

Dong Wook Lee<sup>1</sup> and Melissa Rogers<sup>2</sup>

<sup>1</sup> Department of Political Science, Adelphi University, Garden City, NY 11530, USA. Email: [dzl119@gmail.com](mailto:dzl119@gmail.com)

<sup>2</sup> Department of International Studies, Claremont Graduate University, Claremont, CA 91711, USA.  
Email: [melissa.rogers@cgu.edu](mailto:melissa.rogers@cgu.edu)

## Abstract

Political scientists are increasingly interested in the geographic distribution of political and economic phenomena. Unlike distribution measures at the individual level, geographic distributions depend on the “unit question” in which researchers choose the appropriate political subdivision to analyze, such as nations, subnational regions, urban and rural areas, or electoral districts. We identify concerns with measuring geographic distribution and comparing distributions within and across political units. In particular, we highlight the potential for threats to inference based on the modifiable areal unit problem (MAUP), whereby measuring concepts at different unit aggregations alters the observed value. We offer tangible options for researchers to improve their research design and data analysis to limit the MAUP. To help manage the measurement error when the unit of observation is unclear or appropriate data are not available, we introduce a new measure of geographic distribution that accounts for fluctuations in the scale and number of political units considered. We demonstrate using Monte Carlo simulations that our measure is more reliable and stable across political units than commonly used indicators because it reduces measurement fluctuations associated with the MAUP.

*Keywords:* measurement error, ecological inference, Monte Carlo methods, ecological fallacy

## 1 Introduction

Until recently, geography has been “a blind spot for political scientists” (Rodden 2010, 322). Conceptual insights, empirical innovations, and breakthroughs in data collection have all brought geography into focus in the field. In particular, recent research has reignited interest in the geographic variation of political phenomena. We focus, in this article, on the empirical concerns of measuring geographic distribution and offer a novel methodological approach to manage issues of scale.

In most measures of distribution among individuals in a population, the unit of observation is uncontroversial or can be standardized using straightforward algebraic transformations. For example, measures of income inequality address disparities among individuals or households, with the number of household members included in the calculation. The measurement of distribution across geography, in contrast, is not so easily specified. Political and economic endowments such as votes, income, population, and employment are unevenly distributed across the geography of every nation. The varying number, size, and scope of geographic units available create inherent ambiguity in how to most accurately capture their geographic distribution.

Studies of the geographic distribution of economic and political phenomena must be centrally concerned with the unit question of what the appropriate geographic level of aggregation is and whether existing data match the chosen jurisdiction. Ambiguity about the unit and mismatch in

*Authors' note:* The authors thank Daniel Bochsler for help with construction of the geographic distribution measure. The authors greatly benefited from the feedback from Editor Jeff Gill, three anonymous reviewers, Ryan Bakker, Scott Desposato, Indridi Indridiason, Stephanie Rickard, Guillermo Rosas, and participants at SCCPI at UC San Diego, Faculty Research Colloquium at CSU Long Beach, and MPSA 2017 and 2018. Replication data are available at the Harvard Dataverse: <https://doi.org/10.7910/DVN/DKY4DY>

*Political Analysis* (2019)  
vol. 27:263–280  
DOI: 10.1017/pan.2019.14

Published  
18 March 2019

Corresponding author  
Melissa Rogers

Edited by  
Jeff Gill

© The Author(s) 2019. Published  
by Cambridge University Press  
on behalf of the Society for  
Political Methodology.

the data are characteristic of a large percentage of geography-oriented research (Kwan 2012b). Without a well-specified unit, scholars measuring geographic distribution must contend with the modifiable areal unit problem (MAUP) in which different aggregations of the same data directly influence the outcome measure (Openshaw 1984; Fotheringham 1989). In this article, we focus on the “scale” problem of the MAUP and its application to geographic distribution measures.

Different scale approaches to the same data may result in different observed values and, therefore, the inferences we may draw from it (Guo and Bhat 2007). For example, when measured at the county level, Pennsylvania’s highest rate of poverty appears in Philadelphia County. Once disaggregated to the census tract level, however, Philadelphia County is shown to have both some of the highest poverty areas in the state and some of the lowest (Hayward and Parent 2009).<sup>1</sup> The distribution of poverty in Pennsylvania differs dramatically depending on the political unit. When assessing how US states manage poverty within their borders, what is the correct unit to employ when policy is concurrently made at the state, county, and local levels? The MAUP is a pervasive concern because assessing political phenomena often requires specifying a political unit.

In the following analysis, we describe the theoretical and empirical challenges of measuring geographic distributions of politically relevant phenomena. We begin our discussion by identifying the problems of unit selection and the MAUP in empirical research with geographic data. Building upon the research of geographers, we offer advice on unit selection and detail the prominent solutions to the MAUP tailored to the most common scenarios in political data. Our primary purpose is to help researchers understand how unit selection may affect measurement of geographic distribution and to offer concrete ways to limit the impact of idiosyncratic empirical choices.

In the second part of the paper, we illustrate the MAUP in commonly used measures of geographic distribution. We show fluctuations across nested units in directly comparable data from the European Union (EU) member countries. We demonstrate considerable volatility in existing distribution measures through a replication of Stephanie Rickard’s (2012) study. We argue that when the unit is ambiguous or precisely matched data are unavailable, the consistency of the chosen indicators becomes critical (Hammersley 1987, p.78). We adopt the approach of Hay *et al.* (2001) to offer a scalable indicator to reduce fluctuations across units, using a formula adapted from Bochsler (2010). We compare the consistency of our scalable measure to existing indicators through Monte Carlo simulation experiments. Finally, we suggest areas for further research.

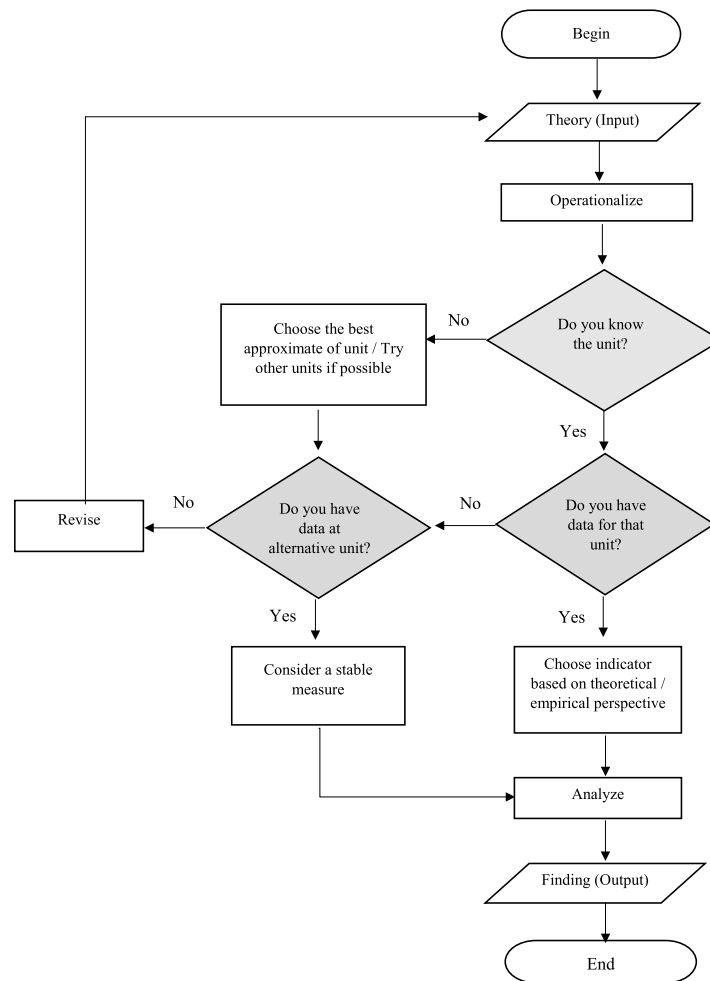
## 2 Considering the Unit Question in Empirical Research

The choice of unit is crucial to empirical research featuring questions of geographic variation. The most studied political phenomena, including elections, political representation, government distribution, and economic productivity, are most often structured by geography. Governments typically deliver resources according to administrative geography (such as transfers to subnational regions) or according to the geographic specificity of government goods (such as infrastructure).

