Appendix: Regression Analysis

Making inferences from simple comparisons of recreation and other nonmetro county means can be misleading because it is possible that much of the observed socioeconomic difference between the two groups could be coincidental and not directly related to the extent of recreation.

For example, during the 1990s, many recreation counties in the Rocky Mountains benefited from an unusual regional phenomenon associated with the outflow of population from metropolitan California. This raises a question: How much of the difference in growth that we observed between recreation and other nonmetro counties nationwide was region-specific, associated with this one-time outflow of population?

Similarly, the decade of the 1990s was one of rapid economic improvement, which may have particularly benefited places with high poverty rates, providing job opportunities to many who, under normal conditions, would have had a hard time finding jobs. Many of these high-poverty rural areas are in the South in other nonmetro counties. This largely regional phenomenon could have led to our finding that recreation counties nationwide benefited less from poverty rate reduction than did other nonmetro counties. But would we find the same thing if we looked at each region separately?

Other factors unrelated to recreation might also be expected to differentially affect recreation and other nonmetro areas and lead to a potential bias in the differences observed between the two types of counties. For example, counties that are more urban in nature may have had developmental advantages over more rural and isolated areas. While recreation is expected to add to the level of urbanization, recreation counties are still less urban than other nonmetro counties on average, so this potential bias could mask the beneficial impact of recreation in simple comparisons.

Regression Methodology

In an attempt to overcome potential biases, we narrowed our analysis to recreation counties and conducted a regression analysis to see how a recreation county's extent of recreation dependency might affect the socioeconomic indicators examined in this report. Our measure of recreation dependency is the weighted average of a county's Z-scores covering tourism-related employment and income shares of the local economy and the recreational home share of total county homes, as developed by Johnson and Beale (2002): the larger the average, the more dependent a county is on recreation and tourism.[19] In addition, we included 10 dichotomous variables reflecting the Johnson and Beale recreation county types (for statistical reasons, we excluded the miscellaneous recreation county type). This allows for significant socioeconomic variations by type of recreation county (but it assumes that impacts associated with changes in recreation dependency do not vary with recreation type).

Following the approach of English et al. (2000), we also included several control variables that were not highly correlated with recreation dependency but that might be expected to affect local socioeconomic conditions. For example, we included eight dichotomous (0,1) variables identifying the

[19]Among the recreation counties we included in our analysis, recreation dependency ranged from a minimum of 0.12 to a maximum of 8.60, with a mean of 1.56 and a standard deviation of 1.23.

Census regional subdivisions. We did not include a dichotomous variable for one of the nine subdivisions—the Southeast—to avoid statistical problems.

We also included several demographic measures related to urbanization that are often included in empirical studies explaining regional socioeconomic variations. One was a dichotomous variable indicating whether the county was influenced by a nearby metropolitan area (based on adjacency as defined in the ERS 1993 Beale Codes, which requires both physical adjacency and significant commuting to the metro area). The other two demographic measures were county population density and percentage of county population residing in the rural portion of the county.

Ideally, an attempt to explain cross-county variations in socioeconomic indicators would involve separate models for each indicator, using theory to identify the explanatory variables and the form of the regression most relevant for a particular indicator. Given the large number of indicators in this study, we decided a simpler approach was expedient, so we followed English et al. in using just one set of explanatory variables for all of the indicators examined in our study. This results in some imprecision.

One of the ways our analysis differed from that of English and his colleagues was that our regressions only explained variations among our 311 recreation counties (rather than including all nonmetro counties as English did). In addition, we ran two ordinary least-squares regressions explaining intercounty variations rather than one. One of our regressions explained intercounty variations in the year 2000 (or the most recent year the data were available). The other regression explained intercounty variations in the change in the indicator over the previous 10 years. The change regression, which used the identical set of explanatory variables, may be viewed as a check on the year 2000 regression. In most cases, the regressions produced similar results: if recreation dependency was significant in the 2000 regression, it usually had the same sign and was significant in the change regression.

We also ran additional regressions for each indicator, adding a "squared" version of the recreation dependency variable to allow for a curvilinear relationship. We do not show the results of these additional regressions because in most cases they did not affect our results—the squared variable either explained little or no additional variation, or it only replaced the non-squared recreation dependency variable in significance with the same sign. In discussing our findings, however, we mention two cases where these curvilinear recreation factor regressions provided interesting results.

Regression Findings

Space limitations prevent us from showing the complete regression results here, including estimated coefficients for the many control variables we used in our regressions.[20] However, we can summarize our findings by showing only the regression coefficients for the recreation dependency variable in the linear regressions we ran to explain variations for each of the socioeconomic variables of interest. For example, each horizontal row in table 7 summarizes the results of one or two regressions covering a particular socioeconomic variable. Results for the 2000 regression refer to regres-

[20]Detailed regression results are available from the authors upon request.

sions that explain socioeconomic variations in the year 2000 (or in the next-closest year available). Results for the 1990s change regression refer to regressions that explain variations in the change in socioeconomic variables during the 1990s. Thus, table 7 summarizes the results for 29 regressions. In addition, the regression statistics shown are unstandardized, and one should not attempt to draw inferences about their relative importance based on their magnitudes.

These regression coefficients are generally consistent with what we previously found when comparing simple means for recreation and other nonmetro counties (tables 2 and 3). Dependency on recreation was significantly related to most of our economic indicators, and the recreation dependency regression coefficients were also generally consistent with most of our prior findings with regard to social indicators.

