3B.5. ANALYSIS OF TORNADOES CASUALTIES USING THE CENSUS TRACT TORNADO PATH DATASET

Kevin M. Simmons and Daniel Sutter*
Austin College, Sherman, Texas, and University of Oklahoma, Norman, Oklahoma

1. INTRODUCTION

We report in this paper about a new data set available for tornado research. The data set features a mapping of tornadoes in the Storm Prediction Center's (SPC) national archive on to census tracts. The SPC archive currently reports the county (or counties) in the path of each storm so analysis of tornado casualties and damages can employ only county level economic and demographic variables. Many counties are quite large (1000 square miles or more), so county level variables may not the storm path. Census tracts offer greater geographic resolution and thus more accurate depiction of the population density or housing characteristics of the storm path. The greater precision of census tract based variables should improve tornado casualty models.

We describe the data set and the census tract level economic and demographic variables available for analysis. We illustrate the potential use of the data set by estimating tornado casualty models for Kansas, Nebraska and Oklahoma between 1990 and 1998 using multiple regression and variables constructed from county and census tract storm paths. The results indicate an improvement in the predictive power of the injury model using census tract level variables.

2. THE CENSUS TRACT STORM PATH DATASET

The latitude and longitude of the beginning and end points of tornado tracks from the SPC archive were plotted on a map of census tracts using Geographic Information System (GIS) software. Two problems arose during this process. First, a number of tornado tracks were too short to be mapped to census tracts. Although more than half of the storm paths could not be mapped, these storms tended to be short track and the mapped storms certainly represent the majority of tornado damage paths. Second, the latitude and longitude of tornado paths recorded in the SPC archive were in some cases either incorrectly or imprecisely recorded, yielding storm paths that were insufficiently accurate to be of use. We will describe how these issues affect the data set below.

All tornadoes in the SPC archive between 1950 and 1999 were plotted against both 1990 and 2000 Census tracts. Census tracts are intended to include a population of 2000 to 8000 and thus must be redrawn with each census in counties experience population growth or decline. We have census tract paths for about 12,900 tornadoes, approximately one third of all tornadoes nationally during this period. The data set identifies for each mapped storm the census tracts along the storm path as well as the portion of the tornado's path within each tract. The portion of the storm path in each census tract can be used to construct weighted average economic and demographic census tract level variables, as we illustrate below.

Census tracts offer considerably greater geographical resolution than counties. The number of tracts per county varies greatly since census tracts are drawn to have approximately equal populations. Thus in 2000 Oklahoma and Tulsa counties had 227 and 172 tracts each, while Roger Mills county had only one tract. Overall Oklahoma has 77 counties but 991 2000 tracts, while Kansas has 105 counties and 727 census tracts and Nebraska 93 counties and 503 census tracts. Census tracts offer about ten times the resolution of counties, although this depends on state population. Census tracts naturally exhibit a greater range of values of economic variables as well. For instance, the population density of Oklahoma counties in 2000 ranged from 1.71 persons per square mile to 986.5, while census tract population density ranged from .88 to 8909.

As mentioned above, quality issues arose with many of the storm paths. We illustrate for the case of Oklahoma in the 1990s. Mapping of SPC storms produced census tract tracks for 166 tornadoes for 1990 through 1998 and 75 tracks for 1999. There were 685 tornadoes in the state during

* *Corresponding author address:* Kevin M. Simmons Department of Economics and Business, Austin College, 9000 Grand Avenue, Sherman, TX 75090-4400; e-mail: ksimmons@austincollege.edu.

the decade, so census tract paths are available for 35% of state tornadoes in this case. We examined the quality of the mapping by comparing the counties of the census tracts identified from the storm path to the counties of the storm path from the SPC archive. Our mapping procedure assumes that tornadoes follow a straight path between its beginning and end points and consequently nonlinear storm paths could produce differences in the counties of the census tracts. Examination revealed that the quality of the matches for 1999 storms, however, was poor. The counties of the census tracts identified on the path matched exactly the counties listed for the storm in the SPC archive for only 9 of 75 1999 Oklahoma tornadoes and there was no agreement between the counties of the census tracts for 34 storms. By contrast, there was no match between the counties identified in the census tract map and the SPC records for only 3 of 166 tornadoes from 1990 through 1998. The poor quality of the paths of 1999 tornadoes is due to the less precise recording of latitude and longitude in the SPC archive. Visual comparison of maps of the paths of the 3 May 1999 tornadoes from the SPC archive to a map of storm paths available on the Norman National Weather Service Weather Forecast Office's website confirmed the inaccuracy of paths mapped on reported latitude and longitude. The data set of census tract paths available for research includes 1999 tornado paths, but we must warn that they be used with caution.