The central role of geography in politics often necessitates data collection by geographic units and requires researchers to choose among many options in complex politics. We emphasize, in this article, that when we seek to compare across geography, the results obtained will depend on the unit selected. Thus, geographic distribution measures are sensitive to unit selection.

Figure 1 lays out the choices for researchers interested in questions related to variation or distribution across geographic (sub)units. The choice of political unit is in some cases obvious. For outcomes in the US House of Representatives, the appropriate unit is the Congressional district. For the US Senate, the relevant geography is the state. For US presidents, scholars should focus

<sup>1</sup> See examples in Wong (2009) and others visually demonstrating how different aggregations of data result in different evaluations of the same data. Simpson’s paradox is another concern with aggregation of unevenly sized units that may bear upon the MAUP.



**Figure 1.** The Location of Unit Issues within the Process of Empirical Research.

upon state-based aggregations of votes weighted by their representation in the Electoral College. In any of those cases, the unit is clear, and scholars often have access to data at those “units” of government. Empirical research of that type is straightforward and does not suffer from problems related to the MAUP. It follows the downward path on the right side of Figure 1.

Consider, however, if the scholar wanted to examine the relationship between variation in subnational characteristics, say race, income, population density, or partisanship, on the ultimate outcomes of the policy process, such as laws passed or budgets allocated in the United States. For example, a scholar may wonder whether US federal resources are allocated according to the partisanship of the recipients (Levitt and Snyder 1995). What would then be the appropriate political unit? With all three constituency units (Congressional districts, states, and Electoral College weighted states) involved in the decision-making process, how do we match the distribution of federal resources to the geographic distribution of partisanship? In this situation, the existing theory cannot confirm the most appropriate unit and we are forced to veer off the cleanest path on the right of the decision tree in Figure 1 by answering “No” to the question, “Do you know the unit?”

Ambiguity about the unit and mismatch in the data are characteristic of a large percentage of geography-oriented research. Typically, scholars select the best “approximate” unit. In the case of federal allocations on the basis of partisanship, Levitt and Snyder (1995) select House Congressional districts as the closest unit to the theory. Data on partisanship characteristics

and allocation of resources to US Congressional districts are available, and so they calculate coefficients of variation of district level receipts of federal programs, which they predict with indicators of partisan support. The authors provide reasonable justification of their chosen unit on theoretical and empirical grounds, but this choice clearly makes assumptions about the policy process (the House is dominant) and, therefore, the appropriate political unit (the Congressional district).

Even scholars who are able to isolate the theoretical unit may find that they lack the appropriate data. This is a challenge for anyone interested in topics requiring more geographic specificity than the nation state.<sup>2</sup> For example, many political phenomena may be affected by voters' "neighborhood"—an unclear designation, and one for which we likely lack data for nearly all relevant variables (Fortunato, Swift, and Williams 2016). Using data from the closest available unit to the neighborhood, the US Census block, is a regular practice throughout the social sciences. Choosing an approximate unit such as the Census block is often necessary, but it brings methodological challenges driven by the MAUP. Put simply, the level at which we aggregate data and the way we determine their aggregation may influence the outcome that we measure. In the next section, we describe the MAUP with examples from commonly used EU subnational data.

### 3 Attributes of the Modifiable Areal Unit Problem

The MAUP is a serious methodological concern for scholars examining geographic units. The logic of the MAUP is straightforward—unless individual observations are identically distributed on all characteristics, if we group the same data differently, those groupings will have different means, different standard errors, and accordingly, regression analysis will provide different results using the same underlying individual-level data (Amrhein 1995; Wong 2009).<sup>3</sup>

Characteristics across geographic units are rarely uniformly or identically distributed. Typically, factors are clustered in space, such as the number of voters that support a party, household incomes, and businesses in the same industry.<sup>4</sup> Given clustering, how we establish units is even more important to assess distribution because drawing lines around those clusters, within those clusters, or through those clusters will give us very different pictures of the cross-unit distribution. Importantly, we are also fundamentally interested in the reasons for and effects of the uneven distribution of these characteristics. An assumption of uniform distribution would not only raise concerns about the MAUP, it would also obscure many important political phenomena.

The MAUP features two central problems—scale and zoning. The scale problem relates to the choice of how big or small to define the geographic unit. The zoning problem refers to where in space we chose to draw the boundary lines that separate the units. A politics-relevant example of the MAUP zoning issue is electoral gerrymandering. Gerrymandering does not change the underlying distribution of votes, but it may alter the electoral outcome by grouping voters into favorable political districts. Our paper addresses the zoning subproblem of the MAUP only to a limited extent because the zoning problem is primarily a theoretical concern. Consider, for example, that political borders may be endogenous to favorable geography (Beramendi *et al.* 2018) or ethnicity (Michalopoulos 2012). If the borders are drawn for the purpose of clustering certain characteristics, we may not be able to assess the true impact of the distribution of that endogenous characteristic on the outcome of interest. Therefore, the zoning problem of the MAUP sits outside the scope of this article, which is focused on managing the scale problem of the MAUP.

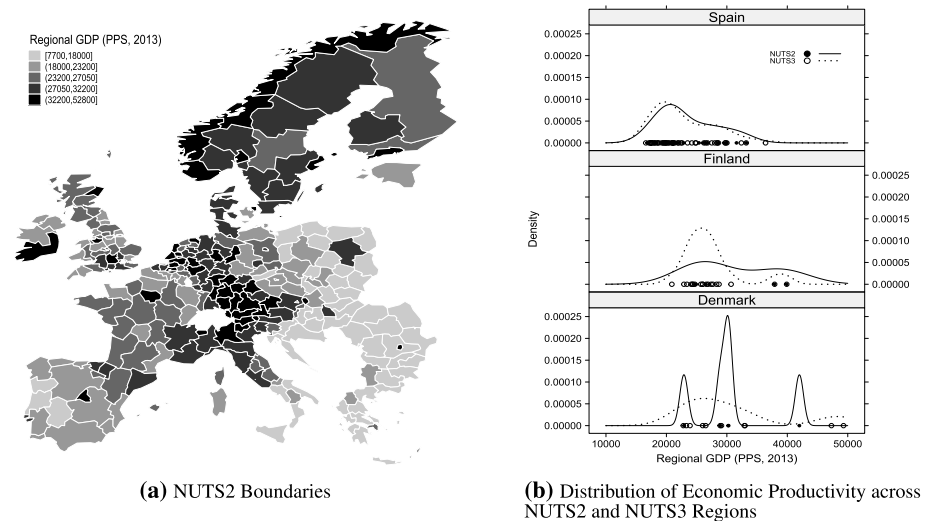
To illustrate potential problems with the scale and number of units, we use data and examples from subnational units within EU countries.<sup>5</sup> As part of their coordination with the EU, member

2 This includes GIS-created grid cells, which are aggregated to an arbitrary unit.

3 The MAUP can also affect regression results through controls (Soifer and Alvarez 2017).

4 See, for example, research identifying spatial distribution in campaign contributions (Tam Cho and Gimpel 2007) and ideology (Monogan and Gill 2016).

