

***Demographic and Socioeconomic Change
in Appalachia Series--#11***

**DEFINING SUBREGIONS IN APPALACHIA:
ARE THERE BETTER ALTERNATIVES?**

by

**Kelvin M. Pollard
Population Reference Bureau**

June 2005



**POPULATION
REFERENCE
BUREAU**

*Celebrating 75 Years
1929-2004*



**APPALACHIAN
REGIONAL
COMMISSION**

DEFINING SUBREGIONS IN APPALACHIA: ARE THERE BETTER ALTERNATIVES? ¹

Kelvin Pollard
Population Reference Bureau
June 2005

The Issue

Since its creation in 1965, the Appalachian Regional Commission (ARC) has recognized the need to plan for the region's development through the creation of easily identifiable subregions. After several years of research, the ARC in 1975 settled on three such areas—northern, central, and southern subregions. The ARC added an overlaying “Highlands” subregion designed for the funding of recreation and conservation projects.²

Much has changed in Appalachia since the mid-1970s. In many socioeconomic dimensions, the region bears a significantly closer resemblance to the rest of the United States. Specifically, the number of counties classified by the ARC as “Distressed” has fallen from 161 in 1970 to 82 in 2004 (Fiscal Year 2005).³ Given these changes, does the current configuration of the Appalachian subregions still provide an accurate reflection of areas that are relatively homogenous in their economic status and development?

To answer the above question, Population Reference Bureau staff developed several alternative subregional configurations within Appalachia. For each of these alternatives, we examined the degree to which variation between subregions accounted for, or “explained,” the overall variation found among several economic variables from the 2000 census. Then, to see if what we found was a longstanding phenomenon, we also performed these tests among selected economic variables from the 1960 and 1970 censuses.

¹ The author wishes to acknowledge the work of John Haaga and XiaoHan Hu, who submitted an earlier version of this report to the Appalachian Regional Commission on June 30, 2004.

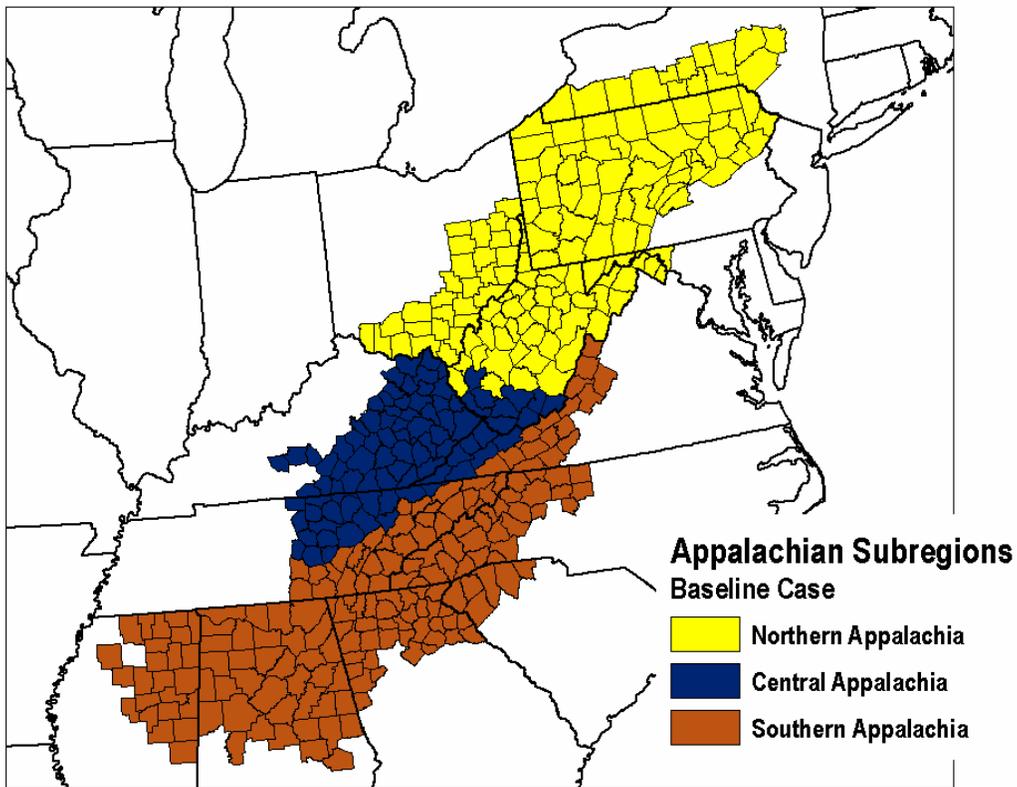
² Appalachian Regional Commission, “The New Appalachian Subregions and the Development Strategies,” *Appalachia* (Aug.-Sept. 1974): 10-26.

