Getting "Rural" Right: Poverty Disparities Across Two Dimensions of Rurality

Jonathan Schroeder, Jose Pacas, David Van Riper
Minnesota Population Center & IPUMS, University of Minnesota

Introduction

To study rural areas in the U.S., researchers commonly use metropolitan/non-metropolitan (i.e., metro/nonmetro) classifications rather than the Census Bureau's urban/rural classifications, often mixing the "metro/nonmetro" and "urban/rural" terminology interchangeably. This practice is problematic. For example, as Figure 1 illustrates, under the metro/nonmetro classification, nonmetro areas have higher poverty rates than metro areas, but under the urban/rural classification, the relationship is reversed; the "rural" areas have the lowest poverty rates overall. This sharp discrepancy indicates the importance of how "rural" is defined, certainly for studies of rural poverty but also for the many other settings where discrepancies may occur. This example gives renewed weight to the warning of Isserman (2005): "getting rural right is in the national interest. When we get rural wrong, we reach incorrect research conclusions and fail to reach the people, places, and businesses our governmental programs are meant to serve." (p. 466).

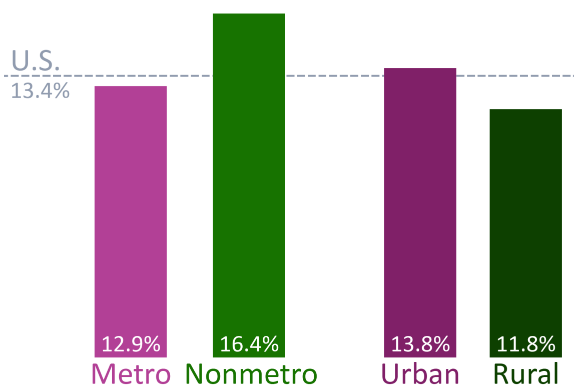

Figure 1. Poverty rates using standard metropolitan/non-metropolitan and urban/rural classifications. *2017 American Community Survey 1-Year Summary File, via Manson et al. 2018.*

In this paper, we present a more robust framework for distinguishing populations across the urban/rural spectrum, and we use the framework to examine how poverty relates to rurality. Instead of relying on any single, existing classification scheme, we compute two continuous indicators—*average tract density* and *average CBSA (core-based statistical area) population*—which correspond to two distinct dimensions of settlement patterns: "concentration" (the local intensity of settlement) and "metropolitanness" (the size of the commuting system). We compute both indicators for all U.S. Public Use Microdata Areas (PUMAs), which are the smallest geographic units identified in public-use census microdata. This enables us to investigate how poverty status relates to two dimensions of rurality while controlling for the demographic characteristics of individuals across the urban/rural spectrum.

We find that poverty rates vary across multiple dimensions of settlement patterns in nonlinear ways: poverty rates are lowest in moderately dense parts of major metropolitan areas, and they are high in both low-density and high-density areas, as well as in moderately dense areas with low "metropolitanness." We also find that the correlation between poverty and different demographic characteristics vary considerably across these different areas. The nonlinearity across multiple dimensions helps to explain the discrepancy between the standard one-dimensional classifications illustrated in Figure 1. Our findings also demonstrate the importance of developing models that explicitly account for continuous degrees of rurality across multiple dimensions.

Limitations of Existing Indices

When "nonmetro" and "rural" are used interchangeably, researchers are conflating two concepts that, while strongly related, differ in important ways. The official rural definition encompasses all population residing outside of urban areas, where urban areas are defined as concentrated settlements with at least 2,500 residents. Metro areas (officially, "metropolitan statistical areas") are defined by the Office of Management and Budget (OMB) as one of two types of "core-based statistical areas" (CBSAs) along with micropolitan statistical areas (micro areas). Each CBSA is defined as a set of counties where a substantial share of workers commute to the same core urban area(s). The urban cores of metro areas have at least 50,000 residents, while the urban cores of micro areas have between 10,000 and 50,000 residents. Urban areas include many "nonmetro" small cities, and large portions of metropolitan areas lie outside of urban areas and are therefore officially "rural" (Figure 2).