5 Replication data are available in Rogers and Lee (2019).



**Figure 2.** Distribution of Regional GDP across Territorial Boundaries.

countries comply with statistical standards to calculate economic variables at four hierarchical structural levels: NUTS3 (Parish, Canton, Oblast, City & Regency, County, or Municipality), NUTS2 (Region, State, Province, or Prefecture), NUTS1 (Region, Group of NUTS2), and NUTS0 (Country). Importantly, we use the NUTS because they are nested subsets of the same data and, thus, are directly comparable across levels. These data are widely used in EU research.

To provide an overview, Figure 2(a) maps the subnational distribution of per capita economic productivity in the EU NUTS2 regions in 2013. This map reveals clear variation within and across the EU countries in subnational economic productivity.

The complexity of the MAUP scale problem is apparent in the comparison of NUTS2 and NUTS3 regions in three countries, shown in Figure 2(b). Recall that NUTS3 is a partition of NUTS2; so they comprise different aggregations of the same data. In the case of Spain, at the top of the figure, the subnational distribution appears very similar in the NUTS2 and NUTS3 data. In contrast, Finland and Denmark show clear evidence of the scale problem. In the case of Finland, the NUTS2 (more aggregated) measure is much smoother and shows lower subnational dispersion than the smaller unit, NUTS3. This is the most common effect of MAUP, whereby subunit variation is reduced through aggregation (Caramani 2004). In the case of Denmark, however, the lower level, NUTS3, is smoother and less distributed. Denmark shows considerable spatial clustering of NUTS3 units by productivity but wide variation across NUTS2 units. Figure 2(b) also reveals the potential confounding effect of variation in the number of subnational units across countries. The higher the number of NUTS3 regions in a given NUTS2 region, the more the NUTS3 distribution tends to converge with that of its encasing units (NUTS2). Spain has a relatively large number of NUTS3 units, while Finland and Denmark have fewer, making their data more lumpy.<sup>6</sup>

This simple demonstration in EU data reveals the potential for serious instability in empirical findings in geographically oriented research. In the next section, we describe the best approaches to manage the MAUP, including by incorporating scale-relevant information into the indicator. In the sections that follow it, we show how the MAUP affects measurement of geographic distribution.

<sup>6</sup> Online Appendix (OA) Table A1 lists the number of NUTS across 24 European countries.

## 4 Choosing the Best Unit and Managing the MAUP

### 4.1 Choosing the best “approximate” unit

Choosing the best approximate unit is primarily a question of theory but often also depends on the availability of data. The first approach should be to reassess the theory and consult the associated literature to find the theoretical unit and thereby avoid approximation. We consider how to do this using the examples of political neighborhoods and the US budget allocation process.

Research on neighborhoods is common in political behavior and other social sciences. For example, many studies suggest that political engagement (e.g., voting and volunteering) is affected by individuals’ interaction with neighbors. Scholars of the United States typically use the census block to represent a neighborhood. At the same time, most acknowledge the approximation—that individuals do not see their neighborhood as a census block, that different phenomena have different neighborhood scales, and that different individuals have different scales.<sup>7</sup> The neighborhood unit is unclear in most research.

Yet there are theoretical and empirical ways to manage the unit ambiguity of a neighborhood. Scholars can examine literature about what constitutes a neighborhood for the question at hand. Scholars can treat the unit of the neighborhood as an empirical question, tracking individual behavior to gain a sense of its scope (Kwan 2012a). If possible, scholars can employ a neighborhood unit to match the one agreed upon by foundational research and can construct measures at the appropriate level and weight according to characteristics of the individual. With individual data, scholars can also test the hypotheses with multiple constructions of the neighborhood based, for instance, on increasing co-centric circles or network bands of data (Guo and Bhat 2007).

The selection of an approximate unit is perhaps more intractable in the example of federal outlays. Levitt and Snyder (1995) use House districts as the approximate unit to represent the budget allocation process. Nonetheless, we know that certain policies such as federal housing subsidies are distributed according to geographic criteria (such as states, counties, or metropolitan areas) that do not align with the Congressional district. Studies testing this hypothesis at other units of analysis have not been able to confirm a clear partisan allocation (Hoover and Pecorino 2005). Unlike the neighborhood question, we cannot simply broaden or contract the scope of the data to approximate the appropriate unit. Across a range of policy issues, such as policy-making and budget allocation processes characterized by multiple decision makers with distinct geographic constituencies, we cannot straightforwardly identify the political geographic unit on theoretical grounds. In such (common) cases, scholars are advised to consider the four main recommendations to manage the MAUP, which we discuss in the next section.

### 4.2 Four approaches to the MAUP

When the appropriate unit is unclear or data are not available for the theoretically relevant unit, geographers advise four primary approaches to manage the MAUP (Openshaw 1984; Dark and Bram 2007). The first option, and by far the most common, is to ignore it. While not ideal, the MAUP can more safely be ignored where the units are equivalent (in terms of size and shape) and the units lack spatial autocorrelation (Arbia and Petrarca 2011). These conditions are almost never realized in the social sciences, making this option dubious for political researchers.

The second suggestion is to identify the theoretical unit and use the data at that unit. This is the best option, if possible, because it eliminates the MAUP by rendering alternative unit levels theoretically irrelevant. What many scholars find, however, is that political data are not available at their preferred unit of analysis. Politically relevant data are often collected at standardized, stable administrative units that cannot fully capture the complexity of human systems. Political scientists

<sup>7</sup> See Kwan (2012b), “The uncertain geographic context problem”—specifying contextual effects of geography when the underlying spatial and temporal variation for individuals is unknown.



have long recognized that different phenomena have different scales, whether by design (such as electoral constituencies) or decentralized social processes (such as neighborhoods). Greater efforts by our fields to classify the geographic scale of political phenomena would help scholars to better identify their theoretical units and collect data at those levels.

The third recommendation is to collect more data to get as close as possible to the individual level of analysis (Murguía and Villaseñor 2000). Having data at the individual level allows for transformations of the data into different unit aggregations. With multiple unit levels available, scholars can test for the extent of the MAUP in their data. Crucially, they can conduct sensitivity analysis by testing their hypothesis at different units of analysis (Jelinski and Wu 1996; Beramendi and Rogers 2018).

In some cases, scholars choose an approximate unit that is more aggregated than their preferred level of analysis to reduce volatility. It may appear obvious that this choice would lead to ecological fallacy. In many cases of political research, however, the unit is not clearly specified enough to determine the nature of the data mismatch. The trade-off of this choice is that the data are smoothed, thus obscuring potentially relevant variation within smaller subunits. The other alternative is to use a lower level of analysis, which may be highly unrepresentative of the theoretical unit. Thus, the cost of this choice is (perhaps dramatic) mischaracterization of the theoretical unit. Both choices are commonly employed despite clear drawbacks, but scholars can strengthen their work by justifying their choice and acknowledging those drawbacks.

Identifying the unit and collecting more data are important ways to mitigate the MAUP in political research. Yet, these options are not always available to or feasible for scholars. As referenced earlier, we do not yet have theoretical tools to understand the unit even in well-studied processes such as US policy making. The fourth recommendation is one that we pursue in this article, which is weighting the subunits of the aggregation with relevant theoretical features (such as population, land area, and economic or political endowments) within the measure (Hay *et al.* 2001; Bochsler 2010). Our scale and scope-corrected Gini (SSGINI) measure, described in detail below, reduces the scale problem of the MAUP by explicitly controlling for the features of the unit that render them difficult to compare directly and bringing them closer to uniform units (Arbia and Petrarca 2011). This reduces the MAUP by accounting for factors that cause variability in scale and makes the measure more consistent across units.

Thus, we articulate a path through the middle column of Figure 1. When the researcher faces an ambiguous unit or imprecise data and does not have a theoretically driven interest in instability across units, we argue that stability of the measure is preferable to the volatility driven by the unit choice. A preferred indicator would produce similar outputs irrespective of data measurement at lower or higher levels of aggregation, reducing the concern that the practical choice of using an approximate unit influences the conclusions we draw from those results. Therefore, a solution to the scale problem of the MAUP requires a statistical property that is invariant to the level of aggregation (King 2013).