In addition, we found statistically significant relationships that were not apparent from comparisons of means for recreation and other nonmetro

Table 7

Linear regression analysis measuring the effect of recreation dependency on economic and social indicators

	2000 regression		1990s change regression	
Dependent variables	Recreation dependency B estimate	Regression's explanatory power[1]	Recreation dependency B estimate	Regression's explanatory power[1]
Economic indicators:				
Job growth rate	NA	NA	5.50**	0.184
Employment-populaton ratio:				
Ages 16-24	1.13**	0.209	0.56**	0.115
Ages 25-64	0.92**	0.211	0.48**	0.139
Ages 65 and over	1.04**	0.364	0.30	0.013
Earnings per job	-7.95	0.396	482.77**	0.265
Earnings per worker[2]	846.49**	0.317	NA	NA
Income per capita	1,044.52**	0.265	487.73**	0.207
Median household income[2]	1,474.40**	0.393	907.59**	0.339
Median rent	32.59**	0.516	10.74**	0.377
Social indicators:				
Population growth rate	4.59**	0.282	2.85**	0.245
Travel time to work	-0.25	0.327	-0.44**	0.157
Poverty rate[2]	-0.84**	0.249	-0.43**	0.242
Percent without HS diploma	-1.37**	0.468	0.22	0.341
Percent with bachelor's degree	2.24**	0.491	0.65**	0.211
Physicians per 100,000 population[3]	0.69	0.280	NA	NA
Age-adjusted death rate per 100,000 population[4]	-24.20**	0.290	NA	NA
Crime rate[2]	0.68**	0.264	NA	NA

NA=Not applicable.

* The coefficient is statistically different from zero at the .05 level.

** The coefficient is statistically different from zero at the .01 level.

[1]Adjusted R-square statistic (fraction of variation explained by regression).

[2]Data are reported for 1999

[3]Data are reported for 2003.

[4]Data are reported for 2000-02

Source: ERS calculations, based on data from U.S. Census Bureau and Bureau of Economic Analysis, U.S. Department of Commerce, and Bureau of Labor Statistics, U.S. Department of Labor.

counties. For example, the regression analysis showed significant positive relationships between recreation and the employment-population ratios for all three age groups studied, whereas there was little or no difference in the means for these ratios.

In some cases, the regression analysis raises questions about previously observed statistical differences. For example, we earlier found that recreation counties were statistically different from other nonmetro counties with respect to number of physicians per 100,000 residents, but the regression analysis found no statistically significant relationship between this indicator and recreation dependency.

For travel time to work, we had previously found no statistically significant difference between recreation and other nonmetro counties, either for the year 2000 or for the trend during the 1990s. However, the regression analysis revealed a statistically significant negative relationship between recreation dependence and change in travel time to work during the 1990s.

One of the more interesting findings was recreation dependency's negative and statistically significant relationship with the change in poverty rate. This means that the more recreation dependent a county is, the bigger its decline in poverty rate during the 1990s, controlling for other factors. The finding contrasts with our simple descriptive analysis, which found that recreation counties had, on average, a smaller decline in poverty than other nonmetro counties during the 1990s. This suggests that, as we suspected, the smaller average decline in poverty for recreation counties may have been simply a geographic coincidence, because when we controlled for regional differences and other factors in our regression analysis we found that the higher a county's recreation dependency, the more its poverty was reduced during this decade.

Another interesting finding involved earnings per job. We initially found that recreation dependency had a negative but statistically insignificant coefficient for earnings per job (in the 2000 model). When we ran the curvilinear version of the first regression (the 2000 model), we found a significant negative coefficient for recreation dependency and a significant positive coefficient for recreation dependency squared.[21] This implies that the recreation counties with moderate degrees of recreation dependency had relatively lower earnings per job, while those with higher or lower recreation dependency had higher earnings. Taken together, these findings present a somewhat muddled picture with respect to recreation impacts on earnings per job—there is no clear indication that recreation hurts a county in this regard. We got a clearer regression finding regarding the change in earnings per job during the 1990s, which revealed a positive and significant relationship between recreation dependency and the growth in earnings per job.

Two other indicators had different results for the 2000 regressions and the 1990s change regression: the employment population ratio for the elderly and the percent of adult (ages 25 and older) residents without high school diplomas. In both cases, the regressions explaining the change in the indicator produced insignificant coefficients for recreation dependency. For the employment-population ratio for ages 65 and up, the change regression performed very poorly, explaining less than 6 percent of the variation—less

[21]The nonlinear version of the change regression did not produce a similar significant relationship.

than any other regression in our analysis. This suggests that we might find a significant relationship if we were to improve the model to explain the behavior of the elderly. For the other indicator, the percentage without high school diplomas, we may need to find some other explanation, since the regression explaining change for this indicator performed better in terms of explaining variation than all of our other change-form regressions. Perhaps something unusual was going on in the 1990s that kept places with higher recreation dependencies from experiencing more significant declines in the percentage lacking high school degrees.[22]

We have already mentioned recreation's curvilinear relationship with earnings per job. The other case where we found a curvilinear relationship involved recreation's effects on population growth rates in the 1990s. The linear regression explaining population growth rate had a statistically significant positive coefficient for recreation dependency. The curvilinear regression had a statistically significant positive coefficient for recreation dependency and a statistically significant negative coefficient for recreation dependency squared. This implies that counties with moderate recreation dependencies have higher growth rates than counties with smaller or larger recreation dependencies.

[22]For example, it may be that during the 1990s, higher educated retirees began to move to a wider array of recreation areas, whereas before they may have concentrated in the most recreation-dependent areas.