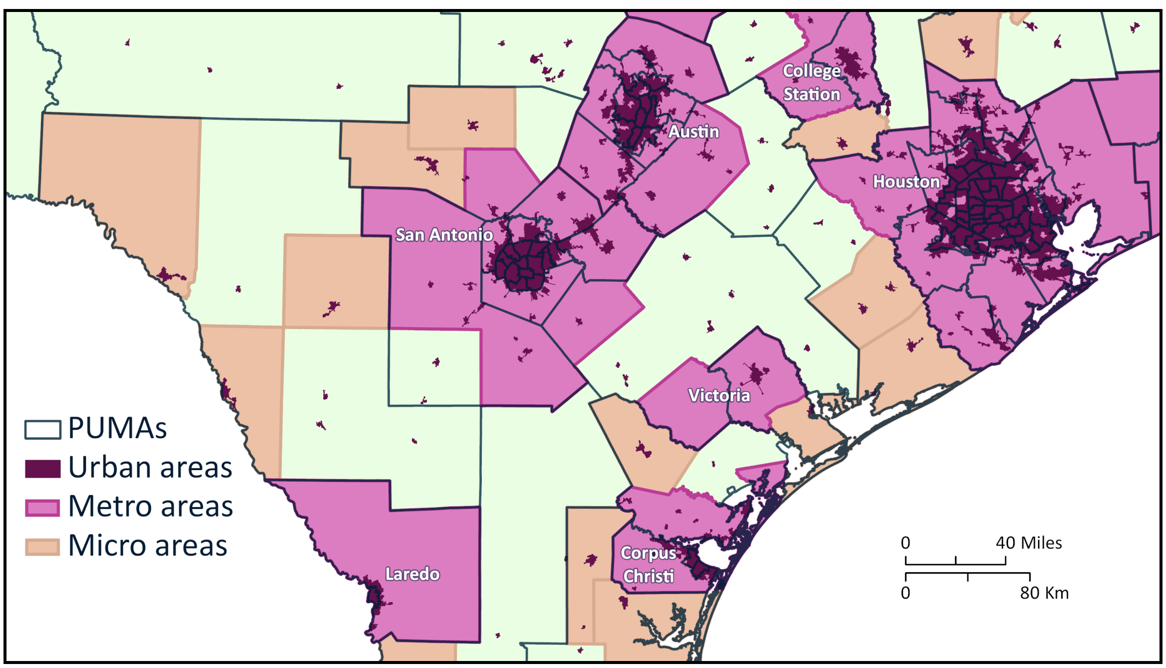

Figure 2. Official definitions of 2010 urban areas and 2013 CBSAs (metropolitan and micropolitan areas) in a section of south-central Texas.

In many settings, it is easy to imagine reasons why one or the other classification should be preferable. For example, nonmetro populations—even those in small urban areas—are generally more remote from major services and employment hubs than rural metro populations are, so in settings where such "remoteness" is a primary concern, it is understandable that "nonmetro" would be used as a stand-in for "rural." Conversely, the official urban/rural definition should generally be preferable in cases where the local population density is more important than remoteness.

Even when there are good reasons to select one classification, however, restricting an analysis to only one of the two standard classifications can be problematic for three major reasons:

1. As Isserman (2005) noted, these classifications correspond to two related but distinct dimensions of rurality, and findings that are true for one dimension need not be true for the other, as in the case of rural poverty (Figure 1). Researchers should therefore consider both dimensions separately before making general assertions about rural populations or at least carefully qualify any assertions that might not hold true across both dimensions.
2. As Waldorf (2006) noted, the standard classifications—because they are *classifications*—mask the underlying *continuous* variation in rurality and impose inflexible thresholds that may be ill-suited for any given analysis. The *degree* to which a place is remote or sparsely populated may be important, and there may be significant associations between rurality and other phenomena that occur *within* the standard classes.
3. As both Isserman (2005) and Waldorf (2006) noted, the standard classifications are based on specific geographic units, which restricts both the scope and quality of information available for them. CBSAs are county-based, which is useful given the relatively large variety of data available for counties, but counties are also spatially coarser than actual commuting systems, so CBSAs are often over-bounded and sometimes under-bounded. Urban and rural areas on the other hand have uniquely complex extents that do not generally align with other units, resulting in a relative paucity of data on "officially rural" populations.

We are not the first to offer alternatives to the standard classifications. Isserman (2005) specified a county-based 4-class rural-urban density typology to complement the metro/nonmetro classification. Waldorf (2006) specified a continuous, multi-dimensional Index of Relative Rurality (IRR), which averages 4 measures that can be computed for any geographic units. The Economic Research Service (ERS) of the USDA has produced a variety of alternative rural classifications, mainly for counties but also for census tracts and ZIP Code Tabulation Areas (U.S. Deparment of Agriculture 2018[1]).

