
Richard J. Cebula

Shirley and Philip Solomons Eminent Scholar, Economics Department, Armstrong Atlantic State University, Savannah, GA 31419; email: cebulari@mail.armstrong.edu

Sherri Todd

Undergraduate Economics Major, Economics Department, Armstrong Atlantic State University, Savannah, GA 31419

Abstract

This study empirically seeks to identify determinants of geographic living-cost differentials among counties in the state of Florida for the year 2003. The heteroskedasticity-corrected ordinary least squares estimates reveal that the cost of living in those counties is an increasing function of population size, per capita income or the percentage of households with an annual income in excess of $100,000, coastal location, and the square of the population density while being unaffected by the unemployment rate.

Keywords: Living-cost differentials; Population; Per capita income; Coastal location; Population density

JEL classification: R10; R11; R12

The authors are indebted to Dan Rickman and three anonymous referees for suggestions that improved the quality of this study.
1. INTRODUCTION

A number of studies have investigated determinants of geographic living-cost differentials in the U.S., including Cebula (1980, 1989); Cebula, Alexander, and Koch (1992); Cobas (1978); Ostrosky (1983); McMahon (1991); Nord (2000); Toma and Cebula (2000); and Kurre (2003). This study provides updated empirical evidence as to determinants of geographic living-cost differentials (L-CDs). Most of the related published research to date has tended to be national in scope, although there are exceptions, e.g., Kurre (2003). In contrast to most of the previous literature, this study focuses on a single state, the state of Florida. In addition, most of the existing published research has focused on metropolitan areas. By contrast, a potentially important aspect of the present study is its use of county data because by focusing on all of the counties in Florida, this study includes both rural areas and urban areas. Since the databases typically used for L-CD research do not include rural areas, this study seeks to add this dimension to L-CD research. In addition, in the interest of relevance, the study year is very recent, namely, 2003.

Two brief observations are in order before proceeding. First, studies of geographic L-CDs have typically dealt with metropolitan areas from across the nation. Consequently, many of them have included a variable intended to reflect the differential strength of labor unions as between the states where those metropolitan areas are located. This variable has taken the form of either a dummy variable to reflect states with or without “right-to-work” laws or a variable indicating the percentage of each state’s labor force that was unionized. Since the present study deals with simply one state, indeed a right-to-work law state, naturally no such variable is included. Second, previous related studies have typically dealt with a limited number of metropolitan areas, generally no more than 39 and sometimes as few as 24. Indeed, the sole exceptions to the smaller number of observations appear to be the studies by Kurre (2003), who deals with 67 Pennsylvania counties, and McMahon (1991), who deals with the 50 states. Thus, in contrast to nearly all of the related literature to date, this study deals with a larger data set, data for some 67 counties; hence, the present study carries with it the virtues accompanying a larger number of observations.

2. THE FRAMEWORK

Following previous research, including the recent study by Kurre (2003), the perspective underlying this analysis is that factors tending to elevate demand in a geographic area tend to elevate the overall level of prices in that area, whereas factors tending to elevate supply or reduce production costs in that area tend to lead to lower prices. We would expect that the greater the population (POP) in an area, the greater the overall demand for goods and services in that area and hence the higher the overall level of prices in that area, ceteris paribus. As observed by Kurre (2003), however, a greater effective demand may permit firms in some industries to benefit from scale economies that reduce production costs and – potentially – prices. Clearly, the latter effect may partially offset the former. Following Cebula (1980, 1989); Cebula, Alexander, and Koch (1992); Cobas
(1978); Ostrosky (1983); and Kurre (2003), however, it is expected that the former effect will dominate, i.e., that the cost of living in county $j$, $\text{COL}_j$, is an increasing function of $\text{POP}_j$.

