Estimating Tourism's Share of Local Income from Secondary Data Sources

John Leatherman and Dave Marcouiller

Extension Report 97.4
May, 1997

* This article was originally published in The Review of Regional Studies 26 (3): 317 - 339. This reprint as an Extension Report is intended for educational purposes only and is not for sale or private distribution.
ESTIMATING TOURISM'S SHARE OF LOCAL INCOME FROM SECONDARY DATA SOURCES

John C. Leatherman and David W. Marcouiller*

Abstract—Typically, primary survey data is necessary to gauge the scale of local tourism. This paper proposes an alternative method for generating county-level estimates of employee compensation attributable to tourism based on secondary data sources. The procedure used principal components and cluster analyses to establish regions matched by tourism structure. Minimum requirements was then used to estimate the share of employee compensation attributable to non-local demand. The procedure was applied to Wisconsin counties to estimate tourism shares attributable to travelers and recreational home owners. The principal components analysis showed that Wisconsin tourism is driven by variable combinations of three components: urban tourism, outdoor-based activities, and natural parks/specialty tourism. Minimum requirements generated county-level estimates of non-local demand for disaggregated tourism-sensitive business sectors. Tourism shares ranged from approximately 3 percent of total employee compensation within the urban clusters to about 10 percent in the nature and parks clusters.

I. INTRODUCTION

Although tourism is an increasingly popular economic development strategy in many communities (Fredrick 1992; Gibson 1993), as a strategy to promote economic growth it is a relatively new phenomenon. Currently, tourism planning lacks an integrative perspective (Marcouiller 1997), an approach that considers tourism development within the context of broader regional development goals and applies greater rigor to the identification of economic, environmental, and/or sociocultural impacts. Coupling the techniques of regional science with an integrative planning perspective will enhance the ability to deal with questions related to tourism and create better local development policy.

Economic development practitioners and specialists are frequently asked to estimate the scope and scale of local tourism activities, but in the absence of primary data collection, it is difficult to estimate the importance of tourism at the local level. In addition, the use of state/provincial or national traveler survey data fails to take into account local economic structure and tourism activities, and local primary data gathering is time-consuming and expensive. Existing techniques

*Assistant Professor/Extension Specialist, Department of Agricultural Economics, Kansas State University; and Assistant Professor/Extension Specialist, Department of Urban and Regional Planning, University of Wisconsin-Madison. The authors would like to thank Jack Huddleston and three anonymous reviewers for helpful comments on earlier drafts of this paper. Any remaining errors or omissions are the sole responsibility of the authors.
employing secondary data sources to estimate tourism activity generally cannot make important distinctions related to the nature of tourism demand. What is needed is a method for making detailed estimates of local tourism activity based on secondary data sources that can serve as a starting point for discussion related to tourism development strategies.

This research proposes a method for estimating the scale of county-level tourism activity based on secondary data sources. To estimate income generated from tourism for counties in the State of Wisconsin (USA) a four-step approach was used. The first step gathered tourism-related data for each county in the state. Second, a principal components analysis was used to summarize the data, permitting classification of each county by the "structure" of its tourism activities. The third step clustered counties based on the similarity of tourism characteristics, providing relevant reference groups for use in the final step. Finally, a minimum requirements technique was applied to estimate the share of local business activity attributable to non-local demand for discrete business sectors.

Tourism Measurement Techniques

Given the broad scope of activities included in "tourism," its measurement begins by establishing basic definitions. One perspective in tourism research focuses on demand-related issues (Johnson and Thomas 1992; Ryan 1991), e.g. trends and forecasts, travelers' motivation and perceptions, tourists' characteristics, or visitors' social and cultural impacts on the host community. Pursuing a supply-side perspective, Smith (1988) suggests that tourism should be viewed as an industry composed of businesses providing goods and services to travelers and visitors. Although criticized as an oversimplification of the complex and partially industrialized structure of tourism supply (Leiper 1990), this perspective emphasizes the measurement of the scale, performance, and economic impacts for locality-specific businesses catering to travelers and visitors (Eadington and Redman 1991; Watson and Cordell 1990; Gibson 1993). Others have suggested a broader definition extending beyond travelers and visitors to include the impacts associated with cottage and second-home owners (Jordan 1980; Jackson 1986; Girard and Gartner 1993).

Following the supply-side perspective, a large body of research has assessed the economic scale and impact of tourism activities. Most tourism impact studies measure either expenditures or employment associated with visitor demand. Expenditure studies typically survey businesses catering to visitors (Johnson et al. 1989), survey the visitors directly (Johnson and Moore 1993), or use regional/national surveys of visitor expenditures (Bergstrom et al. 1990; Smith 1988). While surveys are generally considered to provide reasonably accurate estimates of
tourism activity, the difficulty is the time and cost of local data gathering or the applicability of non-local survey results to the local situation. Furthermore, primary data-gathering efforts suffer from non-comparability with published secondary data sources.

Many studies employ input-output techniques to estimate the indirect impacts associated with tourism spending (Fletcher 1989; Johnson and Moore 1993; Propst 1985). These studies extend understanding about the broader economic impacts of tourism, but the initial spending stimulus must still be provided by survey data from visitors or businesses.

Other studies estimated the employment generated by tourism demand. Smith (1988) identified two tiers of tourism businesses: those businesses that could not exist without travelers (tourism-reliant), and those that serve a mix of local and non-local demand (tourism-enabling). He used national expenditure surveys to identify the "tourism ratio" for these businesses. The application of national ratios to local or regional areas, however, cannot take into consideration variation in tourism amenities and accommodations across places.

