THE CENTER FOR
Rural Pennsylvania
A Legislative Agency of the Pennsylvania General Assembly

July 2000

Differences in the Cost of Living Across Pennsylvania's 67 Counties

200 North Third St., Suite 600, Harrisburg, PA 17101 • Phone (717) 787-9555 • Fax (717) 772-3587 • www.ruralpa.org

Differences in the Cost of Living Across Pennsylvania's 67 Counties

A report by
James A. Kurre, Ph.D.
Pennsylvania State University School of Business
Penn State Erie

This project was sponsored by a grant from the Center for Rural Pennsylvania, a legislative agency of the Pennsylvania General Assembly.

The Center for Rural Pennsylvania is a bipartisan, bicameral legislative agency that serves as a resource for rural policy within the Pennsylvania General Assembly. It was created in 1987 under Act 16, the Rural Revitalization Act, to promote and sustain the vitality of Pennsylvania's rural and small communities.

For more information, contact the Center for Rural Pennsylvania at 200 North Third St., Suite 600, Harrisburg, PA 17101, telephone (717) 787-9555, fax (717) 772-3587, or visit our website at www.ruralpa.org.

Table of Contents

List of Tables

List of Maps

Executive Summary

This study develops spatial cost of living estimates for each of the 67 counties of Pennsylvania for 1997. In addition to the overall cost of living, it generates indexes for each of the six component subindexes: groceries, housing, utilities, transportation, health care and miscellaneous goods and services. The indexes allow identification of high and low cost locations in the state, and permit measurement of the extent to which some areas are more expensive than others.

The major focus of the study is comparison of costs in urban counties with those in rural counties. A key finding of this study is that the cost of living (COL) continues to be lower in Pennsylvania's rural areas than in its urban areas. A typical urban Pennsylvania resident faced a cost that was approximately six percent higher than a typical rural resident in 1997. The cost advantage of rural areas applied to all of the subindexes, although to varying degrees. The report presents the actual index numbers for all counties for all components, as well as maps of the data.

The econometric model that forms the heart of this work allows identification of the determinants of higher cost. For the overall cost of living, these include higher density of population, population growth, higher government costs and higher electricity prices. The study also finds that the determinants of COL have not changed dramatically since the previous study based on 1989 data. Moreover, the general patterns of COL through space have continued, although the differential between urban and rural areas declined over the period.

I. INTRODUCTION

A. Why Do This Study?

Are Pennsylvania's rural areas low-cost places to live and work, compared to urban areas? Might lower costs of living represent a significant competitive advantage for rural areas? Are all types of goods and services less expensive in rural areas, or are some actually more costly there? If rural areas have a cost advantage, has it been consistent over time or not? And what causes costs to be high or low at various places? To answer these important questions, the Center for Rural Pennsylvania and researchers from Penn State Erie initiated this cost of living study.

This project builds on two key foundations: 1) data from the *Cost of Living Index (COLI)* of the American Chamber of Commerce Researchers Association (ACCRA); and 2) *The Cost of Living in Rural Pennsylvania* published by the Center for Rural Pennsylvania in 1992.

The *COLI* estimates the cost of a basket of 59 goods and services in over 300 urban areas each quarter. Unfortunately, rural areas typically are too small to qualify for participation in the ACCRA project. Therefore, estimation of the cost of living (COL) in rural areas is not as straightforward as it might seem. The Center for Rural Pennsylvania's 1992 study *The Cost of Living in Rural Pennsylvania* established an alternative approach to estimating cost of living in rural areas. Specifically, it created an econometric (economic-statistical) model that identified important determinants of the cost of living in a place, based on the *COLI* data for urban areas. The resulting model was then used to estimate COL numbers for each county of Pennsylvania. (That report also explains in more detail why the "direct pricing" approach–actually pricing the items in the market basket–doesn't work well in rural areas. See section I.B of that report, pp. 2-3.)

It has been over seven years since those estimates were made. Moreover, the estimates were based, for the most part, on 1989 data. They are in need of updating, but this is not as simple as plugging new data for Pennsylvania's counties into the model that was estimated in the 1992 study. Given the time that has passed, it is necessary to test whether the structure of that model is still applicable today. The world has changed in many important ways since then. Are the variables that were significant in the older study still the determinants of COL differences today? Even if the same variables are relevant, have their magnitudes (coefficients) changed over the period? A new model may be necessary to answer these questions.

A second round of estimates will also let us see if the cost of living patterns that were found in the initial report are consistent over time, or were merely a deviation applicable at one particular time point. Do that study's findings–that costs tend to be lower in rural areas–still hold up? Are rural Pennsylvania's lower costs a temporary aberration or a lasting advantage?

B. What Have Other Researchers Found Concerning the Cost Of Living?

There has been relatively little work done on spatial cost of living differences, at least compared to the amount of research done on many other topics. But there are some generally recognized findings in the literature. This section will review them before proceeding to the methodology for this project. Since the 1992 study reviewed the literature to about 1991, the current report will focus on findings since that time.

1) The cost of living (COL) is widely recognized as varying from place to place within the nation, but good data about COL are scarce.

Virtually every study that looks at spatial cost of living finds significant differences between places. For example, Walden (1997, p. 237) says: "It is now well-established that prices vary between states." McMahon (1991, p. 426) says: "Significant differences in the cost of living exist among different parts of the country, as well as among different rural and urban counties of the same state." Johnston, McKinney and Stark (1996, p. 568) say: "There is without doubt a need for data on regional variations in prices or costs of living." While some data are available, there is currently no official government program to provide information on this important topic. Deller, Shields and Tomberlin (1997, p. 110) say: "our findings…are… limited by the availability of good regional price data. …regional scientists need to develop a research program to address the shortcomings of our data."

The private sector has responded to some extent, though, to this lack of data. Runzheimer International, a management consulting firm with employment of 180, makes a substantial part of its revenues from estimating living and travel costs in different locations worldwide. Over 2,000 clients pay for this kind of spatial COL data from them. According to Runzheimer, over half of U.S. companies pay salary differentials based on geographic COL differences. (Runzheimer, 5/18/98) The Federal Government also adjusts salaries for some of its employees based on cost of living differences within the country. (See U.S. Office of Personnel Management, 1997 and 1999.)

The American Chamber of Commerce Researchers Association (ACCRA) has been publishing data on cost of living differences in American urban areas since 1968. The ACCRA data serve as the basis for this study and the previous work done for the Center for Rural Pennsylvania (Kurre, 1992). In addition, other firms are beginning to provide COL data. HomeFair (http://www2.homefair.com) offers comparison of costs in pairs of cities, including some foreign cities, in their "Salary Calculator." DataMasters (http://www.datamasters.com) offers similar COL comparisons for pairs of cities, based on ACCRA data supplemented with information on taxes, which are conspicuously absent from the ACCRA database. And ReloSmart at VirtualRelocation.com provides a similar service. These websites typically provide comparisons for pairs of cities, rather than data for all places at one time.

While ACCRA has been very open about its techniques and methodology, it is not clear if the other private firms are willing to share their information. Runzheimer has been unwilling to do so in the past.

2) Much of the COL research work is done as part of studies of other topics.

A key aspect of the literature in the field of spatial cost of living research is that relatively few researchers focus on measurement of COL for its own sake. Rather, they are working on other topics but find that geographical cost of living differences play a role in their area. They then attempt to add the issue into their work, which requires measurement of COL somehow. As a result, in order to find the literature on COL it is necessary to explore other areas of research. There are several research threads that include COL findings, including migration, poverty, convergence/divergence of incomes across regions, housing, and quality of life.

Since the 1992 study there have been a few projects that actually estimate spatial COL indexes. These include works by Voicu and Lahr (1999); Walden (1997); Johnston, McKinney and Stark (1996); Borooah et al (1996); Deller, Shields and Tomberlin (1996); Eberts and Schweitzer (1994); Cebula (1993); and Cebula, Alexander and Koch (1992). An important piece by McMahon was published late in 1991 and was not included in the previous work, so it will also be included here.

3) COL has an important impact on decisions by people, and on economic research findings.

Studies done in those related areas typically show that COL affects peoples' decisions. For example, migration research shows that migration decisions are based on salaries *adjusted for COL differences* between areas. In other words, people are not subject to "money illusion," and realize that a 10% higher salary in an area with a 10% higher cost of living does not represent a higher real income. Cebula (1993) found that this was true both for those over age 55[1] and those in the 20-40 age brackets.

Similarly, several studies including Wojan and Maung (1998), Deller, Shields and Tomberlin (1996), Bishop, Formby and Thistle (1994), and Eberts and Schweitzer (1994) all found that the analysis of interregional convergence of incomes, a key topic in regional economics, is affected in an important way by the inclusion of COL considerations. And Walden (1997) found that income differences among the 100 counties of North Carolina are quite a bit less after adjusting for cost of living differences. Similarly, DuMond, Hirsch and Macpherson (1998) found that differentials in wages across 185 metro areas were much smaller after (partial) adjustment for cost of living differences. (This study used the ACCRA data as its measure of cost of living.)

Walden and Newmark (1995) found that COL is an important factor accounting for differences in teachers' salaries across states. These differences are widely publicized, and could lead to incorrect policy choices if COL differences are not factored into decision-making. And COL differences are taken into account in allocating education funds to Florida counties (McMahon 1991, Florida Department of Education 1998).

In the poverty field, Pearce, Brooks and Outtz (1997) and Zimmerman and Garkovich (1998) have pointed out that the cost of self-sufficiency varies significantly by location, and thus COL should be taken into account in determining welfare benefits and other aid programs.

---

[1] This also ties into economic development since it implies that low COL places may be successful in attracting retirees who are looking for a place to make their fixed incomes go farther. This approach was suggested in the 1992 CRP report (Kurre, 1992, p. 58.) Although they only mention COL in passing, Fagan and Longino (1993) recommend retirees as a source of economic development for smaller communities. They also report that Pennsylvania is one of the states that suffers the most serious income losses as a result of retirees moving out of the state (#47 of the 50 states plus DC.)

4) There have been several different approaches to measuring COL, not all equally successful.

The literature identifies several different techniques for measuring spatial COL differences. These might be grouped into the following categories.

a) Primary price data collection.

The most obvious approach to constructing a cost of living index is to actually collect price data from all sites to be included in the study, for a representative basket of goods and services that a typical household would consume. This is the approach taken by the Bureau of Labor Statistics in calculating the Consumer Price Index.

Collection of such data by a private group (the Rewards Group) in Britain allowed Borooah et al. (1996) and Johnston, McKinney and Stark (1996) to construct spatial COL indexes for twelve regions of the United Kingdom. In the U.S., the American Chamber of Commerce Researchers Association also uses this approach, through an extensive network of volunteers who collect prices quarterly. The ACCRA data were the basis for the 1992 CRP report and are described there. (Kurre, 1992, pp. 17ff.)

Another American example is the Florida Price Level Index, generated annually by the Florida Department of Education (Denslow 1999, Florida Department of Education). The index uses 117 items that are priced regularly in all counties of the state. Data collection for this project is quite expensive, running hundreds of thousands of dollars per year. As mentioned above, McMahon (1991, p.430) notes that the resulting cost of living index has been used since 1974 to adjust state aid to Florida schools.

One popular approach has been to build on or update the last spatial COL indexes that were estimated by the U.S. Government, through its Bureau of Labor Statistics' Family Budget program in 1982 (U.S. Bureau of Labor Statistics, 1982). The Family Budgets are the most recent effort by the federal government to provide an official spatial cost of living index. Indexes were provided for 24 large metro areas, as well as an aggregate index for all metro areas and one for all nonmetro areas. The urban United States was the standard of reference at an index value of 100. Some researchers have updated the Family Budget data using price data from the Consumer Price Index for selected metro areas. This approach was used by McMahon (1991) and Eberts and Schweitzer (1994). However, caution is warranted in basing a current COL index on the Family Budget data for 1981. In fact, the last release of the Family Budget explains the termination of the program in this way: "The expenditure data on which the budgets are based are now 20 years old and continuation of the program would require a revision of concepts, more current expenditure data, and extensive collection of price data, for which funding was not available." (U.S. Bureau of Labor Statistics, 1982, p. 44). Clearly the BLS had some misgivings about the accuracy of the program at the time that it was discontinued. (See also Watts, 1980.)

A related approach, pursued primarily by sociologists, is to calculate the cost of self-sufficiency for families of various types in various places. This approach typically relates to 1996 changes in welfare legislation aimed at moving people off the welfare rolls and into jobs. The key question in these studies is "how much would a household have to earn in order to attain self-sufficiency?" Examples include Pearce, Brooks and Outtz (1997) and Zimmerman and Garkovich (1997). They vary from the studies mentioned above in that they define their market basket to focus on lower-income households, and use published data rather than actually collecting the data themselves. This results in assuming that some costs, such as food, are invariant over space.

As detailed in Kurre (1992, pp. 2-3), there are typically immense logistical problems and high costs to conducting a program of primary price collection, especially in rural areas. Unless such a data collection program already exists for some reason, the direct pricing technique is not a practical approach for most areas.

b) Estimation of a complete set of demand equations for all commodities in all places.

This approach is based solidly in microeconomic theory, and starts with an extensive set of demand equations, one for each commodity consumed. Theoretically, the equations would have interaction variables to allow for substitution and complementarity among all commodities consumed. However, this approach gets very complicated very quickly, and typically is not operational due to data requirements (Crawford, 1996).

c) Estimation of a regression model of the factors that explain (predict) COL in an area.

This technique involves identifying the factors that cause–or at least are correlated with–the cost of living differences between places. Starting with a COL database for a set of areas, the first step is to fit a structural, explanatory equation to those data. Then data for areas which did not participate in the initial study can be plugged into this equation to estimate their COL values.

This is the approach used in the 1992 COL study for the Center for Rural Pennsylvania (Kurre 1992). That study cited a number of other researchers who had used this approach, although typically not with the ACCRA data. Many of those who currently work in the field of spatial COL use a variation of this approach. This includes Walden (1997), Cebula (1993), Cebula, Alexander and Koch (1992), and McMahon (1991).

This approach relies on the pre-existence of a COL database, generated by one of the other techniques. Since there are few of these, relatively little work has been done in this field. However, there are three databases that have been used: the ACCRA database, the Florida Price Level Index, and an updated version of the BLS's Family Budgets. Among these three, the ACCRA database seems clearly superior. It is obviously more current than the Family Budget studies (which ended in 1982), and has broader coverage than the Florida data (which applies, of course, only to counties in Florida.) Given the frequency and currency of the ACCRA database, it is surprising that more researchers have not made use of it. The fact that it is published by a private firm may make it less widely known in the research community, or it may be the case that the cost of the data deters some researchers from using it. In any case, it has proven itself to be a solid foundation upon which to build a COL study such as ours.

d) Estimation of COL data from expenditure data.

This novel approach has been recently attempted by Voicu and Lahr (1999). It is based on the premise that variations in expenditures (i.e., prices times quantities) can be used to approximate prices alone. The problem is that quantities will probably vary as prices do. Economics 101’s fundamental Law of Demand states that when the price of a good goes up, people tend to buy less of it. As a result, changes in expenditures tend to reflect *both* price and quantity changes, rendering use of expenditure data suspect. Voicu and Lahr attempt to address this issue cleverly though use of commodity weights based on elasticities, but it is not clear that their technique measures price variation alone across regions. It may be measuring, at least partially, differences in the quantities of goods purchased, and therefore differences in the standard of living between areas.

e) Conclusions

A review of the recent literature has not exposed any blockbuster, breakthrough findings that have dramatically changed the direction of research in this field. On the contrary, it has tended to confirm that the approach of the 1992 CRP report was sound and remains the best technique for the task of estimating the cost of living for all 67 counties of the state.

C. Objectives of the Project

There are four key issues that this project seeks to address:

1) Rural vs. urban counties.

Are rural counties less expensive places to live than urban counties in Pennsylvania? If so, by how much?

2) Overall index vs. subindexes.

Do the same patterns exist for each of the subindexes (groceries, housing, utilities, transportation, health care, and miscellaneous goods and services) as for the overall cost of living? Do some types of products vary more in price across the state than others?

3) Consistency of the spatial patterns of COL through time.

Are the spatial patterns found in the 1992 study still applicable, or has something important changed?

4) Consistency of the determinants of COL through time.

Are the factors that determine cost of living differences across space stable over time? That is, are the structure and coefficients of the model in 1999 similar to those of 1992, or has something important changed?

D. Research Methodology

The basic premise underlying this project is that the cost of living in an area is caused by (or at least is related to) a relatively small number of determining factors. If those factors can be identified (using theory, previous research and current data) and measured (using statistical techniques), then it would be possible to specify an equation that will tell how COL is related to each of the relevant variables.

Relatively simple economic theory can be used to identify the determining factors, and these typically relate to supply and demand. The basic idea is relatively simple: factors that increase demand in an area will tend to cause local prices to be higher; those that tend to increase supply will cause prices to be lower.

This project uses the *ACCRA Cost of Living Index* data to identify the relevant factors and measure their impact. This analysis results in an equation or formula that will quantify the impact of each determinant, telling how much a larger population, or a faster growth rate will impact an area's cost of living, on average. With this knowledge, it is possible to estimate the COL for an area that did not participate in the *ACCRA Cost of Living Index* program. Given values for the identified determinant variables for any county of Pennsylvania will allow estimates of the COL to be made for that county.

The steps involved in the process are as follows:

1) Identify important determinants of the cost of living on a theoretical level.

The literature review and the previous study have helped to determine variables that might be expected to have an impact on a place's cost of living. These include the following:

a) Population

Population is considered to be a measure of demand. If there are more people competing to purchase the supply of a good or service, the price will be driven higher. If two areas are identical in every way except that one has more people, the COL would probably be higher in the area with more people. (But that assumption of "everything else equal" or *ceteris paribus* is hard to accommodate in practice. Fortunately, there are statistical techniques that let us account for the effects of different variables, and these will be explained below.)

On the other hand, a larger demand may mean that firms producing for local consumption can attain the scale of operations necessary to make use of large-scale production processes which are important in some industries. Larger scale may allow greater specialization in the production process with resulting lower costs per unit. This concept, economies of scale, is very important in some production processes, but certainly not in all. In fact, larger scale may lead to higher costs in some industries. An area's distinctive mix of industries—and the sizes of firms within those industries—will affect its cost of living, then.

Given the offsetting nature of these two factors, it is not immediately clear which would predominate. Previous research has tended to find that larger population has tended to be associated with higher costs of living, and that will be our working hypothesis.

b) Income

Income is expected to affect COL in much the same way as population. If two cities have the same population but one has a higher income per capita, the richer city would experience greater demand for most goods, with concomitant upward pressure on prices. Of course, the "economies of scale" effect could have the same impact here, as well. It is not immediately clear that all prices would be affected equally, however. Higher incomes may result in greater demand for luxuries rather than necessities, so that the price of shoes and canned peas might not be affected as much as the price of champagne and facials.

The impact of these two variables, population and income per capita, might be better measured as an aggregate—the total income of residents of the area. This study will consider the effect both ways—population and income per capita separately and in the form of aggregate income (population times income per capita.) One or the other of these approaches may be more fruitful, especially if population size and one of the income measures are so highly correlated that it is hard to distinguish between them statistically.

c) Density

Density of population—as distinct from sheer numbers of people—may also have an effect on COL. If two cities each have a million inhabitants, but one has them concentrated into a land area that is only a fourth of the other's, we may expect to see that city have greater congestion and resulting transportation problems, higher land costs, and worse problems with environmental issues. As a result, cost of living may be expected to be higher in that place.

d) Growth

Aside from sheer size—either in terms of people or the money they have to spend—the rate of change in that size may have an important effect on the area's COL. The logic of the market goes like this: as demand increases, price is driven up. This higher price means a higher profit for suppliers of the product involved. This may lead current producers to supply more of the good, and also induce new producers to enter the market. The resulting increase in supply can eventually bring price back down to its previous level. However, this won't happen overnight for most products; it takes some time for the supply response to occur. This means that cost of living may be higher in an area during a growth spurt than it will be after some time for adjustment has been allowed to pass. Alternatively, COL may be higher for areas with recent or more severe increases in population and income than in areas without them.

What time period is necessary for adjustment? That depends on the type of product we're talking about. The supply of milk or toilet paper in an area can be increased quickly by shipping more in from other areas. On the other hand, it takes significantly longer for increases in an area's housing stock. Since our key goal is to identify a formula to let us estimate COL for areas, it makes sense to experiment with various period lengths and see which seems most closely related.

Of course, although this discussion is cast in terms of growth, it applies equally well in reverse for decline in population or income. Forces that cause demand to decline should exert a negative influence on prices.

e) Utilities

The price of utilities in a local area can certainly have an impact on local cost of living. Electric and gas prices play their part in determining a homeowner's budget. However these prices are a little different from most local prices in that they have traditionally been heavily regulated in this country. Unlike typical market prices that are determined by supply and demand, electricity and gas prices reflect political influences in an important way. Political appointments to state public utility commissions can affect the cost of heating or cooling the homestead.

Utility costs can also affect local COL in an indirect way since they will have an impact on the cost of production of goods and services produced locally. Higher utility costs will mean higher costs of haircuts, education and car repairs.

Given their special status, it makes sense to include the price of local utilities as separate determinants, unlike other prices. We include the actual prices (per kilowatt hour or per thousand cubic feet of gas) rather than a measure like "average gas bill" since the latter would also include the quantity of the good purchased, which is not our intent.

f) Government

Amount, type and quality of services provided by local governments varies dramatically in this country. Like utility costs, these have both direct and indirect effects on the cost of living in a locality. A government that provides excellent education and efficient garbage collection saves residents the costs of providing similar services out of their own budgets. Similarly, effective police protection and local street maintenance helps keep costs low for local producers, resulting in lower prices for locally produced goods and services.

Of course, there's no such thing as a free lunch, and those services have to be paid for somehow. Measurement of local government efficiency must involve two components: the amount and quality of services provided, and the cost to local residents in terms of taxes and other charges by local government. Actually measuring these, especially the services provided, is a bit tricky, but the idea is simple conceptually: governments that provide better service for a dollar of cost (or more and better services for the same cost) are more efficient, and contribute to a lower cost of living.

g) Unemployment rate

Like income, an area's unemployment rate can provide a measure of the local economy. Higher unemployment in a locality can be expected to mean lower demand from area residents for most goods and services, and concomitant lower prices than in an area with an economy that is humming along near full employment. On the supply side, higher unemployment would mean less upward pressure on wages, keeping costs of production lower. To the extent that residents buy locally produced goods and services, this will contribute to a lower cost of living. Bartik (1991) explains the effects on COL in the short run and the long run.

h) Accessibility

It has been suggested that residents of rural areas incur greater travel costs than their urban counterparts. Except for those who work at home—say, on a family farm—the commute for rural workers will typically be longer, as will trips for shopping, to see a movie and to get health care. Money and time spent on travel cannot be spent for other things, and thus would seem to imply a higher cost of living in rural areas. A potential offset to this might be the increased congestion that plagues the urban resident. To take this into account, travel costs should be measured in time rather than miles. (McKean, Johnson and Walsh 1995)

i) Regional dummy variables

Aside from all the factors discussed above, there are certainly others that affect the COL in various places. It would be surprising indeed if a relative handful of variables, such as we've been discussing, would explain everything there is to know about cost of living differences from place to place. Although we expect to have identified the most important determinants, there are surely others that we haven't captured, or that affect some areas but not others.