Next, a higher per capita income in an area (PCI) is expected to raise the overall demand for goods and services in the area and hence the overall level of prices in the area, *ceteris paribus*. Paralleling Kurre (2003), we again note that to the extent that demand rises and increases the opportunity for at least some firms to benefit from scale economies that reduce costs and – potentially – prices, an offset to rising prices may occur. Nevertheless, once again based on Cebula (1980, 1989); Cebula, Alexander, and Koch (1992); Cobas (1978); Ostrosky (1983); and Kurre (2003), we would expect that $\text{COL}_j$ is directly a function of $\text{PCI}_j$. In the interest of possibly providing further insight into the determinants of L-CDs among Florida counties, a second, alternative measure of income is also considered in this study, although in a different (separate) estimate, namely, $\text{INCOVER}_j$. This income variable is defined as the percentage of households in county $j$ with an income in excess of $100,000 per year. Naturally, we would expect that $\text{COL}_j$ is an increasing function of $\text{INCOVER}_j$, *ceteris paribus*.

While $\text{POP}_j$ and $\text{PCI}_j$ (or $\text{INCOVER}_j$) may jointly help to explain L-CDs for counties in Florida, other factors very likely play roles as well. For example, a higher unemployment rate (UR) might act to reduce demand for goods and services, especially if the duration of unemployment is long. Thus, as in Kurre (2003), we would expect that the higher the unemployment rate in county $j$ ($\text{UR}_j$), the lower the cost of living in area $j$ ($\text{COL}_j$), *ceteris paribus*.

On the other hand, coastal location is deemed by many to be highly desirable. Indeed, we hypothesize here that for many households choosing to reside in Florida, there is an amenity value in closer proximity to “big water,” i.e., in this case, either the Gulf of Mexico or the Atlantic Ocean. Accordingly, we expect that coastal location ($\text{COAST}_j$) exercises a positive impact on $\text{COL}_j$ because many people may be willing to pay a premium for living in a coastal area, *ceteris paribus*. Nevertheless, it is interesting to note that for the U.S., Rappaport and Sachs (2003) conclude that coastal location induces more of a productivity effect (resulting, e.g., from agglomeration economies) than a household amenity effect. In this study, however, which is restricted to the state of Florida, whose population and economic growth have largely been based on its appeal as a place of retirement and/or as a place of residence for non-retired, we argue that the household amenity effect would be much stronger than in the nation as a whole. Accordingly, we expect that the conclusion in Rappaport and Sachs (2003) would have only limited applicability to the case of Florida.

Another possible factor in explaining geographic L-CDs is population density. It could be argued that increased population density in an area might increase productivity in that area through several avenues. Indeed, Ciccone and Hall (1996) found that about half of productivity differentials across the U.S. could be attributed to density differentials. Furthermore, the effect of population density on the cost of living in an area might
well be nonlinear. That is, over some range, increased density may reduce transit and other costs; whereas eventually, as density increases, the resulting increased congestion leads to elevated costs (and prices). These possibilities notwithstanding, however, this study follows the hypothesis proffered in Cebula (1980) and Kurre (2003) that the greater the population density in an area, \textit{ceteris paribus}, the greater the congestion, the higher transportation costs and labor costs, and perhaps even the less the opportunity for producing under conditions most conducive to scale economies, and thus the higher the cost of living in that area. Following Kurre (2003), the density variable adopted here is the population density in an area squared, \textit{DENSSQ}, which is expected, \textit{ceteris paribus}, to exercise a positive impact on the \textit{COL} in area \textit{j}. This specification is consistent with the hypothesis in Cebula (1980) and Kurre (2003) but at the same time is nonlinear.

3. EMPIRICAL MODEL AND ESTIMATES

Based on the eclectic model developed above, the reduced-form equations to be estimated are given as follows.