³ The number of Distressed counties in 1970 (which used similar criteria to what is used today) comes from Wood and Bischak. The 2004 number comes from the Appalachian Regional Commission, Online Resource Center, accessed online at www.arc.gov/index.do?nodeId=56, on June 27, 2005.

Appalachia's Current Subregional Configuration and the Five Alternatives

Baseline case. Under Appalachia's current subregional configuration (see Figure 1), the Appalachian counties of New York, Pennsylvania, Maryland, and Ohio, along with most of West Virginia, comprise northern Appalachia. Included among the counties in central Appalachia are the Mountaineer State's nine southernmost counties, as well as eastern Kentucky, Virginia's southwestern tip, and parts of Tennessee. Southern Appalachia contains western Virginia and East Tennessee, as well as the western Carolinas and the northern parts of Georgia, Alabama, and Mississippi.

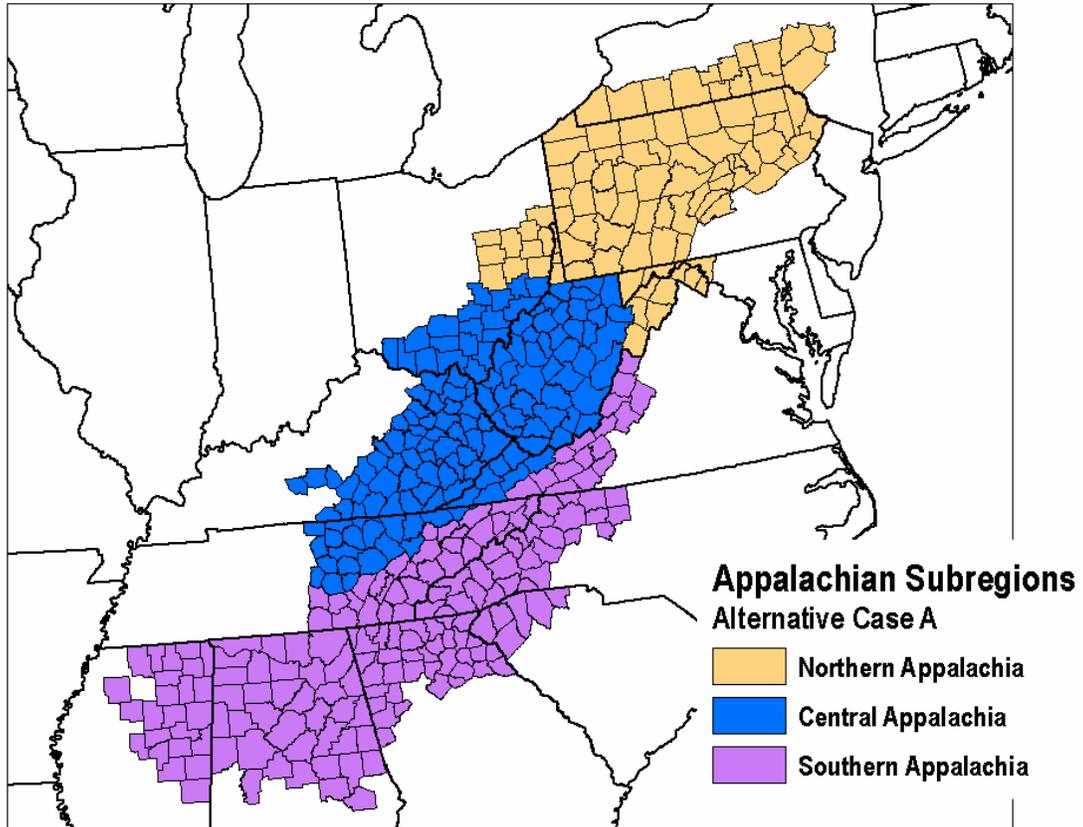
Figure 1
Map of Appalachian Subregions, As Currently Defined (Baseline Case)