However, we must also note the appropriate use and limitations of scalable indicators such as SSGINI. When the theory suggests that the measure should not be scaled, SSGINI would not be preferred to existing, unscaled indicators. For instance, it would not make sense to use a population-scaled distribution indicator to predict US Senate voting outcomes because the states have uniform representation. Moreover, the MAUP may contain valuable information or itself become the subject of a theory. As Jelinski and Wu (1996, p. 138) argue, “The MAUP is not really a ‘problem,’ per se; rather, it may reflect the ‘nature’ of the real systems that are hierarchically structured . . . it carries critical information we need to understand the structure, function, and dynamics of the complex systems in real world.” SSGINI will mask these differences by scaling units with theoretically relevant variables. Therefore, SSGINI is appropriate when scholars seek to minimize differences across comparison units, not to highlight the differences across those units.

## 5 Comparing Measures of Geographic Distribution

In this section, we describe common measures of geographic distribution and compare their features to our scalable measure, SSGINI. Geographic distribution measures can be employed to account for variation in income, population, political resources, democracy (Giraudy and Pribble 2018), or other endowments that are unequal across space. For example, we can measure the extent of regional convergence in income, uneven party support, divergence in military capabilities, and population or economic agglomeration within or across nations. Existing research provides two main geographic distribution concepts: dispersion and concentration. We describe the most widely used measures of these two concepts below. All mathematical properties of the measures are shown in Table 1. The most simple, easy to interpret, and commonly used measure of geographic distribution is the coefficient of variation (COV). Coefficient of variation is a dispersion measure without analytic weights, which is widely employed in the literature on regional economic growth and convergence (Barro and Sala-i Martin 1992). Use of COV implies the direct comparability of the units. This may be appropriate, for example, in the case of US Congressional districts that are apportioned according to population. COV does not account for the contextual differences (e.g., the share of the national population) that can be substantially meaningful. For instance, the voting weight applied to a densely populated region would be considered equal to that of a sparsely populated region in the COV. This property of COV makes it particularly vulnerable to measurement fluctuations dependent on the scale and number of territorial units.

The weighted coefficient of variation (WCOV) is widely used in the regional disparity literature in economics (Rodríguez-Pose and Ezcurra 2009). By assigning a (typically, population) weight parameter ( $p_i$ ), WCOV is robust against single extreme observations (Lessmann 2009). Similarly, the region-adjusted Gini coefficient (RDGINI) retains meaningful information about the extent of relative deprivation, not merely spread (Lessmann 2009). In RDGINI, additional weight is given to a region's per capita value as it veers farther away from the mean of the inter-regional distribution. This weighted value makes RDGINI sensitive to changes in the upper or lower tail of the distribution. The strength of WCOV and RDGINI, therefore, lies in their features controlling within the formula for important substantive characteristics.

Yet, neither WCOV nor RDGINI fully copes with the scale problem of the MAUP. When analyzing variation within and across countries, the number and scale of territorial units may make a marked difference to data inference. Users of WCOV and RDGINI implicitly assume that the variance in the number of units has no effect on the indicators. This assumption is misleading: it would require that there is no within-unit variation (Bochsler 2010). Indeed, nations vary considerably in the number, size, and scale of regional units across the different aggregation levels.<sup>8</sup>

To construct a dispersion measure that scales across an unequal number of unevenly sized territorial units across countries, we build upon existing research on the measurement of party system nationalization, which has similar unit number and scale concerns. We calculate a new unit SSGINI indicator of regional dispersion based on the formula developed by Bochslers (2010, 163).<sup>9</sup> The basic intuition of SSGINI is to weight for the varying sizes of the territorial units by accounting for their share of the national value of population, productivity, or land held by each unit. It also controls for the number of territorial units on which the calculation relies. Thus, it accounts for scope and scale within the calculation. This adjustment to the theoretical properties of the units is akin to “object-specific upscaling,” which reduces the MAUP by making the units more comparable (Hay *et al.* 2001).

<sup>8</sup> See OA Table A2 for cross-unit variation across NUTS2 and NUT3 in EU countries.

<sup>9</sup> The user-written macro functions in Excel and Stata are available at [melissazrogers.com](http://melissazrogers.com).



**Table 1.** Indices for the Geographic Distribution of Economic Productivity.

Indicators	Mathematical Properties	Fluctuates with scale?	Fluctuates with number of units?	Features and Sources
COV	$\frac{1}{\bar{y}} \left( \sum_{i=1}^n (\bar{y} - y_i)^2 \right)^{1/2}$	Yes	Yes	Simple dispersion, Lessmann (2009)
WCOV	$\frac{1}{\bar{y}} \left( \sum_{i=1}^n p_i (\bar{y} - y_i)^2 \right)^{1/2}$	Yes	No	Dispersion adjusted for the population sizes of the regions, Lessmann (2009)
RDGINI	$\frac{2 \sum_{i=1}^n i y_i}{n \sum_{i=1}^n y_i} - \frac{n+1}{n}$	Yes	No	Dispersion sensitive to deviation from the mean, Lessmann (2009)
SSGINI	$1 - \left( \frac{\sum_{i=1}^{n-1} J_i \sum_{j=2}^n Y_j - \sum_{j=2}^n J_j \sum_{i=1}^{n-1} Y_i}{\sum_{i=1}^n J_i \sum_{j=1}^n Y_j} \right)^{1/\log \left( \frac{\sum_{i=1}^n J_i^2}{\sum_{i=1}^n J_i^2} \right)}$	No	No	Dispersion corrected for the unequal number of unevenly sized territorial units, authors' modified version of Bochsler (2010)
AGC	$\frac{\sum_{i=1}^n  g_i - a_i }{2(1 - a_{\min})}$	Yes	No	Concentration sensitive to clustered locations, Spiezia (2002)

Notation for mathematical properties:

*i*: An index for the geographic unit while *n* is the number of units.

*Y*: The GDP of region *i*.

*y*: The GDP per capita of region *i*.

$\bar{y}$ : The country's average GDP per capita.

*J*: Population in region *i*.

*p*: The share of the country's total population in region *i*.

*g*: The share of the country's total GDP by region *i*.

*a*: The share of the country's total geographic area of region *i*.

*a*<sub>min</sub>: The relative area of the smallest region.

SSGINI is constructed similarly to a traditional Gini coefficient.<sup>10</sup> If there were a theoretical country with only two regions that share identical portions of the country's per capita GDP, SSGINI would take a value of zero, capturing a perfectly even distribution. The score of SSGINI would become bigger along with the increasing number of unevenly sized regions. At its extreme, if all GDPs were held in one region, SSGINI would take a value of 1. It ranges from 0 to 1, in which the larger the value, the higher the level of dispersion in the regional distribution.

SSGINI also incorporates weights for territorial units according to their different effective size and number. Size could represent many values, such as land or resource endowments, but for political questions, we most commonly weight by population—the relative population distribution measured in  $(\sum_{i=1}^n J_i)^2$  divided by  $\sum_{i=1}^n J_i^2$ —along with the standardization (Laakso

10 We adapt Bochsler's indicator as follows, using the example of GDP: *y<sub>i</sub>* is for a region's GDP and *y<sub>j</sub>* is thus the accumulated proportion of regional GDP. In the equation, the value subtracted from 1 represents the portion of regional inequality, as in the measure of "party nationalization" developed by Bochsler. This subtraction obtains the residual, equivalent to regional dispersion.

and Taagepera 1979; Bochsler 2010). This component is added as an integrated part of the exponential function of a traditional Gini coefficient and should thus be robust to major fluctuations in the unequal number of subunits across groups.<sup>11</sup> It corrects for and thus standardizes the convex effects of granularity of territorial data that are associated with the incremental number of unequal sized territorial units. If variance is calculated among the characteristics of  $n$  unequal parts of a territory, then variance will increase as the number of parts  $n$ , in which the territory is divided, increases. In application, the larger the number of territorial units, the less within-unit differences in productivity will be factored in by the calculation. Extreme patterns that are only visible in fine-grained regional-level data are averaged out with larger numbers of territorial units. Thus, at least in theory, SSGINI should not be sensitive to the size or number of territorial units involved.