Another approach has indirectly assessed tourism employment using secondary data sources. Johnson and Thomas (1990) used the "employment count method." This method uses published employment data to count the number of jobs from predetermined Standard Industrial Classification (SIC) sectors known to be associated with the tourism industry (e.g., lodging, restaurants, amusement and recreation services, etc.). While making use of standard reporting sources, an aggregate employment count cannot make the important distinction between tourism-enabling and tourism-reliant businesses. Brown and Connelly (1986) utilized state unemployment compensation records (ES-202), together with the employment count method to track seasonal variation in tourism employment. While offering the additional benefit of tracking industry performance over time, detailed reports from this data source are often considered proprietary and unavailable to many practitioners and researchers.

Thus, existing methods for estimating the scale of local tourism activities requires either local primary data collection with the associated time, expense, and effort, or the use of secondary data with overly general assumptions regarding the nature of tourism demand. This research contributes a method that can (1) make use of standardized, readily available secondary data sources, (2) provide local-level information for discrete business sectors, and (3) provide insight into the relative proportion of local business activity attributable to local versus non-local demand.
Estimating Tourism Export Shares

Tourism has been characterized as an export sector in the regional economy (Eadington and Redman 1991). Just as traditional export industries such as agriculture, mining, forestry, and manufacturing deliver locally produced products to non-local markets in exchange for income, tourism also attracts income to a region. The income generated from tourism export sales is then available to support other non-export sectors in the region. Conceptualizing tourism as an export activity permits the use of a variety of techniques developed by regional scientists to estimate the level of non-local demand from secondary data sources (Isserman, 1980; Richardson, 1985).

One method for estimating export activity developed over 30 years ago is the minimum requirements approach (Ullman and Dacey 1960; Ullman et al. 1971). This technique is an adaptation of the location quotient with the primary difference being that the employment characteristics of a region are compared to similar regions rather than a larger reference economy (state or nation). The assumption is that similar-sized regions exhibit similar consumption patterns and export propensities. Sectoral export-based employment is assumed to be the excess in the proportion of sector to total employment in a region as compared to a minimum in a group of similar regions.

The primary criticism of the minimum requirements technique is the assumption that the minimum region sets the standard for consumption. The implication is that all regions except the minimum are exporters, and none are importing (Pratt 1968; Pfister 1980; Keil 1995). An alternative interpretation is that the employment measured by the technique is associated with the consumption from regional production rather than total consumption (Isserman 1980). Thus, since consumption consists of regionally produced output plus imports, minimum requirements would measure gross exports. Each region can be both importer and exporter.

As applied to tourism, these criticisms are minimized by the nature of tourism sales. Tourism consists of the combined sales of a large number of enterprises, each more or less catering to non-local demand. Any region has a number of traveler/visitor-sensitive businesses. In aggregate, every region would be expected to have some tourist-related export sales.

A second concern in the use of minimum requirements is the level of disaggregation employed in the analysis. Tourism exports have been estimated at a 4-digit SIC level in this study. As originally developed, minimum employment was estimated for a 14-industry sector disaggregation, roughly at the SIC Division level (Ullman and Dacey 1960; Ullman 1968; Ullman et al. 1971; Moore and Jacobsen 1984). Pratt (1968) suggested that using minimum requirements at a
more disaggregate level may result in overestimates of export shares. More recently, Keil (1995) echoed this concern, suggesting that the use of minimum requirements at any level of disaggregation below a 1-digit level can lead to overestimates of export shares due to the absence of activities in the minimum regions.

Indirect evidence in support of using minimum requirements at a finer level of disaggregation is provided by Gibson and Worden (1981). They compared minimum requirements estimates, sample surveys, and location quotient estimates of export activity to a comprehensive census of businesses for 20 smaller communities (2,000 to 15,000 population). They found that minimum requirements came closest to the comprehensive census results, performing even better than survey results. While they report only aggregate multipliers, it could be assumed that in the smaller communities, many sectors probably consisted of only a few firms.

Our approach presents an addition to the traditional tourism impact assessment methods. It focuses on the unique tourism-related characteristics of regions. Prior to using minimum requirements to estimate tourism activity, a number of related issues must be addressed. The first is the identification of appropriate comparison regions. The second issue is specification of the economic activity to be included in a measure of tourism.

Tourism Regions

An important consideration in the use of minimum requirements is the specification of comparison regions. Generally the pool of comparison regions should be of approximately equal population, based on the assumption that similarly sized regions exhibit similar consumption and export patterns. In the case of tourism, it is argued that an additional requirement be that the regions also have similar tourism structures. Tourism patterns are expected to differ between regions containing differing tourism resources. For this reason, results using minimum requirements are assumed sensitive to regional specification.

Several researchers have used principal components analysis and county-level data to describe local tourism resources and to specify substate regions based on the similarity of tourism resources. Principal components analysis was originally developed by Hotelling (1933) as a technique whereby a large number of independent variables are combined into a smaller set of coherent, yet statistically distinct "principal components." These components describe the basic structure of the data (Dunteman, 1989). Counties can be assigned principal component scores and grouped into regions based on the similarity of their principal components.
Smith (1987) used principal components analysis to identify four basic types of tourism for the Province of Ontario: urban tourism, outdoor recreation, cottaging/boating and urban fringe tourism. After classifying Ontario counties according to the degree they reflected these basic types, he used a cluster analysis to identify tourism regions. The resulting regions were structurally similar, though not necessarily contiguous. Replication of his work has occurred in South Carolina (Lovin-good and Mitchell 1989) and Wisconsin (Lystad 1991).