(1) \[
\text{COL}_j = a + b \text{POP}_j + c \text{PCI}_j + d \text{UR}_j + e \text{COAST}_j + f \text{DENSSQ}_j + u
\]

(2) \[
\text{COL}_j = g + h \text{POP}_j + i \text{INCOVER}_j + k \text{UR}_j + l \text{COAST}_j + m \text{DENSSQ}_j + u'
\]

where:

\[
\text{COL}_j = \text{the average cost of living for a four-person family in county } j \text{ in 2003, expressed as an index (100.00 = average)};
\]

\[
a, g = \text{constants};
\]

\[
\text{POP}_j = \text{the estimated total population in county } j, 2003;
\]

\[
\text{PCI}_j = \text{the 2001 per capita income in county } j;
\]

\[
\text{INCOVER}_j = \text{the percentage of households in county } j \text{ with an annual income in year 2000 in excess of $100,000};
\]

\[
\text{UR}_j = \text{the average percentage unemployment rate in county } j \text{ in the year 2002};
\]

\[
\text{COAST}_j = \text{a binary (dummy) variable for counties located on either the Gulf of Mexico or the Atlantic Ocean; for counties so located, COAST}_j = 1, and for other counties, COAST}_j = 0;
\]

\[
\text{DENSSQ}_j = \text{the squared value of the number of persons per square mile in county } j, 2003;
\]

\[
u, u' = \text{stochastic error terms}.
\]

Based on Section 2 above, it is expected that:
(3) \(b, h > 0, c, i > 0, d, k < 0, e, l > 0, f, m < 0\).

The data sources are as follows. For variables \(\text{COL}_j, \text{POP}_j, \text{UR}_j, \) and \(\text{PCI}_j\), the Florida Department of Education (2003): http://www.eflorida.com/countyprofiles/ranking.asp?level1=3&level2=127&level3=339; for variables \(\text{DENSSQ}_j\) and \(\text{INCOVER}_j\), the U.S. Department of Commerce (2003): http://quickfacts.census.gov/qfd/states/12000.html; and, finally, the variable \(\text{COAST}_j\) is a binary (0,1) variable. The \(\text{COL}_j\) data are generated annually by the Florida Department of Education. Construction of the \(\text{COL}_j\) index involves some 117 items that are priced regularly in all 67 of Florida’s counties. These data are used to adjust state aid to public schools in Florida (McMahon 1991; Kurre 2003). For a fuller discussion of and a critique of the \(\text{COL}_j\) series, see Denslow (1996).

Table 1 provides the correlation coefficients among the explanatory variables shown in equations (1) and (2). As shown, arguably only the correlation between \(\text{PCI}_j\) and \(\text{INCOVER}_j\) is problematic (+0.87); however, these two variables are treated as substitutes for one another and are therefore found in two separate estimations [(4) and (5)] below.

Estimating equations (1) and (2) by OLS using the White (1980) correction for heteroskedasticity yields equations (4) and (5), respectively:

(4) \[\text{COL}_j = 89.0 + 0.000006 \text{POP}_j + 0.0002 \text{PCI}_j + 0.2 \text{UR}_j + 0.4 \text{COAST}_j + 0.0006 \text{DENSSQ}_j \]
\[\text{(6.62)} \quad \text{(3.59)} \quad \text{(1.20)} \quad \text{(2.11)} \quad \text{(2.24)}\]
\[R^2 = 0.69, \text{adj}R^2 = 0.65, F = 21.6\]

(5) \[\text{COL}_j = 91.2 + 0.000006 \text{POP}_j + 0.72 \text{INCOVER}_j + 0.11 \text{UR}_j + 1.2 \text{COAST}_j + 0.0005 \text{DENSSQ}_j \]
\[\text{(9.54)} \quad \text{(3.95)} \quad \text{(0.65)} \quad \text{(2.38)} \quad \text{(2.60)}\]
\[R^2 = 0.65, \text{adj}R^2 = 0.61, F = 17.9\]

where terms in parentheses beneath coefficients are t-values.