None of these options address all three of the problems identified above. Isserman's typology addresses the first problem by indexing an urban/rural dimension distinct from the metro/nonmetro classification, but it remains a coarse classification with fixed thresholds (problem #2). Waldorf's IRR is helpfully continuous, but it collapses multiple dimensions of rurality into one index, preventing separate analysis of each dimension. The ERS classifications include a broader range of categories than the standard classifications, and they are available for more geographic levels, but they are still fixed classifications and available only for a fixed set of units.

[1] https://www.ers.usda.gov/topics/rural-economy-population/rural-classifications/

Existing classifications and indices offer especially little help for researchers analyzing census microdata. The only sub-state geographic units identified in U.S. public-use microdata are PUMAs, which are required to include at least 100,000 residents each in order to protect confidentiality. This population limit enables the delineation of many PUMAs within large metro areas, but elsewhere, PUMAs commonly span multiple counties. As a result, many PUMAs include a mix of metro and nonmetro counties, and even more PUMAs contain both urban and rural areas (Figure 2). This makes it impossible to determine the metro/nonmetro or urban/rural status of all individuals in microdata. IPUMS USA (https://usa.ipums.org), which harmonizes and redistributes U.S. census microdata, provides a "METRO" variable that distinguishes three classes of metro status—metro, nonmetro, and indeterminable (mixed)—based on how each PUMA corresponds to metropolitan areas (Figure 3). A similar ERS scheme identifies all PUMAs as either metro or nonmetro, allocating each "mixed" PUMA to one of these classes based on where the majority of PUMA residents live (U.S. Department of Agriculture 2018).

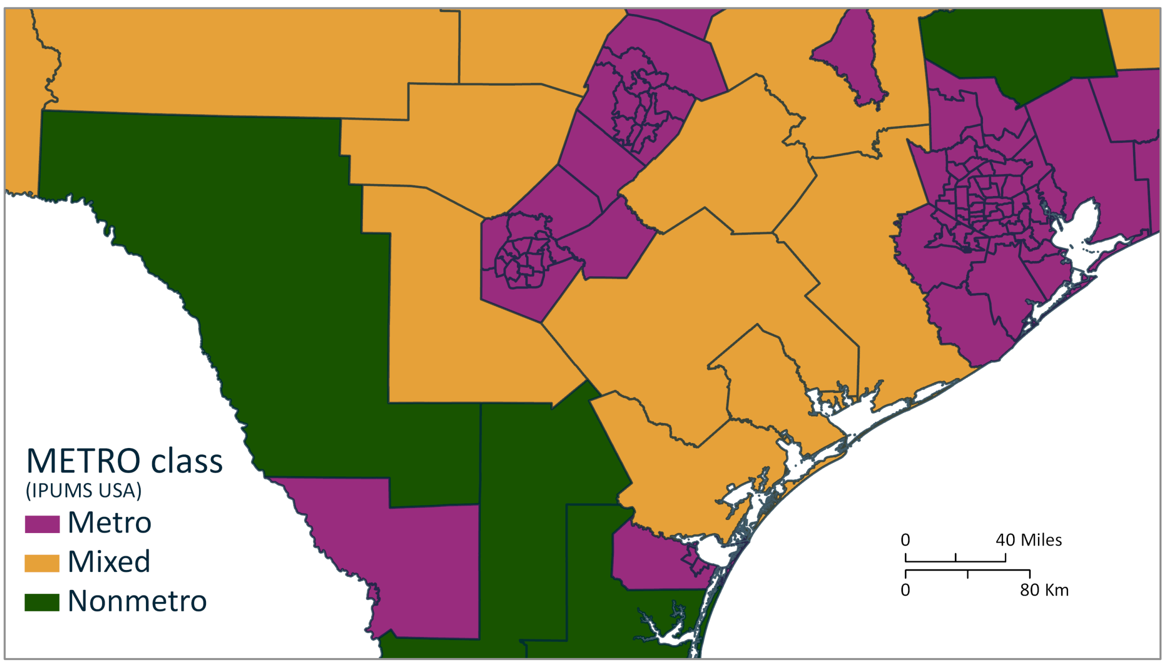

Figure 3. IPUMS-USA METRO classes for PUMAs in a section of south-central Texas.