Source: Appalachian Regional Commission.

Alternative Case A. The first of the five alternate configurations (see Figure 2) maintains the three subregions, but shrinks northern Appalachia by reclassifying most counties in Ohio and West Virginia that lie south of the extended Mason-Dixon line into the central portion. Counties in the Potomac Valley (western Maryland and the West Virginia panhandle) remain in northern Appalachia. Southern Appalachia's boundaries remain unchanged from the baseline case.

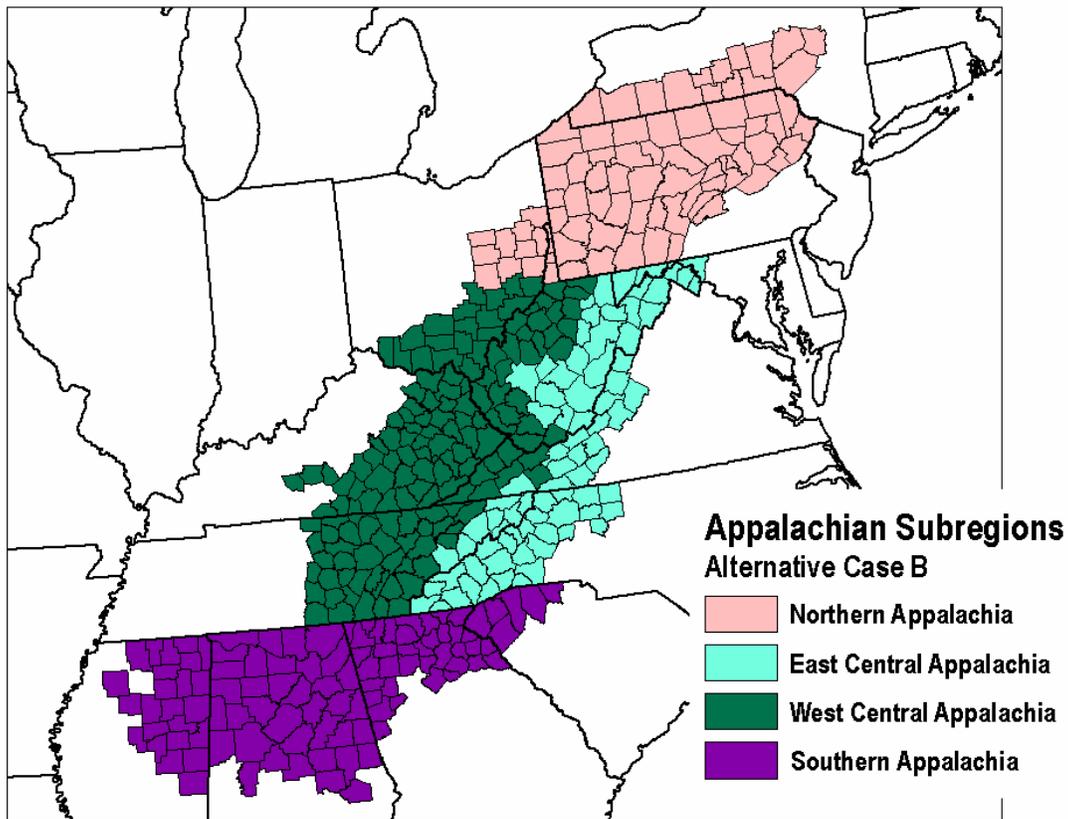
Figure 2
Map of Appalachian Subregions, Alternative Case A



Source: Population Reference Bureau, analysis of data from Appalachian Regional Commission.