One way to try to account for some of these effects is to introduce variables for each region. To the extent that there are factors unique to a region which are not included above, a "New England" or a "Rocky Mountain" variable may capture some of their effect, and help make our estimating process a little more accurate.

2) Specify the econometric model

After identifying the important determinants of COL based on the theory explained above and on the empirical work done by others on this topic, the second step in the process is to specify a model of the determinants of the cost of living. The model takes the form:

$$COL_i = f\left(\overset{-/+}{POP_i},\ \overset{+/-}{DENSITY_i},\ \overset{+}{INCOME\ PER\ CAP_i},\ \overset{+}{GROWTH\ RATE_i},\ \overset{+}{UTILITIES_i},\ \overset{+}{GCOST_i},\ \overset{-}{UNEMPT_i},\ \overset{-}{ACCESS_i},\ \overset{+/-}{REGION_i}\right)$$

where:

COL_i:	overall cost of living in area i;
POP_i:	population of area i;
$DENSITY_i$:	people per square mile in area i;
$INCOME\ PER\ CAP_i$:	income per capita of residents in area i;
$GROWTH\ RATE_i$:	rate of growth of population or aggregate income in area i;
$UTILITIES_i$:	utility rates (prices) in area i;
$GCOST_i$:	government cost per unit of service in area i;
$UNEMPT_i$:	unemployment rate of area i;
$ACCESS_i$:	accessibility of area i; and
$REGION_i$:	the region in which area i is located.

This model says that the overall cost of living in a community is a function of the community's population, population density, income per capita, growth rate, utility rates, efficiency of the government sector, unemployment rate, accessibility, and region of location. The sign above each variable indicates the type of effect it is expected to have on cost of living. For example, the positive sign above DENSITY means that a greater density is expected to cause the COL to be higher.

But we must also recognize that variables may not exhibit simple linear relationships with COL. For example, a doubling of population may not lead to a doubling of COL. A variable may have a weaker effect at first and then a stronger one, or vice versa. It could even be the case that a variable would have a negative effect at first, but then a positive one. Such would occur if increases in population at first result in lower costs, but successive population increments cause the beneficial effects to gradually diminish until they actually result in higher costs.

Our statistical techniques can capture these nonlinear effects if we include in the model both the variable (e.g., POP) and its square (POPSQ.) A negative sign on POP would indicate a decrease in COL as population increases, while a positive sign on POPSQ would mean a gradual diminution or reversal of the effect. Opposite signs would yield an opposite scenario, of course. To allow for nonlinear relationships, we will test squared versions of appropriate variables in the model.

3) Gather necessary data

The third step in the process is to gather all the data necessary for specifying and calibrating the model. Of course, lack of availability of key data can hamper our efforts.

a) Geographical level

Since the key goal of this project is to estimate COL for the counties of Pennsylvania, the county will be the basic geographical unit of this analysis. Whenever possible, data are gathered at the county level. Since the model will be calibrated using ACCRA *COLI* data from areas all over the country, it is necessary to have data for all variables in the model for all areas that participated in *COLI*, and not just for the counties of Pennsylvania.

In some cases, data are not available at the county level and some other level will be used as a proxy for county data. For example, electric and gas prices are only available at the state level. In such cases, we have to weigh the inaccuracy from using state-level (rather than county-level) data with the error that would be introduced by not having these variables in the model at all. At the other end of the geographic spectrum, the ACCRA data typically apply to cities or urban areas, which are usually smaller than counties. In these cases, we assume that the data apply to the whole county that contains that city or urban area.

Some areas of the country do not make use of counties for political units. Some of these have equivalent units, such as the parishes of Louisiana. In other areas, there are independent towns or cities that can be treated as equivalent to counties; Philadelphia is an example. In yet other places, there are multiple independent cities or towns that are much smaller than counties. Given their small size, they were considered not to be comparable to the counties that are the basis for our study. Including them would result in trying to compare apples with oranges—or maybe kumquats. As a result, some areas of the country had to be omitted from this study, typically in parts of New England and Virginia.

b) Time period

Ideally, we would like to estimate the model for the most recent time period available. ACCRA data are available quarterly, and the most recent issue at the time when the data were required was for the third quarter of 1999. Data for the explanatory variables were not always available for that recent a period, however. For several of the explanatory variables, 1997 was the most recent data available, and that is the year used to estimate the cost of living equation. In some cases it was necessary to use older data, but 1997 provides a good compromise between having the data for the explanatory variables match the period for the COL data, and using data for as recent a period as possible.

Given the determinants of cost of living patterns, it would not be expected that COL would vary dramatically from year to year, at least over relatively short periods. If this is the case, then results found based on 1997 data should still be relevant in 2000. Comparison of 1997 and 1999 data can shed some light on this issue. There were 260 areas that participated in the ACCRA database in both the third quarters of 1997 and 1999. Correlation between them for the overall cost of living index was .96 (out of a possible 1.00.) This means that areas that had a high COL in 1997 tended to still have a high COL in 1999, and use of 1997 data for this research should be acceptable.

c) Data used

This section presents basic facts about the data used in this study, along with other relevant information when it is called for. In cases in which the available data are somewhat less than ideal, more than one measure will be considered for the variable in an effort to increase the accuracy of our estimation process.

ACCRA Cost of Living Data

Unit of measure:	index; 100 = average of 321 urban areas participating in the study.
Source:	*ACCRA Cost of Living Index*, v. 30, #3.
Geographic level:	urban areas
Date:	Third quarter, 1997 (published January 1998)

The ACCRA *Cost of Living Index* measures the cost of buying a specific basket of goods and services in a large number of urban areas around North America each quarter. Volunteers from each area, often from Chambers of Commerce, agree to price the 59 or so items that the ACCRA team has identified as representing the standard of living of a midmanagement executive household. (Appendix B presents all the items and their relative weights in the market basket.) The midmanagement executive is a salaried employee, typically in the top quintile of income, earning perhaps double the average household income in the area. The total basket of goods and services to be priced is divided into six subcategories: groceries, housing, utilities, health care, transportation and miscellaneous. The ACCRA team chooses items to represent each category, based on the Bureau of Labor Statistics' 1992 Consumer Expenditure Survey, which is also the basis for the Consumer Price Index. These items are priced in each area, and the ACCRA home office checks the data and calculates the indexes.[2] The index for each quarter is unique, since a different group of areas typically participates in each edition. The index value of 100.0 represents the average cost of the specified basket for that quarter's particular group of participants.

The third quarter 1997 issue of the *Cost of Living Index,* which was used for this study, contained data for 321 urban areas. Not all were usable, however. Data for Saskatoon SK and four cities in Alaska[3] were discarded as being different in kind from the typical American areas. These may be subject to COL factors that are far different from the normal American county, and their inclusion would not help us understand the factors at work in Pennsylvania counties. It was necessary to eliminate eight other participants from the study since they were independent cities or units that could not be matched up with counties in the data.[4] Finally, there were five cases in which two different cities from the same county participated in the survey.[5] In these cases a weighted average of their ACCRA COL index numbers was employed, using the populations of the areas as weights.

This leaves 303 areas, or 94% of the survey's 321 participants, as the basis for the study. These 303 counties represent approximately 10% of the 3,142 counties in the nation, and 96,742,262 people in

[2] Complete details on the market basket and pricing procedures are presented in the *ACCRA Cost of Living Index Manual.*

[3] Anchorage, Fairbanks, Juneau and Kodiak.

[4] These were: Fitchburg-Leominster MA, Hampton Roads/Southeast VA, the Virginia peninsula, Richmond VA, Roanoke VA, Fredericksburg, Bristol VA, and Carson City NV.

[5] Phoenix and Scottsdale in Maricopa County, AZ; Palm Springs and Riverside City in Riverside County, CA; Fort Collins and Loveland in Larimer County, CO: Kansas City and Lee's Summit in Jackson County, MO; and Beaufort city and Hilton Head Island in Beaufort County, SC.

1997, or 36% of the total national population. This compares favorably with the 248 areas used in the 1992 Pennsylvania study for the Center for Rural Pennsylvania. Those 248 counties included less than 79 million people in 1989 or 31% of the national population.

Third quarter data are used because that quarter is closest to the midpoint of the year, which is most appropriate for use of annual data for the explanatory variables.

Population

Unit of measure: number of people (residents)

Source: Regional Economic Information System (REIS), Bureau of Economic Analysis, U.S. Department of Commerce. These data are originally from the Census Bureau, and represent midyear estimates.

Geographic level: counties

Date: 1997 and various other years

Income

A) Income per capita

Unit of measure: dollars

Source: Regional Economic Information System (REIS), Bureau of Economic Analysis, U.S. Department of Commerce.

Geographic level: counties

Date: 1997 and various other years

B) Aggregate Income

Unit of measure: dollars

Source: Regional Economic Information System (REIS), Bureau of Economic Analysis, U.S. Department of Commerce.

Geographic level: counties

Date: 1997 and various other years

Growth

A) Growth of Population

Unit of measure: percentage change

Source: calculated from Population data.

Geographic level: counties

Date: various periods, 1988 to 1997

B) Growth of Income

Unit of measure: percentage change

Source: calculated from Income data.

Geographic level: counties

Date: 1992 to 1997

While growth rates will be calculated for a five-year period as was done for the previous study, it may be useful to explore shorter periods of growth. Comparison of the effects of a short-run growth spurt with a more sustained rate of growth may expose some interesting COL effects. Sustained high growth may either cause more severe shortages of crucial goods like housing, leading to higher cost of living levels, or it could encourage more rapid expansion of the supply side of the market, helping to soften the price effects. The literature has not yet explored the impact of different periods of growth on COL. We will examine the growth rate of population from 1988 to 1997, and from 1996 to 1997, as well as the 1992 to 1997 period used in the previous study. Of course, "growth" rates may be negative as well as positive.

It should be mentioned that the "growth of income" variables are not adjusted for inflation. If the same rate of inflation (i.e., the national rate) affects all areas, inflation will not have a differential effect on some areas compared to others. For example, if incomes in all areas were to double over a period of years, they would not affect our spatial COL estimates; incomes that are twice as high would lead to a coefficient on the income variable that is half as great, yielding no differential effect on the COL estimates.

Density

Unit of measure: people per square mile

Source: calculated from Population and Land Area data.

Geographic level: counties

Date: 1997

Density was calculated by dividing the county's population for 1997 by its land area.

Land Area

Unit of measure: square miles

Source: ArcView GIS, Environmental Systems Research Institute, Inc.

Geographic level: counties

Date: "current"

Unemployment Rate

Unit of measure: percentage of labor force officially unemployed

Source: Local Area Unemployment Statistics program, Bureau of Labor Statistics, Department of Labor.

Geographic level: counties

Date: 1997 (annual average)

Accessibility

A) Average Travel Time to Work

Unit of measure: minutes, from home to work (one way)

Source: Calculated from data in the 1990 U.S. Census of Population and Housing, Summary Tape File 3, U.S. Census Bureau.

Geographic level: counties

Date: 1990 (released 1992)

This variable was calculated for each county in the nation from data available in the 1990 Census. Average travel time to work was derived by dividing aggregate travel time to work by the total number of workers in the county who were 16 years of age or older. Note that the total number of workers *includes* those who worked at home and thus had no travel time. It was decided to include these workers, since accessibility is not a problem for them and that should be reflected in the data.

Data on commuting patterns at the county level are gathered in the decennial censuses, meaning that the data are rather old. Unfortunately, there is sometimes a tradeoff between currency of the data and geographical detail. Can data from 1990 be trusted in a study of COL patterns in 1997? If commuting patterns have remained relatively stable over that period, the 1990 data would be useful. It is unclear if commuting patterns change rapidly over time, but it might be suspected that the incidence of telecommuting has increased significantly since 1990, due to advances in both telecommunications and computers. To deal with the issue of age of these data, we examined another accessibility variable.

B) Vehicle-miles Traveled per Person

Unit of measure:	Annual vehicle-miles of travel per person
Source:	*Highway Statistics 1997*, Office of Highway Information Management, Federal Highway Administration, U.S. Department of Transportation.
Geographic level:	states, with rural and urban breakdown
Date:	1997 (published October 1998)

This variable provides another measure of travel. The data are for 1997, seven years more recent than the Census data, but they are available by state and not by county. This measure has the benefit of including all travel, and not just travel to work like the previous variable.

To give a more accurate measure of accessibility, the number of vehicle-miles of travel can be adjusted by population for each state, to yield vehicle-miles traveled per person. Otherwise, states with a larger population would register a larger number of vehicle-miles simply because there are more people to travel. This would not provide the measure of personal accessibility that would help us explain cost of living differentials.

But these data do not include information on the number of people per vehicle, so they do not measure actual number of person-miles traveled, which is a more appropriate measure for our purposes. For example, if two people travel ten miles to work separately, that would amount to twenty vehicle-miles and twenty person-miles of travel. However, if two of their co-workers carpool, they would report twenty person-miles but only ten vehicle-miles. The vehicle-miles data would give an accurate picture of travel across the states IF the number of people per vehicle on a typical trip is the same across states. Given the differing availability of public transit such as buses and subways, this is unlikely.

Another problem with these data is that they cannot distinguish between travel by residents and those just passing through. The total number of miles reflects those of all vehicles on the highway, so they are not a good measure of travel by those who live in a location. This is an especially troublesome problem for the small states, like New Hampshire, where the ratio of pass-throughs to residents is high. As a result of these problems, it was decided that this database is not useful for the project at hand.

Utilities

A) Price of Electricity

Unit of measure: average revenue, cents per kilowatt hour, delivered to residential consumers

Source: *Electric Sales and Revenue 1997.* Energy Information Administration, U.S. Department of Energy.

Geographic level: states

Date: 1997 (published October 1998)

B) Price of Gas

Unit of measure: dollars per thousand cubic feet, average price of natural gas delivered to residential or commercial consumers (two variables)

Source: *Natural Gas Annual, 1997.* Energy Information Administration, U.S. Department of Energy.

Geographic level: states

Date: 1997 (October 1998)

The Energy Information Administration (EIA) of the U.S. Department of Energy provides a wealth of energy information, but it is typically at the national and state level. Below that level, they provide information by company (utility). But company service areas do not often coincide with county boundaries, and several companies may provide service in any one county. Data on rates by company are available, as well as lists of which companies sell their energy in each county. The missing piece of the puzzle is the number of customers each company has in a county, or its total sales of energy by county. If one of these were available, it would be possible to estimate a weighted average rate for electricity or gas in each county. Discussions with EIA officials and others in the industry (such as the National Energy Information Center) have led to the conclusion that, unfortunately, they are not available. As a result, state-level data will be used. The state average price of gas and electricity will be used for each county in the state.

It is anticipated that this variable will become less important in the future as competition continues to enter the market for production of energy. Economic development officials have commented that the price of power has become a non-issue in location choice for major industrial companies because of the deregulation of energy markets. To the extent that deregulation and competition have not yet reached many markets around the country, we can expect it to play a role in COL patterns in our data.

<u>Government Sector</u>

A) <u>Taxes and fees</u>

Unit of measure: dollars of revenue from "own" sources.

Source: *Compendium of Government Finances*, v.4, #5 of 1992 Census of Government, U.S. Census Bureau. (Table 50, Local Government Finances for Individual County Areas by State: 1991-92).

Geographic level: counties (i.e., all local governments in the county)

Date: 1991-92 (Published February 1997)

B) <u>Services</u>

Unit of measure: Full-time equivalent employment

Source: *Compendium of Public Employment*, vol. 3, #2 of 1992 Census of Government, U.S. Census Bureau. (Table 18, Local Government Employment and Payrolls in Individual County Areas, October).

Geographic level: counties (i.e., all local governments in the county)

Date: October 1992 (Published February 1997)

The variable for the government sector attempts to measure the efficiency of local governments in a county by comparing revenues raised from citizens with the employment by local government (as a proxy for services provided.) Note that, although the "geographical level" is the county, the data include information for *all* local governments in the county, including counties, municipalities, townships, special districts and school districts.

The revenue portion of this variable counts revenue raised from "own" sources by governments in a county. This includes all general tax revenues such as income, property and sales taxes as well as revenues for liquor stores, insurance and utility revenues, and charges and miscellaneous revenues. It attempts to measure dollars paid to local governments in the county for all purposes. Thus, if residents of a county pay fees or taxes to local governments for water, sewage treatment or garbage collection, those dollars will be included in the total, along with their income and property taxes. It does not include intergovernmental revenue—funds transferred to the local governments by states or the federal government. It is appropriate to exclude those funds since we are trying to measure the amount that local taxpayers have to pay for locally provided government goods and services. If a local government is successful in attracting lots of federal money to pay for local government employment and services, that government is more efficient from the point of view of a local resident. Its cost per worker will be lower, which may lead to lower costs of living generally in the local area.

Unfortunately, the most recent data available for this purpose is the 1992 Census of Government, which covers data for fiscal 1991-92. While the Census Bureau publishes an annual *Survey of State and Local Government Finances*, those data are for *levels* of government, not for areas. This means that the survey provides data for county governments, but not for all governments *in* a county. As a result, we must rely on the Census of Governments data, which provide the coverage and detail that we need, but are older than we'd like. This is, of course, preferable to simply excluding the variable from our analysis, which is the other option.

What do taxpayers get in return for their money? Ideally, we would like to include a measure of the quantity and quality of public services. Unfortunately, there are no generally accepted measures of this complicated bundle of goods and services. As a result, we use government employment as a proxy for the quantity and quality of government services. The underlying assumption is that areas with more workers provide more and better services. Of course, this is not a perfect measure since it implicitly assumes that government workers are equally productive across all areas.

In measuring government employment it is necessary to recognize that some governments may use a higher proportion of part-time workers than others. We adjust for this by using "full-time equivalent" (FTE) employment data. In this measure, two workers who work 20 hours per week each would be counted as a single full-time employee. Similarly, four workers who work ten hours per week would be counted as one full-time equivalent employee.

The variable used in the analysis is the ratio of the revenue and the employment variables, or government cost per (FTE) employee.

<u>Regional Dummy Variables</u>

Unit of measure: 0/1 (yes/no)

Source: calculated. Regional definitions come from U.S. Bureau of the Census; a recent map is given in *County and City Data Book 1994.*

Geographic level: counties (each is assigned to a region)

Date: Current regional definitions.

The regional dummy variables simply identify the Census region in which a county is located. There is one variable for each region: New England, Middle Atlantic, South Atlantic, East South Central, West South Central, East North Central, West North Central, Mountain, and Pacific. (Pennsylvania is in the Middle Atlantic region.) The variables themselves simply consist of zeroes and ones. For each county in the database, the regional dummy has a value of one if the county is in that region, and a zero otherwise.

Table 1 presents descriptive statistics for each of the basic independent variables, along with the county that exhibited the minimum and maximum value for each.

Table 1

Descriptive Statistics for Independent Variables for the 303 Counties in the Database

Independent Variable		Date	Unit	Mean	Standard Deviation	Minimum Value (Area)	Maximum Value (Area)
POP	Population	1997	number of residents	319,281	654,981	12,263 Gunnison CO	9,116,506 Los Angeles CA
POPGTH	% change in population	1988-97	percent	10.4	12.0	-16.0 Washington DC	78.0 Washington UT
		1992-97	percent	5.4	6.7	-9.4 Washington DC	42.2 Washington UT
		1996-97	percent	0.8	1.3	-2.5 Curry NM	6.2 Williamson TX
DEN	Density	1997	people per square mile	711	4,235	3.8 Gunnison CO	71,690 New York NY
INC	Income per capita	1997	dollars	22,929	5,020	12,005 Hidalgo TX	68,686 New York NY
INCGTH	% change in income per capita	1992-97	percent	23.9	4.4	11.2 Tulare CA	38.7 Benton OR
AGGINC	Aggregate income	1997	thousands of dollars	8,318,569	18,231,072	234,831 Gunnison CO	234,469,261 Los Angeles CA
AGGGTH	% change in aggregate income	1992-97	percent	23.9	4.4	11.2 Broome NY	82.2 Washington UT
ELEC	Electric rate	1997	cents per kilowatt-hour	7.9	1.8	5.0 WA	14.1 NY
GAS	Gas rate	1997	$ per thousand cubic feet	7.04	1.55	4.58 WY	11.90 FL
GCST	Government cost per worker	1991-92	thousands of dollars	45.9	217.5	13.2 Mayes OK	3,815.0 Washoe NV
UNEMR	Unemploymen t rate	1997	percent	4.8	2.5	1.2 Clay SD	27.9 Yuma AZ
ATT	Average travel time to work	1990	minutes, one way	17.9	3.2	11.0 Ford KS	30.7 Nassau NY

4) Estimate the model.

The final estimates for the overall cost of living were created by testing different models which used various combinations of the hypothesized variables listed above. Many different models were tested, using data for the 303 areas around the country for which ACCRA COL data and complete data for the independent (causal) variables were available. Those areas form the pool that lets us measure the effect of each variable. Once the model is calibrated using data for these 303 areas, data for each of the 67 counties of the state can be applied to the same model to get the COL estimates for each county.

The calibration process started with "all" variables in the model, and then winnowed and adjusted to get to the best model. In each case, the decision to keep or eliminate a variable was based on the sign of the coefficient compared to its expected sign, and statistical significance or lack thereof. Variables were added and dropped until the model that gave the best result with respect to overall fit (adjusted coefficient of determination or R^2) and economic logic was identified. The issue of inter-correlation of the variables (multicollinearity) was also taken into account in this process, ensuring that an equation did not contain two variables that are so closely related that statistical techniques are unable to distinguish their separate effects. (See the discussion of Multicollinearity in Appendix C for details.) Variables such as population and total income tend to exhibit this problem. In those cases, it is necessary to include one or the other of the variables in the equation, but not both.

