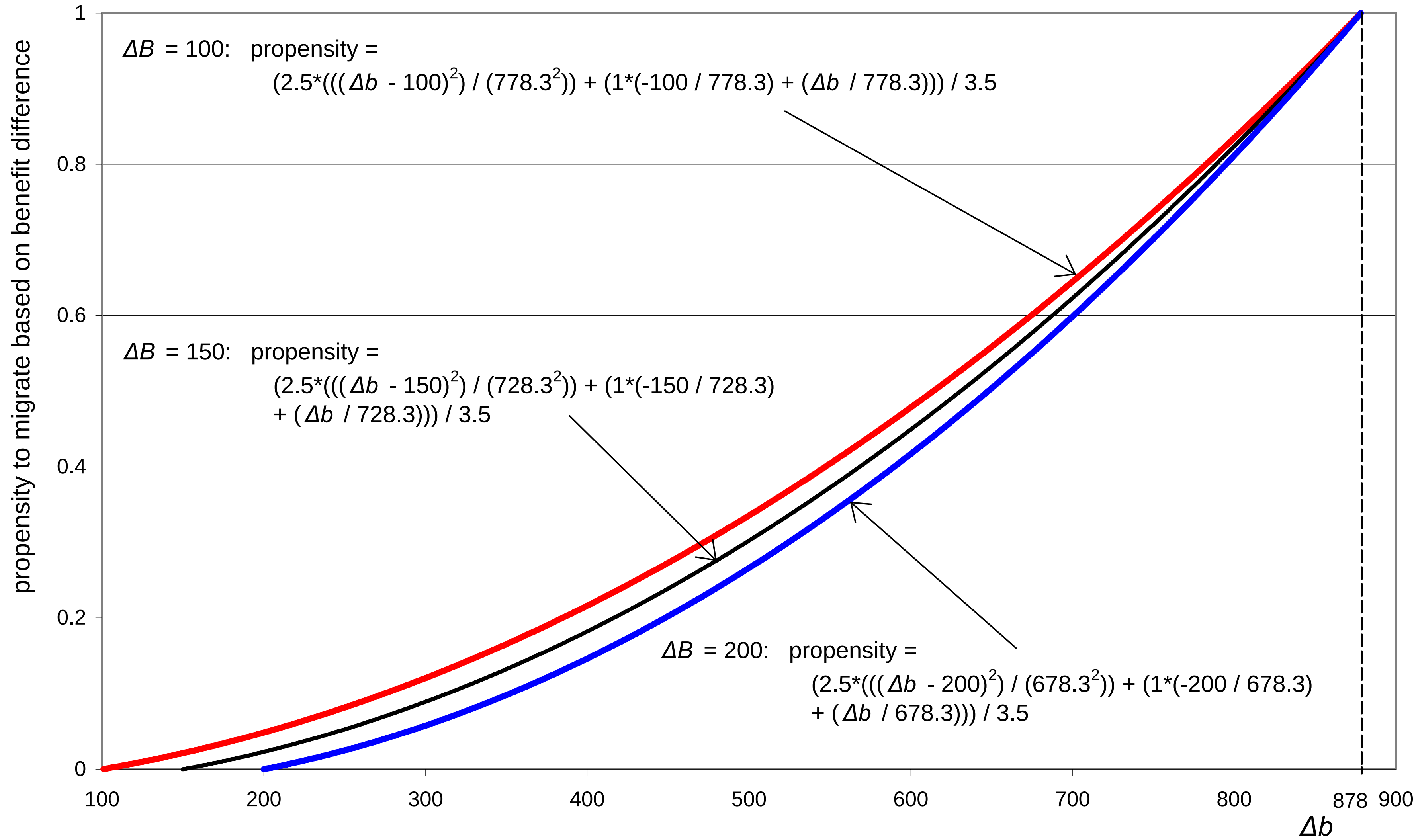