Alternative Case B. This configuration divides Appalachia into four subregions (see Figure 3). Northern Appalachia includes only those counties lying north of the extended Mason-Dixon line, while southern Appalachia includes only the Appalachian counties in South Carolina, Georgia, Alabama, and Mississippi. A now-extended central Appalachia is then split into east-central and west-central subregions—using interstate highways (Interstates 77 and 79 in West Virginia, then Interstate 81 south of the I-77 intersection) to form the boundary. For those counties through which the highways pass, the preponderance of their land area determines whether they end up in east-central Appalachia or west-central Appalachia. The result separates western North Carolina and the Great Valley (which includes many recreational and retirement magnets, plus several university towns) from the West Virginia counties where mining and manufacturing predominate.

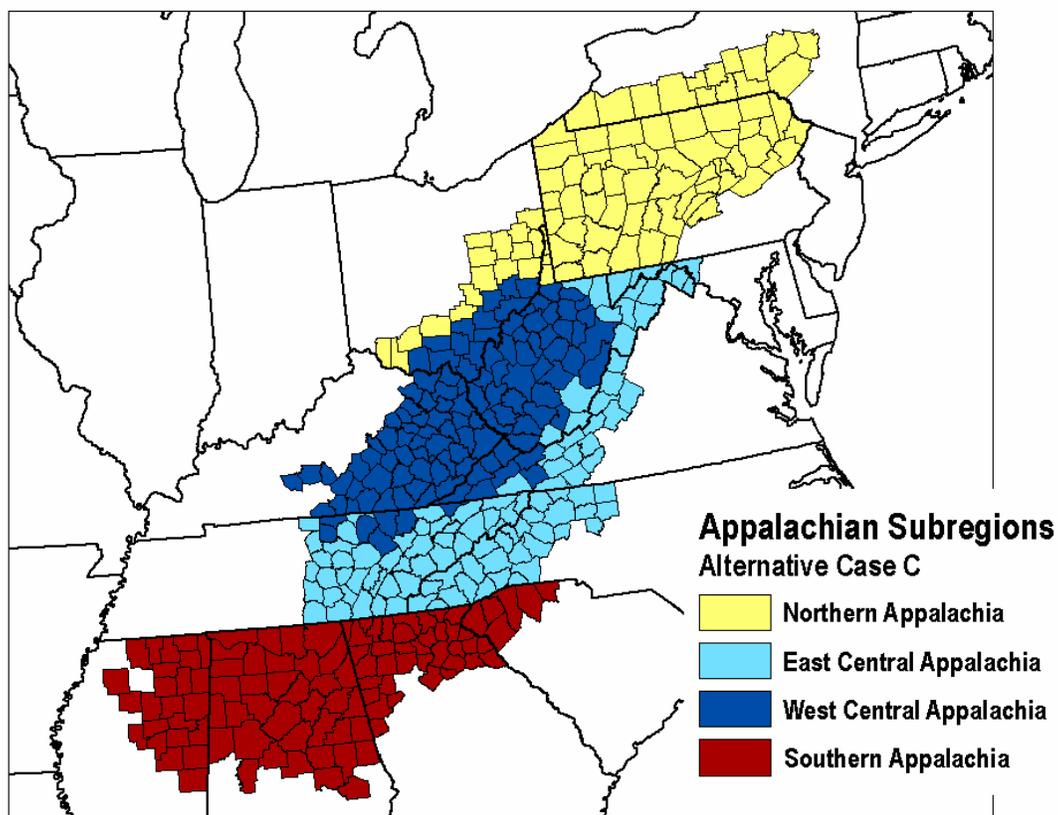
Figure 3
Map of Appalachian Subregions, Alternative Case B



Source: Population Reference Bureau, analysis of data from Appalachian Regional Commission.