Neither IPUMS USA nor the ERS supply any PUMA-based indicators of urban/rural status, so researchers studying rural populations face the following dilemma. In order to find data on "officially rural" areas, the researchers must rely on Census Bureau summary files, which limits the researcher to Census-defined tables. To produce custom cross-tabulations or to measure associations at the individual or household level, a researcher may turn to microdata but is then hampered by public-use microdata's restricted geographic detail. Existing PUMA classifications enable the identification of metro/nonmetro status in microdata, but the match with official metro areas is inexact, and even if it were not, the classification is still limited to only 2 or 3 classes along one dimension of rurality. Meanwhile, there are currently no readily available sources of urban/rural status information for microdata.

Continuous Measures of Settlement Patterns

Figure 3 illustrates how we conceptualize "rurality" as a condition that varies continuously in two dimensions. This model resembles a previous model first proposed by Coombes and Raybould (2001), which identifies three dimensions of settlement patterns: *settlement size* (from hamlet to metropolitan), *concentration* (from sparse to dense), and *accessibility* (from remote to central). Here, we re-use the *concentration* dimension, but essentially collapse Coombes and Raybould's other two dimensions into one: *metropolitanness*. The *concentration* dimension corresponds to local population density and, approximately, to the Census's urban/rural classification: areas with dense concentrations of population are more urban and areas with sparse populations are more rural. The second dimension, *metropolitanness*, corresponds to the functional size of the commuting system in a given location, which is also the basis for how metropolitan areas are defined: places where most workers commute to (or within) a large urban area are more metropolitan, and places where people work in rural areas or small urban areas are more remote and non-metropolitan.

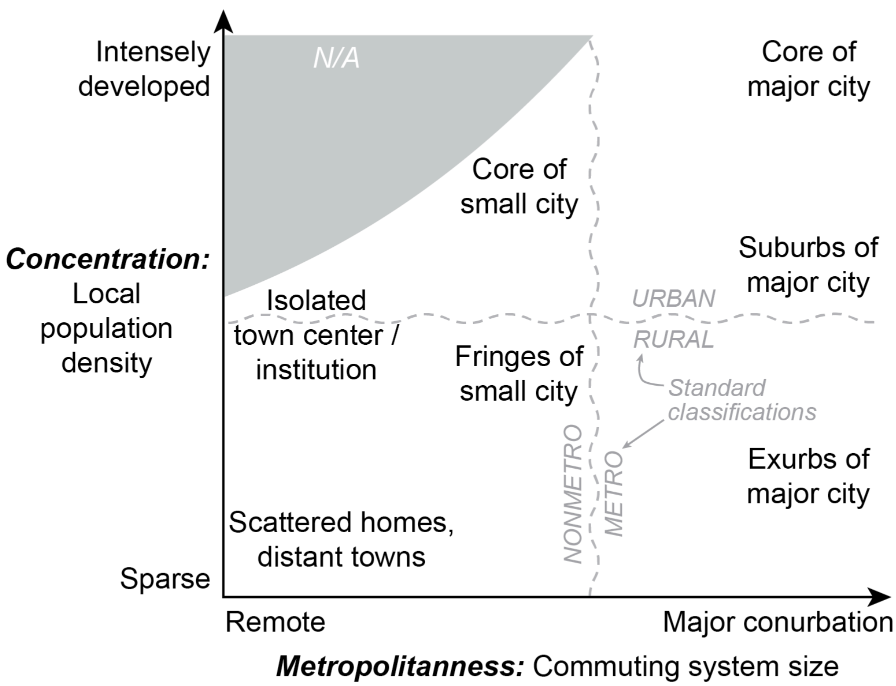

Figure 3. Conceptual model of two continuous dimensions of rurality.

Scattered homes between distant towns (bottom-left quadrant) are unambiguously rural along both dimensions. An isolated town center (left-hand side) is less rural (more urban) along the concentration dimension but is no more metropolitan than are scattered homes. Exurban large-lot developments (lower-right quadrant) are also less "rural" than scattered, remote homes, but mainly in their metropolitanness and less so in their concentration. We expect the upper-left corner to be empty because high local population densities require substantially large populations, which can occur only in at least moderately large commuting systems.

To measure concentration at the PUMA level, we use the *average tract density*—specifically, the population-weighted average of the population densities for all census tracts in each PUMA. Following Craig (1984), we use a geometric mean to compute the average, as is appropriate for a log-normally distributed variable like population density. The exact formula is

$$\log ATD = \frac{\sum p_t \log d_t}{\sum p_t} \tag{1}$$

where *ATD* is the average tract density, p_t is the population of tract t, and d_t is the population density in tract t. Compared to the standard population density measure (total population divided by land area), the average tract density is less affected by large, unpopulated open spaces. For example, the population of a PUMA in southern Florida may reside mostly in dense developments near the coast, but if the PUMA is mainly comprised of unpopulated interior wetlands, the PUMA's population density would still be relatively low. This PUMA's average tract density, however, would be large, indicating that PUMA residents live "on average" in densely populated tracts, which is more relevant for the "concentration" dimension of rurality.