The model that best measures total cost of living is:

$$\begin{aligned} TOTCOL = {} & 96.548 + 0.835\ POPGTH9697^{*} + 0.00299\ DEN^{*} - 1.81(e\text{-}8)\ DENSQ^{*} \\ & + 1.221\ ELEC^{*} + 0.00312\ GCOST^{*} + 0.330\ NE - 8.504\ MA^{*} - 9.500\ SA^{*} \\ & - 11.291\ ESC^{*} - 15.069\ WSC^{*} - 9.049\ ENC^{*} - 9.800\ WNC^{*} - 6.469\ MTN^{*} \end{aligned}$$

adjusted $R^2 = .787$ *statistically significant at the 5% level

This equation says that the total cost of living in an area (TOTCOL) is positively affected by population growth in the previous year (POPGTH9697) and the area's density of population (DEN and DENSQ). The density effect is nonlinear; higher density means a higher cost of living, but the effect increases at a decreasing rate with higher densities, reflecting the negative sign on the "density squared" variable.[6] Electric rates (ELEC) and higher government costs (GCOST) also cause higher costs of living, as had been hypothesized.

[6] Theoretically, a high enough density could actually mean a lower COL according to this formulation. However, this would not occur until a density of over 23 million people per square mile, which is significantly higher than in any area in the sample. For comparison, the highest density in the 303 areas of the study occurs in New York county at 71,690 people per square mile and the next highest is 10,958 in Suffolk MA. The average density for these 303 urban areas is just 711 people per square mile, and for Pennsylvania's 67 counties it is 428.

The last eight variables are so-called dummy variables for the regions of the country.[7] These are intended to capture the effect of other factors that vary by region but are not included explicitly in the model. The Pacific region is used as the standard of comparison (and so is omitted from the model), and all other areas except New England had lower costs than the Pacific region on average, as reflected in the negative signs on all their variables.

The asterisks indicate that all variables in this equation were statistically significant at the 5% level of significance (a typical standard used in statistical tests of this kind); in fact all are significant at the more stringent 2% level. The only exception is the dummy variable for the New England area; this implies that the average costs in New England are relatively the same as the costs in the Pacific region. The adjusted R^2, or coefficient of determination, tells us that the variables in this equation explain 78.7% of the variation of cost of living from place to place in the 303 areas included in the sample. Given the complexity of the forces that are involved to cause costs to be high in some places and low in other places, it is somewhat surprising that a relative handful of variables can capture so much of the process. The F-statistic (reported in Table 3) tests whether the variables in the model tell anything about the dependent variable, the COL in our case. A high value of the F-statistic, with a low probability value, means that the variables DO provide useful information about the dependent variable, and help explain its variation over space.[8]

There were ten areas of Pennsylvania that participated in the ACCRA Cost of Living study for the sample period. Comparison of the estimated COL values with their actual data will provide some sense of the accuracy of the estimates. The actual and estimated values are presented in Table 2 below. The means for the estimated and actual values are quite close, and the two sets of variables show a .940 correlation. While five of the ten estimated values are within 2% of the actual COL index, three others are off by more than 4%.

[7] The regions are the U.S. Census Bureau's standard regions: NE New England, MA Middle Atlantic (including Pennsylvania), SA South Atlantic, ESC East South Central, WSC West South Central, ENC East North Central, WNC West North Central, MTN Mountain, and PAC Pacific (which is excluded from the equation since it is the standard against which the others are compared.)

[8] Technically, the "low probability" means that there is a low probability that the null hypothesis—that all the coefficients are actually zero and the independent variables have no effect on COL—is true.

Table 2

Comparison of Estimates with ACCRA Data

ACCRA Area	County	Actual	Estimated
Altoona	Blair	103.5	100.6
Harrisburg	Dauphin	99.2	101.5
Erie	Erie	102.0	101.0
Chambersburg-Franklin County	Franklin	96.6	101.2
Indiana	Indiana	100.4	100.0
Lancaster	Lancaster	106.6	102.3
Wilkes-Barre	Luzerne	100.0	100.4
Williamsport-Lycoming County	Lycoming	100.5	100.1
Philadelphia	Philadelphia	122.5	127.6
Hanover	York	101.7	102.1
	Mean	103.3	103.7
	Std deviation	7.2	8.4
	Minimum	96.6	100.0
	Maximum	122.5	127.6
	Range	25.9	27.6

Table 3 provides the best estimating equation for each of the six subindexes of the cost of living. In each case, various models were tested to identify the one that works best for that particular subindex. Coefficients on all the included independent variables are statistically significant at the 5% level (except for a regional dummy occasionally), and R^2s range from .34 to .77 for the subindexes.

Table 3

Regression Results: Cost of Living Estimating Equations

Variable	Expected sign	**Total**	**Groceries**	**Housing**	**Utilities**	**Trans-portation**	**Health Care**	**Miscell-aneous**
Intercept		96.548*	97.145*	123.170*	-27.787	87.093*	104.292*	97.929*
Population 97	(+)						4.59E-06*	
Pop squared							-7.06E-13*	
Pop growth 92-97	+							
96-97	+	0.835*						0.508*
Density 97	(+)	0.00299*	0.000423*	0.00492*		0.00214*		
Density squared		-1.81E-08*				-2.72E-08*		
Income per cap 97	(+)							
Income per cap squared							2.03E-08*	6.38E-09*
Income per cap gth	+							-0.168*
Aggregate income	(+)		3.48E-08*		8.37E-08*			
Agg inc squared								
Agg inc growth 92-97	+							
Gov't cost per worker	(+)	0.00312*		0.00689*		0.208*		
Gov't cost squared						-5.33E-05*		
Electric rate 97	(+)	1.221*	0.928*		4.750*	1.241*		0.707*
Electric rate squared								
Gas rate 97	(+)				19.336*		1.350*	
Gas rate squared					-1.107*			
Mean travel time 1990	+							
Unemployment rate	-			-0.976*				
New England		0.330	0.289	1.943	36.369*	-5.546	-12.426*	0.012
Middle Atlantic		-8.504*	-5.045*	-18.207*	17.252*	-7.859*	-29.569*	-4.699*
South Atlantic		-9.498*	-4.784*	-24.971*	16.491*	-9.072*	-32.815*	-2.839*
East South Central		-11.291*	-5.721*	-30.786*	9.363*	-7.397*	-34.749*	-3.254*
West South Central		-15.069*	-12.643*	-34.968*	8.650*	-5.326*	-31.300*	-4.872*
East North Central		-9.049*	-4.336*	-22.006*	9.788*	-5.533*	-29.130*	-4.928*
West North Central		-9.801*	-6.828*	-25.815*	6.761*	-4.951*	-27.972*	-4.122*
Mountain		-6.469*	-1.501	-13.952*	5.554	0.840	-16.901*	-4.482*
Adjusted R-squared		0.787	0.570	0.765	0.606	0.434	0.688	0.341
F-statistic		87.0	37.4	90.6	39.7	18.8	56.4	14.0
Probability (F-statistic)		0.000	0.000	0.000	0.000	0.000	0.000	0.000

* Statistically significant at the 5% level.

Expected signs: (+) indicates that the variable is expected to exert a positive effect, although the combination of signs may vary when both the linear and squared variables are included in the model. See text for more information.

5) Use the equations to estimate COL for each county of Pennsylvania

Given these equations, it is possible to derive estimates of the COL for each county of Pennsylvania by plugging values for the independent variables for each county into each equation. The next sections present the results for the overall cost of living and each of the six subindexes.

Before presenting the data, however, we should offer a word of caution. These cost of living numbers are estimates, and like any estimates, they include some error. Error can enter through a number of doors. First of all, the estimates are based on other data and errors in any of those data will, of course, be reflected in the estimates. For this study, the ACCRA data are crucial. Any errors in the data collection procedures of the ACCRA field agents, or in calculation or publication by the ACCRA offices would obviously affect our estimates. The regression estimates make use of a large number of explanatory variables, all with their own data. All of these data come from reliable sources, but there are many thousands of individual data points involved and it would not be surprising if a few contained errors of some type. All of the data have been checked to rule out obvious errors, though.

Also, remember that the ACCRA data are based on the standard of living of a midmanagement executive household. This standard reflects a higher than average standard of living, so the estimates cannot claim to represent the costs faced by an average consumer in a county. Single consumers, childless couples and those with higher or lower incomes than the midmanagement family would presumably buy a different mix of goods and services than that family, and thus would experience somewhat different costs than those reported here.

Finally, the regression models estimated here do not account for all of the factors that affect COL; no single model could. To the extent that there are other important factors not accounted for in our models, the estimates presented below will be inaccurate. It is encouraging to note, however, that the model for the overall cost of living accounted for 78% of the variation in COL among the 303 places in the sample.

Despite these cautions, the numbers presented below should serve as useful guides to differences in the cost of living across Pennsylvania's counties. Because of the potential error, however, it is best not to assign great important to small differences in COL indexes between places.

a) Total expenditures

Table 5 presents the estimates for Total cost of living for each of Pennsylvania's 67 counties, in alphabetical order. Table 6 shows the same data, sorted from lowest to highest COL value, and Map 1 maps the results.

For the total budget, the cost of living averaged 101.5 for Pennsylvania's 67 counties, which is about 1.5% above the average for the 321 urban areas that participated in the *ACCRA Cost of Living Index* for the third quarter of 1997. The COL index numbers ranged from a low of 99.7 in McKean, Clarion, Elk, Warren and Cameron counties to a high of 127.6 in Philadelphia. This means that it cost about 28% more to live in Philadelphia, overall, than in the state's least expensive counties in 1997.

Map 1 shows that COL tended to be highest in the eastern part of the state, especially in the southeastern portion. Philadelphia and its three neighboring counties comprise four of the five highest-cost counties of the state, with Allegheny county rounding out the top five. The western frontier counties also tended to be higher cost, although not nearly as high cost as the southeast. The northern tier of the state, with the exceptions of the far east and west, were all low cost places to live, as were most of the rural counties of central Pennsylvania.

These data also confirm the results of the earlier study, that the cost of living tends to be lower in Pennsylvania's rural counties than in its urban counties. The average for Pennsylvania's 25 urban counties was 103.0, 2.4% more expensive than the average of its 42 rural counties.

However, the averages cited above are simple averages of the index values for the counties; they simply sum the index values for all 67 counties and divide by 67. These simple averages do not reflect the fact that the high-cost urban areas are much more heavily populated than low-cost rural areas. If we want to determine the COL experienced by the average urban resident as compared to the average rural resident, we need to factor in the number of people in each county. This can be done by calculating a weighted average COL, with the weights being the population in each county. Table 4 presents the results.

Table 4

Comparison of Simple and Population-Weighted Average

Cost of Living Index Values for Rural and Urban Counties

	Simple Average	Population Weighted Average
All counties	101.5	105.5
42 rural counties	100.6	100.7
25 urban counties	103.0	106.7
Difference between urban and rural	2.4%	6.0%

These numbers indicate that the average resident of Pennsylvania experiences a COL index of approximately 105.5, or about 5.5% higher than the average of the 321 urban areas that participated in the ACCRA Cost of Living Index for the third quarter of 1997. It is not surprising that the weighted average is higher than the simple average, since it reflects the fact that the very high COL of Philadelphia applies to nearly a million and a half people, while the lower costs of the state's rural areas apply to much smaller numbers of people. Since the higher cost counties generally tend to be more populous than the lower cost counties, it is not surprising that the weighted averages are higher than the simple averages, even for the rural areas.

A key conclusion of this study is that the average urban resident in Pennsylvania experiences a cost of living that is about six percent higher than that of the average rural resident.

Table 5
Cost of Living Estimates for Pennsylvania Counties: Total Expenditures

County	Rural	COL Index
Adams	R	101.9
Allegheny		104.6
Armstrong	R	100.2
Beaver		101.0
Bedford	R	100.4
Berks		102.0
Blair		100.6
Bradford	R	100.5
Bucks		103.5
Butler	R	101.7
Cambria		100.2
Cameron	R	99.7
Carbon		101.0
Centre		101.1
Chester		103.1
Clarion	R	99.7
Clearfield	R	100.4
Clinton	R	100.1
Columbia	R	100.3
Crawford	R	100.6
Cumberland		101.7
Dauphin		101.5
Delaware		108.4
Elk	R	99.7
Erie		101.0
Fayette	R	100.5
Forest	R	101.0
Franklin	R	101.2
Fulton	R	101.2
Greene	R	100.1
Huntingdon	R	100.2
Indiana	R	100.0
Jefferson	R	100.2
Juniata	R	100.9

County	Rural	COL Index
Lackawanna		100.7
Lancaster		102.3
Lawrence	R	100.6
Lebanon	R	101.4
Lehigh		103.1
Luzerne		100.4
Lycoming		100.1
McKean	R	99.7
Mercer		100.6
Mifflin	R	100.7
Monroe	R	103.2
Montgomery		105.0
Montour	R	100.6
Northampton		102.5
Northumberland	R	100.0
Perry	R	101.3
Philadelphia		127.6
Pike	R	103.2
Potter	R	100.2
Schuylkill	R	100.0
Snyder	R	100.6
Somerset	R	100.5
Sullivan	R	100.1
Susquehanna	R	100.5
Tioga	R	100.4
Union	R	100.5
Venango	R	99.9
Warren	R	99.7
Washington		100.7
Wayne	R	101.2
Westmoreland		100.9
Wyoming	R	99.8
York		102.1

Descriptive Statistics

	Total	Rural	Urban
Mean:	101.5	100.6	103.0
Standard deviation:	3.6	0.8	5.4
Minimum:	99.7	99.7	100.1
Maximum:	127.6	103.2	127.6

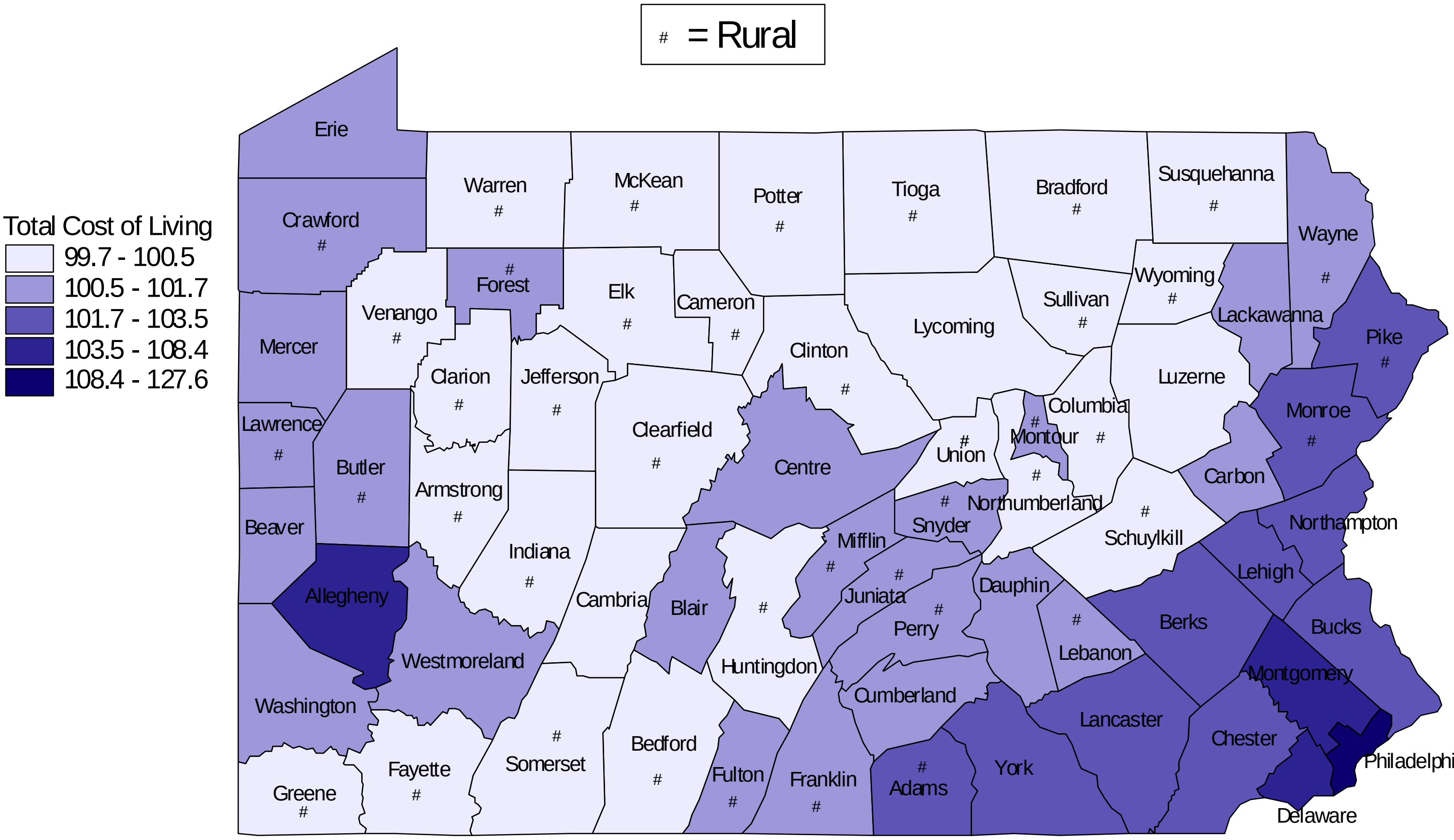
100.0 = average of 321 areas participating in ACCRA *Cost of Living Index*, third quarter, 1997.

Table 6
Total Cost of Living Ranked from Lowest to Highest

County	Rural	COL Index
1 McKean	R	99.7
2 Clarion	R	99.7
3 Elk	R	99.7
4 Warren	R	99.7
5 Cameron	R	99.7
6 Wyoming	R	99.8
7 Venango	R	99.9
8 Indiana	R	100.0
9 Schuylkill	R	100.0
10 Northumberland	R	100.0
11 Clinton	R	100.1
12 Lycoming		100.1
13 Greene	R	100.1
14 Sullivan	R	100.1
15 Armstrong	R	100.2
16 Huntingdon	R	100.2
17 Cambria		100.2
18 Potter	R	100.2
19 Jefferson	R	100.2
20 Columbia	R	100.3
21 Tioga	R	100.4
22 Luzerne		100.4
23 Bedford	R	100.4
24 Clearfield	R	100.4
25 Somerset	R	100.5
26 Bradford	R	100.5
27 Union	R	100.5
28 Susquehanna	R	100.5
29 Fayette	R	100.5
30 Crawford	R	100.6
31 Blair		100.6
32 Montour	R	100.6
33 Snyder	R	100.6
34 Mercer		100.6

County	Rural	COL Index
35 Lawrence	R	100.6
36 Washington		100.7
37 Mifflin	R	100.7
38 Lackawanna		100.7
39 Juniata	R	100.9
40 Westmoreland		100.9
41 Forest	R	101.0
42 Carbon		101.0
43 Erie		101.0
44 Beaver		101.0
45 Centre		101.1
46 Wayne	R	101.2
47 Franklin	R	101.2
48 Fulton	R	101.2
49 Perry	R	101.3
50 Lebanon	R	101.4
51 Dauphin		101.5
52 Butler	R	101.7
53 Cumberland		101.7
54 Adams	R	101.9
55 Berks		102.0
56 York		102.1
57 Lancaster		102.3
58 Northampton		102.5
59 Lehigh		103.1
60 Chester		103.1
61 Monroe	R	103.2
62 Pike	R	103.2
63 Bucks		103.5
64 Allegheny		104.6
65 Montgomery		105.0
66 Delaware		108.4
67 Philadelphia		127.6

100.0 = average of 321 areas participating in ACCRA *Cost of Living Index*, third quarter, 1997.



Map 1: Total Cost of Living

b) Subindexes

This section presents information for each of the six subindexes of the overall cost of living. Table 8 presents data for each of the six, as well as the overall cost of living in each of Pennsylvania's 67 counties.

The subindexes tended to follow the general spatial pattern of the overall cost of living. Table 7 shows that all of the subindexes were positively correlated with the overall cost of living, meaning that counties with a high overall cost of living also tended to have high costs in each subindex. (Recall that a correlation coefficient of 1.000 between two indexes would indicate that they vary together perfectly across counties; a coefficient of 0.000 would indicate no relationship between them whatsoever.) Within the subindexes, housing and groceries tended to most closely follow the overall pattern, with miscellaneous goods and health care being less closely related.

Table 7
Correlation of Estimated Subindexes

	Total	Groceries	Housing	Utilities	Transptn	Health Care	Misc.
Total	1.000						
Groceries	.958	1.000					
Housing	.963	.965	1.000				
Utilities	.703	.851	.757	1.000			
Transportation	.802	.825	.843	.746	1.000		
Health Care	.384	.522	.488	.786	.655	1.000	
Miscellaneous	.285	.327	.321	.575	.535	.830	1.000

1) Groceries Subindex

The groceries subindex exhibited relatively little variation across counties in Pennsylvania. This is also true for the nation as a whole; the groceries and miscellaneous goods subindexes had the least amount of variation across the 321 areas in the *ACCRA Cost of Living Index* for the third quarter of 1997. Given the easy portability of these kinds of goods, it is not surprising that they would be quickly transported from lower to higher cost places, tending to reduce spatial price differences.

The most expensive county in the state, Philadelphia, had grocery costs that were only about 5.4% higher than those of the lowest cost counties. In fact, the 24 lowest cost counties had index numbers that were within a range from 100.83 to 100.93, a single percentage point. Map 2 shows that the higher costs tended to occur in the southeastern part of the state as well as Allegheny county.

Even with the relatively small variation in grocery costs across counties, the rural counties had a slightly lower average grocery cost than the urban counties. Rural counties averaged 100.9 versus 101.6 in urban counties, a difference of about 0.7%.