Alternative Case C. Like Alternative Case B, this is a four-subregion configuration that maintains the southern boundaries of Alternative B (see Figure 4). Under this case, however, the dividing line between east-central Appalachia and west-central Appalachia are more freely drawn, placing a greater weight on a county's economic status in Fiscal Year 2003 (which had been used in the report series *Demographic and Socioeconomic Change in Appalachia*). This resulted in west-central Appalachia encompassing blocs of Distressed counties in West Virginia and northern Tennessee. The greater emphasis on economic status also comes into effect for the Ohio counties south of the extended Mason-Dixon line, as a handful of counties that were *not* classified as Distressed are returned to northern Appalachia, leaving a bloc of mostly Distressed counties in the west-central subregion.

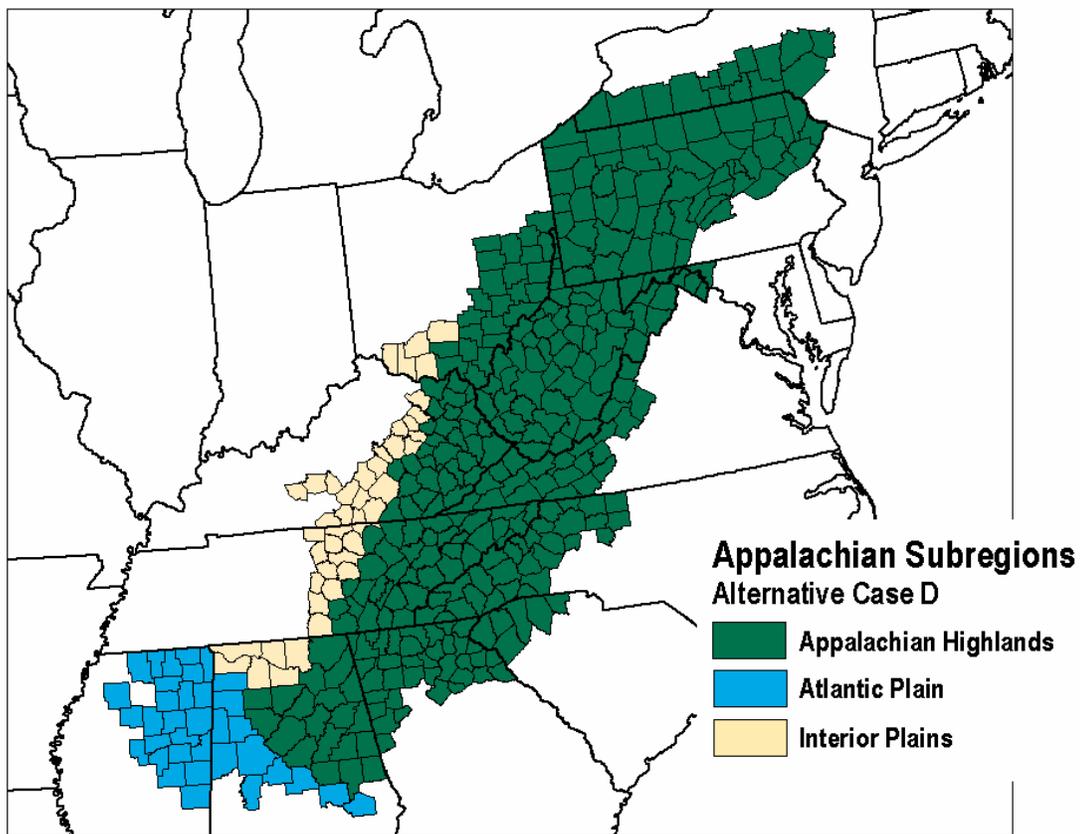
Figure 4
Map of Appalachian Subregions, Alternative Case C



Source: Population Reference Bureau, analysis of data from Appalachian Regional Commission.

Alternative Case D. Under this configuration (see Figure 5), Appalachia again is divided into three subregions—only in this case the boundaries are determined by the major physiographic divisions designated by the U.S. Geological Service (USGS). The USGS bases their system—used throughout the country—on geology type, surface texture, elevation, and other factors. The Appalachian region covers three physiographic divisions: the Appalachian Highlands (containing 331 of the 410 counties), the Atlantic Plain (which includes eastern Mississippi and central Alabama), and the Interior Plains (containing parts of Ohio, Kentucky, Tennessee, and northwestern Alabama).