Some distribution questions may be less concerned with dispersion and more interested in how values are concentrated within territorial units. Concentration measures the shape of the distribution, while dispersion measures its spread. Importantly, dispersion and concentration are distinct—countries with the same value of dispersion may have differently shaped distributions, reflecting different degrees of concentration (Lee and Rogers 2019). If scholars believe that concentration is an important phenomenon aside from dispersion, they require distinct, appropriate measurement concepts (Chen and Rodden 2013; Jurado and Leon 2017).

One option, the index of adjusted geographic concentration (AGC), is shown in the last row of Table 1. AGC captures whether economic productivity (or any other attribute) is disproportionately held in one or a small number of regions (Spiezia 2002). AGC ranges from 0 to 1, with higher values reflecting more concentration of the national total in certain regions.<sup>12</sup> It incorporates weights for geographic concentration of population and productivity across all regions within a country.

In the next section, we document the properties of these distribution measures and how they may be impacted by the MAUP. We demonstrate in descriptive data (Section 6), a replication study (Section 7), and a Monte Carlo simulation (Section 8) that SSGINI is the most reliable and stable choice of available measures.

## 6 MAUP in EU Regional Productivity Data

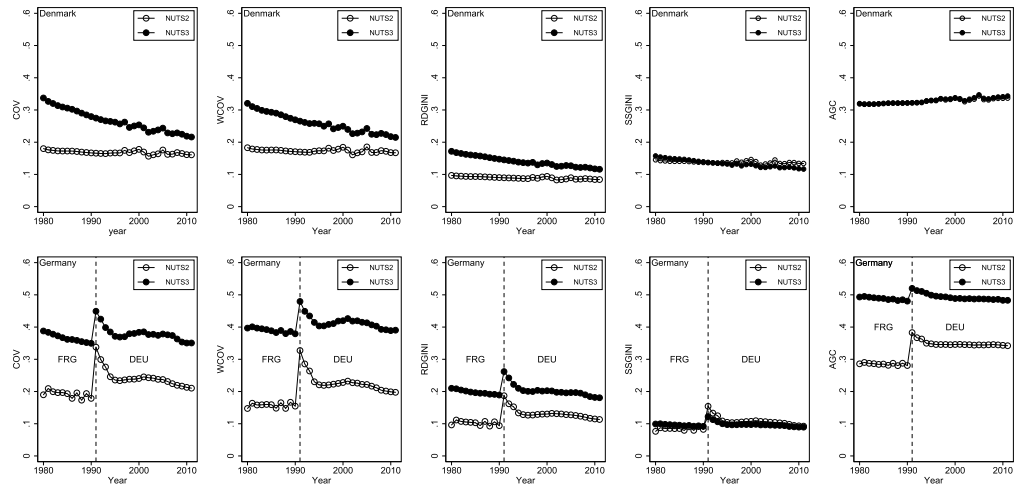
Measures of geographic distribution are often assumed to be reliable without examining whether they are, in fact, consistent across different unit scales and stable over time. The problem we highlight in Figure 3 suggests that the value of the distribution calculated at NUTS2 may be very different from the value obtained with the same data at NUTS3. Yet, if the difference lies only in the size of the measured value (consistency) and not in the patterns in the data (stability), the impact may be understated in statistical analysis. This difference will only affect the magnitude of the effect, not the statistical significance.

If we find, however, that the measures vary in how they capture patterns in the data, they would also impact the direction and significance of the effect. Both Type I and Type II errors are possible, especially if particular measures or unit levels are chosen because they have a data pattern preferred by the researcher.

One way to see whether the data patterns are stable across NUTS2 and NUTS3 is to plot them in comparison over time. If the data are consistently trending with each other at the NUTS2 and NUTS3 values, we can be more assured that the choice of the unit will not affect regression analysis. If the patterns become more divergent or convergent, however, we can assume that the choice of the indicator or unit level will affect significance (and magnitude of the effect).

<sup>11</sup> Thus, information about the size of the unit is required to calculate SSGINI. This may limit its application where the size of the unit cannot be assessed.

<sup>12</sup> AGC is similar to entropy measures such as the Theil index, discussed below.



Notes: Germany (DEU) has 38 NUTS2 regions and 396 NUTS3 regions. Denmark has 5 NUTS2 regions and 11 NUTS3 regions. Data Source: Cambridge Econometrics.

**Figure 3.** Case Comparison: Denmark and Germany (FRG and DEU).

Figure 3 compares the cross-unit difference in the measurement of inter-regional distribution in productivity for Germany (FRG prior to 1991, DEU after 1991) and Denmark over 30 years (1980–2010).<sup>13</sup> Using the same data, we calculate COV, WCOV, RDGINI, SSGINI, and AGC. Conceptually, the geographic distribution trend among NUTS2 regions should be directly comparable with that of NUTS3 regions because they are partitions. So long as these values are calculated using the same measure of geographic distribution, their values and trends should be similar. A reliable measure would minimize scale differences in the values and show parallel trends.

Among several instruments used to measure the geographic distribution of economic productivity, patterns in the COV and WCOV measures are meaningfully affected by the choice of the subnational unit. These measures suffer from a great deal of inconsistency (for Denmark and Germany) and instability (for Denmark). Consistency diminishes when the number of regions differs greatly between NUTS2 and NUTS3 (as in the case of Germany).<sup>14</sup> Even with a small difference in the number of units, as seen in Denmark, we see very different patterns in the data over time. The NUTS2 and NUTS3 values converge in the observed period, indicating that the patterns in the data vary over time. This would have a substantial effect on regression estimates. On the other hand, RDGINI, AGC, and SSGINI reduce this cross-unit difference problem considerably. However, RDGINI is still less consistent in both country measures (indicated by the values’ distance on the y-axis). AGC shows considerable inconsistency in its measure for Germany, suggesting that it is sensitive to difference in the number of units across cases. SSGINI is clearly the most consistent in values and also shows more stability in measurement over time. Thus, we should expect more consistent output in regression estimates with SSGINI.

## 7 The MAUP in a Replication Study

Much of the concern with the MAUP is focused on Type I errors (false positives), but the MAUP is equally likely to invoke Type II errors (false negatives), leading us to dismiss valid theories. To demonstrate the properties of the common geographic distribution measures and how

<sup>13</sup> These countries are selected to provide the variation in the number of regions at NUTS2 and NUTS3. See Column 4 of OA Table A1 for details. The total number of NUTS2 regions in Germany (DEU) is nearly 19 times larger than that in the countries with the fewest regions, such as Ireland and Slovenia. The unit differences across Germany’s NUTS2 and NUTS3 is also very large (358).

<sup>14</sup> We also chose Germany because of the break in geographic distribution at Reunification in 1991.

results may be influenced by the choice of geographic unit, we replicate results from a recently published article focused on economic geography. The intuition of this exercise is that if the MAUP is negligible in the distribution measures, the replication should yield consistent results, independent of geographic unit choices. Alternatively, if the MAUP is relevant, once we change the geographic unit within the measures, the results may differ in direction or significance without a clear theoretical justification. Replication also offers a chance to examine how the distribution measures perform in comparison to each other in an established theoretical framework. The results show that failure to consider the MAUP may lead us to dismiss theories with strong justification and results simply because of our choice of indicator.