Table 8
Cost of Living Estimates, Total and All Subindexes, for Pennsylvania Counties, 1997

County	Rural	Total	Groceries	Housing	Utilities	Transportation	Health Care	Miscellaneous
Adams	R	101.9	101.0	101.4	120.9	100.0	95.8	100.8
Allegheny		104.6	102.9	109.5	123.9	105.2	108.2	101.3
Armstrong	R	100.2	100.9	98.3	120.9	98.6	94.5	98.6
Beaver		101.0	101.1	102.3	121.1	100.7	96.4	98.5
Bedford	R	100.4	100.9	98.2	120.8	96.8	92.2	98.4
Berks		102.0	101.3	103.1	121.5	101.3	101.3	101.1
Blair		100.6	101.0	100.9	121.0	98.0	95.1	98.7
Bradford	R	100.5	100.9	100.0	120.8	99.0	93.3	99.4
Bucks		103.5	101.8	105.9	122.2	105.0	107.6	102.3
Butler	R	101.7	101.0	101.5	121.1	100.2	96.9	100.9
Cambria		100.2	101.0	98.5	121.0	98.1	94.6	98.8
Cameron	R	99.7	100.8	98.7	120.8	97.9	96.1	98.2
Carbon		101.0	100.9	98.8	120.8	98.6	94.9	99.2
Centre		101.1	101.0	102.9	121.0	100.4	95.5	99.4
Chester		103.1	101.6	104.9	122.1	104.4	118.2	105.7
Clarion	R	99.7	100.9	99.1	120.8	96.8	93.8	97.8
Clearfield	R	100.4	100.9	97.5	120.9	96.9	93.7	99.1
Clinton	R	100.1	100.9	98.0	120.8	101.7	93.0	98.6
Columbia	R	100.3	100.9	98.7	120.9	98.0	94.1	99.0
Crawford	R	100.6	100.9	100.1	120.9	99.3	93.9	99.0
Cumberland		101.7	101.2	104.3	121.2	101.2	102.0	101.6
Dauphin		101.5	101.2	104.0	121.3	101.9	102.1	100.7
Delaware		108.4	102.6	115.3	122.1	108.5	107.1	102.0
Elk	R	99.7	100.9	99.5	120.8	99.1	97.2	99.0
Erie		101.0	101.2	101.2	121.3	100.2	97.1	99.6
Fayette	R	100.5	101.0	97.8	121.0	96.8	93.7	98.7
Forest	R	101.0	100.8	95.3	120.8	97.8	91.3	98.8
Franklin	R	101.2	101.0	101.3	121.0	98.4	96.6	99.7
Fulton	R	101.2	100.8	97.7	120.8	96.4	92.7	97.7
Greene	R	100.1	100.9	96.8	120.8	98.1	92.0	98.7
Huntingdon	R	100.2	100.9	95.2	120.8	97.3	91.5	98.6
Indiana	R	100.0	100.9	97.8	120.9	97.4	93.8	99.2
Jefferson	R	100.2	100.9	98.6	120.8	98.3	94.2	99.2
Juniata	R	100.9	100.9	97.6	120.8	96.6	92.7	100.3
Lackawanna		100.7	101.2	100.6	121.1	101.6	97.4	99.3
Lancaster		102.3	101.4	104.6	121.7	102.3	100.3	101.2
Lawrence	R	100.6	101.0	100.3	120.9	98.6	94.6	98.7
Lebanon	R	101.4	101.0	103.5	121.0	99.1	97.3	99.8
Lehigh		103.1	101.5	105.0	121.4	103.9	102.8	101.4
Luzerne		100.4	101.2	99.9	121.3	99.5	97.7	99.0
Lycoming		100.1	100.9	99.3	120.9	99.4	94.7	99.4
McKean	R	99.7	100.9	98.9	120.8	97.4	94.8	98.4
Mercer		100.6	101.0	100.9	120.9	98.3	94.6	99.2
Mifflin	R	100.7	100.9	98.8	120.8	99.7	92.4	98.9
Monroe	R	103.2	101.0	98.5	121.0	101.5	95.5	101.4
Montgomery		105.0	102.4	109.0	123.1	107.0	121.8	106.8
Montour	R	100.6	100.9	102.1	120.8	111.8	101.8	101.3
Northampton		102.5	101.3	103.7	121.3	101.9	99.5	100.6
Northumberland	R	100.0	101.0	99.3	120.9	97.7	94.4	98.9
Perry	R	101.3	100.9	101.8	120.8	98.0	94.0	99.8
Philadelphia		127.6	106.3	148.6	123.5	119.2	101.7	99.7
Pike	R	103.2	100.9	100.4	120.8	101.0	94.2	102.0
Potter	R	100.2	100.8	98.0	120.8	97.5	94.0	98.3
Schuylkill	R	100.0	101.0	98.7	121.0	98.4	95.3	98.9
Snyder	R	100.6	100.9	101.0	120.8	98.5	97.4	100.0
Somerset	R	100.5	100.9	98.3	120.9	97.1	93.6	100.2
Sullivan	R	100.1	100.8	98.4	120.8	97.1	92.4	100.0
Susquehanna	R	100.5	100.9	98.5	120.8	96.8	93.2	100.4
Tioga	R	100.4	100.9	98.8	120.8	96.7	93.2	98.6
Union	R	100.5	100.9	101.5	120.8	99.3	94.1	101.4
Venango	R	99.9	100.9	99.0	120.9	98.1	97.8	99.1
Warren	R	99.7	100.9	100.2	120.8	98.5	96.1	99.9
Washington		100.7	101.1	101.1	121.2	99.4	99.3	99.1
Wayne	R	101.2	100.9	97.4	120.8	100.1	93.5	100.4
Westmoreland		100.9	101.3	101.5	121.5	101.2	98.8	99.8
Wyoming	R	99.8	100.9	97.4	120.8	99.0	93.9	99.5
York		102.1	101.3	103.3	121.5	101.6	99.4	101.5

2) Housing Subindex

Housing was the subindex that varied the most across the state, from a low of 95.2 in Huntingdon County to 148.6 in Philadelphia. Map 3 and Table 10 present the data. Again the southeastern portion of the state exhibited the highest costs, along with Allegheny County. But the western edge of the state had higher housing costs than the central portion of the state. Centre County was an exception, with a relatively high cost (102.9) compared to nearby counties. The presence of Penn State University in Centre County goes a long way toward explaining this, of course.

Again, urban counties had a higher cost than rural counties. Housing costs in Pennsylvania's urban counties were 6.2% higher than in its rural counties, on average. In fact, the housing index for rural counties (99.0) was just below the average (100.0) for the 321 areas in the *ACCRA Cost of Living Index* that is the standard of comparison for this study.

3) Utilities Subindex

With a value of 121.1, the utilities subindex bears the dubious distinction of having the highest average cost relative to other areas of the country. The estimates in Table 11 indicate that counties in Pennsylvania paid approximately twenty percent more for their utilities than the other areas around the nation that comprise our basis for comparison. This subindex showed the least amount of variation across counties, as well, being relatively high in all of them. To give a more accurate feel for these facts, Map 4 takes a somewhat different approach than the other maps, showing no counties with a relatively light color.

The utilities subindex still exhibits the typical pattern with costs being highest in the southeastern portion of the state, although Allegheny county topped the category. Costs in rural areas were slightly lower than in urban areas, but there was very little difference between the two.

4) Transportation Subindex

Transportation costs in the 67 counties of the state averaged 100.0, the same as for the 321 counties nationwide in 1997. Table 12 and Map 5 show that this subindex exhibited the typical pattern of higher costs in the Philadelphia and Pittsburgh areas, although Montour county weighed in with the highest estimated costs, surprisingly. Urban counties averaged 102.3, or 3.7% higher than the rural average of 98.7.

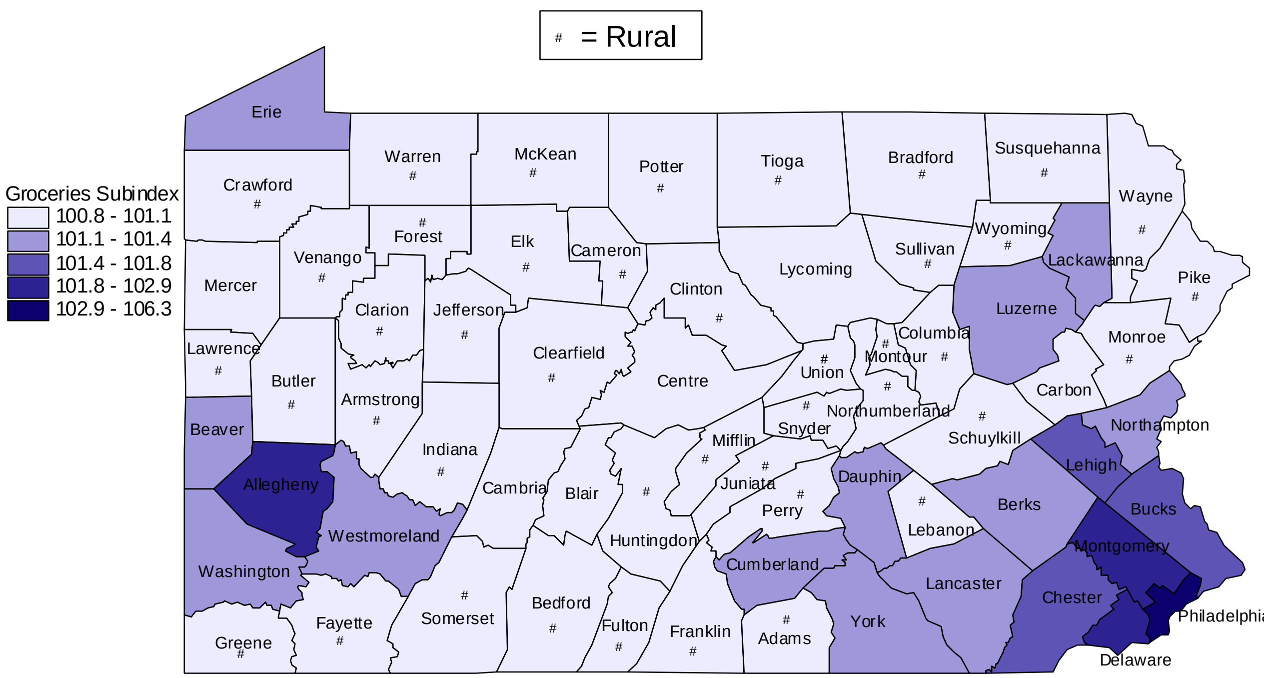
5) Health Care Subindex

Health care is the subindex with the lowest average COL at 96.9, over 3% less than the average for the 321 counties in the ACCRA database. Table 13 shows that 55 of the state's 67 counties had estimated health care indexes less than the national average. Health care would seem to be a bit of a bargain in Pennsylvania, according to these estimates.

This was also the subindex that showed the greatest difference between rural and urban averages, with urban counties being 7.1% more expensive on average than rural counties. Map 6 shows that the highest costs for health care occurred in suburban Philadelphia—although not in the city itself—and Allegheny County. Again, rural Montour County shows up in the company of expensive counties for health care.

6) Miscellaneous Goods and Services

Table 14 shows that the cost of miscellaneous goods and services ranged from 97.7 to 106.8, with and average of 99.9 for the state. Urban areas were about 1.2% more expensive than rural areas, and the southeastern counties tended to be most expensive, with the exception of Philadelphia itself. Map 7 shows the spatial patterns.



Map 2: Groceries Subindex

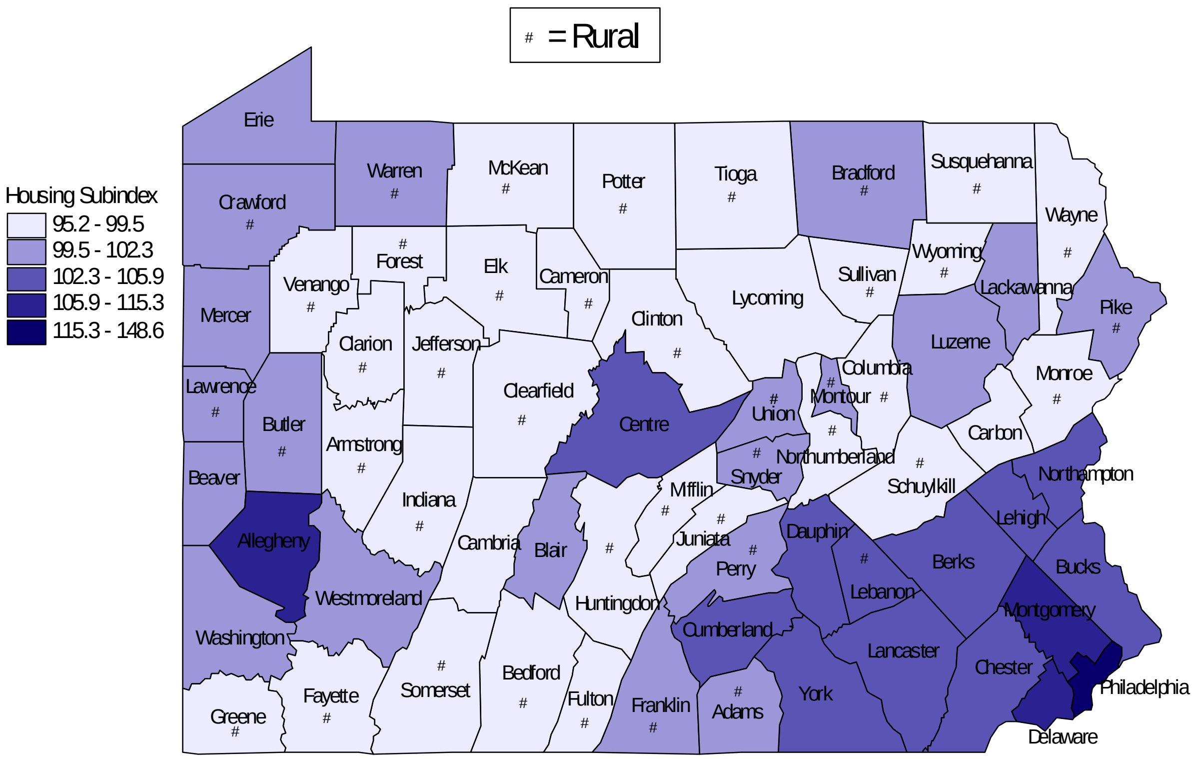
Table 9
Groceries Subindex Ranked from Lowest to Highest

Rank	County	Rural	Index
1	Cameron	R	100.8
2	Forest	R	100.8
3	Fulton	R	100.8
4	Potter	R	100.8
5	Sullivan	R	100.8
6	Armstrong	R	100.9
7	Bedford	R	100.9
8	Bradford	R	100.9
9	Carbon		100.9
10	Clarion	R	100.9
11	Clearfield	R	100.9
12	Clinton	R	100.9
13	Columbia	R	100.9
14	Crawford	R	100.9
15	Elk	R	100.9
16	Greene	R	100.9
17	Huntingdon	R	100.9
18	Indiana	R	100.9
19	Jefferson	R	100.9
20	Juniata	R	100.9
21	Lycoming		100.9
22	McKean	R	100.9
23	Mifflin	R	100.9
24	Montour	R	100.9
25	Perry	R	100.9
26	Pike	R	100.9
27	Snyder	R	100.9
28	Somerset	R	100.9
29	Susquehanna	R	100.9
30	Tioga	R	100.9
31	Union	R	100.9
32	Venango	R	100.9
33	Warren	R	100.9
34	Wayne	R	100.9
35	Wyoming	R	100.9
36	Adams	R	101.0
37	Blair		101.0
38	Butler	R	101.0
39	Cambria		101.0
40	Centre		101.0
41	Fayette	R	101.0
42	Franklin	R	101.0
43	Lawrence	R	101.0
44	Lebanon	R	101.0
45	Mercer		101.0
46	Monroe	R	101.0
47	Northumberland	R	101.0
48	Schuylkill	R	101.0
49	Beaver		101.1
50	Washington		101.1
51	Cumberland		101.2
52	Dauphin		101.2
53	Erie		101.2
54	Lackawanna		101.2
55	Luzerne		101.2
56	Berks		101.3
57	Northampton		101.3
58	Westmoreland		101.3
59	York		101.3
60	Lancaster		101.4
61	Lehigh		101.5
62	Chester		101.6
63	Bucks		101.8
64	Montgomery		102.4
65	Delaware		102.6
66	Allegheny		102.9
67	Philadelphia		106.3

Descriptive Statistics

	Total	Rural	Urban
Mean:	101.2	100.4	101.6
Standard deviation:	0.9	0.1	1.1
Minimum:	100.8	100.8	100.9
Maximum:	106.3	101.0	106.3
Range:	5.5	0.2	5.3

100.0 = average of 321 areas participating in ACCRA *Cost of Living Index,* third quarter, 1997.



Map 3: Housing Subindex

Table 10
Housing Subindex
Ranked from Lowest to Highest

Rank	County	Rural	Index	Rank	County	Rural	Index
1	Huntingdon	R	95.2	35	Crawford	R	100.1
2	Forest	R	95.3	36	Warren	R	100.2
3	Greene	R	96.8	37	Lawrence	R	100.3
4	Wyoming	R	97.4	38	Pike	R	100.4
5	Wayne	R	97.4	39	Lackawanna		100.6
6	Clearfield	R	97.5	40	Mercer		100.9
7	Juniata	R	97.6	41	Blair		100.9
8	Fulton	R	97.7	42	Snyder	R	101.0
9	Indiana	R	97.8	43	Washington		101.1
10	Fayette	R	97.8	44	Erie		101.2
11	Clinton	R	98.0	45	Franklin	R	101.3
12	Potter	R	98.0	46	Adams	R	101.4
13	Bedford	R	98.2	47	Union	R	101.5
14	Somerset	R	98.3	48	Butler	R	101.5
15	Armstrong	R	98.3	49	Westmoreland		101.5
16	Sullivan	R	98.4	50	Perry	R	101.8
17	Susquehanna	R	98.5	51	Montour	R	102.1
18	Monroe	R	98.5	52	Beaver		102.3
19	Cambria		98.5	53	Centre		102.9
20	Jefferson	R	98.6	54	Berks		103.1
21	Schuylkill	R	98.7	55	York		103.3
22	Columbia	R	98.7	56	Lebanon	R	103.5
23	Cameron	R	98.7	57	Northampton		103.7
24	Carbon		98.8	58	Dauphin		104.0
25	Tioga	R	98.8	59	Cumberland		104.3
26	Mifflin	R	98.8	60	Lancaster		104.6
27	McKean	R	98.9	61	Chester		104.9
28	Venango	R	99.0	62	Lehigh		105.0
29	Clarion	R	99.1	63	Bucks		105.9
30	Northumberland	R	99.3	64	Montgomery		109.0
31	Lycoming		99.3	65	Allegheny		109.5
32	Elk	R	99.5	66	Delaware		115.3
33	Luzerne		99.9	67	Philadelphia		148.6
34	Bradford	R	100.0				

Descriptive Statistics

	Total	Rural	Urban
Mean:	101.3	99.0	105.2
Standard deviation:	6.8	1.7	9.8
Minimum:	95.2	95.2	98.5
Maximum:	148.6	103.5	148.6
Range:	53.4	8.3	50.1

100.0 = average of 321 areas participating in ACCRA *Cost of Living Index*, third quarter, 1997

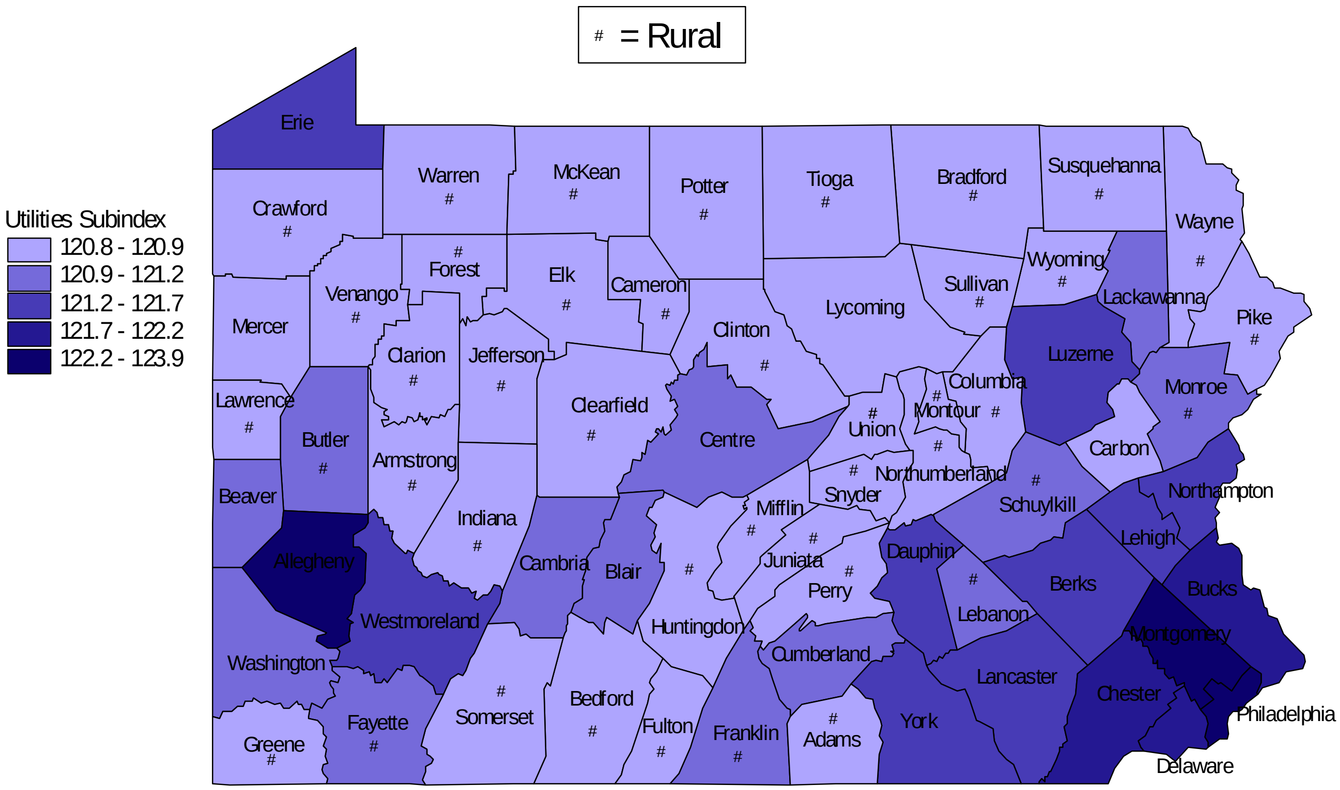
Table 11
Utilities Subindex
Ranked from Lowest to Highest

Rank	County	Rural	Index
1	Forest	R	120.8
2	Sullivan	R	120.8
3	Cameron	R	120.8
4	Fulton	R	120.8
5	Potter	R	120.8
6	Juniata	R	120.8
7	Montour	R	120.8
8	Wyoming	R	120.8
9	Clinton	R	120.8
10	Greene	R	120.8
11	Huntingdon	R	120.8
12	Tioga	R	120.8
13	Pike	R	120.8
14	Susquehanna	R	120.8
15	Clarion	R	120.8
16	Elk	R	120.8
17	Union	R	120.8
18	Mifflin	R	120.8
19	Bedford	R	120.8
20	Wayne	R	120.8
21	Perry	R	120.8
22	Snyder	R	120.8
23	Jefferson	R	120.8
24	McKean	R	120.8
25	Warren	R	120.8
26	Bradford	R	120.8
27	Carbon		120.8
28	Columbia	R	120.9
29	Venango	R	120.9
30	Armstrong	R	120.9
31	Somerset	R	120.9
32	Clearfield	R	120.9
33	Indiana	R	120.9
34	Crawford	R	120.9
35	Adams	R	120.9
36	Northumberland	R	120.9
37	Lawrence	R	120.9
38	Lycoming		120.9
39	Mercer		120.9
40	Monroe	R	121.0
41	Blair		121.0
42	Lebanon	R	121.0
43	Fayette	R	121.0
44	Centre		121.0
45	Franklin	R	121.0
46	Schuylkill	R	121.0
47	Cambria		121.0
48	Butler	R	121.1
49	Beaver		121.1
50	Lackawanna		121.1
51	Washington		121.2
52	Cumberland		121.2
53	Erie		121.3
54	Northampton		121.3
55	Dauphin		121.3
56	Luzerne		121.3
57	Lehigh		121.4
58	Westmoreland		121.5
59	York		121.5
60	Berks		121.5
61	Lancaster		121.7
62	Chester		122.1
63	Delaware		122.1
64	Bucks		122.2
65	Montgomery		123.1
66	Philadelphia		123.5
67	Allegheny		123.9

Descriptive Statistics

	Total	Rural	Urban
Mean:	121.1	120.8	121.6
Standard deviation:	0.6	0.1	0.8
Minimum:	120.8	120.8	120.8
Maximum:	123.9	121.1	123.9
Range:	3.1	0.3	3.0

100.0 = average of 321 areas participating in ACCRA *Cost of Living Index*, third quarter, 1997.