3. AN ILLUSTRATIVE ANALYSIS

We proceed now with an analysis of tornadoes between 1990 to 1998 in Kansas, Nebraska and Oklahoma, three states at the heart of "Tornado Alley," to illustrate the research potential of this data set. There is no reason to expect that tornadoes in these states and these years offer particular insight into casualties so our analysis is mainly for illustration. We did not employ 1999 due to the low quality of the matches of the tornado paths as discussed above. After omitting the storms for which the counties of the census tracts did not match any of the counties reported in the SPC archive, we were left with a total of 466 storms. Table 1 displays a breakdown of the tornadoes for which we have census tract paths against all tornadoes in these states in these years. Overall 25.7% of tornadoes from the SPC archive have a usable census tract map. The census tract data set, however, is more valuable for research on tornado casualties or damages than this percentage suggests. The majority of unmapped tornadoes had short paths and were rated F0 or F1 on the Fujita scale; the census tract data set includes most longer track, more dangerous tornadoes. While 11.7% of F0 and 36.6% of F1 tornadoes have census tract paths, 66.9% F2 tornadoes, 89.2% F3 tornadoes, all 13 F4 and three F5 tornadoes in these states during the period were mapped. The data set includes all 11 tornadoes producing one or more fatality and 77 of 101 injury generating tornadoes. The average path length of the storms with census tracts was 8.21 miles, while the average path for tornadoes in the state during the period was 2.86 miles, so the average path length of the storms that were not mapped was .998 miles. Overall tornadoes in the census tract path data set account for all fatalities and 90% of the injuries in these states and years.

We use three economic/demographic control variables in our calculations here, population density, median family income (in 1999 dollars), and mobile homes as a percentage of housing units. For each storm we constructed a weighted average value of these variables using both the 1990 and 2000 census values and storm paths with use the ratio of the tornado's path within a census tract as the weight. For example, suppose a tornado with a total path length of 5 miles tracked through three census tracts, with a path of 2.5 miles in tract A, 2.0 miles in tract B, and 0.5 miles in tract C. The ratio of the storm's path in each tract is .5, .4 and .1 respectively. If the population density in the three tracts were 40, 20 and 100 persons/mi^2 respectively, the weighted average density for this storm would be 38. Estimation of the annual values of the variables for tornadoes in different years uses linear interpolation from the decennial values. Annual county level variables were also interpolated from the decennial values.

Table 2 compares the range, mean and standard deviation of the three variables based on county and census tract storm paths. The census tract storm paths exhibit a greater range and variation for all three variables. The most populated tornado path based on counties has a density of 915.7 persons/mi^2, while the most populated path based on census tracts has a density of 5177 persons/mi^2. The maximum income with census tracts is twice as large as with counties, and the maximum percent mobile homes with census tracts is 40% compared to only 30% with county values. The standard deviation for each variable is higher with census tract paths as well. The values of the variables for individual storms differed considerably across the two types of paths. The difference in county and census tract path values exceeded 100 persons/mi^2 for 50 storms (and exceeded 1000 for 10 storms), $10,000 for medium income for 28 storms, and 10% for the percentage of mobile homes for 44 storms.

Table 3 compares the county and census tract path variables for the 11 tornadoes producing fatalities and 76 tornadoes producing injuries. Differences in the mean values are particularly apparent for population density and mobile homes. The population density of the storm paths of fatality and injury producing tornadoes are respectively 47% and 124% higher based on census tract data than county data, while the percentage of mobile homes in the path of fatality and injury producing tornadoes are 44% and 30% higher respectively. Since we expect more casualties when tornadoes strike densely populated areas with a large proportion of mobile homes, the census tract storm path characteristics provide a better indication of why these storms produced casualties.

Examination of population densities of the paths of F3 tornadoes provides a further indication of the value of census tract paths. Our sample included 54 F3 tornadoes, of which three produced fatalities and 21 produced injuries. The mean county population density of the fatality producing storms was 43.3 persons/mi^2, compared to 38.8 for the storms producing no fatalities. By contrast, the mean census tract population density of the fatality producing storms was 175 compared to 34.6 for the storms which did not produce fatalities. Injury and noninjury F3 tornadoes tell a similar story. The county population density was 51.8 and 31.0 for injury and noninjury tornadoes, compared to census tract densities of 61.0 and 30.7. The F3 tornadoes which produced casualties hit more populated areas but only at the census tract level do we see how populated these storms' paths were than indicated by county level paths.