To measure a PUMA's metropolitanness, we use the *average CBSA population*—specifically, the population-weighted average of CBSA (metro or micro area) populations for all residents in each PUMA. For PUMA residents who live outside of any CBSA, we assign a CBSA population of zero. For example, if 10% of a PUMA's population resides in part of a CBSA that has a total population of 1 million, and the other 90% resides outside of any CBSA, the PUMA's average CBSA population is 100,000 (1 million * 10% + 0 * 90%). PUMA-level values of average CBSA population range from zero (where no PUMA residents reside in a CBSA) to 19.6 million for all PUMAs lying within the New York City metro area.

Both the average tract density and average CBSA population have approximately log-normal distributions (long positive tails) among PUMAs, so it is generally useful to apply a log transformation to the averages in figures and models. The log of a zero value, however, is infinitely negative, so we inflate all zero values—which occur only for average CBSA populations—to 1,000.[2]

Figure 4 illustrates the two-dimensional spread of average tract densities and average CBSA populations for all U.S. PUMAs using 2010 populations, 2010 tract definitions, and 2013 CBSA delineations. The point colors also indicate the metro/nonmetro class of each PUMA. The overall distribution mirrors closely the

[2] In a follow-up analysis, we plan to investigate two modifications to our average CBSA population measure. First, to prevent zero values altogether, we will use county populations in place of CBSA populations for residents who live outside of CBSAs. The county population is not an ideal indicator of a remote population's "commuting system size," but it may nevertheless be more appropriate than assigning all such areas a size of zero. Second, as with our average tract density, we plan to use a geometric mean for the average CBSA population instead of the arithmetic mean we currently use. We believe a geometric mean may be more suitable because of the log-normal distribution of CBSA populations. For example, in a PUMA where only a small portion of residents live in a very large metropolitan area, the arithmetic mean can be very large, which does not seem to indicate well the "typical" metropolitanness for PUMA residents.

conceptual model in Figure 3: the upper right contains PUMAs with high densities in large metro areas; the lower right contains PUMAs with low densities in large metro areas; the lower left contains PUMAs with low densities and outside of any CBSA; and as expected, the upper left is empty, indicating that PUMAs with dense populations occur only in or around medium-to-large CBSAs.