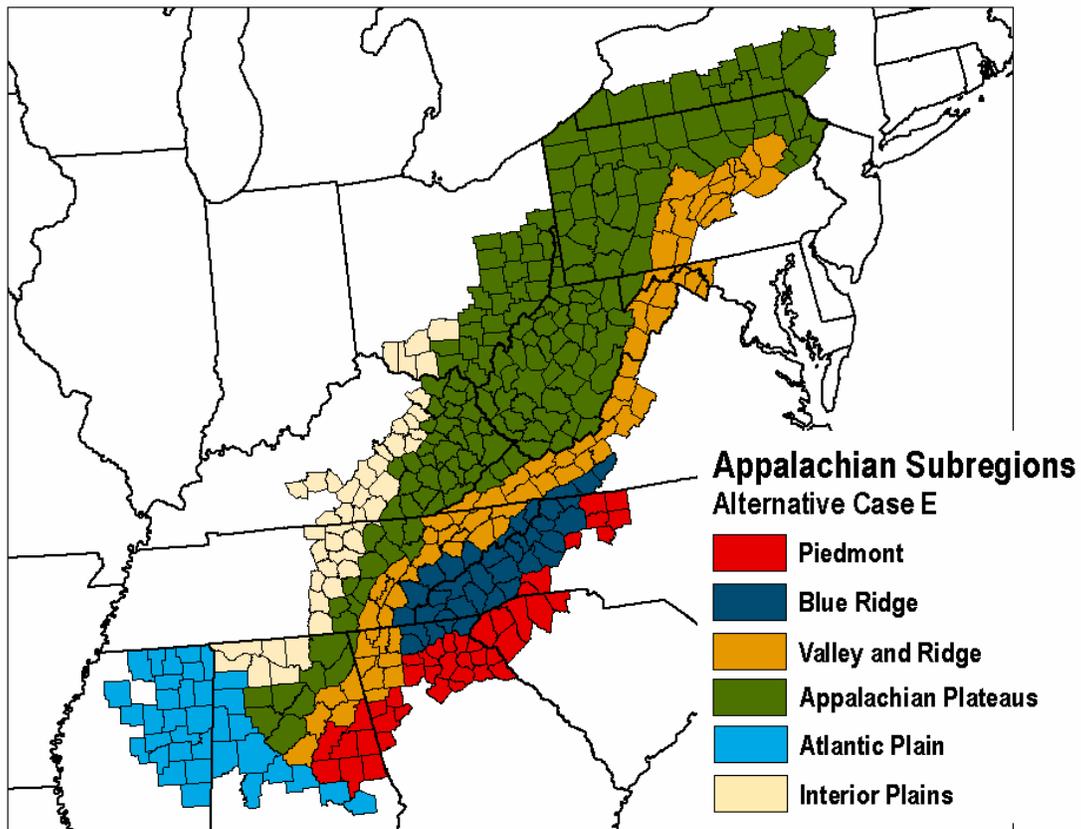
Figure 5
Map of Appalachian Subregions, Alternative Case D



Source: Population Reference Bureau, analysis of data from U.S. Geological Service.

Alternative Case E. As in Alternative D, this final case uses the USGS physiographic classifications in determining the subregions, but divides the Appalachian Highlands into its four physiographic “provinces”—Piedmont, Blue Ridge, Valley and Ridge, and Appalachian Plateaus (see Figure 6). The resulting configuration produces six subregions, reducing the imbalance that exists in Alternative D. (Yet it does not completely eliminate the imbalance: the Appalachian Plateaus subregion consists of 177 counties, while another 74 are in the Valley and Ridge subregion. None of the other subregions, by contrast, contains more than 43 counties.)

Figure 6
Map of Appalachian Subregions, Alternative Case E



Source: Population Reference Bureau, analysis of data U.S. Geological Service.