4. ESTIMATING TORNADO CASUALTY MODELS WITH CENSUS TRACT DATA

We now estimate regression models of casualties with both county and census tract level control variables to investigate whether we obtain better fitting models with the more precise census tract tornado paths. We will describe the procedure briefly here; the interested reader can find additional details in Merrell, Simmons and Sutter (forthcoming) or Simmons and Sutter (2004). The number of casualties from a tornado take on integer values with a large proportion of zero observations and economists typically employ a Poisson regression model to analyze such count data (Greene 2000, pp.880-886). The model assumes that the expected number of casualties is a random variable with a Poisson distribution, with the parameter λ assumed to depend on independent variables. The log of the conditional mean of the dependent variable is modeled as a linear function of the independent variables and the model assumes equality of the conditional mean and variance of the dependent variable. A Pearson Chi-Square test indicated that this assumption was violated for injuries but not fatalities. Consequently we estimate a Poisson model of fatalities and a negative binomial regression, a generalization of the Poisson model appropriate in the presence of overdispersion (Greene 2000, pp.886-888), for injuries.

The independent variables of our model include the population density, income and mobile homes variables defined above. In addition we use the F-scale rating of the tornado, F-scale, entered as an integer value, and four variables describing the time of day, day of week, and year of tornadoes. Season is a dummy variable which the tornado occurs which equals one for tornadoes in the months of April, May or June and zero for tornadoes in any other month. Weekend is a dummy variable which equals one for tornadoes which occur on Saturdays or Sundays. Tornadoes at different times of day pose different types of threats, so we divided the storms into three time categories. Day, Evening and Night are dummy variables which equal one if the tornado struck between 6AM and 6PM, 6PM and 10PM or between 10PM and 6AM respectively, and zero otherwise. Day tornadoes are the omitted category in our regressions, so the coefficients on Evening and Night measure the effect of storms at these times relative to a storm occurring the day.

Table 4 presents the estimation of a Poisson model of the determinants of fatalities. Note that because no fatalities occurred at Night in our sample, we had to omit Night as a control variable. The first column presents the results with county level variables and the second column the results with census tract variables. In this case the county level variables perform better. F-scale is a positive and highly significant (at better than the .01 level) determinant of fatalities in both specifications. As expected, a tornado rated higher on the Fujita scale is expected to result in more fatalities. Population density is a positive and significant (at the .10 level) determinant of fatalities in the county specification; in the census tract specification; in the census tract specification density has a positive coefficient which is not statistically different from zero. A tornado of given strength hitting a more populated storm path is expected to be more deadly, and we consistently find this in other studies of tornado fatalities (Merrell, Simmons and Sutter forthcoming, Simmons and Sutter 2004). The insignificance of census tract density is disappointing. None of the other county or census tract variables are significant. Also of relevance are measures of goodness-of-fit of the models. We present three in Table 4, the log of the likelihood function, the deviance, and the Pearson Chi-Square statistic divided by model degrees of freedom; the higher the value of the likelihood function, a smaller value for the deviance (with a value of zero representing a perfect fit), and a scaled Pearson Chi-Square statistic closer to 1.0 indicate a better fit of the model. Two of the three measures show the model with county variables performing better, so at least for fatalities in this sample census tract variables did not produce a better fitting casualty model.

Table 5 presents the injury models. Again the first column presents the model estimated with county level variables and the second column the results with census tract variables. F-scale is positive and significant (at better than the .01 level) determinant of injuries in both specifications, as expected. This signs of the time variables are consistent across the two specifications, with Day significant and negative at the .10 level in the census tract specification and Weekend significant and positive in each specification. Population Density has a positive sign but is significant only in the census tract specification, which is an indication of the better performance of this model for injuries. On the other hand, income is positive and significant in the county specification but insignificant with a negative point estimate in the census tract specification. All three measures of goodness-of-fit, the log of the likelihood function, the deviance, and the Pearson Chi Square divided by degrees of freedom, indicate a better fit for the census tract specification. Thus we see that for injuries the census tract variables do improve the fit of the model.

5. CONCLUSION

Census tracts offer considerably greater geographic resolution for tornado paths than counties, which is the resolution of paths in the SPC tornado archive. A more accurate depiction of paths should allow the estimation of better models of tornado casualties, which could be crucial in attempting to evaluate the effectiveness of improved tornado warnings or disaster preparedness planning in reducing casualties. By better describing the area actually struck by a tornado, census tracts could allow effects of a policy which reduces casualties by 5 or 10 percent to be established statistically.

We have illustrated the potential improvement in the analysis of tornado casualties using census tract variables through an examination of tornadoes in Kansas, Nebraska and Oklahoma in the 1990s. The census tract paths of fatality and injury producing tornadoes exhibited higher population densities and a larger percentage of mobile homes than the county level paths. A regression model of tornado injuries produced a better fit with census tract variables, and the census tract population density was a significant and positive determinant of casualties whereas county population density was insignificant. The sample of tornadoes examined here is too small to draw any general conclusions, but our analysis illustrates the potential of the new data set. A model of tornado fatalities, however, actually fit better when estimated with county level variables than census tract variables. This is not totally unexpected, since there were about seven times more injury tornadoes than fatality tornadoes in our sample, and so th number of fatality tornadoes might be too small for the advantage of census tract paths to be evident.