II. METHODS

Our analysis consisted of four steps. First, data related to tourism was gathered for each county in Wisconsin, and principal components analysis was used to transform the data into broad indices of tourism. Each county was then assigned a principal component score based on the extent to which it reflected a given type of tourism activity. The county scores were then used in a cluster analysis to establish regions composed of counties with similar tourism attributes. The fourth step of the analysis used these regional specifications to identify labor income attributable to tourism and travel by applying a minimum requirements technique to employee compensation data for tourism-sensitive industries.

Wisconsin County Tourism Data

Data reflecting tourism activities were gathered for 70 Wisconsin counties.\(^1\) Data variables were selected to reflect natural and cultural amenities, tourism infrastructure and other factors thought to influence the level and type of tourism.\(^2\) Natural resource variables included forested acreage, inland water acreage, and cropland acreage. Cultural attractions included events (festivals, competitions, etc.), historic sites, and museums. Tourism facilities included numbers of restaurants, lodging rooms, camp sites, snowmobile mileage, downhill ski runs, and seasonal housing. Finally, several other variables thought to reflect the nature of tourism were included, such as population density, remoteness from large population centers, and the proportion of total employment in sectors known to cater to tourists (restaurants, hotels/motels, and amusements). The data were "structured" to reflect the importance of tourism in a county.\(^3\) The resulting data set consisted of a large number of measures related to tourism activity for each county in the analysis.
Creating a Tourism Typology using Principal Components Analysis

In the second step, principal components analysis (PCA) was used to summarize the data by combining them into a smaller set of comprehensive, yet distinct variables. A principal component is a summary measure that combines a larger number of related variables into one conceptually coherent variable. In this case, a number of variables thought to reflect tourism were combined into a smaller set of "summary" variables, statistically independent of each other and interpreted as representing distinct types of tourism.

Principal components analysis begins by creating a correlation matrix of the variables. Highly correlated variables are linearly combined into a new set of variables called "components." The first principal component ascribes a line of best fit in geometric space across all the data. It accounts for as much of the total variance in the correlation matrix across all of the variables as possible. The first principal component is the best summary of the entire data set.

Once the first component is determined, the explained variance is removed from the correlation matrix and a second principal component is calculated. This component is geometrically orthogonal to the first line and attempts to explain as much of the remaining variance in the correlation matrix as possible. The second component is then interpreted as contrasting with the first component. For example, if the first component represents "urban" tourism, the second component may represent "rural" tourism.

A third component orthogonal to the first two is calculated next. This decomposition of the correlation matrix continues until the total variance in the correlation matrix is explained. Since the resulting components are uncorrelated to each other, each makes a unique contribution in accounting for the variance in the correlation matrix. When complete, the principal components are independent of each other and provide a concise, comprehensive description of the original data.

There can be as many principal components as there are original variables in the data set. However, the objective was to classify each county using a general typology of tourism. Each principal component represents a different type of tourism. The number of principal components to retain was determined by the eigenvalue, a summary measure of the amount of variance in the correlation matrix accounted for by a principal component. Generally, components with eigenvalues above 0.7 suggest sufficient explanatory power to be retained in the analysis (Dunteman 1989). Principal components with eigenvalues of less than 0.7 were discarded, because such a component is not considered to strongly represent the concept being measured by the original data.

An eigenvector (a column of weighted scores) decomposes the contribution of each of the original variables to the principal component. The eigenvector is
transformed into a principal component loading score by multiplying each element of the vector by the square root of the eigenvalue. Variables with loading scores above 0.5 are considered to contribute to the meaning of the principal component (ibid.). A variable that never "loads" strongly to any of the components suggests that it is unrelated to the other variables. The unrelated variable offering little explanatory power is generally removed and the analysis repeated. The variables with strong loading scores are then used to interpret the meaning of the principal component.

Principal components analysis is an iterative process. Several combinations of variables and variations in data format are assessed until a series of components providing the clearest explanation of the total variance in the correlation matrix are obtained. The number of components ultimately retained depends on the statistical values calculated in combination with the ability to coherently interpret the results. In this analysis, three principal components were retained for use, and a score was generated for every county. These scores were used to generate reference groups of counties with similar tourism structure.

Matching Counties with Cluster Analysis

Cluster analysis is a generic term referring to a number of methods used to generate classifications of similar entities (Aldenderfer and Blashfield 1984). The resulting clusters are intended to minimize variation within groups while maximizing differences between groups. The clustering method used was a hierarchical algorithm known as Ward's Minimum-Variance Method (Ward 1963). A point representing each county's principal component scores is set in geometric space. Using a generalized Pythagorean theorem, Ward's Method calculates the distance between each point. The closest points are joined and replaced by a new point representing the two. The process of combining points continues until only a single point remains.

At each step, the error sum of squares is calculated to measure the amount of information loss associated with the joining of points. By plotting the cluster output, it is possible to observe where large jumps occur in the error term. These jumps indicate a boundary between relatively distinct clusters. Some clusters are quite distinct, while others show only subtle differences. Depending on the purpose of the analysis, any number of clusters can be distinguished up to the original number of entities in the analysis.