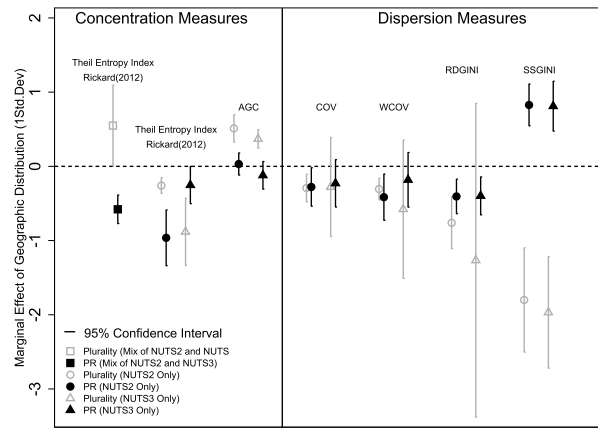
To demonstrate the impact of the unit choice in an existing study, we replicate Stephanie Rickard's (2012) "Electoral Systems, Voters' Interests, and Geographic Dispersion" in the *British Journal of Political Science*. We chose Rickard's study because the theoretical focus is on geographic distribution of manufacturing employment, it draws from NUTS data, and it offers a clear and compelling theoretical framework through which we can examine the performance of geographic distribution measures. Rickard argues that whether governments with proportional representation electoral systems (PR) or plurality electoral rules (non-PR) allocate more manufacturing subsidies depends on whether economic interests are dispersed or concentrated across the geography of the nation. Specifically, governments with PR are expected to deliver more subsidies when economic interests are more broadly dispersed because their constituencies are broader. When manufacturing employment is geographically concentrated, governments with plurality electoral rules are expected to give more subsidies in response to district-specific constituency interests.

We designed the replication to examine the difference across the unit levels (NUTS2 and NUTS3) and across the distribution measures. For the closest replication, we use the concentration index of manufacturing employment used by Rickard (the Theil entropy measure), with samples drawn from NUTS2 and NUTS3 regions.<sup>15</sup> We also test the concentration measure AGC at NUTS2 and NUTS3. Additionally, we examine the reciprocal phenomenon within Rickard's theory, geographic dispersion, using COV, WCOV, RDGINI, and SSGINI.

The relevant theoretical unit of the study is the political constituency level. For this, Rickard employs data at a combination of the NUTS2 and NUTS3 levels from Brühlhart and Traeger (2005). As is common in studies of political and economic geography, the theoretical unit is not obviously knowable in all of Rickard's sample cases, and data are not made available for all countries at constituency-level units. NUTS2 and NUTS3, in some cases, match the constituency level of (at least one of the) decision-making bodies in the countries in the sample, but in certain cases (such as the Netherlands), they do not. In other cases, they match one decision-making body (such as an upper house) but not the lower house. With firm-level data, it may be possible to construct the ideal data, but the theoretical unit would remain elusive. Thus, a perfect match of the data to the theoretical unit in this study is neither available nor can it be conceptually identified in all of the country cases. We stress that these challenges are pervasive in research on political geography and should be met with the best approaches available to identify the unit and manage the MAUP.

A few points should be noted in evaluating our replication. First, we cannot fully recreate Rickard's sample due to missing data prior to 1980 (her study starts in 1975). Rickard employed Brühlhart and Traeger's (2005) calculation of the Theil entropy index that used data from Cambridge Econometrics. We also use Cambridge Econometrics sectoral employment data, but they no longer make data prior to 1980 available. Nonetheless, we do not have obvious reasons to expect that the pre-1980 data should alter the results of the study, apart from sample size. Second, we employ dispersion and concentration measures to the study with some theoretical caution.

<sup>15</sup> We focus on AGC in this article as the measure more commonly used in political geography.



Notes: The dependent variable of the study is manufacturing budget shares. Each measure of geographic distribution is interacted with PR or its alternative (plurality including a range of systems classified as non-PR) to plot reciprocal marginal effects. The first column is the closest replication of Rickard’s (2012) model interacting the decomposed “within” Theil entropy index of concentration with PR and plurality. The geographic distribution measures are calculated as the degree of the manufacturing sector concentration or dispersion relative to the geographic distribution of aggregate employment within a country. Point estimates are average marginal effects; bands are the 95% confidence interval.

**Figure 4.** Marginal Effect of Geographic Distribution on Subsidy Budget Shares.

AGC closely approximates Rickard’s Theil index and, thus, may be seen as a reliable substitute. COV, WCOV, RDGINI, and SSGINI are theoretically relevant because they capture dispersion in employment. According to the theory, dispersion should be associated with more spending under PR and lower spending under plurality systems. However, concentration and dispersion are not necessarily opposite properties (Lee and Rogers 2019). A country may simultaneously have highly dispersed manufacturing, yet high concentration in particular regions (Germany is such a case).

Figure 4 summarizes the results of the replication. We report the marginal effects of employment concentration and dispersion under PR and plurality systems on government subsidy budget shares.<sup>16</sup> Full results are shown in OA Table A3. The figure is organized, from left to right, to show the concentration measures first: the Theil index as calculated in Rickard’s study with a combination of NUTS2 and NUTS3 values (rectangular symbol), the Theil index at NUTS2 (circle symbol) and NUTS3 (triangle symbol), and AGC calculated at NUTS2 and NUTS3.

Next to the concentration measures, we show the dispersion measures, COV, WCOV, RDGINI, and SSGINI, calculated at NUTS2 and NUTS3 levels. Estimates and confidence intervals for PR are dark-colored and plurality are light-colored. Beginning in the first column, we first show strong support for Rickard’s theory. A country with PR reduces subsidies significantly when economic interests are geographically concentrated, and, reciprocally, a country with a plurality system increases manufacturing subsidies as manufacturing employment grows more geographically concentrated. The nearby columns for concentration (Theil and AGC) show a more mixed story based on both the direction of the effect and the significance. The marginal effects of the Theil index (separating NUTS2 from NUTS3) under PR are significantly different from the corresponding effects under plurality across NUTS2 regions. However, these differences are not statistically significant, and in the case of the NUTS3 regions, they are in the opposite direction from what is predicted by the theory. The estimates for PR interacted with concentration at NUTS3 are in the opposite direction from the theory. The results for AGC under PR are significantly different from plurality with AGC, but the magnitude of the effect for PR systems is minimal for both NUTS2 and NUTS3.

<sup>16</sup> We plot the symmetrical interaction results, showing how both the PR and Non-PR variables interact with geographic distribution of manufacturing (Berry, Golder, and Milton 2012).

The dispersion measures, with the exception of SSGINI, provide a convoluted set of results with respect to the theory. With COV, WCOV, and RDGINI, the results of the interaction term are in many cases not significantly different from zero. In the case of plurality systems, the results are more consistently in the expected direction, at least at the NUTS2 level, but the estimates are not significantly different between PR and plurality systems (shown in overlapping confidence intervals) either at the NUTS2 or NUTS3 level. Only in the case of the SSGINI, among the dispersion measures, do we see results consistent with the theoretical expectations across both the NUTS2- and NUTS3-level data. Once we scale the manufacturing employment to the population share of the region, and the number of regions, as the theory would warrant, we obtain a result in the SSGINI indicator that is highly supportive of Rickard's study: subsidy budgets should be higher under PR and lower under plurality systems when manufacturing is dispersed. These conditional effects are robust across NUTS2 and NUTS3.