Results Using Variables from the 2000 Census

We first tested the six Appalachian configurations—the baseline case and the five alternatives—through the use of ordinary least-squares regression. Using one observation for each of Appalachia’s 410 counties, we generated R-square statistics for various economic variables from the 2000 census. (The R-square statistic is meant to show the portion of the total variation “explained” by the independent variables in a given model.) As dichotomous independent variables, we used the various subregions in each configuration—omitting one subregion as the reference variable.

We analyzed the following five variables from the 2000 census:

- Percent of persons in poverty, 1999
- Percent of men who are employed, 2000
- Percent of women who are employed, 2000
- Median household income, 1999
- Median family income, 1999

The results in Table 1 show that the baseline case explained more of the variation than three of the five alternative cases on each of the five variables. Under the current subregional classifications, regional differences accounted for nearly one-third of the variation of the poverty indicator and about one-fourth of the variability of most of the other items.

*Table 1
Percentage of Overall Variation Accounted for by Subregional Classifications of Appalachian Counties: Variables from the 2000 Census*

Variable	Baseline Case	Alternative Case A	Alternative Case B	Alternative Case C	Alternative Case D	Alternative Case E
Percent of persons in poverty, 1999	32.2	33.9	25.8	40.6	3.6	17.9
Percent of men who are employed, 2000	26.6	30.2	18.9	37.1	1.1	25.3
Percent of women who are employed, 2000	23.3	29.7	18.9	36.5	0.8	22.7
Median household income, 1999	26.8	26.8	20.4	31.4	2.1	20.5
Median family income, 1999	28.8	29.4	22.5	34.5	1.7	17.8

N=410.

Source: Population Reference Bureau, analysis of data from 2000 census.

Yet as well as the baseline case performed, two of the five alternatives did even better. Alternative Case A, the three-area configuration that has a smaller northern subregion, explained more of the total variation for four of the five variables and the same variation on the fifth (median household income). Performing even better was Alternative Case C (four subregions, with economic status a factor in separating an expanded central Appalachia into east-central and west-central portions). On all five variables, subregional differences accounted for more overall variation under Case C than under any other configuration. For Case C, in fact, subregional differences explained more than one-third of the variation for four variables and more than two-fifths of the region's overall poverty variation. Given that economic status was one criterion for Case C's regional boundaries, the above development probably is not surprising.

Performing more poorly than the other configurations (including Alternative Case B, which performed noticeably worse than the baseline or Alternative Cases A and C) were the two alternatives based on physiographic boundaries. Alternative Case D fared particularly badly—subregional boundaries accounted for less than 5 percent of the total variation of any of the five variables. One factor might be the fact that in Case D, 331 of the 410 counties were in the Appalachian Highlands subregion. Indeed, Case E, which split the Appalachian Highlands four ways (albeit with 177 counties in the Appalachian Plateaus), performed appreciably better. In Case E, subregional differences explained between 18 percent and 25 percent of the overall variation on any one variable. Still, those figures remain noticeably lower than either the baseline or the other three alternative cases.

Results Using Historical Variables

Based on the above analysis, Alternative C yielded noticeably better results than the baseline case, while Alternative A also yielded solid results. But has this always been the case? Would the use of variables from earlier censuses—variables from the period of the ARC’s creation—have yielded similar results? Again using ordinary least-squares regression, we tested for subregional variations among the following six variables from the 1960 and 1970 decennial censuses:⁴

- Percent of persons in poverty, 1959 (1960 census)
- Percent of employed persons in agriculture, forestry, and fisheries, 1960
- Percent of employed persons in mining, 1960
- Percent of employed persons in manufacturing, 1960
- Percent of employed persons in service industries, 1960
- Percent of persons in poverty, 1969 (1970 census)

As Table 2 shows, the baseline case explained more of the variation than any of the alternative cases on the 1959 poverty variable. It also performed well on the 1969 poverty and 1960 manufacturing variables, and explained more of the variation than at least three of the alternatives on each of other items.