Table 1

<table>
<thead>
<tr>
<th></th>
<th>POP</th>
<th>PCI</th>
<th>UR</th>
<th>COAST</th>
<th>DENSSQ</th>
<th>INCOVER</th>
<th>PCLPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>POP</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCI</td>
<td>0.41</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UR</td>
<td>0.08</td>
<td>-0.07</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COAST</td>
<td>0.34</td>
<td>0.49</td>
<td>-0.07</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DENSSQ</td>
<td>0.50</td>
<td>0.41</td>
<td>-0.05</td>
<td>0.08</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INCOVER</td>
<td>0.46</td>
<td>0.87</td>
<td>-0.17</td>
<td>0.40</td>
<td>0.43</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>PCLPT</td>
<td>0.33</td>
<td>0.63</td>
<td>0.25</td>
<td>0.59</td>
<td>0.27</td>
<td>0.57</td>
<td>1.00</td>
</tr>
</tbody>
</table>
In equations (4) and (5), eight of the 10 estimated coefficients are statistically significant at the five percent level or beyond with the expected signs. Only the unemployment rate fails to be statistically significant in the two equations; indeed, the sign on UR\textsubscript{j} is actually positive in both cases. In both cases, the F-statistic is significant at the one percent level, attesting to the overall strength of the models. The unadjusted coefficients of determination in equations (4) and (5) are approximately 0.69 and 0.65, respectively. Thus, the variables in equation (4) jointly explain nearly seven-tenths of the variation in the county cost of living, whereas nearly two-thirds of that variation is explained by the variables in equation (5).

The coefficients on variable POP\textsubscript{j} are both positive and significant at the one percent level. Thus, there is strong evidence that the greater the population in a county, the higher its overall cost of living, presumably because the greater population size implies a higher demand for goods and services within the county. The coefficients on variables PCI\textsubscript{j} and INCOVER\textsubscript{j} are also both positive and significant at the one percent level, implying that the higher the per capita income in a county or the higher the percentage of households having an annual income in excess of $100,000 in a county, the higher the demand for goods and services and hence the higher the overall level of prices in the county. The coefficient for variable COAST\textsubscript{j} is positive and significant at the five percent level in equation (4) and at the 2.5 percent level in equation (5), implying that the desirability of coastal location among Florida counties results in higher overall price levels in the coastal counties. Plausibly, this may very well be largely reflected in the higher price of housing in coastal areas as people capitalize a premium for the amenity value of coastal location into housing prices. Finally, the coefficients on the density variable, DENSSQ\textsubscript{j}, are positive and significant at the three and two percent levels, respectively, in equations (4) and (5). Thus, it appears that as the square of the population density rises in a Florida county, so too does the overall cost of living in that county.

4. CONCLUSION

This empirical note finds that in the state of Florida, the overall cost of living in county \( j \) in 2003 was an increasing function of the population in county \( j \), per capita income in county \( j \) or the percent of households in county \( j \) with an annual income in excess of $100,000, location of county \( j \) on either the Gulf of Mexico or the Atlantic Ocean, and the square of the population density in county \( j \).

In closing, there are two final comments. First, it may be of interest to observe that additional potential explanatory factors for the COL\textsubscript{j} among Florida counties have been investigated, including the per capita local property tax, PCLPT\textsubscript{j}. Harriss (1974), Tullock (1971), and Cebula (1980) have argued in ways that would support introducing such a variable into the analysis on theoretical grounds. However, although experimentation with just such a variable for COL\textsubscript{j} reveals that in many estimations it has a positive and statistically significant coefficient, its introduction introduces as many as three multicollinearity problems (as shown in the last row in Table 1). In all of these estimations,
although the population and income variables retain much (but not all) of their statistical significance, the coastal dummy becomes statistically insignificant.

Second, a comparison of the actual $\text{COL}_j$ and predicted $\text{COL}_j$ values for all 67 counties in the state was made. It was found that there were only three cases where the two numbers were noticeably different, those for Indian River County (with a residual of -2.925), Monroe County (with a residual of +10.236), and Washington County (with a residual of -4.905). In all 64 of the remaining cases, no residual of greater than the absolute value of (2.3) was found.

REFERENCES