The key takeaway in Figure 4 is that there are considerable differences in marginal effect estimates across geographic levels and across indicators that likely indicate problems of measurement, not theory. Among the measures, SSGINI is least susceptible to (atheoretical) inconsistency driven by a choice of regional units at NUTS2 or NUTS3, in other words a threat to inference from the scale subproblem of the MAUP. Rickard's theory is affirmed with this measure, which is appropriately scaled to the theoretical units in the study. This stability of SSGINI results from its incorporation of the number of regions and their population share within the measure.

## 8 Monte Carlo Simulations

To illustrate the relative consistency and stability of SSGINI under different MAUP conditions, we conduct a Monte Carlo integration using simulations (Openshaw 1984). The intent of this simulation is to create a function to sort individual-level geo-referenced data to emulate unknown distributions. This is an important confirmatory methodological analysis to evaluate, using a foundation of real-world data, whether switching the scale or the number of units has a meaningful impact on the value obtained (Rey and Janikas 2005).

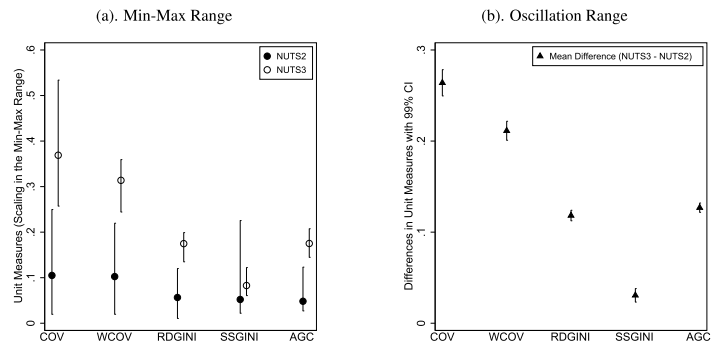
Our Monte Carlo experiment is constructed to simulate the data generation process of subnational characteristics. We focus on three common factors of distribution: productivity, population, and land. We start with individual-level measures, such as a firm, individual, or parcel of land, that are sorted into a subregion (equivalent to a NUTS3 unit). Because subregions are hierarchically nested, sorting this individual into a subregion also sorts it into its encasing larger region (equivalent to NUTS2).<sup>17</sup> In our base specification, we simulate this hierarchically nested sorting process 100 times with a beta distribution of GDP, population, and land size. We draw samples from the standard beta distribution whose shape is determined by the random combination of the two shape parameters,  $\alpha > 0$  and  $\beta > 0$ . The beta distribution enables us to explore a flexible probability density taking  $\alpha$  and  $\beta$ , given that we do not assume what that probability may be and thus its potential distribution. We use a user-defined function that provides random values for the mean and variance of the probability distribution so that we can obtain random parts of  $\alpha$  and  $\beta$ .<sup>18</sup>

The results of our simulations are the beta distribution values at random for the two nested subunits (equivalent to NUTS3 and NUTS2) for simulated nations. For example, for the 100 simulations, if a pair of region counts is 20 for NUTS2 and 35 for NUTS3, then we simulate 20 random values for regional GDP, population, and land area. Simultaneously, we generate 35 random values for regional GDP, population, and land area. Once we have these simulated values,

17 Our algorithm for region counts simulation is detailed in the OA: flowchart (Figure A1), R code (Figure A2), and simulation results (Figure A3).

18 The procedural explanation for the beta distribution can be found in our OA: algorithm flowchart (Figure A4), R code (Figure A5), and evaluation (Figure A6).





Notes: To calibrate inter-regional distribution (0-1 standardized index), we used simulated regional-level data on GDP, population, and land area. The X-axes show our five measures of geographic distribution. Values on Y-axes are comparative scores between distribution measured at NUTS2 and NUTS3 by their arithmetic differences. Confidence intervals calculated using 100 simulated samples.

Figure 5. The Cross-Unit Differences in Measuring Geographic Distribution.

we calculate our five geographic distribution measures at the two subunit levels and compare the values across the two levels for each measure. Our expectation is that we should see volatility in COV, WCOV, RDGINI, and AGC that comes from the aggregation process. We anticipate that SSGINI will show similar values measured at the two different levels because it incorporates relevant features into the calculation (i.e., productivity, population, or land share).

As depicted in Figure 5, SSGINI is the most reliable measure in the simulation. We refer to measurement consistency as the proximity between the NUTS2-level and NUTS3-level value. Measurement stability is the minimization of the oscillation range between the measures at NUTS2 and NUTS3. Figure 5(a) compares the average (with scaling in the min–max range) in values obtained in the distribution measures at NUTS3 (black circles) and the corresponding data measured at NUTS2 (hollow circles). As expected, projected estimates of regional distribution fluctuate more across NUTS3 than NUTS2.

The first takeaway from Figure 5 is that the most widely used measure of geographic distribution, COV, shows the most drastic difference between data measured at NUTS2 and NUTS3.<sup>19</sup> Thus, the choice of unit in COV is highly consequential to the measured value of the distribution. With the exception of SSGINI, each measure shows a significantly different value when measured at NUTS3 versus NUTS2. SSGINI’s overlapped data distribution at NUTS2 and NUTS3 indicates that they produce statistically indistinguishable values. Figure 5(a) thus highlights the consistency of SSGINI relative to commonly used measures in economic geography.

Similarly, Figure 5(b) shows the relative merit of using SSGINI to manage scale and unit number-driven volatility. The lower the value on the y-axis (representing the differences across units), the greater the level of stability is expected. SSGINI shows the least fluctuation across NUTS2 and NUTS3 of the analyzed measures. Thus, the simulation results indicate that SSGINI is the most consistent and stable aggregate indicator when dealing with differences in scale.

Importantly, as we emphasize in Section 4.2, the difference in value across NUTS2 and NUTS3 may be substantively meaningful. If the researcher has a clear theory to identify the appropriate unit of analysis at the NUTS2 or NUTS3 region, the scholar may use existing measures at that level with reasonable confidence. However, when the researcher is not able to make a theoretical distinction between units, SSGINI offers a more reliable representation of both levels of data.

19 We demonstrate robustness by setting different seeds for random draws (see OA Section 8).

## 9 Conclusions and Research Implications

Recent research has made considerable advances in measuring geographic variation within and across countries. These indicators include calculations of dispersion and concentration of geographic distribution. However, these works have not been centrally concerned with the selection of units and the implications for data comparability across units and measurement reliability across countries. In this article, we have offered advice on how to manage common research challenges in measuring geographic distribution and delved deeply into the unit question to offer approaches to avoid idiosyncratic choices that may impact researchers' results.

We stress, again, that the best way to manage concerns with the measure of geographic distribution lies in the articulation of a clear theory linking the unit of geography to a particular characteristic of interest. The best unit to choose is always the most theoretically appropriate unit. Even where researchers have followed the optimal path outlined in Figure 1, SSGINI can serve as a useful comparison measure to be sure that the differing number of units across countries is not increasing instability in one of the traditional measures of geographic distribution.

However, when the unit is unclear or appropriate data are not available, SSGINI should be considered because it is scalable. SSGINI has a potentially broad application to studies of economic and political geography, including measuring distributions of people, economic resources, and political support. Even if we know the best unit and have data for it, SSGINI helps to mitigate fluctuations that arise from the variation in the number of units across country cases. We hope that researchers draw from this discussion a need to take seriously the scale of the geographic unit in their study and to justify the choice of unit on theoretical grounds.

As we see it, potential problems in statistical analysis come in the consistency of the measures, affecting the magnitude of the results, and the stability of the measures, affecting the direction and significance of the measures. Figure 5, using simulated data, shows that both the magnitude and direction of statistical results may be affected by the choice of NUTS2 and NUTS3 data. Figure 3, using actual data from Denmark and Germany, shows how these differences might bear out. Not only are the magnitudes different across NUTS2 and NUTS3 data on the commonly used existing measures (COV, WCOV, RDGINI, and AGC), but the slope also differs in many cases. Using Danish COV or WCOV measures would give different significance levels, and perhaps direction of the effect, when used on NUTS2 versus the NUTS3 level. Our replication in Figure 4 shows that the choice of a measure not scaled to theoretical properties of the units could lead to incorrect inferences about a scholar's theory. What researchers need to assess is whether this difference is meaningful or spurious, given their theoretical approach.