Table 2
Percentage of Overall Variation Accounted for by Subregional Classifications of Appalachian Counties: Historical Census Variables

Variable	Baseline Case	Alternative Case A	Alternative Case B	Alternative Case C	Alternative Case D	Alternative Case E
Percent of persons in poverty, 1959	48.8	42.4	44.8	48.0	9.2	12.6
Percent of persons in agriculture, 1960	15.8	7.6	9.7	8.8	23.5	25.9
Percent of employed persons in mining, 1960	18.2	20.4	11.9	23.5	4.2	17.5
Percent of employed persons in manufacturing, 1960	28.0	29.6	20.6	32.6	4.3	19.0
Percent of persons in service industries, 1960	12.2	7.0	4.7	6.2	3.1	12.9
Percent of persons in poverty, 1969	37.6	33.9	33.4	39.9	10.3	13.1

N=410 (N=399 for 1959 poverty variable).

Source: Population Reference Bureau, analysis of data from 1960 and 1970 censuses, as provided by Appalachian Regional Commission.

⁴ The author thanks the Appalachian Regional Commission for providing the county-level data from the 1960 and 1970 censuses for 399 counties. Where available, PRB staff compiled data for the remaining counties.

As for the alternative configurations:

- Alternative Case C performed the best on three of the variables—1969 poverty, 1960 manufacturing, and 1960 mining (where subregional boundaries accounted for nearly one-fourth of the overall variation). And even though it was slightly less homogeneous than the baseline on the 1959 poverty variable, subregions accounted for nearly half of the variation. It performed significantly less well on the agricultural and service variables, where subregional differences accounted for less than 10 percent of the overall variation in the Appalachian region.
- Although Alternative Case A did not perform the best of any configuration, it did slightly better than the baseline on the manufacturing and mining variables.
- Alternative Cases D and E performed much better than the baseline on the 1960 agricultural variable; in each case, subregional differences explained roughly one-fourth of the variation for that indicator. Alternative E also performed best on the service variable. Just as with the 2000 variables, however, these two cases performed far worse than either the baseline or any of the other alternatives for most of the variables—reinforcing the problematic nature of using physiographic boundaries to determine subregions.

Discussion and Suggested Next Steps

Our analyses show that Alternative Case C—which divides Appalachia into four subregions (northern, east-central, and west-central, and southern)—performed best overall. That this turned out to be the result probably is not surprising. After all, economic status was one of the criteria used to determine the regional boundaries, leaving blocs of economically distressed counties in West Virginia, northern Tennessee, and southern Ohio in the west-central subregion. Keeping that in mind, one should not be surprised of its performance on the economic variables tested here—particularly the ones from the 2000 census.

At the other end of the spectrum, the poor performances of Alternative Cases D and E—and the relatively poor performance of Alternative Case B—suggest that physical boundaries (whether physiographic features or interstate highways) often do not make good boundaries for subregions that are designed to be economically homogenous. Physiographic features follow natural geologic features, while interstates often follow historic trade routes. Neither type correlates well with postindustrial economic characteristics.

So what's next? The above results suggest that Alternative C might prove a worthy alternative to the current configuration of Appalachian subregions. But they also show that the 1975 classification (i.e., the baseline case) has held up pretty well through the 2000 census. Therefore, before deciding if Alternative C would be enough of an improvement to justify changing from the baseline, further tests should be made comparing the two configurations.

These tests might include any (or all) of the following:

- Using weighted least squares, with county populations as weights.
- Using adjusted R-squares as the statistic for comparison (although preliminary tests suggest that it would not have made much of a difference).
- Adding more county-level economic variables—not necessarily from the decennial census. Some variables might include the poverty, per capita income, and unemployment data that the ARC uses to classify a county's economic development status. Others could involve the share of a county's workforce involved in such activities as agriculture, manufacturing, or specific service industries (e.g., recreation-based services).
- Performing spatially-based statistical analyses. For example, a cluster analysis (using ArcGIS software) could tell us if counties with similar values on a variety of indicators cluster in a way similar to any of the configurations examined here.