Map 4: Utilities Subindex

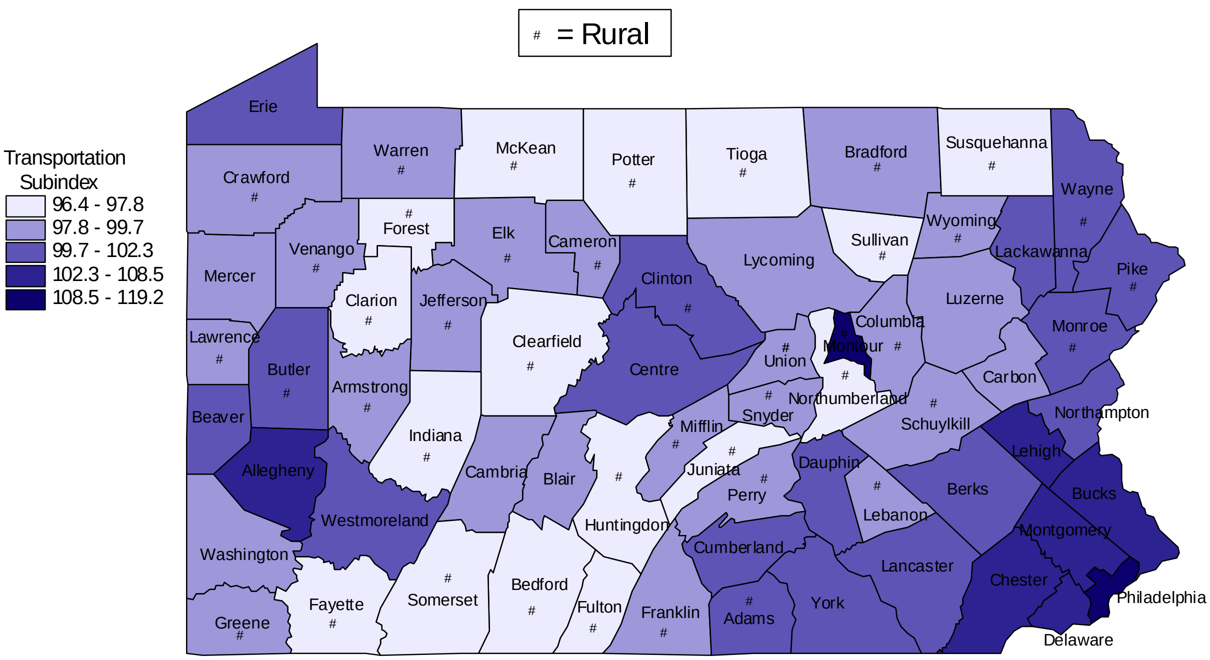
Table 12
Transportation Subindex
Ranked from Lowest to Highest

Rank	County	Rural	Index	Rank	County	Rural	Index
1	Fulton	R	96.4	35	Elk	R	99.1
2	Juniata	R	96.6	36	Lebanon	R	99.1
3	Tioga	R	96.7	37	Union	R	99.3
4	Clarion	R	96.8	38	Crawford	R	99.3
5	Fayette	R	96.8	39	Washington		99.4
6	Bedford	R	96.8	40	Lycoming		99.4
7	Susquehanna	R	96.8	41	Luzerne		99.5
8	Clearfield	R	96.9	42	Mifflin	R	99.7
9	Sullivan	R	97.1	43	Adams	R	100.0
10	Somerset	R	97.1	44	Wayne	R	100.1
11	Huntingdon	R	97.3	45	Erie		100.2
12	McKean	R	97.4	46	Butler	R	100.2
13	Indiana	R	97.4	47	Centre		100.4
14	Potter	R	97.5	48	Beaver		100.7
15	Northumberland	R	97.7	49	Pike	R	101.0
16	Forest	R	97.8	50	Westmoreland		101.2
17	Cameron	R	97.9	51	Cumberland		101.2
18	Perry	R	98.0	52	Berks		101.3
19	Blair		98.0	53	Monroe	R	101.5
20	Columbia	R	98.0	54	York		101.6
21	Cambria		98.1	55	Lackawanna		101.6
22	Venango	R	98.1	56	Clinton	R	101.7
23	Greene	R	98.1	57	Dauphin		101.9
24	Mercer		98.3	58	Northampton		101.9
25	Jefferson	R	98.3	59	Lancaster		102.3
26	Schuylkill	R	98.4	60	Lehigh		103.9
27	Franklin	R	98.4	61	Chester		104.4
28	Warren	R	98.5	62	Bucks		105.0
29	Snyder	R	98.5	63	Allegheny		105.2
30	Lawrence	R	98.6	64	Montgomery		107.0
31	Armstrong	R	98.6	65	Delaware		108.5
32	Carbon		98.6	66	Montour	R	111.8
33	Bradford	R	99.0	67	Philadelphia		119.2
34	Wyoming	R	99.0				

Descriptive Statistics

	Total	Rural	Urban
Mean:	100.0	98.7	102.3
Standard deviation:	3.8	2.5	4.5
Minimum:	96.4	96.4	98.0
Maximum:	119.2	111.8	119.2
Range:	22.7	15.3	21.2

100.0 = average of 321 areas participating in ACCRA *Cost of Living Index,* third quarter, 1997.



Map 5: Transportation Subindex

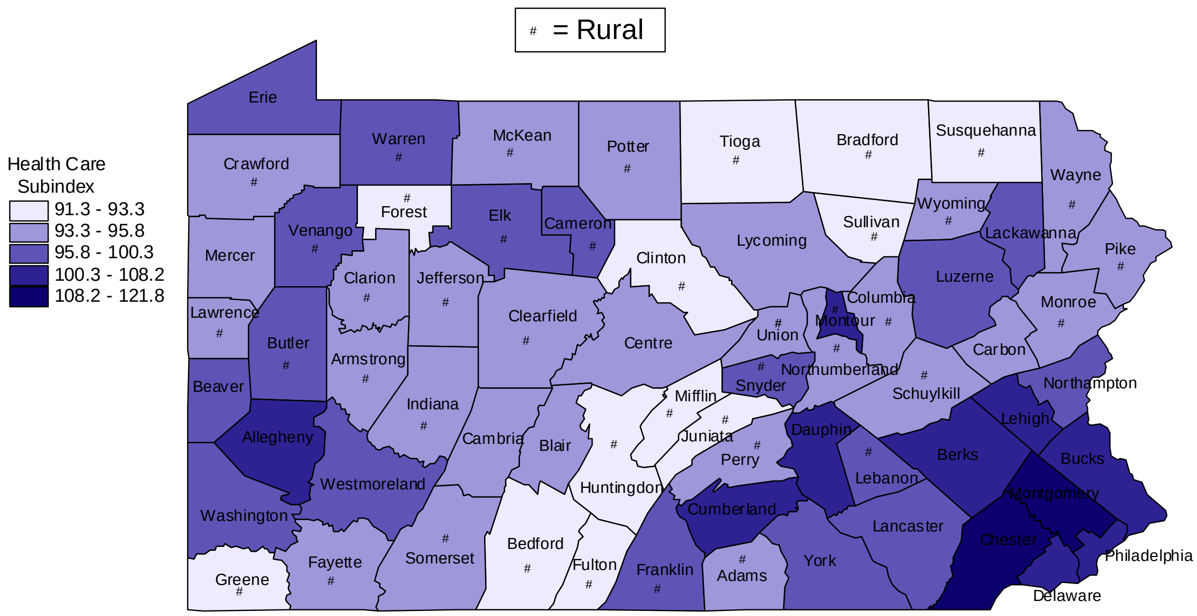
Table 13
Health Care Subindex
Ranked from Lowest to Highest

Rank	County	Rural	Index	Rank	County	Rural	Index
1	Forest	R	91.3	35	Blair		95.1
2	Huntingdon	R	91.5	36	Schuylkill	R	95.3
3	Greene	R	92.0	37	Monroe	R	95.5
4	Bedford	R	92.2	38	Centre		95.5
5	Mifflin	R	92.4	39	Adams	R	95.8
6	Sullivan	R	92.4	40	Cameron	R	96.1
7	Fulton	R	92.7	41	Warren	R	96.1
8	Juniata	R	92.7	42	Beaver		96.4
9	Clinton	R	93.0	43	Franklin	R	96.6
10	Susquehanna	R	93.2	44	Butler	R	96.9
11	Tioga	R	93.2	45	Erie		97.1
12	Bradford	R	93.3	46	Elk	R	97.2
13	Wayne	R	93.5	47	Lebanon	R	97.3
14	Somerset	R	93.6	48	Snyder	R	97.4
15	Fayette	R	93.7	49	Lackawanna		97.4
16	Clearfield	R	93.7	50	Luzerne		97.7
17	Clarion	R	93.8	51	Venango	R	97.8
18	Indiana	R	93.8	52	Westmoreland		98.8
19	Crawford	R	93.9	53	Washington		99.3
20	Wyoming	R	93.9	54	York		99.4
21	Potter	R	94.0	55	Northampton		99.5
22	Perry	R	94.0	56	Lancaster		100.3
23	Union	R	94.1	57	Berks		101.3
24	Columbia	R	94.1	58	Philadelphia		101.7
25	Jefferson	R	94.2	59	Montour	R	101.8
26	Pike	R	94.2	60	Cumberland		102.0
27	Northumberland	R	94.4	61	Dauphin		102.1
28	Armstrong	R	94.5	62	Lehigh		102.8
29	Lawrence	R	94.6	63	Delaware		107.1
30	Mercer		94.6	64	Bucks		107.6
31	Cambria		94.6	65	Allegheny		108.2
32	Lycoming		94.7	66	Chester		118.2
33	McKean	R	94.8	67	Montgomery		121.8
34	Carbon		94.9				

Descriptive Statistics

	Total	Rural	Urban
Mean:	96.9	94.4	101.1
Standard deviation:	5.5	2.0	7.0
Minimum:	91.3	91.3	94.6
Maximum:	121.8	101.8	121.8
Range:	30.5	10.5	27.2

100.0 = average of 321 areas participating in ACCRA *Cost of Living Index*, third quarter, 1997.



Map 6: Health Care Subindex

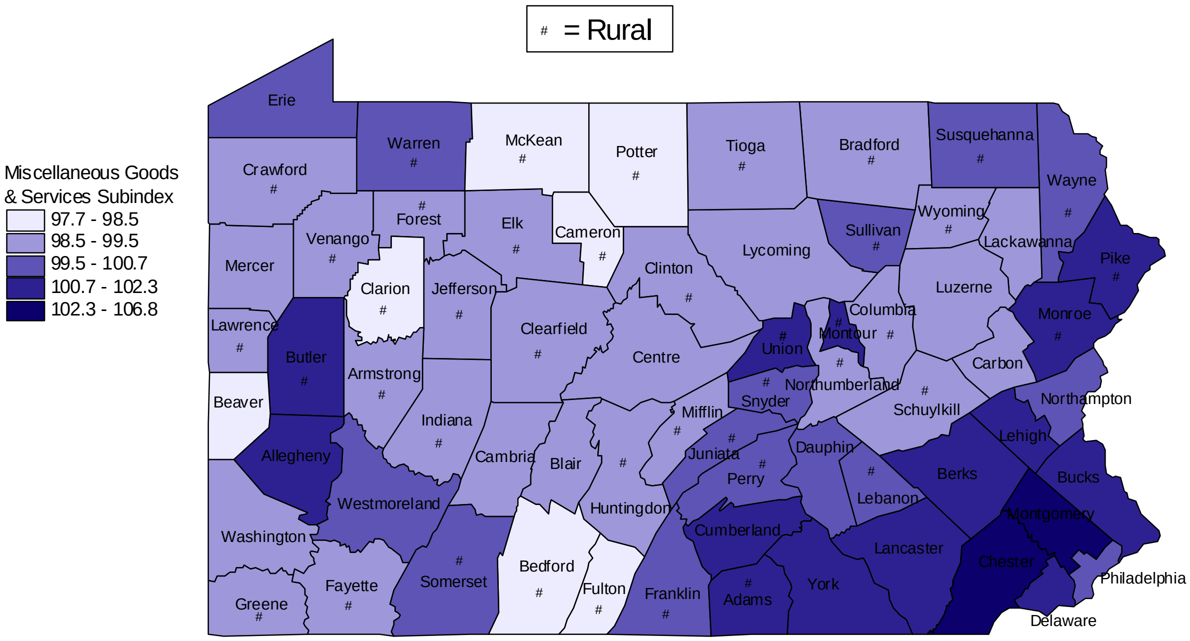
Table 14
Miscellaneous Goods and Services Subindex
Ranked from Lowest to Highest

Rank	County	Rural	Index	Rank	County	Rural	Index
1	Fulton	R	97.7	35	Bradford	R	99.4
2	Clarion	R	97.8	36	Wyoming	R	99.5
3	Cameron	R	98.2	37	Erie		99.6
4	Potter	R	98.3	38	Philadelphia		99.7
5	McKean	R	98.4	39	Franklin	R	99.7
6	Bedford	R	98.4	40	Westmoreland		99.8
7	Beaver		98.5	41	Perry	R	99.8
8	Clinton	R	98.6	42	Lebanon	R	99.8
9	Tioga	R	98.6	43	Warren	R	99.9
10	Huntingdon	R	98.6	44	Snyder	R	100.0
11	Armstrong	R	98.6	45	Sullivan	R	100.0
12	Lawrence	R	98.7	46	Somerset	R	100.2
13	Greene	R	98.7	47	Juniata	R	100.3
14	Blair		98.7	48	Susquehanna	R	100.4
15	Fayette	R	98.7	49	Wayne	R	100.4
16	Cambria		98.8	50	Northampton		100.6
17	Forest	R	98.8	51	Dauphin		100.7
18	Schuylkill	R	98.9	52	Adams	R	100.8
19	Northumberland	R	98.9	53	Butler	R	100.9
20	Mifflin	R	98.9	54	Berks		101.1
21	Elk	R	99.0	55	Lancaster		101.2
22	Luzerne		99.0	56	Montour	R	101.3
23	Columbia	R	99.0	57	Allegheny		101.3
24	Crawford	R	99.0	58	Union	R	101.4
25	Clearfield	R	99.1	59	Monroe	R	101.4
26	Venango	R	99.1	60	Lehigh		101.4
27	Washington		99.1	61	York		101.5
28	Indiana	R	99.2	62	Cumberland		101.6
29	Jefferson	R	99.2	63	Pike	R	102.0
30	Mercer		99.2	64	Delaware		102.0
31	Carbon		99.2	65	Bucks		102.3
32	Lackawanna		99.3	66	Chester		105.7
33	Centre		99.4	67	Montgomery		106.8
34	Lycoming		99.4				

Descriptive Statistics

	Total	Rural	Urban
Mean:	99.9	94.4	100.6
Standard deviation:	1.6	1.0	2.0
Minimum:	97.7	97.7	98.5
Maximum:	106.8	102.0	106.8
Range:	9.1	4.3	8.3

100.0 = average of 321 areas participating in ACCRA *Cost of Living Index*, third quarter, 1997.



Map 7: Miscellaneous Goods and Services Subindex

II. RESULTS AND CONCLUSIONS

As explained above, this project seeks to address four key issues. Here are the findings for each.

A. Rural vs. Urban Counties.

Are rural counties less expensive places to live than urban counties in Pennsylvania? The simple answer is "yes." Rural counties in Pennsylvania tend to have lower costs of living overall and for each of the six subindexes than urban counties.

However, not all rural places are less expensive than all urban places. For example, urban Lycoming (Williamsport), Cambria (Johnstown) and Luzerne (Wilkes-Barre) counties tended to have lower costs than many rural counties, and Pike and Monroe counties tended to have higher costs despite their rural classification.

The general cost of living patterns in the state include the following:

- Urban areas tend to be more expensive than rural areas.
- Eastern-tier counties tend to be more expensive than others.
- The southeastern counties, especially Philadelphia and its near suburbs, tend to be especially high-cost.
- Counties on the western tier of the state and those around Pittsburgh also tend to exhibit higher costs than northern tier counties.

How much more expensive are urban counties than rural counties? Table 15 summarizes the results. Using simple averages of the counties' COL data, urban areas are about 2.4% more expensive overall than rural areas. But as explained earlier, more people live in urban than rural areas, so if we wish to compare the cost of the average urban resident with that of the average rural resident, it is necessary to calculate population-weighted averages. Table 16 does this. Virtually all of the numbers in Table 16 are higher than in Table 15, since the higher COL values tend to apply to counties which have higher populations; a bigger segment of the population is subject to the higher urban COL numbers. Using the population-weighted averages, the cost of living differences between urban and rural residents are larger than for the unadjusted averages. For the overall cost of living, urban residents of the state pay about 6.0% more on average than do rural residents. Similar patterns apply for all of the subindexes, although to differing degrees.

B. Overall Index vs. Subindexes.

For this key issue we asked: Do the same patterns exist for each of the subindexes (groceries, housing, utilities, transportation, health care, and miscellaneous goods and services) as for the overall cost of living? Do some types of products vary more in price across the state than others?

As Tables 15 and 16 show, the answer to both questions is "yes". Urban areas tend to be more expensive for all of the subindexes, as well as for the overall cost of living. The maps presented earlier in the report show that there are some differences in the spatial patterns, however. While the southeastern part of the state was the high-cost areas for most items, Philadelphia was not the highest-cost location for health care and for miscellaneous goods and services. And some of the rural counties climbed up the cost list for selected categories of goods.

Table 15
Rural vs. Urban Costs, Unadjusted for Population

	Overall	Rural	Urban	Urban % higher
Total	101.5	100.6	103.0	2.4
Groceries	101.2	100.9	101.6	0.7
Housing	101.3	99.0	105.2	6.3
Utilities	121.1	120.8	121.6	0.7
Transportn	100.0	98.7	102.3	3.6
Health Care	96.9	94.4	101.1	7.1
Misc.	99.9	99.4	100.6	1.2

Table 16
Rural vs. Urban Costs, Adjusted for Population

	Overall	Rural	Urban	Urban % higher
Total	105.5	100.7	106.7	6.0
Groceries	102.1	100.9	102.5	1.5
Housing	109.3	99.4	112.0	12.7
Utilities	122.0	120.9	122.3	1.1
Transportn	104.1	98.6	105.6	7.1
Health Care	102.2	94.7	104.1	9.9
Misc.	100.9	99.5	101.1	1.7

Housing and health care were the two categories of goods that varied the most across counties, and utilities, groceries and miscellaneous goods varied least.

C. Consistency of the Spatial Patterns of COL through Time.

Are the spatial patterns found in the 1992 study still applicable, or has something important changed?

The central finding related to this issue is that similar patterns occur in the 1997 as in the 1989 data. Eastern and urban counties were more expensive then, and still are. The urban counties have, in many cases, seen a reduction in their COLs over the period, resulting in a smaller spread between rural and urban costs over the period.

D. Consistency of the Determinants of COL through Time.

Are the factors that determine cost of living differences across space stable over time? Are the structure and coefficients of the model in 1999 similar to those of 1992, or has something important changed?

For comparison purposes, Table 17 presents the equations that were found in the 1992 study of 1989 data, along with the current equations based on 1997 data. In general, the current models (for 1997 data) give better fits, as evidenced by their higher adjusted R^2s. Similar, if not identical, variables apply in most cases. For example, aggregate income may enter the equation instead of population, but since these two variables are highly correlated, they may be considered as measuring very similar forces.

Consider first the models for total cost of living:

1989TOT = 83.939* + 0.00362 DEN* - 1.647(e-8) DENSQ* + 7.427(e-8) AGGINC*
+ 0.109 AGGGTH* + 1.784 ELEC* + 0.175 GCST* + 11.025 NE* - 5.019 MA*
- 9.104 SA* - 10.262 ESC* - 9.743 WSC* - 9.034 ENC* - 11.430 WNC* - 7.967 MTN*
adjusted R^2 = .636 *statistically significant at the 5% level

1997TOT = 96.548* + 0.00299 DEN* - 1.81(e-8) DENSQ* + 0.835 POPGTH9697
+ 1.221 ELEC* + 0.0031 GCST* + 0.330 NE – 8.504 MA* – 9.498 SA*
– 11.291 ESC* – 15.069 WSC* – 9.049 ENC* – 9.801 WNC* – 6.469 MTN*
adjusted R^2 = .787 *statistically significant at the 5% level

Both the 1997 and 1989 equations include the density variable as a key determinant (both density and density squared are included, to allow for nonlinear effects), as well as electric rates, the government cost variable and the eight regional dummy variables. All variables have the same signs in both studies (i.e., the variables affect COL in the same way.) The 1989 data's equation is different in that it included growth of aggregate income (AGGGTH) while the 1997 data included growth of population (POPGTH9697.) As mentioned earlier, the two growth variables were closely related; when an area's population rises, its aggregate income rises as well. Although AGGGTH gave very similar results to the POPGTH variable, the latter variable was found to have a slightly higher level of significance. The 1989 variable also included a variable for current level of aggregate income of the area (AGGINC). Aggregate income was tested in the current study, but not found to be significant once the other variables were included.