The resulting clusters can be mapped to display the dispersion of similar counties and to identify larger regions. Each can be described by calculating the means of the principal component scores and the original data variables. A group of counties clustered by the similarity of their tourism structure serves as the ref-
Estimating Tourism's Share of Local Income

Employee compensation estimates were obtained from Micro-IMPLAN (ver. 91-F) 1991 county data files (Minnesota IMPLAN Group 1993). The IMPLAN data was selected for use because (1) it is readily available, (2) it provides annualized estimates, (3) it contains necessary sectoral detail, (4) it estimates non-disclosed sectors, and (5) the output can be incorporated into additional model building. Employee compensation was used to estimate tourism shares. Given the seasonal and part-time nature of the employment associated with tourism, an income measure may be a better general indicator of tourism share than employment.

In our research, two types of tourism were distinguished by economic sector. The first was a group that focused on sectors sensitive to traveler expenditures. The sectors in this category included overnight accommodations, eating/drinking establishments, recreation/amusements, and miscellaneous retail. The second category focused on recreational homeowners. Sectors in this category included real estate, new construction, building maintenance, and food stores.

The minimum requirements technique established tourism shares by industry sector. The calculation used the form:

\[ ES_i^t = \left[ (ec_i^t/ec_T^t) - (ec_{im}^t/ec_{m}^t) \right] ec_T^t \]

where \( ES_i^t \) represents the export share of employee compensation in the industry \( i \) in time period \( t \). The ratio \( ec_i^t/ec_T^t \) represents the employee compensation in industry \( i \) in time period \( t \) to the total employee compensation in the county. This ratio was calculated for all counties and a minimum ratio was identified by industry. This minimum ratio \( (ec_{im}^t/ec_{m}^t) \) of employee compensation for industry \( i \) to total employee compensation for the county during time period \( t \) was identified by sector among all counties within a region consisting of similar counties. In the minimum ratio county, all production within the minimum sector was assumed to meet local demand, i.e., there were no "gross exports" of tourism within that sector for that county. In all other comparison counties, the excess above the minimum ratio was used to support out-of-region (tourism) demand.

Total export share of employee compensation attributable to out-of-region sources can be shown as
\[ \sum E_{S_i}^T = \sum \left( \frac{ec_i^T}{ec_T^T} \right) ec_T^T - \sum \left( \frac{ec_{im}^T}{ec_{m}^T} \right) ec_i^T = ec_i^T \left( 1 - \sum \left( \frac{ec_{im}^T}{ec_{m}^T} \right) \right) \] (2)

The clusters identified in the previous step form the regions for which the minimum ratios were calculated. This procedure provided relevant comparisons by matching counties with similar tourism structure. Among Wisconsin counties, 12 clusters were identified.

All counties are arrayed by principal component 1, the line of best fit across all the original tourism variables. The minimum ratios are calculated from within the cluster plus selected counties of similar population in immediately adjacent clusters. Selection from adjacent clusters creates a "floating region." Given the relatively fine detail of the clustering scheme, this set-up permits selection of minimum ratios from a deeper pool of counties.9

III. RESULTS

Wisconsin's Tourism Regions

In the initial principal components analysis, some variables had low component loading scores, indicating little explanatory value in the data set. These variables (cropland, downhill ski runs, and historic sites) were removed.10 The three principal components selected explained 75 percent of the total variance in the correlation matrix generated from the original data set. The eigenvalue of the fourth component suggested inclusion, but none of the variables approached the minimum 0.5 value on the loading scores necessary to interpret the component. The high component loading scores for the variables in the three component solution are shown in Table 1.

The loading scores indicated a basic distinction between outdoor tourism and urban tourism. At one extreme were counties distant from population centers with access to forest and water resources. Counties that exemplify this type also had the infrastructure necessary to accommodate tourists, e.g., camping, lodging, restaurants and seasonal housing. In contrast, the second component showed high scores for overall tourism-related employment, number of museums, and number of tourist events. In addition, this component was weighted toward large population centers in the state. The third component accounted for the impact of state and national parks and recreation areas.

This interpretation was supported by mapping Wisconsin counties by their scores on the first two components. Counties scoring at the extreme on component 1 tended to be located in northern Wisconsin's lakes regions. These are
Table 1: Higher Principal Component Loading Scores for Wisconsin Tourism Data, 1991

<table>
<thead>
<tr>
<th>Variable</th>
<th>Component 1</th>
<th>Component 2</th>
<th>Component 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camp Sites per 1,000</td>
<td>0.718</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forest Acres per Capita</td>
<td>0.799</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lodging Units per 1,000</td>
<td>0.827</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Events</td>
<td></td>
<td>0.781</td>
<td></td>
</tr>
<tr>
<td>Number of Museums</td>
<td></td>
<td>0.719</td>
<td></td>
</tr>
<tr>
<td>HRT Empl. per 1,000</td>
<td></td>
<td>0.803</td>
<td></td>
</tr>
<tr>
<td>Park Acres</td>
<td></td>
<td></td>
<td>0.772</td>
</tr>
<tr>
<td>Population/Square Mile</td>
<td></td>
<td>0.549</td>
<td></td>
</tr>
<tr>
<td>Restaurants per 1,000</td>
<td>0.936</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remoteness (distance)</td>
<td></td>
<td>0.790</td>
<td></td>
</tr>
<tr>
<td>Seasonal Housing/1,000</td>
<td></td>
<td>0.903</td>
<td></td>
</tr>
<tr>
<td>Water Acres per Capita</td>
<td></td>
<td>0.870</td>
<td></td>
</tr>
</tbody>
</table>

Explained Variance: 44.9% + 21.4% + 8.8% = 75.1%

1 A principal component loading score above 0.7 is considered significant in explaining the variance in a set of data. Scores between 0.5 and 0.7 are supportive in explaining the variance. Below 0.5, the variable does not contribute to the component.