References

Brooks, H. E., and C. A. Doswell III, 2002: Deaths in the 3 May 1999 Oklahoma City Tornado from a Historical Perspective. Weather and *Forecasting*, **17**, 354-361.

Greene, W. H., 2000: *Econometric Analysis*, 4th ed. Prentice Hall, 1004 pp.

Merrell, D., K. M. Simmons, and D. Sutter. The Determinants of Tornado Casualties and the Benefits of Tornado Shelters. *Land Economics*, forthcoming (expected publication date, February 2005). [Available currently online at http://www.ou.edu/cas/econ/Sutter/shelters.pdf]

Simmons, K. M., and D. Sutter. 2004. WSR-88D Radar, Tornado Warnings, and Tornado Casualties. Unpublished working paper. [Available from Department of Economics, University of Oklahoma and online at http://www.ou.edu/cas/econ/Sutter/radar%20paper.pdf]

Acknowlegment: The census tract paths data set was constructed under the NOAA grant "A Proposal to Match Tornado Tracks and Census Tracts" to the Cooperative Institute for Mesoscale Meteorological Studies at the University of Oklahoma. We thank May Yuan and Christie Clabot for their work in performing the GIS mapping of tornado paths. Xiaoyi Mu and Ying Xiao provided assistance in gathering and matching the economic variables to the tornado records.

Table 1

Comparison of Tornadoes with Census Tract Path Maps to all Tornadoes

	Census Tract Paths Mapped	All Oklahoma Tornadoes
F0	136	1157
F1	155	424
F2	101	151
F3	58	65
F4	13	13
F5	3	3
Fatality Tornadoes	11	11
Total Fatalities	35	35
Injury Tornadoes	77	101
Total Injuries	738	815

Table 2

Comparison of Storm Path Characteristics, Counties vs. Census Tract Data

	County	Census Tract
Population Density		
Minimum	.349	.678
Maximum	915.7	5177
Mean	43.48	93.44
Standard Deviation	85.52	363.8
Median Income		
Minimum	23798	9349
Maximum	60893	124009
Mean	38564	38105
Standard Deviation	5536	7846
Percent Mobile Homes		
Minimum	2.23	0
Maximum	30.01	40.30
Mean	10.90	13.51
Standard Deviation	5.06	7.50

Table 3

Comparison of County and Census Tract Paths of Tornadoes Generating Casualties

	Mean of County Variables	Mean of Census Tract Variables
Fatality Tornadoes		
Density	100.9	148.8
Income	$41,015	$40,256
Mobile Homes	10.84	15.65
Injury Tornadoes		
Density	75.4	169.0
Income	$39,639	$39,295
Mobile Homes	10.93	14.16

Table 4

Poisson Estimation of Tornado Fatalities

Independent Variable	County Variables	Census Tract Variables
F-scale	2.14**	2.36**
	(.350)	(.350)
Population Density	.0069*	.0045
	(.0024)	(.0085)
Income	-.0350	.0070
	(.0720)	(.0326)
Mobile Homes	8.53	1.38
	(6.58)	(3.76)
Day	1.07	1.13*
	(.657)	(.628)
Weekend	-.858*	1.18*
	(.848)	(.541)
Season	1.92*	2.49**
	(.788)	(.827)
Intercept	-11.4**	-13.4**
	(2.27)	(1.88)
Log Likelihood	7.631	3.057
Pearson Chi-Square/DF	.687	.889
Deviance	44.9	54.1

Standard errors are in parentheses. * and ** indicate that the estimated coefficient is significantly different from zero at the .10 and .01 levels respectively.

Table 5

Negative Binomial Estimation of Tornado Injuries

Independent Variable	County Variables	Census Tract Variables
F-scale	1.37**	1.46**
	(.169)	(.172)
Population Density	.0007	.0015**
	(.0021)	(.0009)
Income	.0567*	-.0124
	(.0328)	(.0218)
Mobile Homes	.467	-2.11
	(3.98)	(2.42)
Day	-.480	-1.12*
	(1.84)	(.675)
Evening	.0881	-.466
	(.668)	(.642)
Weekend	.767*	.740
	(.436)	(.408)
Season	.105	.303
	(.360)	(.371)
Intercept	-5.24**	-2.07**
	(1.70)	(1.13)
Log Likelihood	1669.2	1669.9
Pearson Chi-Square/DF	1.438	0.971
Deviance	186.5	182.3

Standard errors are in parentheses. * and ** indicate that the estimated coefficient is significantly different from zero at the .10 and .01 levels respectively.