Again for purposes of comparison, a model was estimated using the variables from the final model of the 1992 study of 1989 data but with the current (1997) data. The results are very similar:

TOT = 94.650* + 0.0029 DEN* - 1.68(e-8) DENSQ* + 6.37(e-9) AGGINC
+ 0.0822 AGGGTH* + 1.291 ELEC* + 0.0034 GCST* - 0.390 NE – 9.378 MA*
– 9.867 SA* – 11.765 ESC* – 15.435 WSC* – 9.761 ENC* – 10.480 WNC*
– 7.068 MTN*

adjusted R^2 = .784 *statistically significant at the 5% level

Table 17
Comparison of 1997 and 1989 Results

	Expected	**Total**		Groceries		Housing		Utilities		Transportation		Health Care		Miscellaneous	
Variable	sign	1997	1989	1997	1989	1997	1989	1997	1989	1997	1989	1997	1989	1997	1989
Intercept		96.548*	83.939*	97.145*	123.981*	123.170*	56.524*	-27.787	106.833*	87.093*	111.201*	104.292*	127.755*	97.929*	103.026*
Population 97	(+)											4.59E-06*			
Pop squared												-7.06E-13*			
Pop growth 92-97	+												0.232*		0.0837*
Pop growth 96-97	+	0.835*												0.508	
Density 97	(+)	0.00299*	0.00362*	0.000423*		0.00492*				0.00214*	0.00191*		0.00126*		0.00169*
Density squared		-1.81E-08*	-1.647(e-7)*							-2.72E-08*	-1.233(e-7)*				-1.095(e-7)*
Income per cap 97	(+)				-.00265*						-.00299*				-0.00110
Income per cap squared					8.417(e-8)*						1.031(e-7)*	2.03E-08*		6.38E-09*	4.528(e-8)*
Income per cap growth 92-97	+													-0.168*	
Aggregate income 97	(+)		7.427(e-8)*	3.48E-08*			5.346(e-7)*	8.37(e-8)*	3.104(e-7)*				4.215(e-7)*		
Agg inc squared									-2.268(e-15)				-2.272(e-15)*		
Agg inc growth 92-97	+		0.109*		-.0519*		0.369*		-.143*						
Gov't cost per worker FY92	(+)	0.00312*	0.175*			0.00689*	0.628*			0.208*	0.806*		0.296*		
Gov't cost squared										-5.33E-05*	-.00986*				
Electric rate 97	(+)	1.221*	1.784*	0.928*			4.281*	4.750*	4.829*	1.241*			-5.316*	0.707*	0.824*
Electric rate squared													0.378*		
Gas rate 97	(+)							19.336*	-26.446*			1.350*			
Gas rate squared								-1.107*	3.249*						
Mean travel time 1990	+														
Unemployment rate 97	-					-0.976*									
NE		0.330	11.025*	0.289	6.664*	1.943	38.447*	36.369*	6.810	-5.546	4.533	-12.426*	-8.030	0.012	-2.890
MA		-8.504*	-5.019*	-5.045*	1.943	-18.207*	-15.421*	17.252*	25.609*	-7.859*	-5.412*	-29.569*	-26.200*	-4.699*	-4.792*
SA		-9.498*	-9.104*	-4.784*	-4.012*	-24.971*	-22.620*	16.491*	17.581*	-9.072*	-7.149*	-32.815*	-24.818*	-2.839*	-3.799*
ESC		-11.291*	-10.262*	-5.721*	-4.329*	-30.786*	-26.606*	9.363*	19.858*	-7.397*	-8.359*	-34.749*	-31.349*	-3.254*	-4.226*
WSC		-15.069*	-9.743*	-12.643*	-2.298	-34.968*	-32.115*	8.650*	17.291*	-5.326*	-6.534*	-31.300*	-29.067*	-4.872*	-4.754*
ENC		-9.049*	-9.034*	-4.336*	-2.455	-22.006*	-20.488*	9.788*	12.702*	-5.533*	-6.715*	-29.130*	-24.380*	-4.928*	-7.013*
WNC		-9.801*	-11.430*	-6.828*	-5.325*	-25.815*	-27.357*	6.761*	13.721*	-4.951*	-5.968*	-27.972*	-28.747*	-4.122*	-7.351*
MTN		-6.469*	-7.967*	-1.501	-2.113	-13.952*	-21.237*	5.554	3.496	0.840	-2.656	-16.901*	-16.072*	-4.482*	-5.887*
Adjusted R-squared		0.787	0.636	0.570	0.257	0.765	0.514	0.606	0.522	0.434	0.361	0.688	0.657	0.341	0.399
Sample size		303	248	303	248	303	248	303	248	303	248	303	248	303	248

The aggregate income variable (AGGINC) was found not to be statistically significant and was dropped from the final model for this year's study. Estimates of the COL index for the 67 counties of Pennsylvania were made based on this model, and compared to the results from the best model given earlier. The results from the two models were very similar, with a correlation of .999 (out of a possible 1.000.) This suggests that the same basic determining forces that were found in the 1989 data were still at work in 1997.

The final models from the 1992 study were also fitted for each of the subindexes, using 1997 data. In most cases, the variables exhibited the same signs as for the 1989 data, although the significance levels of some were lower. Again, this tends to reinforce the idea that the determinants of cost of living have not changed dramatically since the last study. However, it was frequently possible to derive a better model with the new data by substituting different variables in some cases. This is why the final models for the 1997 data differ somewhat from the models for the 1989 data.

E. Other Issues

This study also considered some other issues or variables, which were not included as part of the final estimates. These issues are detailed in this section.

1) Urban/rural transport cost issues

As part of the examination of rural costs compared to urban costs, we considered the issue of whether rural residents incur higher transport costs than urban residents, thus increasing their cost of living relative to urban dwellers. By their very nature, rural areas are low density places. This means that there will typically not be concentrations of economic activity nearby, and that more travel may be necessary.

To examine this issue, we might ask the simple question of whether the average rural resident travels farther than the average urban resident in a typical week or year. Recognizing that urban areas often suffer from congestion, this question should more appropriately be couched in terms of travel time, rather than distance.

Data from the 1990 Census of Population allow an examination of this issue. The 1990 Census asked people about the amount of time it took them to get to their typical place of work. Data were available on travel time for all counties in the country, along with other variables of interest. It was possible to calculate average travel time for all workers in each county, as well as percentages of residents in each county who lived in rural and urban areas.

The first approach to the data was to look at the relationship of a county's "percentage rural" with its average travel time. Does a higher percentage of rural residents mean a higher average travel time to work? Correlation analysis can help answer this question. The correlation coefficient can vary from negative one to positive one. A value of zero would mean that the two variables are completely unrelated. A value of positive one would mean that they were perfectly correlated; every time one variable increases, the other increases in lockstep with it. A value of negative one would mean that they were perfectly related, but inversely, such that every time one variable increases, the other would decrease, again in lockstep.

The actual correlation coefficient in this case was .024—a positive number, but very small. This means that there was a very slight positive relationship nationwide between the percentage of a county's residents that lived in rural areas and the amount of time they spent commuting to work in 1990. On the other hand, there was a -.389 correlation between the percentage of a county's residents who lived on farms and average travel time to work; more farm residents meant less time spent travelling to work. Presumably many of those who live on a farm are only commuting as far as the barn. This implies that the non-farm rural folks must be commuting even farther, then, to make up for the low travel time of the farmers.

To examine the data from a different direction, counties were ranked from highest to lowest percentage of residents who were rural. The average county had 63.4% of its population living in rural areas in 1990. (Note, this is not to say that 63.4% of people lived in rural areas! The urban counties had much larger populations than the rural counties.) 1,657 counties had 63.4% or more of their people in rural areas; we'll call them "rural counties." 1,484 counties had less that that percent in rural areas, and we'll call these "urban counties." The average travel time to work for the rural counties (as we've defined them) was 19.1 minutes one way in 1990, and the average time in the urban counties was 17.9 minutes. Rural residents spent 6.5% more time commuting than urban residents by this measure. 1.2 minutes per day each way translates to 2.4 minutes per day, times 5 days a week times

50 weeks per year translates to 10 hours per year of extra travel time for rural residents for commuting to and from work. However, this tells only part of the story. These travel time data are only for journeys to work, and overlook all the other travel that is necessary in a typical week. Shopping, entertainment, medical care, socializing, and school transport all contribute to a household's total travel costs. We have found no good geographically detailed data for total travel time (as opposed to work travel time.)

However, on the other side of the story it must be mentioned that residents of what we've called "urban" counties had per capita incomes that were about 19% higher than those of the rural counties. This can be interpreted in two ways. First, it means that they have more money to pay their travel costs. And second, the value of their time is higher. The wage that an urban resident gives up when she spends an hour commuting is higher than the amount that her rural cousin foregoes. An economist would say that the urban resident's opportunity cost is higher. Thus, although they may spend 7% less time commuting, their time is 19% more valuable. This approach suggests that it actually costs the urban resident more for travel to work, if we look beyond the out-of-pocket costs to the amount of income foregone due to time spent travelling to work.

2) Unearned income

It has been suggested that "unearned income" may play a role in determining cost of living differences across space. Unearned income includes dividends, interest, rent and transfer payment income. (Transfer payments include income payments to persons for which no current services are performed, and include items such as retirement and disability payments, medical payments, income security payments, unemployment insurance payments, and veterans' benefits.)

Data on unearned income for all counties of the country were derived from variables in the BEA's Regional Economic Information System (REIS) databanks, checked, and aligned with the 303 counties participating in ACCRA's *Cost of Living Index* for the third quarter of 1997. Since unearned income can be expected to vary in value with the size of the region, we calculated the *percent* of each county's income that was received in the form of unearned income.

The percent of unearned income had a near-zero correlation with the total cost of living in the 303 counties. (The actual correlation coefficient was -.001.) It was also not highly correlated with any of the components of the overall cost of living; in fact, it was slightly negatively correlated with three of the six subindexes. The highest correlation (.138) was with the utilities subindex, and the lowest correlation was -.091 with the transportation subindex.

As this implies, inclusion of the “Percent of Income Unearned” in the regression equation resulted in no increase in explanatory power of the equation. The “Percent Unearned” variable was not statistically significant, and actually reduced the adjusted R-squared statistic, which means its inclusion would reduce the goodness of fit of the equation.

3) Housing starts or building permits as possible determinants of COL

It has been suggested that housing starts or building permits may be useful variables in explaining cost of living differences from place to place. Since housing costs are a key component of the overall COL, it makes sense to consider variables related to the housing market. Housing starts and building permits may also be considered proxies for population or income growth in the local economy, since such growth frequently leads to housing construction.

To include these variables in the analysis, data would need to be available for all 303 counties in our sample, in order to fit the estimating equations and measure the impact of the housing starts and/or building permits variables. Data would also need to be available for all 67 of Pennsylvania’s counties, for use in estimating the COL index values from the regression equations.

Unfortunately, we were unable to locate usable data for either housing starts or building permits for the 303 counties in our database. The U.S. Census Bureau releases some data on housing starts, but only at the regional level (for four aggregate regions of the country, but not for their individual components.) As for building permits, townships or boroughs are typically the issuing agents, and the data are published for these units in many cases. Given the large numbers of issuing agents in many counties (there are some 19,000 permit-issuing places nationwide covering 95% of the population in 1990[9], 890 in Pennsylvania alone[10]), aggregation of data for all issuing units into county-level data would be a major undertaking for the 303 counties of this study. The Census Bureau does not

[9] U.S. Census Bureau, Residential Construction Branch, Manufacturing and Construction Division, “Percent of Population in Permit Issuing Places.” On the web at: http://www.census.gov/pub/const/C40/pct_pop.pdf
[10] U.S. Census Bureau, Residential Construction Branch, Manufacturing and Construction Division, “Residential Construction Statistics Ordering Information Package.” On the web at: http://www.census.gov/const/C40/Sample/neworder.pdf

aggregate these into county-level data currently. And since the Census Bureau sells these data rather than posting them for free on the Web, there would also be a substantial expense. It might be noted that the Census Bureau will start posting data for counties on their website later this year, and this variable may be available for future studies. Census officials also indicate that county level data for 1997 may be included in the 1997 Census of Construction, which will be released later in 2000. If this is the case, it may be possible to test this variable in the future.

However, given that this study already includes measures of both population and income growth, the benefit from attempting to construct housing start data by county is estimated to be small relative to the costs which would be involved.

REFERENCES

American Chamber of Commerce Researchers Association, *ACCRA Cost of Living Index*. Alexandria, VA: ACCRA, quarterly.

American Chamber of Commerce Researchers Association, *ACCRA Cost of Living Index Manual*. Revised 1997. Alexandria, VA: ACCRA.

Bartik, Timothy J. *Who Benefits From State and Local Economic Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 1991.

Bishop, John A.; Formby, John P.; and Thistle, Paul D. "Convergence and Divergence of Regional Income Distributions and Welfare." *Review of Economics and Statistics.* v. 76, #2 (May 1994), pp. 228-235.

Borooah, Vani K.; McGregor, Patrick P.L.; McKee, Patricia M. and Mulholland, Gwyenth E. "Cost-of-Living Differences between the Regions of the United Kingdom." pp. 103-132 in Hills, John, ed. *New Inequalities: The Changing Distribution of Income and Wealth in the United Kingdom.* Cambridge: Cambridge University Press, 1996.

Cebula, Richard J. "The Impact of Living Costs on Geographic Mobility." *Quarterly Review of Economics and Finance*, v. 33, #1 (Spring 1993), pp. 101-105.

Cebula, Richard J.; Alexander, Gigi M. and Koch, James V. "A Further Note on Determinants of Geographic Living-Cost Differentials." *Quarterly Review of Economics and Finance*, v. 32, #4 (Winter 1992), pp. 144-147.

Crawford, Ian. "UK Household Cost-of-Living Indices, 1979-92." pp. 76-102 in Hills, John, ed. *New Inequalities: The Changing Distribution of Income and Wealth in the United Kingdom.* Cambridge: Cambridge University Press, 1996.

DataMasters.com *Cost of Living Calculator* online at: http://www.datamasters.com/cgi-bin/col.pl

Deller, Steven C.; Shields, Martin and Tomberlin, David. "Price Differentials and Trends in State Income Levels: A Research Note." *Review of Regional Studies*, v. 26, #1 (Summer 1996), pp. 99-113.

Denslow, David. Personal correspondence to the author. Bureau of Business and Economic Research, University of Florida. March 31, 1999.

Doti, James L and Adibi, Esmael. *The Practice of Econometrics with EViews.* Irvine, CA: Quantitative Micro Software, 1998.

DuMond, J. Michael; Hirsch, Barry T. and Macpherson, David A. "Wage Differentials Across Labor Markets and Workers: Does Cost of Living Matter?" (Paper provided by Hirsch, pre-publication in *Economic Inquiry*.) July 1998, 32 pages.

Eberts, R.W. and M.E. Schweitzer, "Regional Wage Convergence and Divergence: Adjusting Wages for Cost of Living Differences." Federal Reserve Bank of Cleveland's *Economic Review*, v. 30, #2 (1994), pp. 26-37.

Environmental Systems Research Institute, Inc. *ArcView GIS.* Redlands CA: ESRI Inc, 1966.

Fagan, Mark and Longino, Jr., Charles F. "Migrating Retirees: A Source for Economic Development." *Economic Development Quarterly*, v. 7 #1 (Feb 1993), pp. 98-106.

Florida Department of Education. *Florida Price Level Index.* 1998. On the web at: http://www.firn.edu/doe/bin00047/fplifnl.htm

Gujarati, Damodar. *Basic Econometrics.* New York: McGraw-Hill Book Co., 1978.

Homefair.com. *The Salary Calculator*. Online at: http://www2.homefair.com/calc/salcalc.html

Johnston, Richard; McKinney, Martin, and Stark, Tom. "Regional Price Level Variations and Real Household Incomes in the United Kingdom, 1979/80-1993. *Regional Studies*, v. 30, #6 (Oct 1996), pp. 567-578.

Kurre, James A. *The Cost of Living in Rural Pennsylvania.* Erie, PA: Penn State University at Erie, for the Center for Rural Pennsylvania, June 1992.

McKean, John R.; Johnson, Donn M. and Walsh, Richard G. "Valuing Time in Travel Cost Demand Analysis: An Empirical Investigation." *Land Economics*, v. 71, #1 (Feb 1995), pp. 96-105.

McMahon, Walter W. "Geographical Cost of Living Differences: An Update." *AREUEA Journal.* v. 19, #3 (1991) pp. 426-450.

Pearce, Diana, and Brooks, Jennifer (with the assistance of Janice Hamilton Outtz). *The Self-Sufficiency Standard for Pennsylvania.* Washington DC: Wider Opportunities for Women, Inc., 1997.

Quantitative Micro Software. *EViews 3 User's Guide.* Irvine CA: Quantitative Micro Software, 1998.

Runzheimer Inc. "Over Half of U.S. Companies Have Pay Differential Policies; Both Salaried and Hourly Employees Participate." May 18, 1998. http://www.runzheimer.com/051898.HTM

U.S. Bureau of Labor Statistics, "Final Report on Family Budgets: Cost Increases Slowed, Autumn 1981." *Monthly Labor Review*, (July 1982), pp. 44-46.

U.S. Bureau of Labor Statistics. *Local Area Unemployment Statistics*. Online at: http://146.142.4.24/cgi-bin/dsrv?la

U.S. Bureau of Economic Analysis. *Regional Economic Information System 1969-97*. CD-ROM issued May 1999.

U.S. Census Bureau, *Census and You*, March 1993.

U.S. Census Bureau, *1990 Census of Population and Housing*, Summary Tape File 3. Online at: http://venus.census.gov/cdrom/lo

U.S. Census Bureau, *Compendium of Government Finances*. Volume 4, #5 of *1992 Census of Government.* Washington DC: USGPO. February 1997. Online at: http://www.census.gov:80/prod/2/gov/gc92-4/gc924-5.pdf

U.S. Census Bureau, *Compendium of Public Employment*. Volume 3, #2 of *1992 Census of Government.* Washington DC: USGPO. February 1997. Online at: http://www.census.gov:80/prod/2/gov/gc92-3/gc923-2.pdf

U.S. Census Bureau, *County and City Data Book 1994.* Washington DC: USGPO, August 1994.

U.S. Census Bureau, *Poverty Thresholds.* Online at: http://www.census.gov/hhes/poverty/threshld.html

U.S. Census Bureau, *Poverty Thresholds: 1997.* Online at: http://www.census.gov/hhes/poverty/threshld/thresh97.html

U.S. Census Bureau, Residential Construction Branch, Manufacturing and Construction Division, "Percent of Population in Permit Issuing Places." http://www.census.gov/pub/const/C40/pct_pop.pdf

U.S. Census Bureau, Residential Construction Branch, Manufacturing and Construction Division, "Residential Construction Statistics Ordering Information Package." http://www.census.gov/const/C40/Sample/neworder.pdf

U.S. Congressional Research Service, Library of Congress. *Cash and Noncash Benefits for Persons With Limited Income: Eligibility Rules, Recipient and Expenditure Data, FY1996 - FY1998.* (CRS Report for Congress – Order Code RL30401), December 15, 1999.

U.S. Department of Agriculture, "U.S. County Economic Typologies." http://www.econ.ag.gov/epubs/other/typolog/

U.S. Department of Energy, Energy Information Administration. *Electric Sales and Revenue 1997.* Washington DC: EIA, October 1998.

U.S. Department of Energy, Energy Information Administration. *Natural Gas Annual, 1997.* Washington DC: EIA, October 1998.

U.S. Department of Health and Human Services. *The 1997 HHS Poverty Guidelines.* Online at: http://aspe.hhs.gov/poverty/97poverty.htm

U.S. Department of Transportation, Federal Highway Administration, Office of Highway Policy Information. "Highway Statistics Series." http://www.fhwa.dot.gov/ohim/ohimstat.htm

U.S. Department of Transportation, Federal Highway Administration. "Nationwide Personal Transportation Survey." http://www-cta.ornl.gov/npts/1995/doc/index.shtml

U.S. Office of Personnel Management, "Cost of Living Allowances (Nonforeign Areas)" *Federal Register*, v. 62, #54, (March 20, 1997) pp. 13354-13355.

U.S. Office of Personnel Management, "1999 General Schedule and Locality Pay Tables," 1999. http://www.opm.gov/oca/99tables/GSannual/INDEX.HTM)

VirtualRelocation.com. *ReloSmart* online at: http://www.VirtualRelocation.com/Relosmart/

Voicu, Alexander and Lahr, Michael L. "Expenditure Based Inter-Area Cost of Living Index." Paper presented at the 38th Annual Meeting of the Western Regional Science Association, Ojai, CA, Feb 1999.

Walden, Michael L. "How Much Income Variation "Really" Exists Within a State?" *Review of Regional Studies,* v. 27, #3 (1997), pp. 237-150.

Walden, Michael L. and Newmark, Craig M. "Interstate Variation in Teacher Salaries." *Economics of Education Review*, v. 14, #4 (Dec 1995), pp. 395-402.

Watts, Harold W. "Special Panel Suggests Changes in BLS Family Budget Program." *Monthly Labor Review*, (Dec 1980), pp. 3-10.

Wojan, Timothy R. and Maung, Adam C. "The Debate Over State-Level Inequality: Transparent Methods, Rules of Evidence, and Empirical Power." *Review of Regional Studies*, v. 28, #1 (Summer 1998), pp. 63-80.

Zimmerman, Julie N. and Garkovich, Lorraine. "Meeting the Monthly Budget: Earnings and the Cost of Living in Rural Kentucky." Paper presented at the 1998 annual meetings of the Southern Rural Sociological Association, Little Rock AR.

APPENDIX A

DATA AND MAPS FOR INDEPENDENT VARIABLES FOR PENNSYLVANIA COUNTIES

This appendix includes a table of data for the basic independent variables for all 67 counties, as well as maps.

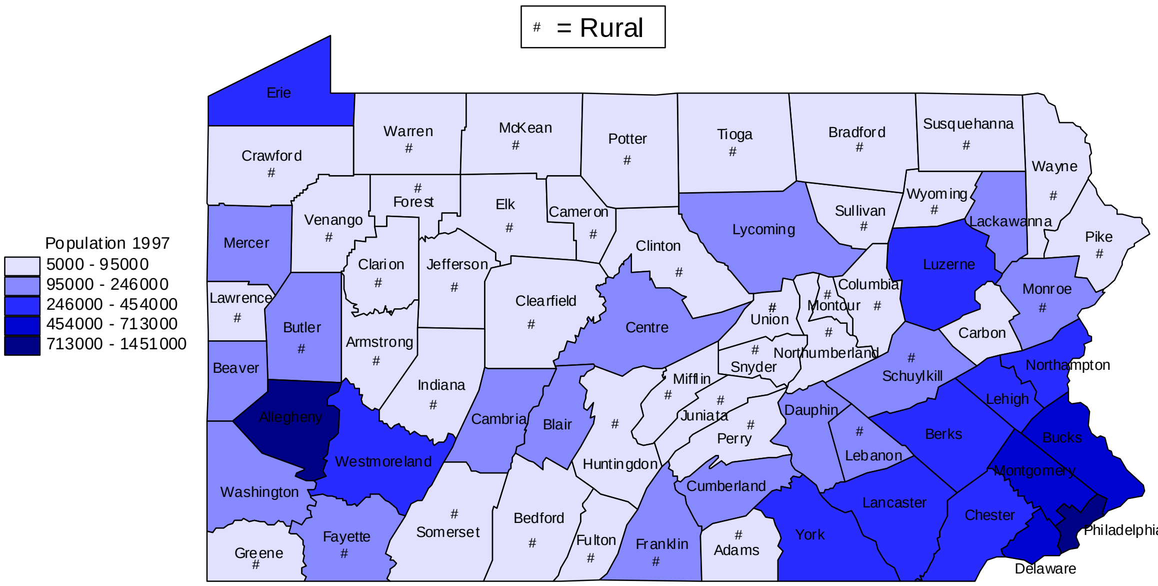
More detailed explanations for the variables, as well as the source of all data, are presented in the body of the text.