2 Data was collected using 1991 as the base year, with minor deviations depending on the availability of published information.

Popular tourist destinations for boating, fishing, and hunting. These areas also tended to have a very large proportion of the housing stock classified as seasonal homes. Counties scoring high on component 2 tended to be concentrated in the more densely populated southeastern corner of the state. Tourism opportunities in these counties were dominated by cultural amenities. The parks component tended to be more randomly distributed throughout the state. The location of parks reflected environmental conditions and were not associated with historical development patterns as was the tendency with components 1 and 2.

Counties' scores on the three principal components were entered into a cluster analysis. Clear cluster breaks occurred at the extremes. Three distinct groups of counties emerged: an urban cluster (16 counties); an amenities-rich outdoor cluster (4 counties); and a tourism specialty/parks cluster (3 counties). The majority of counties fell into a large group (47 counties) that reflected a continuum between the extremes of urban and outdoor groups.
A detailed set of clusters (12 clusters) was chosen based on smaller minimums in the sum of square ratios. While more detailed than previous research employing this method (generally 6-8 clusters), this set is more useful in identifying relevant reference groups for the purpose of estimating tourism shares by business sector. The clusters ranged in size from three to 10 counties. A useful method to characterize clusters and determine their distinctness is to compare the means of the original variables for each cluster. Table 2 shows the average principal component score and variable scores for each cluster, and Figure 1 displays the distribution of Wisconsin tourism clusters.

Referring to Figure 1 and moving along the continuum of urban tourism to outdoor tourism, we see that the counties labeled 1 tend to be the most urban, while the counties labeled 11 tend to exemplify amenities-rich outdoor tourism. Counties labeled 11 and 12 are generally considered to be the most popular outdoor tourism counties. The counties labeled 12 are “super-tourism” destinations with extraordinary parks and tourism infrastructure. The counties labeled 0 include Milwaukee County and the Menominee Indian Nation, both of which have unique economic structures relative to other counties and were excluded from the analysis. Alternately, each of these counties can be considered a unique tourism region consisting of a cluster of one.

Establishing Tourism Export Shares

Employee compensation represents wages and salaries paid to workers in tourism-related businesses. Minimum requirements were used to estimate the proportion of employee compensation attributable to out-of-county sources. The breakdown of tourism shares by type of business for each of the clusters is shown in Tables 3 and 4. Similar estimates can be generated for individual counties.

The urban clusters and the clusters with a larger number of counties obviously had the largest amount of tourism-related employee compensation. The proportions in Table 4 show the importance of tourism as a contributor to local income. In general, the importance of tourism increased as county population decreased. Tourism accounted for 2.4 percent of total employee compensation in the urban counties. As counties become more rural, tourism accounted for between about 3 percent and 9 percent of all employee compensation. Tourism accounted for 8.5 percent and 9.5 percent of total employee compensation in clusters 11 and 12, respectively.

Of course, allocating tourism-sensitive businesses into categories servicing travelers versus recreational homeowners is not completely accurate. All tourism-sensitive business sectors derive significant income from both groups. Acknowledging this crossover, it remains useful to recognize the variable impact these
TABLE 2
Average Values of Principal Components and Original Variables per Wisconsin County\(^1\) by Cluster

<table>
<thead>
<tr>
<th>Variable (^2)</th>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 (n=6)</td>
</tr>
<tr>
<td>Component 1</td>
<td>-2.7</td>
</tr>
<tr>
<td>Component 2</td>
<td>2.5</td>
</tr>
<tr>
<td>Component 3</td>
<td>-1.0</td>
</tr>
<tr>
<td>Camp Sites per 1,000</td>
<td>7.5</td>
</tr>
<tr>
<td>Forest Acres/Per Capita</td>
<td>0.1</td>
</tr>
<tr>
<td>Lodging Units/1,000</td>
<td>10.0</td>
</tr>
<tr>
<td>Number of Events</td>
<td>25.7</td>
</tr>
<tr>
<td>Number of Museums</td>
<td>6.2</td>
</tr>
<tr>
<td>HRT Empl. per 1,000</td>
<td>42.9</td>
</tr>
<tr>
<td>Park Acres</td>
<td>901.5</td>
</tr>
<tr>
<td>Population/Square Mile</td>
<td>424.0</td>
</tr>
<tr>
<td>Restaurants per 1,000</td>
<td>2.2</td>
</tr>
<tr>
<td>Remoteness (distance)</td>
<td>0.0</td>
</tr>
<tr>
<td>Seastional Housing/1,000</td>
<td>6.5</td>
</tr>
<tr>
<td>Water Acres per Capita</td>
<td>0.2</td>
</tr>
</tbody>
</table>

\(^1\) Milwaukee County and the Menominee Indian Nation are excluded.
\(^2\) Data was collected using 1991 as the base year, with minor deviations depending on the availability of published information.
\(^3\) \(n\) refers to the number of counties in a cluster.
FIGURE 1
Wisconsin Counties Clustered by Tourism Structure, 1991