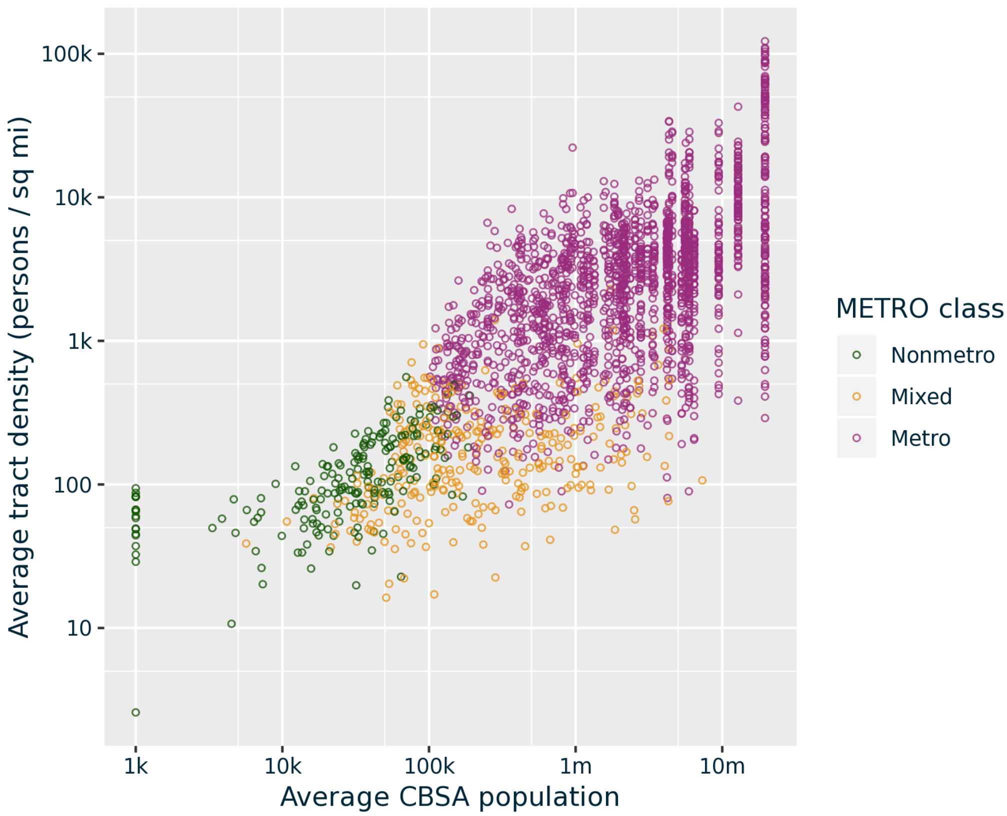

Figure 4. Relationships among three PUMA-level rurality indicators.

The colors in Figure 4 indicate that most PUMAs that lie entirely within metro areas have relatively high average densities, but some have low average densities. Such low-density metro PUMAs may or may not fit our expectations for "rural" areas. They may or may not share characteristics with other low-density PUMAs. Similarly, some nonmetro and mixed PUMAs have relatively high densities. It is possible that such areas have more in common with metro PUMAs at similar densities than with nonmetro and mixed PUMAs at lower densities. We believe that this two-dimensional framework offers great potential as a means to investigate such possibilities and to determine whether "concentration" or "metropolitanness" are both important factors, separately or together, in any study of rural populations. The framework should be useful as well as in broader studies seeking to distinguish populations across all densities or across all levels of the urban hierarchy.

Illustrative Results

The motivating example in this paper is how poverty rates differ across measures of rurality (see Figure 1). This section examines poverty using these two continuous measures and ultimately will illustrate how the distribution of poverty, as well as other key demographic characteristics, is not uniform across these dimensions of rurality. Consequently, we also measure regression models predicting poverty while controlling for these demographic factors. The use of our continuous measures shows that the correlation between poverty, rurality, and different demographic dimnesions, varies in ways that cannot be captured by a simple metro/nonmetro distinction.

Using ACS data from 2012 through 2017, we calculate poverty rates using the metro/nonmetro distinction. We obtain the ACS microdata and PUMA mapping files from IPUMS USA, and we obtain nationwide and tract-level summary data, along with mapping files for metro and urban areas, from IPUMS NHGIS. Table 4 shows that poverty rates are linearly related to metro classification. That is, poverty rates are highest in nonmetro areas (17.6%), lower in mixed metro areas (15.9%), and then lowest in metro areas (14.5%). However, when using the continous measures, we can clearly see that the relationship between poverty and rurality is much more complex. Figure 5 plots PUMAs by average tract density, average CBSA size, and PUMA poverty rates. The non-linearity between poverty and rurality is apparent. The highest and lowest poverty rates are concentrated in areas with higher CBSA population and higher average tract densities. Indeed, the areas with lower CBSA populations and lower densities tend to have relatively high poverty rates but not at the higher rates suggested by the metro/nonmetro distinction.

Figure 5. 2000 Public Use Microdata Areas (PUMAs) classified by 2010 population-weighted density

In Table 4, we take the continuous measures and separate them into different categories to illustrate our results with microdata. We create four categories for average CBSA population (0-50k, 50k-250k, 250k-1m, 1m+) and five categories for average tract density (0-100, 100-400, 400-1600, 1600-6400, 6400+). We then calculate the poverty rates for each of the cells that roughly correspond to Figure 4. Table 4 clearly shows that the highest rates of poverty are found in the most dense PUMAs where CBSA populations average between 250k and 1 million people (23.8%). The lowest poverty rates are found areas with medium density and CBSAs of over 1 million people. We also examine the rates of only those people residing in nonmetro areas. By limiting our analysis to the nonmetro population, we highlight the usefulness of our measures in the nonmetro setting. The nonmetro population resides in PUMAs with densities of 1600 or less and populations of no more than 250k. The poverty rates for the least dense and most dense areas with 50k-250k average CBSA population demonstrate the highest and lowest poverty rates. In all, this analysis highlights the nonlinear variation in poverty rates that is apparent by using continuous measures of rurality.