Table A-1
Data for Independent Variables for Pennsylvania Counties

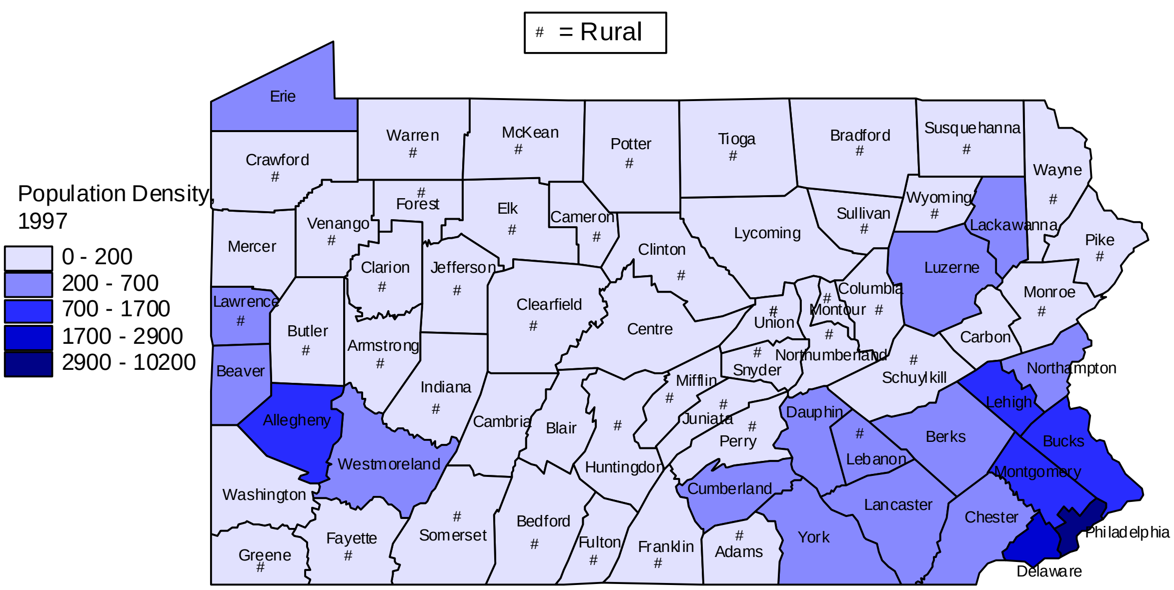
Pennsylvania County	Rural	Population	Density	Personal Income	Personal Income per Capita	Population Growth	Personal Income Growth	Pers Income per Capita Growth
Period:		1997	1997	1997	1997	1992-97	1992-97	1992-97
Unit:		people	people per sq mile	thousands of dollars	dollars	%	%	%
Adams	R	85,612	164	1,844,120	21,540	5.51	24.95	18.43
Allegheny		1,280,873	1,720	37,565,476	29,328	-3.89	18.61	23.40
Armstrong	R	73,436	111	1,479,668	20,149	-0.66	22.79	23.61
Beaver		185,585	418	4,033,082	21,732	-1.21	25.05	26.58
Bedford	R	49,269	48	846,210	17,175	1.57	24.22	22.30
Berks		353,925	409	9,220,126	26,051	3.16	26.10	22.23
Blair		130,678	248	2,676,594	20,482	-0.50	23.10	23.72
Bradford	R	62,366	54	1,157,800	18,565	1.11	19.72	18.41
Bucks		582,511	936	17,921,883	30,767	5.15	31.62	25.17
Butler	R	168,897	213	3,776,972	22,363	7.93	27.22	17.88
Cambria		157,372	227	3,117,885	19,812	-3.05	17.45	21.14
Cameron	R	5,681	14	126,759	22,313	-2.10	26.41	29.12
Carbon		58,878	151	1,214,873	20,634	1.67	25.44	23.38
Centre		132,809	119	2,792,727	21,028	4.45	29.19	23.69
Chester		415,422	547	16,079,962	38,708	7.46	38.62	28.99
Clarion	R	41,813	69	810,479	19,383	-0.76	25.41	26.37
Clearfield	R	80,907	70	1,544,598	19,091	3.26	24.88	20.94
Clinton	R	36,952	41	677,694	18,340	-1.12	19.87	21.22
Columbia	R	64,079	131	1,257,317	19,621	0.04	20.90	20.85
Crawford	R	89,283	86	1,716,648	19,227	2.83	25.00	21.55
Cumberland		207,801	377	5,672,139	27,296	3.50	25.50	21.26
Dauphin		245,713	441	6,681,658	27,193	1.73	27.19	25.03
Delaware		543,560	2,935	16,552,602	30,452	-0.99	21.40	22.61
Elk	R	34,818	42	812,175	23,326	-0.73	25.01	25.93
Erie		277,593	343	6,140,439	22,120	-0.28	21.28	21.62
Fayette	R	145,126	182	2,709,778	18,672	-0.62	20.56	21.31
Forest	R	4,965	12	80,072	16,127	2.50	24.04	21.02
Franklin	R	127,462	165	2,835,858	22,249	2.77	26.95	23.52
Fulton	R	14,522	33	262,844	18,100	2.61	34.45	31.03
Greene	R	40,693	70	693,373	17,039	2.55	22.08	19.04
Huntingdon	R	44,977	51	727,950	16,185	1.35	20.60	18.99
Indiana	R	89,277	107	1,710,512	19,160	-1.05	17.14	18.39
Jefferson	R	46,504	71	922,279	19,832	0.47	20.84	20.27
Juniata	R	21,914	56	395,639	18,054	4.26	18.80	13.95
Lackawanna		210,494	453	4,796,961	22,789	-3.28	18.08	22.08
Lancaster		453,826	461	11,206,930	24,694	4.58	25.19	19.71
Lawrence	R	95,281	263	1,910,497	20,051	-1.20	21.77	23.24
Lebanon	R	117,172	323	2,700,046	23,043	1.43	25.09	23.32
Lehigh		298,287	856	8,253,236	27,669	0.86	24.08	23.02
Luzerne		317,055	350	7,142,149	22,527	-3.69	18.89	23.44
Lycoming		118,212	95	2,377,399	20,111	-1.81	16.62	18.77
McKean	R	46,805	48	962,092	20,555	-1.69	22.59	24.69
Mercer		122,042	179	2,434,743	19,950	0.12	20.75	20.61
Mifflin	R	46,992	114	824,200	17,539	0.83	20.97	19.97
Monroe	R	122,706	199	2,575,073	20,986	17.51	39.02	18.31
Montgomery		713,135	1,463	28,702,998	40,249	3.42	28.06	23.82
Montour	R	17,847	135	496,686	27,830	-1.00	21.09	22.31
Northampton		257,501	682	6,367,005	24,726	2.38	24.93	22.03
Northumberland	R	94,984	199	1,883,452	19,829	-1.07	18.78	20.07
Perry	R	44,149	79	869,602	19,697	4.61	26.07	20.51
Philadelphia		1,450,683	10,171	33,053,641	22,785	-6.35	11.10	18.64
Pike	R	39,129	69	778,514	19,896	20.71	38.21	14.51
Potter	R	17,137	16	338,368	19,745	1.79	28.25	26.00
Schuylkill	R	149,423	191	3,084,930	20,646	-3.00	17.87	21.51
Snyder	R	38,166	115	899,137	23,559	2.31	25.50	22.67
Somerset	R	80,497	74	1,527,103	18,971	2.33	16.62	13.96
Sullivan	R	6,109	14	108,843	17,817	-0.18	12.72	12.92
Susquehanna	R	42,095	51	781,926	18,575	2.55	15.79	12.91
Tioga	R	41,547	37	772,385	18,591	-0.10	23.00	23.12
Union	R	41,650	131	821,586	19,726	14.19	22.70	7.46
Venango	R	58,175	85	1,390,472	23,902	-1.90	24.14	26.54
Warren	R	44,167	49	977,345	22,128	-1.90	15.99	18.23
Washington		205,632	239	5,076,254	24,686	-0.17	28.69	28.91
Wayne	R	45,180	60	859,125	19,016	6.80	23.10	15.26
Westmoreland		373,711	361	8,766,682	23,458	-0.19	22.12	22.35
Wyoming	R	29,304	72	573,782	19,580	0.56	17.39	16.74
York		370,912	407	8,953,241	24,138	5.84	23.75	16.92

Pennsylvania County	Period: Rural	1991-92 Government Revenue per Worker	1997 Electric Rate Residential	1997 Gas Rate Residential	1990 Mean One-way Commute Time	1997 Unemployment Rate
	Unit:	thou dollars per worker	cents per Kilowatt hour	dollars per thou cubic ft.	minutes	%
Adams	R	39.508	9.90	8.33	20.37	4.73
Allegheny		49.190	9.90	8.33	22.65	4.39
Armstrong	R	33.108	9.90	8.33	21.33	7.58
Beaver		40.464	9.90	8.33	21.10	5.09
Bedford	R	25.129	9.90	8.33	20.20	7.33
Berks		43.226	9.90	8.33	18.18	4.28
Blair		28.853	9.90	8.33	15.89	5.61
Bradford	R	35.598	9.90	8.33	16.57	5.58
Bucks		56.179	9.90	8.33	23.63	4.21
Butler	R	39.980	9.90	8.33	20.45	4.93
Cambria		29.513	9.90	8.33	17.83	7.96
Cameron	R	30.642	9.90	8.33	11.29	6.66
Carbon		32.791	9.90	8.33	21.72	7.34
Centre		41.816	9.90	8.33	16.47	3.02
Chester		56.972	9.90	8.33	23.09	3.20
Clarion	R	24.743	9.90	8.33	17.96	6.52
Clearfield	R	25.496	9.90	8.33	18.42	8.19
Clinton	R	49.150	9.90	8.33	18.19	7.73
Columbia	R	30.268	9.90	8.33	17.29	7.31
Crawford	R	37.095	9.90	8.33	17.18	5.67
Cumberland		43.161	9.90	8.33	17.70	2.88
Dauphin		45.731	9.90	8.33	18.58	3.49
Delaware		53.330	9.90	8.33	24.04	4.55
Elk	R	36.379	9.90	8.33	13.84	6.02
Erie		38.415	9.90	8.33	16.15	5.88
Fayette	R	23.722	9.90	8.33	20.67	8.40
Forest	R	30.348	9.90	8.33	19.04	10.16
Franklin	R	31.638	9.90	8.33	17.78	4.84
Fulton	R	23.402	9.90	8.33	24.27	7.82
Greene	R	31.314	9.90	8.33	22.17	8.95
Huntingdon	R	27.210	9.90	8.33	20.79	10.42
Indiana	R	27.562	9.90	8.33	18.26	8.07
Jefferson	R	32.176	9.90	8.33	16.52	7.12
Juniata	R	23.861	9.90	8.33	24.11	8.00
Lackawanna		44.517	9.90	8.33	16.53	7.09
Lancaster		47.488	9.90	8.33	17.49	2.99
Lawrence	R	31.471	9.90	8.33	17.74	6.37
Lebanon	R	33.532	9.90	8.33	18.16	3.33
Lehigh		51.282	9.90	8.33	18.12	4.67
Luzerne		35.280	9.90	8.33	17.81	7.24
Lycoming		37.226	9.90	8.33	16.16	6.51
McKean	R	28.125	9.90	8.33	15.29	6.62
Mercer		30.925	9.90	8.33	16.38	5.32
Mifflin	R	38.710	9.90	8.33	17.52	7.13
Monroe	R	46.418	9.90	8.33	23.27	7.94
Montgomery		60.810	9.90	8.33	21.82	3.71
Montour	R	98.359	9.90	8.33	14.93	4.33
Northampton		43.382	9.90	8.33	20.04	5.04
Northumberland	R	27.667	9.90	8.33	18.64	6.98
Perry	R	30.441	9.90	8.33	29.32	3.90
Philadelphia		42.225	9.90	8.33	26.90	6.84
Pike	R	45.615	9.90	8.33	26.54	5.32
Potter	R	28.826	9.90	8.33	18.01	7.39
Schuylkill	R	31.208	9.90	8.33	19.92	7.64
Snyder	R	32.610	9.90	8.33	16.78	4.90
Somerset	R	26.421	9.90	8.33	17.90	7.43
Sullivan	R	26.874	9.90	8.33	20.08	6.94
Susquehanna	R	25.146	9.90	8.33	20.69	7.08
Tioga	R	24.617	9.90	8.33	18.25	6.67
Union	R	36.304	9.90	8.33	15.44	4.51
Venango	R	31.069	9.90	8.33	16.90	6.74
Warren	R	33.180	9.90	8.33	16.69	5.33
Washington		35.599	9.90	8.33	21.61	5.43
Wayne	R	41.267	9.90	8.33	19.02	8.32
Westmoreland		43.087	9.90	8.33	21.12	5.63
Wyoming	R	35.443	9.90	8.33	20.63	8.40
York		44.874	9.90	8.33	19.67	4.11

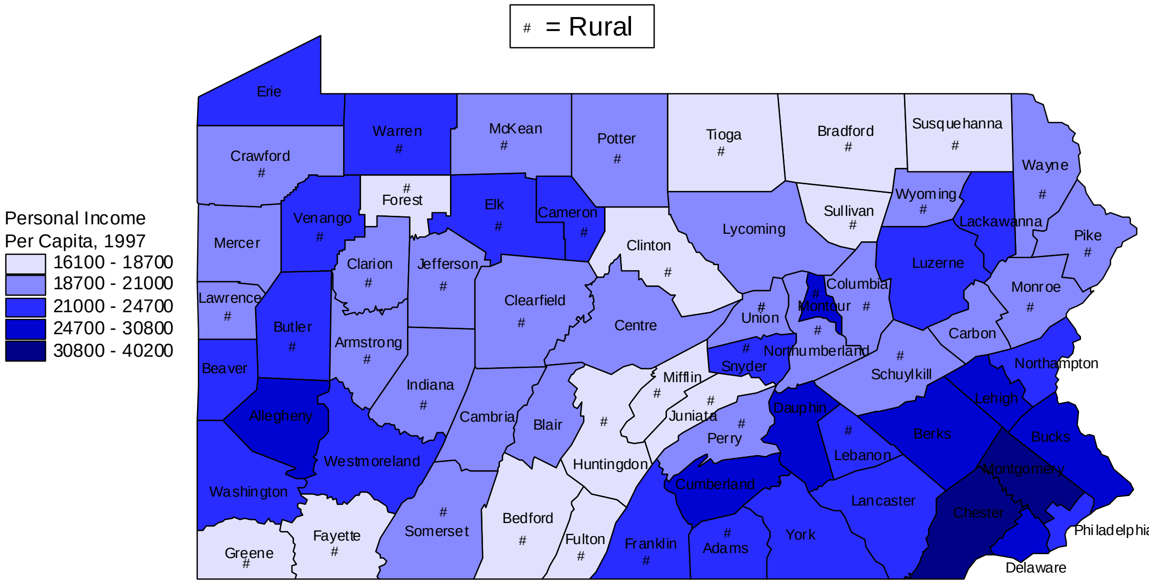
Map A-1
Population, 1997



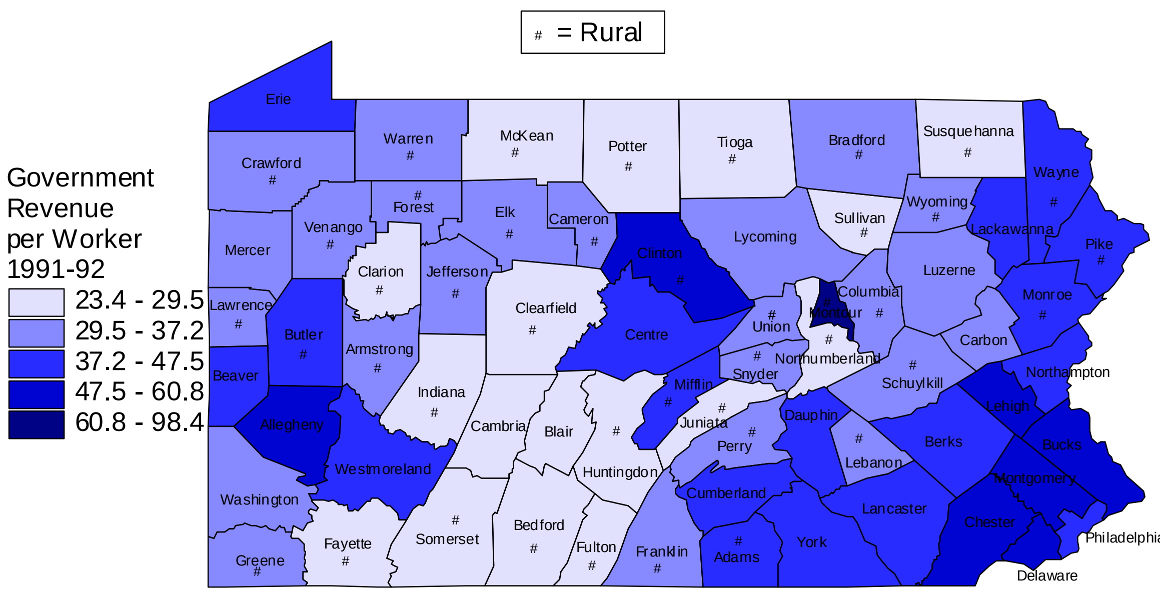
Map A-2
Population Density, 1997



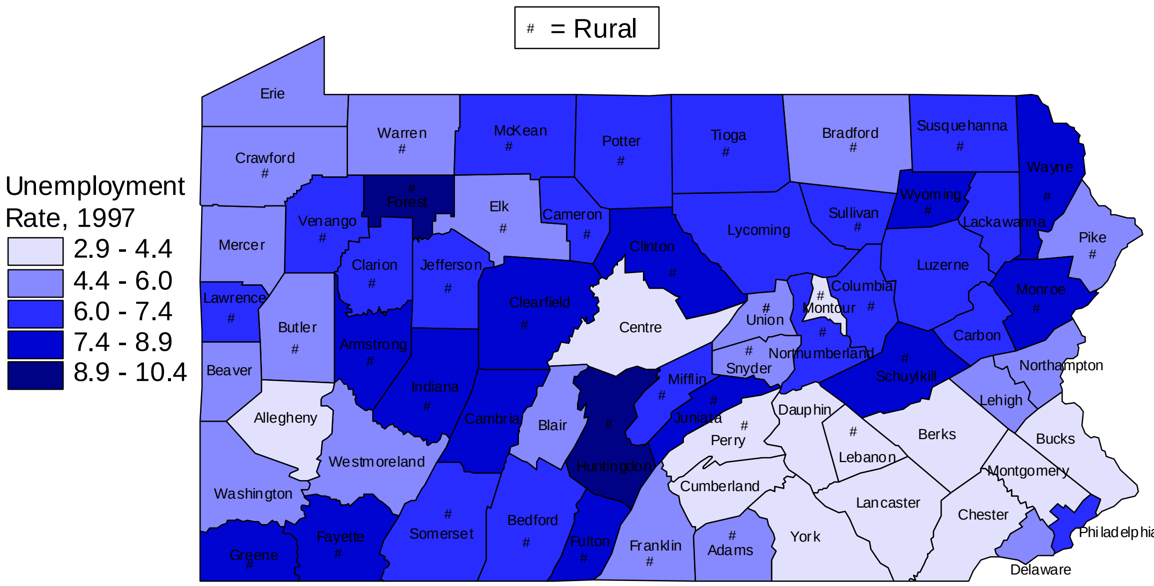
Map A-3
Personal Income Per Capita, 1997



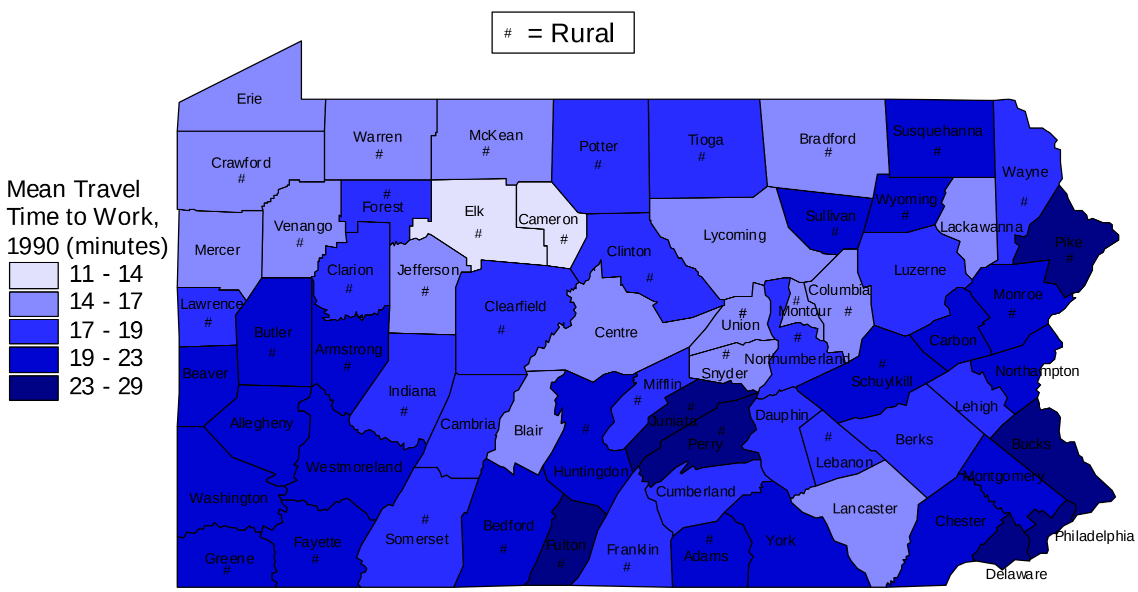
Map A-4
Government Revenue per Worker, 1991-92