Numbers refer to cluster groups.
Milwaukee and Menominee Counties (labeled 0) were excluded from the analysis.
| Source/Sector                | Cluster | Cluster | Cluster | Cluster | Cluster | Cluster | Cluster | Cluster | Cluster | Cluster | Cluster |
|-----------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|                             | 1 (n=6) | 2 (n=10)| 3 (n=6) | 4 (n=5) | 5 (n=4) | 6 (n=6) | 7 (n=9) | 8 (n=9) | 9 (n=5) | 10 (n=4) | 11 (n=4) | 12 (n=3) |
| Travelers                   |         |         |         |         |         |         |         |         |         |         |         |         |
| Eating & Drinking           | $65.2   | $95.8   | $24.8   | $12.5   | $15.4   | $6.5    | $4.7    | $23.7   | $12.5   | $2.8    | $6.2    | $18.3   |
| Miscellaneous Retail        | 86.9    | 51.2    | 50.7    | 6.0     | 7.5     | 4.9     | 1.9     | 44.9    | 27.5    | 3.0     | 5.9     | 5.3     |
| Lodging                     | 64.2    | 50.6    | 15.5    | 5.8     | 10.7    | 2.9     | 4.4     | 5.7     | 3.0     | 1.5     | 14.8    | 24.9    |
| Bowling Alleys              | 4.0     | 4.0     | 1.5     | 1.3     | 0.5     | 0.5     | 0.2     | 1.0     | 1.6     | 0.3     | 0.1     | 0.9     |
| Amusement/Recreation        | 14.6    | 12.1    | 4.5     | 1.2     | 4.5     | 1.1     | 1.5     | 3.1     | 2.7     | 0.4     | 1.7     | 9.5     |
| TOTAL                       | $234.9  | $213.7  | $97.0   | $26.8   | $38.6   | $15.9   | $12.7   | $78.4   | $47.3   | $8.0    | $28.7   | $58.9   |
| Recreational Homes          |         |         |         |         |         |         |         |         |         |         |         |         |
| New Residential Structures  | $47.5   | $37.5   | $11.7   | $3.5    | $4.9    | $1.9    | $6.3    | $11.6   | $7.1    | $1.0    | $3.3    | $9.8    |
| Maintenance, Residential    | 27.0    | 21.3    | 6.7     | 2.0     | 2.8     | 1.1     | 3.6     | 6.6     | 4.0     | 0.6     | 1.9     | 5.6     |
| Food Stores                 | 43.5    | 62.9    | 11.8    | 13.5    | 6.2     | 6.2     | 4.5     | 8.5     | 12.1    | 4.7     | 9.0     | 4.0     |
| Real Estate                 | 49.3    | 28.0    | 5.6     | 2.2     | 1.5     | 1.8     | 1.4     | 3.6     | 1.4     | 1.7     | 3.5     | 2.8     |
| TOTAL                       | $167.3  | $149.7  | $35.8   | $21.2   | $15.4   | $11.0   | $15.8   | $30.3   | $24.6   | $8.0    | $17.7   | $22.2   |