Table 4. Poverty Rates by Local Population Density and Average CBSA Size, 2012-2017

Nonmetro	Mixed metro	Metro
17.6	15.9	14.5

All People

	Average CBSA Population			
Local Population Density	0-50k	50k-250k	250k-1m	1m+
6400+			23.8	18.8
1600-6400		19.0	17.9	13.0
400-1600		16.0	13.5	9.5
100-400	17.5	17.3	15.1	12.3
0-100	17.8	18.5	16.8	14.6

Nonmetro Residents Only

	Average CBSA Population	
Local Population Density	0-50k	50k-250k
400-1600		12.79762
100-400	17.46696	17.66346
0-100	17.82136	18.57174

One of the main advantages of incorporating these continuous measures to microdata is the ability to perform analyses beyond what is available in census summary files. In Table 5, we provide summary statistics by the metro/nonmetro categories as well as the more refined categories for nonmetro areas. In doing so, we illustrate how our measures can uncovers important variation. Focusing on demographic variation first, we notice that nonmetro areas have an average age of about 40 years. However, within nonmetro areas, the largest and most dense PUMAs are, on average, about two years older, proportionally more White (79.7% v. 88.6%), and higher incomes, wages and salaries, and Social Security income. Internet access is an important issues in rural areas. Indeed, we show that about 16.5% of people in nonmetro areas have no access to internet. But by looking at the least dense and least populated areas, we notice that this rate is a full percentage-point higher (17.5%). Meanwhile, the most dense and most populated nonmetro areas have much lower rates of internet inaccessibility (11.2%) While intuitive, the point of this exercise is to highlight the variation that is noticeable within the nonmetro PUMAs.

Table 5. Demographics for Metro Areas and Nonmetro Areas, 2012-2017

	Nonmetro	0-100, 0-50k	0-100, 50k-250k	100-400, 0-50k	100-400, 50k-250k	400-1600, 50k-250k
Demographics						
Age	40.2	40.5	40.5	39.6	40.0	42.0
Female (%)	50.0	49.9	49.6	49.8	50.4	50.0
Latino (%)	7.9	6.8	6.8	10.4	7.9	5.6
White (%)	79.7	79.8	77.8	80.0	79.1	88.6
African American (%)	7.3	8.1	1.9	6.8	7.6	3.0
Native American (%)	2.1	2.8	10.1	0.5	1.3	0.2
Asian (%)	1.0	0.8	1.0	0.7	1.7	1.0
Other Race (%)	0.1	0.1	0.1	0.0	0.1	0.1
Multiple Races (%)	1.9	1.8	2.2	1.6	2.3	1.5
Foreign-born (%)	4.3	3.8	4.6	4.5	4.9	4.7
Non-citizen (%)	2.3	2.1	2.1	2.5	2.6	1.7
Income						
Total Personal Income	28,346	28,113	28,687	27,571	28,754	34,359
Wage and Salary Income	19,138	18,587	18,948	18,958	19,711	24,450
Social Security Income	3,255	3,326	3,296	3,115	3,229	3,502
No Access to Internet	16.5	17.5	16.6	16.8	15.2	11.2
No Health Insurance	12.4	12.6	13.0	13.1	11.6	7.6
Observations	2,006,375	920,904	95,972	438,506	504,903	46,090

These demographic differences highlight the need to control for demographic factors when analyzing poverty, a practice that is impossible without microdata. Table 5 makes the obvious point that controlling for demographics is necessary in a regression modeling poverty. The first regression predicts poverty by controlling for residing in metro PUMAs or mixed PUMAs. The coefficient shows that people in metro PUMAs are less likely to be in poverty than those in nonmetro areas. However, when we include a battery of demographic controls in the second regression, we notice an increase in the coefficient from -0.0311 to -0.0401. Clearly, in order to model predictors of poverty, it is necessary to use microdata, a hopefully obvious yet important point.

Table 6. Regression Results - Predicting Poverty with Different Controls, 2012-2017

	No controls	With Demographic Controls
Metro	-0.0311***	-0.0401***
Mixed metro	-0.0171***	-0.0138***
N	18,120,063	18,120,063

We next incorporate our continuous measures to our regression framework. Table 7 reports the results where the omitted geography is the area with the lowest poverty rate (400-1600 density, 1 million plus population). Our results demonstrate that the areas that are more highly correlated with poverty are the nonmetro areas, the 1600-6400 density and 50k-250k population, and the 6400+ density and 250k-1 million+ population. These results mirror the patterns seen in Table 4 but with one important difference: whereas the densest and most populated areas had among the highest rates of poverty (Table 4: 18.8%), much of this correlation is due to demographic differences. The coefficient is measured at 0.0442 which is substantially lower than the other high poverty areas. This finding is evidence of how incorporating these measures into the microdata can result in more nuanced results regarding poverty.