We also stress that SSGINI addresses problems of measurement that cannot be managed with existing statistical methods. Variations in the scale, number of units, and the appropriate unit of analysis are conceptual and theoretical problems that have statistical implications but cannot be addressed post hoc in regression analysis. In general, we advise researchers to illustrate robustness in their empirical analysis. Simply demonstrating that results hold across different measures and across different units of analysis can go a long way towards assuaging concerns about the unit problem in research into geographic distribution.

Although attention to geographic distribution is growing in political science, very little guidance is available for those scholars interested in issues of concept and measurement to guide their research. While this paper takes on aspects of theory and empirical analysis around the unit question, much remains to be described in a palatable format for researchers. Future research could offer a conceptual map of the different distribution measures that we analyzed above. Few scholars in political science are familiar with these measures, and even fewer are familiar with their different properties. A simple mapping of these measures, their related concepts, and their empirical properties would be enormously useful for researchers keen to jump into the subject.

## Supplementary material

For supplementary material accompanying this paper, please visit

<https://doi.org/10.1017/pan.2019.14>.

## References

- Amrhein, C. G. 1995. "Searching For the Elusive Aggregation Effect." *Environment and Planning A* 27(1):105–119.
- Arbia, G., and F. Petrarca. 2011. "Effects of MAUP on Spatial Econometric Models." *Letters in Spatial and Resource Sciences* 4(3):173.
- Barro, R. J., and X. Sala-i Martin. 1992. "Convergence." *Journal of Political Economy* 100(2):223–251.
- Beramendi, P., D. W. Lee, M. Rogers, and K. Wikstrom. 2018. Endogenous regional borders. Technical report.
- Beramendi, P., and M. Rogers. 2018. "Disparate Geography and the Origins of Tax Capacity." *The Review of International Organizations*, <https://doi.org/10.1007/s11558-018-9319-2>.
- Berry, W. D., M. Golder, and D. Milton. 2012. "Improving Tests of Theories Positing Interaction." *The Journal of Politics* 74(3):653–671.
- Bochsler, D. 2010. "Measuring Party Nationalisation." *Electoral Studies* 29(1):155–168.
- Brühlhart, M., and R. Traeger. 2005. "An Account of Geographic Concentration Patterns in Europe." *Regional Science and Urban Economics* 35(6):597–624.
- Caramani, D. 2004. *The Nationalization of Politics*. New York: Cambridge University Press.
- Chen, J., and J. Rodden. 2013. "Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures." *Quarterly Journal of Political Science* 8(3):239–269.
- Dark, S. J., and D. Bram. 2007. "The Modifiable Areal Unit Problem in Physical Geography." *Progress in Physical Geography* 31(5):471–479.
- Fortunato, D., C. S. Swift, and L. K. Williams. 2016. "All Economics is Local." *Political Science Research and Methods* 6(3):1–21.
- Fotheringham, A. S. 1989. "Scale-Independent Spatial Analysis." In *Accuracy of Spatial Databases*, edited by M. Goodchild and S. Gopal, 221–228. London: Taylor and Francis.
- Giraudy, A., and J. Pribble. 2018. "Rethinking Measures of Democracy and Welfare State Universalism: Lessons from Subnational Research." *Regional and Federal Studies*, <https://doi.org/10.1080/13597566.2018.1473250>.
- Guo, J. Y., and C. R. Bhat. 2007. "Operationalizing the Concept of Neighborhood." *Journal of Transport Geography* 15(1):31–45.
- Hammersley, M. 1987. "Some Notes on the Terms Validity and Reliability." *British Educational Research Journal* 13(1):73–82.
- Hay, G., D. Marceau, P. Dube, and A. Bouchard. 2001. "A Multiscale Framework for Landscape Analysis." *Landscape Ecology* 16(6):471–490.
- Hayward, P., and J. Parent. 2009. "Modeling the Influence of the Modifiable Areal Unit Problem on Poverty in Pennsylvania." *The Pennsylvania Geographer* 47(1):120–135.
- Hoover, G. A., and P. Pecorino. 2005. "The Political Determinants of Federal Expenditure at the State Level." *Public Choice* 123(1-2):95–113.
- Jelinski, D. E., and J. Wu. 1996. "The Modifiable Areal Unit Problem and Implications for Landscape Ecology." *Landscape Ecology* 11(3):129–140.
- Jurado, I., and S. Leon. 2017. "Geography Matters." *British Journal of Political Science* 1–23.
- King, G. 2013. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton: Princeton University Press.
- Kwan, M.-P. 2012a). "How GIS Can Help Address The Uncertain Geographic Context Problem in Social Science Research." *Annals of GIS* 18(4):245–255.
- Kwan, M.-P. 2012b. "The Uncertain Geographic Context Problem." *Annals of the Association of American Geographers* 102(5):958–968.
- Laakso, M., and R. Taagepera. 1979. "Effective" Number of Parties." *Comparative Political Studies* 12(1):3–27.
- Lee, D. W., and M. Rogers. 2019. "Inter-Regional Inequality and the Dynamics of Government Spending." *The Journal of Politics* 81(2):.
- Lessmann, C. 2009. "Fiscal Decentralization and Regional Disparity." *Environment and Planning A* 41(10):2455–2473.
- Levitt, S. D., and J. M. Snyder. 1995. "Political Parties and the Distribution of Federal Outlays." *American Journal of Political Science* 39(4):958–980.
- Michalopoulos, S. 2012. "The Origins of Ethnolinguistic Diversity." *American Economic Review* 102(4):1508–1539.
- Monogan, J. E., and J. Gill. 2016. "Measuring State and District Ideology with Spatial Realignment." *Political Science Research and Methods* 4(1):97–121.

- Murguía, M., and J. L. Villaseñor. 2000. "Estimating the Quality of the Records Used in Quantitative Biogeography with Presence–Absence Matrices." *Annales Botanici Fennici* 37:289–296.
- Openshaw, S. 1984. *The Modifiable Areal Unit Problem*. Norwich, UK: Geo Books.
- Rey, S. J., and M. V. Janikas. 2005. "Regional Convergence, Inequality, and Space." *Journal of Economic Geography* 5(2):155–176.
- Rickard, S. J. 2012. "Electoral Systems, Voters Interests, and Geographic Dispersion." *British Journal of Political Science* 42(4):855–877.
- Rodden, J. 2010. "The Geographic Distribution of Political Preferences." *Annual Review of Political Science* 13:321–340.
- Rodríguez-Pose, A., and R. Ezcurra. 2009. "Does Decentralization Matter for Regional Disparities? A Cross-Country Analysis." *Journal of Economic Geography* 10(5):619–644.
- Rogers, M., and D. W. Lee. 2019. "Measuring Geographic Distribution for Political Research." <https://doi.org/10.7910/DVN/DKY4DY>, Harvard Dataverse. V1.
- Soifer, H. D., and A. M. Alvarez. 2017. "Choosing Units of Analysis in Subnational Research." Working Paper.
- Spiezia, V. 2002. *Geographic Concentration of Production and Unemployment in OECD Countries*. Paris: OECD.
- Tam Cho, W. K., and J. G. Gimpel. 2007. "Prospecting for (Campaign) Gold." *American Journal of Political Science* 51(2):255–268.
- Wong, D. 2009. "The modifiable Areal Unit Problem." In *The SAGE Handbook of Spatial Analysis*, edited by A. S. Fotheringham and P. Rogerson, 105–123. London: SAGE Publications.