1 Milwaukee County and the Menominee Indian Nation are excluded.
2 n refers to the number of counties in a cluster.
<table>
<thead>
<tr>
<th>Source/Sector</th>
<th>Cluster 1 (n=6)</th>
<th>Cluster 2 (n=10)</th>
<th>Cluster 3 (n=6)</th>
<th>Cluster 4 (n=5)</th>
<th>Cluster 5 (n=3)</th>
<th>Cluster 6 (n=4)</th>
<th>Cluster 7 (n=6)</th>
<th>Cluster 8 (n=9)</th>
<th>Cluster 9 (n=9)</th>
<th>Cluster 10 (n=5)</th>
<th>Cluster 11 (n=4)</th>
<th>Cluster 12 (n=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traveler</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eating &amp; Drinking</td>
<td>18.5</td>
<td>36.7</td>
<td>27.8</td>
<td>36.8</td>
<td>37.8</td>
<td>22.4</td>
<td>26.2</td>
<td>47.2</td>
<td>37.6</td>
<td>33.3</td>
<td>30.1</td>
<td>49.4</td>
</tr>
<tr>
<td>Miscellaneous Retail</td>
<td>34.2</td>
<td>36.6</td>
<td>68.5</td>
<td>50.1</td>
<td>51.3</td>
<td>39.1</td>
<td>19.2</td>
<td>78.8</td>
<td>77.0</td>
<td>51.6</td>
<td>56.7</td>
<td>49.4</td>
</tr>
<tr>
<td>Lodging</td>
<td>89.5</td>
<td>84.6</td>
<td>87.4</td>
<td>92.1</td>
<td>85.9</td>
<td>71.3</td>
<td>79.4</td>
<td>82.9</td>
<td>79.8</td>
<td>74.9</td>
<td>85.4</td>
<td>91.8</td>
</tr>
<tr>
<td>Bowling Alleys</td>
<td>40.2</td>
<td>51.8</td>
<td>48.7</td>
<td>79.4</td>
<td>55.8</td>
<td>63.5</td>
<td>87.1</td>
<td>86.4</td>
<td>94.8</td>
<td>100.0</td>
<td>100.0</td>
<td>83.0</td>
</tr>
<tr>
<td>Amusement/Recreation</td>
<td>54.7</td>
<td>72.2</td>
<td>70.9</td>
<td>74.5</td>
<td>90.4</td>
<td>73.6</td>
<td>80.5</td>
<td>98.1</td>
<td>99.9</td>
<td>66.3</td>
<td>74.2</td>
<td>93.8</td>
</tr>
<tr>
<td>% of Total Empl. Comp.</td>
<td>1.3</td>
<td>1.9</td>
<td>2.1</td>
<td>2.0</td>
<td>3.3</td>
<td>1.3</td>
<td>1.4</td>
<td>4.3</td>
<td>3.3</td>
<td>2.6</td>
<td>5.2</td>
<td>6.9</td>
</tr>
<tr>
<td>Recreational Homes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Residential Structures</td>
<td>24.4</td>
<td>30.9</td>
<td>27.7</td>
<td>30.3</td>
<td>40.4</td>
<td>21.4</td>
<td>53.8</td>
<td>52.7</td>
<td>50.4</td>
<td>35.1</td>
<td>44.0</td>
<td>64.6</td>
</tr>
<tr>
<td>Maintenance, Residential</td>
<td>24.4</td>
<td>30.9</td>
<td>27.7</td>
<td>30.3</td>
<td>40.4</td>
<td>21.4</td>
<td>53.8</td>
<td>52.7</td>
<td>50.4</td>
<td>35.1</td>
<td>44.0</td>
<td>64.6</td>
</tr>
<tr>
<td>Food Stores</td>
<td>20.6</td>
<td>36.2</td>
<td>19.8</td>
<td>47.4</td>
<td>29.4</td>
<td>33.0</td>
<td>32.4</td>
<td>29.5</td>
<td>48.7</td>
<td>63.3</td>
<td>49.2</td>
<td>27.1</td>
</tr>
<tr>
<td>Real Estate</td>
<td>52.8</td>
<td>57.0</td>
<td>52.7</td>
<td>70.7</td>
<td>55.9</td>
<td>71.2</td>
<td>72.6</td>
<td>70.7</td>
<td>60.5</td>
<td>82.6</td>
<td>83.0</td>
<td>69.4</td>
</tr>
<tr>
<td>% of Total Empl. Comp.</td>
<td>0.9</td>
<td>1.4</td>
<td>0.8</td>
<td>1.6</td>
<td>1.3</td>
<td>0.9</td>
<td>1.7</td>
<td>1.6</td>
<td>1.7</td>
<td>2.6</td>
<td>3.2</td>
<td>2.6</td>
</tr>
</tbody>
</table>

1 The original benchmark data was used to calculate the percent of employee compensation for sectors and total.
2 Milwaukee County and the Menominee Indian Nation are excluded.
3 n refers to the number of counties in a cluster.
two types of visitors have on the local economy as well as the need for tourism-sensitive businesses to identify the nature of tourism demand.

IV. DISCUSSION

The minimum requirements procedure offers the opportunity to estimate tourism shares of business receipts in the absence of primary expenditure surveys. Once established, shares of employee compensation can serve as a reasonable proxy in estimating the share of local employment or sales attributable to tourism. These procedures allow researchers to pursue questions otherwise unavailable due to the time and expense of primary data gathering.

The major problem in using minimum requirements is that the technique is relatively crude and suffers from a number of theoretical flaws. The assumption that the minimum county in a region establishes the standard for consumption for all comparison counties is especially troublesome.

Unfortunately, it is not possible to directly compare the calculation of employee compensation attributable to tourism directly with another independent measure. In a survey completed under contract to the State of Wisconsin, the consulting firm of Davidson-Peterson Associates, Inc. (1991) estimated total state personal income (wages, salaries, and proprietary income) directly attributable to tourism spending to be $1.77 billion in 1990. This estimate compares to our state total estimate of employee compensation of $1.38 billion in 1991. The two estimates, however, are not strictly comparable in that the income measures were different (total income versus employee compensation) and the definition of tourism sectors also differed.

While not the primary focus of this research, it is worth noting that principal components analysis can be a useful tool in tourism planning and marketing. The tourism clusters generated for this analysis were highly detailed for the purpose of creating reference groups for which the minimum requirements calculations could be completed. However, the ability to classify counties by tourism structure can be very useful. Recognizing that the similarities within the 1-2 clusters, the 3-4-5 clusters, or the 10-11 clusters are much greater than their dissimilarities, opportunities for joint marketing may become evident. The contrasts between counties may also offer marketing opportunities. Many people use an urban center as a base from which to enjoy the outdoor amenities of surrounding counties. Marketing a group of dissimilar counties as a package suggests that an area offers a variety of amenities and experiences.

This information can also be used by the public sector to plan for service and facility needs. While most county officials already have knowledge of local resi-
dents and seasonal visitors, understanding more about the nature of the population is essential to planning the type and timing of service and facilities provision. Knowing something about neighboring counties can also lead to joint planning efforts for expensive items such as waste management facilities.

Knowledge of these clusters offers the opportunity to make valid comparisons between counties. Local officials and tourism organizations frequently evaluate their own performance by making comparisons between themselves and other areas they perceive as similar. Among the important benefits of PCA is establishing a fair, manageable and useful reference group for local development officials on which to model or conduct self-evaluations. This information will help make fair comparisons and encourage looking beyond immediate neighbors.