Table 7. Coefficients on Controls for Local Population Density and Average CBSA Size, 2012-2017

	Average CBSA Population			
Local Population Density	0-50k	50k-250k	250k-1m	1m+
6400+			0.0863	0.0442
1600-6400		0.0761	0.0566	0.0125
400-1600		0.0548	0.0309	0
100-400	0.0657	0.0619	0.0449	0.0231
0-100	0.0668	0.0708	0.0531	0.0364

A final angle we take to this analysis is to focus on the correlation between Latinos, noncitizens, and poverty. We run our regression model on different categories of tract density and CBSA size and report the coefficients in Table 8. Focusing on the Latino population, we notice that the coefficient for Latino in metro areas is 0.733. But when we run the model for different subsamples, we notice wide variation in the coefficient. For example, the coefficient on Latino in the most dense areas are 0.0324 and 0.056, both substantially lower than the 0.0733 from the metro model. Similarly, the noncitizen coefficients in the nonmetro/metro models are between 0.04 and 0.05. However, in breaking these down to different areas, we notice wide variation ranging from 0.0239 and 0.0813. Ultimately, these models demonstrate the substantial variation in results than can be exploited by our continuous measures of rurality.

Table 8a. Regressions by Subsample, 2012-2017

	Nonmetro	Mixed	Metro
Latino	0.0636***	0.0604***	0.0733***
Noncitizen	0.0458***	0.0478***	0.0399***
N	4,831,802	2,921,780	13,288,261

Table 8b. Coefficient on Latino for Subsamples by Local Population Density and Average CBSA Size, 2012-2017

	Average CBSA Population			
Local Population Density	0-50k	50k-250k	250k-1m	1m+
6400+			0.0324	0.056
1600-6400		0.0621	0.0859	0.0608
400-1600		0.0669	0.0835	0.0659
100-400	0.0542	0.0762	0.1044	0.0614
0-100	0.0607	0.0794	0.0467	0.0661

Table 8c. Coefficient on Noncitizen for Subsamples by Local Population Density and Average CBSA Size, 2012-2017

	Average CBSA Population			
Local Population Density	0-50k	50k-250k	250k-1m	1m+
6400+			0.0239	0.0382
1600-6400		0.0457	0.0458	0.0355
400-1600		0.0785	0.0503	0.0333
100-400	0.0524	0.0538	0.0467	0.0404
0-100	0.0382	0.0324	0.026	0.0813

Conclusions

The interchangeability of nonmetro and rural is commonplace but, as our paper shows, the practice is inherently flawed. From a microdata perspective, this choice is understandable since the researcher has practically no other alternatives. We add to the literature of settlement patterns by providing a theoretical framework for thinking about rurality within two different dimensions: “concentration” (the local intensity of settlement) and “metropolitanness” (the size of the commuting system). By creating two continuous measures, we show that the researcher need not be constricted by the binary metro/nonmetro classification. Indeed, as our illustrative examples with poverty shows, there is a substantial variation along these different dimensions. As we move to making these measures publicly available through IPUMS USA, our paper shows how researchers will benefit from a better way of identifying rural areas in microdata.

References

Coombes, Mike, and Simon Raybould. "Public policy and population distribution: developing appropriate indicators of settlement patterns." Environment and Planning C: Government and Policy 19.2 (2001): 223-248.

Craig, J. (1984). Averaging population density. *Demography, 21*(3), 405-412.

Isserman, A. M. (2005). In the national interest: Defining rural and urban correctly in research and public policy. *International regional science review*, *28*(4), 465-499.

U.S. Department of Agriculture (2013). Rural Classifications Overview. U.S. Department of Agriculture, Economic Research Service. Washington, DC. https://www.ers.usda.gov/topics/rural-economy-population/rural-classifications/ [Accessed 2018 September 13].

U.S. Department of Agriculture (2013). Rural-urban continuum codes documentation. U.S. Department of Agriculture, Economic Research Service. Washington, DC. https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation/ [Accessed 2018 September 13].

Waldorf, B. (2006). A continuous multi-dimensional measure of rurality: Moving beyond threshold measures. In *annual meeting of the American Agricultural Economics Association, Long Island, CA*.