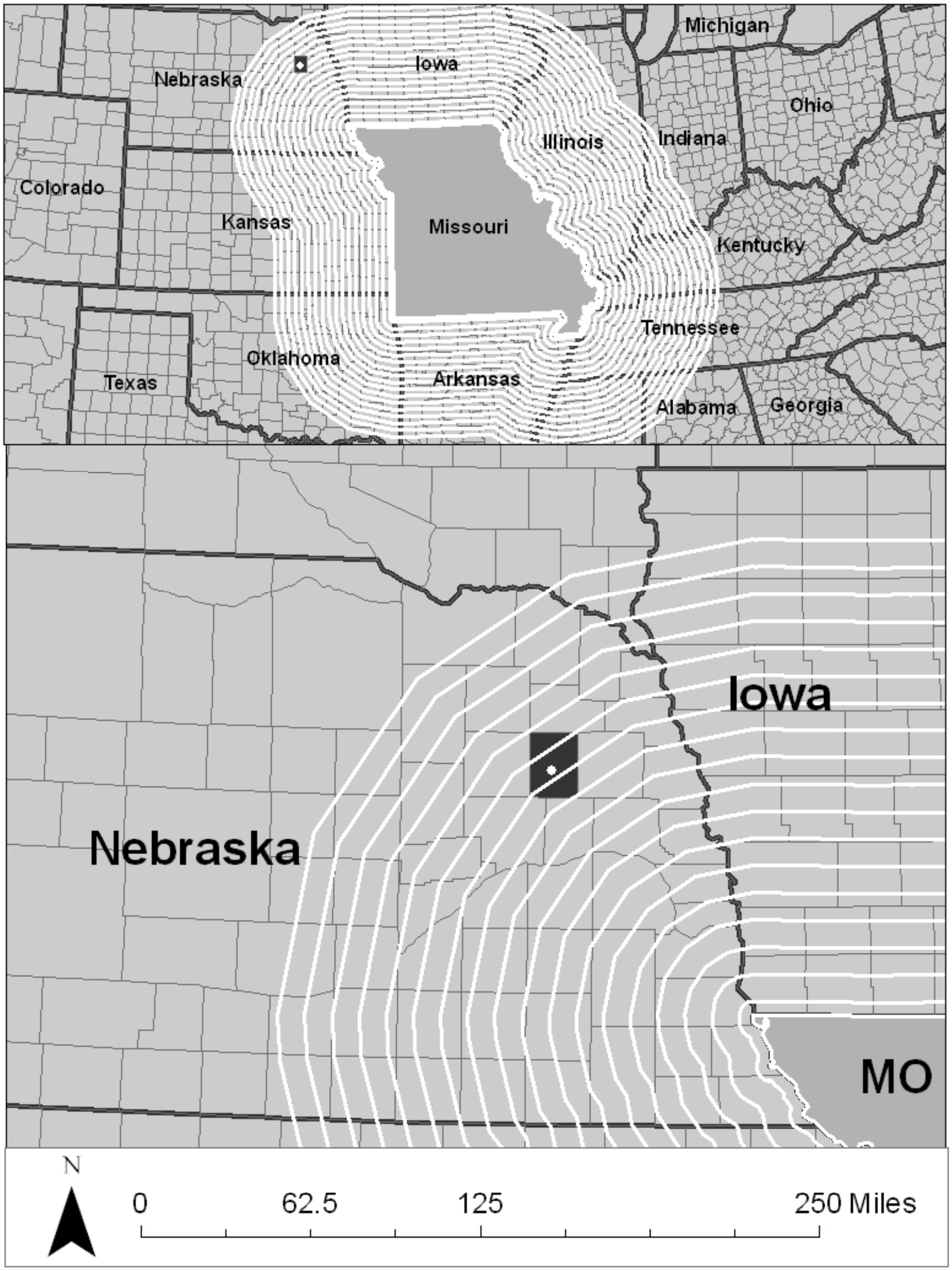