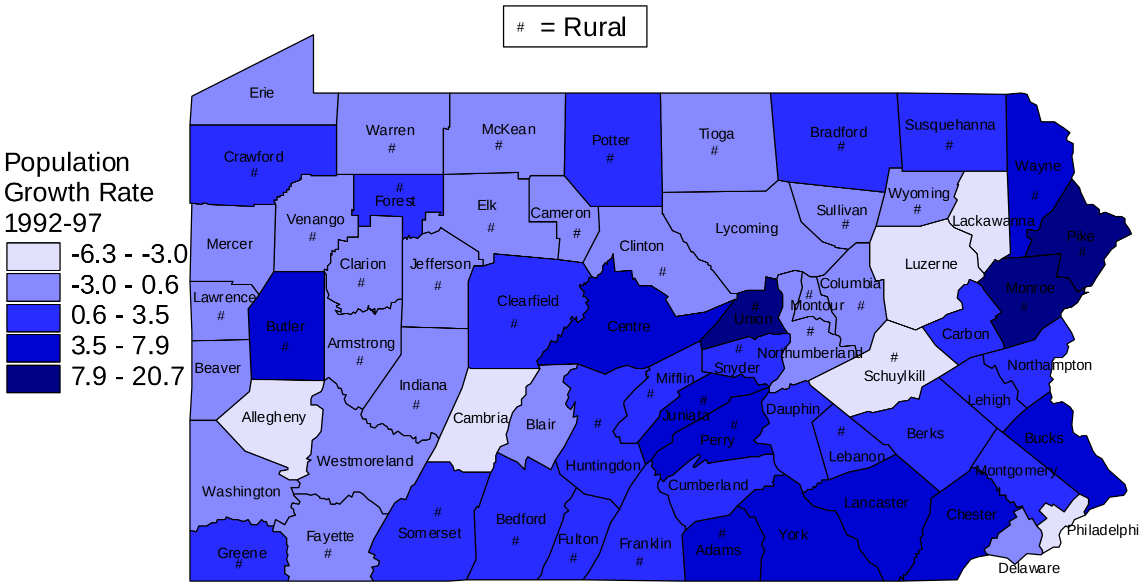
Map A-5
Unemployment Rate, 1997



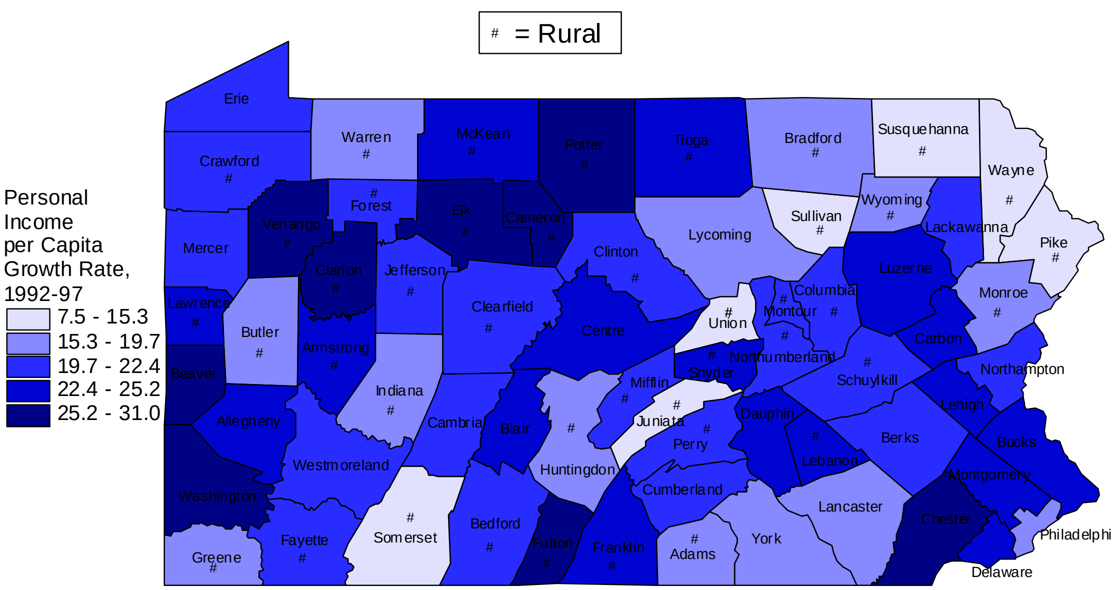
Map A-6
Average Travel Time to Work, 1990



Map A-7
Population Growth Rate, 1992-97



Map A-8
Personal Income per Capita Growth Rate, 1992-97



APPENDIX B
ACCRA MARKET BASKET FOR 1997

This appendix presents a list of the items that comprised the ACCRA market basket for a midmanagement executive household in the third quarter of 1997. These are not the precise item definitions; more exact definitions are given in the pricing manual to ensure that every area gathers price data on the same items.

Weight	Item
16.0000%	**GROCERY ITEMS**
0.5840	T-bone steak, price per pound
0.5840	Ground beef or hamburger, price per pound, lowest price
0.7600	Sausage, price per pound, Jimmy Dean brand, 100% pork
0.7584	Frying chicken, price per pound, whole fryer
0.5648	Tuna, 6.0-6.125 oz. can, chunk light, Starkist or Chicken of the Sea
0.6176	Whole milk, half-gallon carton
0.1280	Eggs, one dozen grade A large
0.4720	Margarine, one pound cubes, Blue Bonnet or Parkay
0.4720	Parmesan cheese, 8 oz., grated, Kraft
0.3472	Potatoes, 10 lb. sack, white or red
0.7056	Bananas, price per pound
0.3472	Iceberg lettuce, approx. 1.25 pound head
1.4608	White bread, 24 oz. loaf or 24-oz. equivalent, lowest price
1.0768	Cigarettes, carton, king size (85 mm.), Winston
1.1024	Coffee, 13 oz. can, vacuum packed, Maxwell House, Hills Brothers or Folgers
0.5328	Sugar, 4 lb., cane or beet, lowest price
0.7328	Corn flakes, 18 oz., Kellogg's or Post Toasties
0.0928	Sweet peas, 15-17 oz. can, Del Monte or Green Giant
0.0928	Tomatoes, 14.5 oz. can, Hunt's or Del Monte
0.2752	Peaches, 29 oz. can, halves or slices, Hunt's, Del Monte or Libby's
0.6048	Facial tissues, 175-count box, Kleenex brand
0.6656	Dishwashing powder, 50 oz., Cascade
0.3456	Shortening, 3 lb. can, all vegetable, Crisco
0.8304	Frozen orange juice, 12 oz. can, Minute Maid
0.1840	Frozen corn, 16 oz., whole kernel, lowest price
0.0689	Baby food, 4.0-4.5 oz. jar, strained vegetables, lowest price
0.5600	Soft drink, 2 liter, Coca Cola, excluding any deposit
28.0000%	**HOUSING**
5.3816	Apartment, monthly rent; two bedroom, unfurnished, excluding all utilities except water, 1½ or 2 baths, approx. 950 sq. ft.
22.6184	Home purchase, consisting of monthly principal and interest payment on a 30 year first mortgage, based on 75% loan with current conventional fixed rate mortgage, on 1,800 sq. ft. living area new house, with approximately 8,000 sq. ft. lot, in appropriate urban area with all utilities

Weight	Item
8.0000%	**UTILITIES**
6.9280	Total energy costs at current rates for average monthly consumption of all types of energy during the previous 12 months for the type of home specified above
1.0720	Telephone, private residential line, customer owns instruments. Price includes: basic monthly rate; additional local user charges, if any, for a family of four; touch tone fee; all other mandatory monthly charges, such as long distance access fee and 911 fee; and all taxes on foregoing.
10.0000%	**TRANSPORTATION**
1.0700	Commuting fare, typical one way, up to 10 miles
3.7000	Auto maintenance, average price for computer or spin balance of one front wheel
5.2300	Gasoline, one gallon, unleaded regular, national brand, cash price at self-service pump, including all taxes
5.0000%	**HEALTH CARE**
0.8750	Hospital room, semi-private, average cost per day
1.7545	Doctor, office visit, general practitioner's routine exam of established patient, average charge
1.7545	Dentist, office visit, adult teeth cleaning and periodic oral exam
0.6160	Antibiotic ointment, ½ oz. Tube, Polysporin
33.0000%	**MISCELLANEOUS GOODS AND SERVICES**
3.0822	Hamburger sandwich, quarter pound patty with cheese, McDonald's
3.0822	Pizza, 11-12" thin crust cheese pizza, Pizza Hut or Pizza Inn
3.0822	Fried chicken, thigh and drumstick, Kentucky Fried Chicken or Church's
0.6171	Man's barbershop haircut, no styling
0.6171	Woman's beauty salon visit, including shampoo, trim and blow-dry
0.6171	Toothpaste, 6-7 oz. tube, Crest or Colgate
0.6171	Shampoo, 15 oz. bottle, Alberto VO5
0.6171	Dry cleaning, man's two-piece suit
4.3131	Man's dress shirt, 100% cotton pinpoint Oxford, long sleeves
1.3629	Boy's underwear, three briefs, size 10-14, cotton, lowest price
4.3131	Man's denim jeans, Levi's brand, 501 or 505, rinsed or washed or bleached, size 28/30-34/36
1.5873	Major appliance repair, home service call, washing machine, minimum labor charge excluding parts
0.9438	Newspaper subscription, daily and Sunday home delivery of large city paper, monthly rate
1.3365	Movie, first run, indoor, evening, no discount
1.3365	Bowling, average price per game, evening rate
2.2638	Tennis balls, can of three extra-duty, yellow, Wilson or Penn brand
1.5213	Child's game, "Monopoly", No. 9 edition
0.5643	Liquor, J&B Scotch, 750 ml. bottle
0.5610	Beer, 6-pack of 12 oz. containers, Miller Lite or Budweiser, excluding deposit
0.5643	Wine, 1.5 liter bottle Chablis blanc, Gallo
100.0000%	**TOTAL**

APPENDIX C
FURTHER STATISTICAL TESTS

The goal of this project is to generate estimates of the cost of living for all counties of the state, using a statistical model that avoids the need for primary data collection. To do this, we use economic theory to identify a number of factors that may be logically expected to have an effect on COL levels from place to place. The estimations presented in the body of the text identify equations that yield the best estimates based on those variables. Moreover, examination of the signs of those variables will tell something about the size and type of impact each one has.

It may be noticed that many of the coefficients in the equations are rather small. This results when the scale of the independent variable is large, as was the case for aggregate income that ran into the hundreds of billions of dollars for some areas. But how can we be sure that these coefficients are not so small as to be negligible--not close enough to zero to ignore? In technical terms, we are asking if the coefficients are *statistically significant*, or *statistically different from zero*. There are standard statistical tests to answer this question. The most common is to build a confidence interval around the coefficient. This approach is based on the idea that we would expect to get slightly different coefficients for the independent variables if we used slightly different samples of counties. Thus, any one estimate may not precisely measure what we might call the "true effect"--the effect that the variable actually exerts on the COL of all areas nationwide. However, we expect that each estimate is reasonably close to that true value; we can be about 95% sure that the real value lies within two standard deviations above or below our estimated coefficient. In other words, 95 times out of 100 samples, a two-standard-deviation band around the estimated coefficient will include the true value.

If the value “zero” is not within that two-standard-deviation band, then we can be 95% sure that the effect of this variable is not zero; i.e., that this variable *does* have an effect on the COL. In such a case, we say that the variable (or more accurately, the coefficient) is statistically significant. A quick test of statistical significance is to divide the coefficient by its standard deviation (yielding a value called a t-statistic). If the result is greater than two in absolute value, then the two-standard-deviation band does not include the value zero, and the coefficient is statistically significant. The asterisks in the equations in Table 3 in the body of the text identify the variables whose coefficients were statistically significant at the 95% level. However, this procedure will give accurate answers only if certain pre-conditions are met. These include two that we need to discuss for our sample.

A. MULTICOLLINEARITY

The first pre-condition is that the independent variables not be correlated with each other. Each coefficient in the equation is supposed to measure the effect of that single variable on the COL, with all other variables held constant. For example, in the estimating equation for the total budget:

$$\begin{aligned}1997TOT = 96.548 + 0.835\ POPG9697 + 0.00299\ DEN - 1.81(e\text{-}8)\ DENSQ \\ + 0.00312\ GCST + 1.221\ ELEC + 0.330\ NE - 8.504\ MA - 9.498\ SA \\ - 11.291\ ESC - 15.069\ WSC - 9.049\ ENC - 9.801\ WNC - 6.469\ MTN\end{aligned}$$

the 0.835 coefficient on population growth between 1996 and 1997 (POPG9697) means that an increase of one percent in population growth in the 1996-97 period would result in a 0.835 increase in the COL index, all other variables held constant, or that the COL would rise one point for every increase of 1.20% (the inverse of 0.835) in the population growth rate. However, if population growth were highly correlated with government cost (GCST) in our sample of areas, such that areas with higher population growth always tended to have higher government cost, then it would be impossible in this sample to disentangle the effects of the two variables. The coefficient of the population growth variable could not be trusted to truly measure the effect of population growth alone, since the government cost and population growth effects can't be separated. The two variables are said to be intercorrelated or, synonymously, multicollinearity is said to exist.

In technical terms, multicollinearity causes the standard error of the coefficient to be greater, resulting in t-statistics that are statistically insignificant, although the overall fit of the equation is relatively high. Individual variables seem to be unimportant, but a set of independent variables together have a significant effect on the dependent variable. The coefficients of individual variables are more likely to have incorrect signs, and to change dramatically with additions or deletions of other independent variables from the model.

It should be pointed out that while multicollinearity may make the individual coefficients suspect, it does *not* affect the ability of the overall equation to provide good estimates of the COL for counties not in the sample. Our goal in this project is to derive such estimates, not to determine exactly how much of the COL effect is due to each individual variable. If the coefficients of POPG9697 and GCST interact to give the appropriate effect on COL, that's all we need for the purposes of this project.

Nevertheless, does multicollinearity exist in our data set? This question can be answered by examining the pairwise correlations among the independent variables, and by looking at changes in the coefficients as variables are added to or deleted from the model. Table C-1 presents correlation coefficients for the independent variables in the model.

Fortunately, most of the independent variables are not closely correlated. Out of the 78 pairs of variables, only four correlation coefficients are high enough to present a concern. These are shaded in table C-1. The high correlations between POP and AGGINC (.970), and between POPGTH9297 and AGGINCGTH9297 (.887) are unsurprising, since they are alternative measures of the same underlying variable, the amount of demand. For that reason they are not used in the same model. There are also high correlations between DEN and INC (.608) and between AGGINCGTH9297 and INCGTH9297 (.573). For this reason, these pairs are not used in the same models. As a result, multicollinearity is not a problem in our data set.

This impression is reinforced by examination of the reactions of estimated coefficients to addition or deletion of variables in the estimating process. The coefficients did not change dramatically in value or sign, again indicating that multicollinearity is not a problem.

Table C-1
Correlation Coefficients for Independent Variables

	POP	DEN	AGGINC	INC	POPGTH 9297	AGGINC GTH9297	INCGTH 9297	GCST	ELEC	GAS	ATT	UNEMR
POP	1.000											
DEN	.192	1.000										
AGGINC	.970	.391	1.000									
INC	.304	.608	.470	1.000								
POPGTH9297	-.006	-.070	-.021	-.081	1.000							
AGGINCGTH9297	-.025	-.012	-.013	.066	.887	1.000						
INCGTH9297	-.047	.100	.004	.283	.131	.573	1.000					
GCST	.013	.001	.020	.111	.076	.040	-.044	1.000				
ELEC	.228	.232	.256	.177	-.244	-.297	-.216	-.025	1.000			
GAS	.084	.134	.109	.194	-.104	-.079	.011	-.019	.395	1.000		
ATT	.449	.274	.478	.352	.124	.136	.067	-.009	.163	.314	1.000	
UNEMR	.039	.060	.019	-.273	-.029	-.214	-.395	-.033	.198	.086	.094	1.000
PctUnearned	-.108	-.060	-.123	-.141	-.211	-.282	-.225	-.005	.140	.308	-.129	.319

N = 303 for all variables.

B. HETEROSKEDASTICITY

A second pre-condition for our hypothesis tests to work is that the error term must have constant variance; it must not vary with the size of any of the independent variables. If the error is greater for areas with a larger population or income, for example, we have a technical problem called heteroskedasticity.

Heteroskedasticity presents a problem since it results in the standard estimating technique (ordinary least squares) underestimating the standard error of the affected coefficient. Since the standard error of the coefficient is used in the t-test of significance, we might be led to draw an incorrect conclusion about the significance of the variable. Since the effect is to underestimate the true standard error, the t-statistics appear larger than they should be, perhaps causing us to accept a coefficient as statistically significant when in fact it is not.

Notice that this problem only affects the standard error of the coefficient and not the actual estimate of the coefficient itself. For this reason, the estimates provided by a data set where heteroskedasticity exists are still valid, accurate estimates. The problem arises only if we attempt to assess the statistical significance of a particular coefficient. The problem of heteroskedasticity is similar to that of multicollinearity in this regard; neither is a problem for the actual estimates that we seek to generate. They are only a problem if we attempt to interpret individual coefficients, or assess the impact of a specific variable in the estimating equation. Despite the fact that the problems of multicollinearity and heteroskedasticity will not affect the accuracy of the estimates, we examined the datasets for their presence to give the reader full information about the estimation process.

Tests for the presence of heteroskedasticity typically involve examination of the errors (residuals) from the regression estimating equation. After the "best" estimating equation is selected, it is used to generate estimates for each of the 303 areas in the sample. These estimates (usually called "fitted" values) are compared to the actual data values for each area, and the difference is the error or residual unaccounted for by the variables in the regression equation.

If the data set is homoskedastic (i.e., if no heteroskedasticity is present), then the residuals should be approximately constant relative to each independent variable. One quick way to check for heteroskedasticity is to sort the residuals by an independent variable, and plot them. If the spread of residuals increases or decreases noticeably, heteroskedasticity is present. For example, we can sort the 303 areas in the sample by their density levels, from smallest to largest, and then examine the residuals. If the residuals tend to increase in absolute value for the denser areas, then heteroskedasticity is present with respect to the density variable. The same exercise can be performed for each independent variable, of course.

We examined the COL estimating equation for Total Expenditures. For this dependent variable, the residual plots look clean for the independent variables (population growth from 1996-97, density, government revenue per worker and electric rate) with the possible exception of the electric rate variable. The range of residuals may increase slightly with higher values of that variable.

A more formal statistical test for heteroskedasticity is the Goldfeld-Quandt test.[11] It sorts the sample by an independent variable (for example, from lowest to highest in terms of density) and then compares the residuals from an equation fitted to the least dense areas with those from the same equation fitted to the most dense areas. If the residuals from the two samples are significantly

[11] See Doti and Adibi, section 5.3.

different (as measured by the sum of squared residuals), then heteroskedasticity is present. The actual test uses an F-statistic computed as the ratio of the two residual sums of squares. A small F-statistic means the two are not statistically different, so heteroskedasticity is not indicated.

The Goldfeld-Quandt test was run once for each independent variable (population growth 1996-97, density, government cost and electric rate.) For each case, the sample was sorted by that independent variable and then divided into three parts. The top and bottom parts each had 101 areas. Regressions including all of the independent variables--except the region dummy variables--were run on each subsample, and the F-statistic was computed as the ratio of the error sum of squares of the two regressions. In each case, the computed F-statistic was less than the critical F value, which was approximately 1.66 for the 1% level of significance. Since the computed F-statistics were less than the critical value, the conclusion to be drawn is that heteroskedasticity is NOT present in the Total Expenditure equation.

Table C-2
Results of Goldfeld-Quandt Tests
for Total Expenditures

Variable	F statistic
Population growth 1996-97	1.38
Density	1.53
Gov't cost per worker	1.23
Electric rate	1.52
Critical F value (1%):	approx 1.66

C. CHOW TESTS: DOES THE ESTIMATING EQUATION WORK FOR SMALL AREAS?

This project has fitted the estimating equation to a sample of 303 areas, but these 303 areas are not necessarily representative of the 67 Pennsylvania counties for which we wish to derive estimates. After all, the average population of the 303 areas in the database is 319,281 which is far greater than the population of most of Pennsylvania's counties. The model tries to account for this by including all relevant independent variables, as well as including dummy variables for regions.

But the question remains as to whether the final estimating equation fits equally well for large and small areas within the sample. A good way to test this is by using the Chow test.[12] This test involves fitting the estimating equation to a subset of the areas, and then comparing the residuals of this subset with the residuals from the whole sample, to see if they are significantly different. If they are not different, then the equation fits as well for the subset as for the whole sample.

To apply this test, we sorted the sample from largest to smallest area, to see if the equation fits small areas as well as it fits the whole sample. We considered the smallest 25 areas in the sample, with 1997 population from 12,263 to 42,135, and the smallest 50 areas, which included areas with population up to 59,033. For the large areas, we looked at the largest 50, with populations from 504,654 to

[12] See Gujarati (1978), pp. 305-6 and Quantitative Micro Software (1998), pp. 347-350.

9,116,506 and the largest 25, which covered areas with populations of 814,286 and up. Table C-3 reports the results from smallest areas to largest areas. The Chow test results in small F-statistics for the smallest 25 areas and the smallest 50 areas. This means that the estimation errors for the smallest areas were relatively small; the estimating equation from the whole sample of areas fit the small areas well.

Table C-3
Chow Test Results

			1997 Population		
	F	Probability	Average	Minimum	Maximum
Smallest 25 areas	1.345	.131	30,171	12,263	42,135
Smallest 50 areas	1.132	.268	40,853	12,263	59,033
Largest 50 areas	1.278	.117	1,183,689	504,654	9,116,506
Largest 25 areas	1.885	.008	1,727,034	814,286	9,116,506

For the largest 50 areas, the same result obtains. The overall equation gives a good fit when applied just to these 50 areas. On the other hand, the fit for the largest 25 area was not as good. The Chow test indicates that the fitting equation from the overall sample gives relatively large estimation errors when used for the 25 areas with the largest populations.

The conclusion that we can draw is that the estimating equation works for the smaller areas as well as for the whole sample. Since we are most concerned with deriving estimates for the rural areas of the state, which have smaller populations, this is reassuring. The estimating equation even works well for the largest 50 areas as a group, which includes areas up to 800,000 in population. It is only when we consider the largest 25 areas, with populations over 800,000 that the equation doesn’t perform as well.

The Center for Rural Pennsylvania
200 North Third St., Suite 600
Harrisburg, PA 17101
phone (717) 787-9555
fax (717) 772-3587
www.ruralpa.org
1P0700 - 250