The limitations of the principal components analysis should be recognized. The components describe only the data provided. A different set of variables could lead to a different conceptualization. Perhaps the biggest deficiency in the data is the lack of information on demand structure. Whereas the tourism data used in the principal components analysis provides an indirect demand indicator, data on visitor usage would permit a weighting of the variables and might lead to more distinct principal components and county clusters. As presented, demand can only be inferred from tourism components and measures of supply.

V. SUMMARY AND CONCLUSION

This research has proposed a method for estimating the share of employee compensation attributable to out-of-region sources for local tourism-sensitive sectors. The procedure incorporates the use of readily available secondary data sources. While primary data is likely a superior method of generating information related to tourism demand, our method may serve as an alternative to conducting expensive and time-consuming primary research. Further, it has been proposed that the measurement of tourism activity should clearly distinguish out-of-region demand attributable to travelers/visitors versus seasonal home-owners.

The procedure consisted of four steps: (1) gathering county-level data related to tourism activity; (2) performing a principal components analysis to summarize the data; (3) conducting a cluster analysis to establish valid tourism-comparison regions; and (4) utilizing a minimum requirements procedure to estimate tourism shares of local employee compensation attributable to non-local demand.

Minimum requirements offers an additional option for estimating tourism activity from secondary data sources. The method provides a relatively quick and easy way to create estimates of tourism's contribution to employment and income generation. Combining the technique with data such as that available in the IM-
PLAN county data files permits very detailed analyses. However, the key to the best utilization of the technique is to make equitable comparisons. In the case of the tourism industry, an equitable comparison includes equivalence in total population and in tourism structure. Techniques such as principal components analysis and cluster analysis offer a viable means for identifying similarities in structure. More research will be necessary to determine what level of sectoral detail is required to obtain valid estimates of non-local tourism demand.

ENDNOTES

1. Principal components analysis is sensitive to data outliers. With counties serving as the unit of analysis, it was assumed that Milwaukee County (by a large margin, the largest metropolitan county in Wisconsin) and the Menominee (County) Indian Nation (a sovereign Native American nation) were two such outliers. Thus, they were not included in the analysis and can be considered tourism regions consisting of one county each.

2. Selection of tourism variables is an important consideration in any analysis attempting to characterize the tourism industry. Unfortunately, there are no standard or commonly accepted procedures for describing tourism. This research included variables used by other researchers and reported in the tourism literature that were otherwise logically thought to influence or represent a county's tourism activities based on familiarity with the state, the availability of published data, and erring on the side of being overly-inclusive. This admittedly ad hoc procedure is common in the literature and will likely continue to be used until researchers achieve consensus as to what constitutes the tourism industry and how it should be measured.

3. In principal components analysis, the structure of the variables matters. Representing a variable by a count or on a per capita basis will change the structure of the principal components. Where logical, the variables were presented on a per capita basis. Large per capita values imply facilities which may be in excess of values necessary to accommodate local demand.

4. The variables entered into the PCA may be in their original form or they may be standardized. Since PCA attempts to explain the total variance of a data set, those individual variables with the largest variance tend to dominate the component loadings. By standardizing the variables, the variance for each is matched to a common scale and each will have an equal chance to load onto a component. Both the raw data and the standardized variables were used in creating the principal components. The components created from the raw data were selected for use in the analysis because they presented a clearer representation of the fun-
damental nature of the component. Since PCA is simply a method for describing a data set, there is no "correct" component to be "discovered." The identification of principal components is a matter of interpretation by the researcher.

5. This three-component solution is more simplified than the results of previous research. Smith (1987) settled on four components, breaking outdoor tourism into outdoor recreation (which included a parks variable) and cottage/boating components. Similarly, he obtained two urban components: urban tourism and urban-fringe tourism. Lovingood and Mitchell (1989) also settled on four components: urban recreation-amenity rich, urban recreation-tourism, boating (fishing) and camping, and outdoor recreation-nature oriented. Finally, Lystad (1991) characterized Wisconsin counties and also settled on a four-component solution: urban tourism, outdoor recreation, nature tourism and rural tourism.

6. IMPLAN is a commercially available input-output analysis software package. The package includes data files containing a variety of county-level information.

7. The primary advantage of using the IMPLAN data over County Business Patterns (which serves as a source of IMPLAN data) are the estimates provided for nondisclosed sectors. In many smaller counties, nondisclosure for numerous sectors can leave large gaps in the analysis.

8. While labor costs make up a large proportion of total expenses for tourism-related businesses, a more complete understanding of the income-generation characteristics of tourism would also take into account other components of value added. Proprietary income, for example, is an important source of income for the many sole proprietorships included in the service and retail enterprises associated with tourism. This type of income could also be looked at using our approach.

9. Ullman (1968) suggested future work include the use of an "overlapping series" of sequential calculations as a means of including a larger range of regional population centers in calculation of minima.

10. Dropping a variable from the principal components analysis does not mean that it is unrelated or unimportant to tourism. Obviously, downhill ski runs and historic sites are important attractions. It simply means these variables do not contribute a great deal to the concepts represented by the principal component. The intent of the research is to characterize some of the fundamental structures associated with tourism in Wisconsin for the purpose of making meaningful comparisons between counties.
REFERENCES


Hotelling, H. "Analysis of a Complex of Statistical Variables into Principal Components." Journal of Educational Psychology. 24, No. 6 (1933): 417-441, and 24, No 7 (1933): 498-520.