Figure S-2. Buffers Around Missouri for Measuring Degree of Concern for the Lottery Analysis (and an Illustrative Nebraska county)



Notes: In both maps, the darkened county in Nebraska is Stanton County. Also, in both maps, the dot in Stanton County denotes the geographic center of the county. Data for the electronic state and county boundaries are from United States Bureau of the Census (2005).

TABLE S-1. Regression Results for Testing the Welfare Learning and Competition Hypotheses when ΔB = 100

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Model with degree of concern substituted for benefit relative to neighbors		Model including both benefit relative to neighbors and degree of concern			
Value assumed for *M*	Slope estimate for degree of concern[a]	*t*-ratio for degree of concern	Slope estimate for degree of concern[a]	*t*-ratio for degree of concern	Slope estimate for benefit relative to neighbors	*t*-ratio for benefit relative to neighbors
Equations Assuming *p* = 0:						
150	-0.596	-0.65	0.032	0.37	-41.24*	-2.44
160	-0.082	-1.36	-0.027	-0.47	-37.10*	-2.18
170	-0.077*	-1.81	-0.042	-1.01	-33.93*	-1.99
220	-0.070*	-2.13	-0.045	-1.40	-31.55*	-1.86
Equations Assuming *p* = 100,000:						
150	-0.138	-0.41	0.050	0.15	-39.97*	-2.33
160	-0.010	-0.52	0.024	0.13	-40.00*	-2.33
170	-0.082	-0.67	-0.000	-0.00	-39.52*	-2.30
220	-0.081	-0.93	-0.023	-0.27	-38.30*	-2.23
Equations Assuming *p* = 200,000:						
150	-1.588	-1.25	-0.928	-0.73	-37.15*	-2.16
160	-0.223	-0.47	0.048	0.10	-39.91*	-2.31
170	-0.076	-0.33	0.058	0.25	-40.55*	-2.35
220	-0.057	-0.39	0.036	0.24	-40.59*	-2.35
Equations Assuming *p* = 500,000:						
150	3.382*	-2.48	-2.877*	-2.12	-35.66*	-2.08
160	-0.933*	-1.69	-0.703	-1.28	-36.11*	-2.10
170	-0.404	-1.51	-0.289	-1.08	-36.34*	-2.10
220	-0.285*	-1.65	-0.204	-1.18	-35.62*	-2.06

Notes: All equations are estimated using the same 1440 observations employed by BFH (2003). Results are obtained using the xtpcse procedure in Stata 8; *t*-ratios are computed using panel corrected standard errors. This table reports coefficient estimates for only those variables that are relevant for testing the welfare learning and competition hypotheses. Among the other variables assumed to influence a state's welfare benefit are the state's poverty rate, its wage levels, several variables reflecting economic conditions in the state and in neighboring states, whether the state has a residency requirement, a set of dummy variables for the states, and a lagged dependent variable. A full set of coefficient estimates for each equation is reported in the appendix of our unpublished supplement.

[a] Coefficients in this column are in 1000s.
* p < .05 (one-tailed)

TABLE S-2. Regression Results for Testing the Welfare Learning and Competition Hypotheses when ΔB = 200

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Model with degree of concern substituted for benefit relative to neighbors		Model including both benefit relative to neighbors and degree of concern			
Value assumed for *M*	Slope estimate for degree of concern[a]	*t*-ratio for degree of concern	Slope estimate for degree of concern[a]	*t*-ratio for degree of concern	Slope estimate for benefit relative to neighbors	*t*-ratio for benefit relative to neighbors
Equations Assuming *p* = 0:						
150	-0.133	-1.05	-0.042	-0.34	-38.30*	-2.23
160	-0.136	-1.61	-0.078	-0.93	-35.97*	-2.09
170	-0.125*	-2.04	-0.086	-1.40	-34.02*	-1.97
220	-0.116*	-2.41	-0.087*	-1.80	-32.24*	-1.87
Equations Assuming *p* = 100,000:						
150	-0.007	-0.02	0.152	0.40	-40.29*	-2.38
160	-0.062	-0.25	0.062	0.26	-40.21*	-2.38
170	-0.085	-0.50	0.008	0.05	-39.68*	-2.35
220	-0.101	-0.83	-0.034	-0.29	-38.51*	-2.27
Equations Assuming *p* = 200,000:						
150	-3.968*	-1.87	-3.035*	-1.43	-35.10*	-2.05
160	-0.921	-1.22	-0.548	-0.72	-36.77*	-2.14
170	-0.391	-1.10	-0.210	-0.59	-37.23*	-2.16
220	-0.265	-1.19	-0.147	-0.66	-36.81*	-2.14
Equations Assuming *p* = 500,000:						
150	-5.454*	-2.25	-4.659*	-1.94	-35.95*	-2.11
160	-1.576*	-1.70	-1.209	-1.31	-35.82*	-2.09
170	-0.702	-1.64	-0.522	-1.23	-35.81*	-2.08
220	-0.489*	-1.85	-0.370	-1.40	-34.92*	-2.03

Notes: All equations are estimated using the same 1440 observations employed by BFH (2003). Results are obtained using the xtpcse procedure in Stata 8; *t*-ratios are computed using panel corrected standard errors. This table reports coefficient estimates for only those variables that are relevant for testing the welfare learning and competition hypotheses. Among the other variables assumed to influence a state's welfare benefit are the state's poverty rate, its wage levels, several variables reflecting economic conditions in the state and in neighboring states, whether the state has a residency requirement, a set of dummy variables for the states, and a lagged dependent variable. A full set of coefficient estimates for each equation is reported in the appendix of our unpublished supplement.

[a] Coefficients in this column are in 1000s.
* p < .05 (one-tailed)