(2017) 47, 289–308

The Review of Regional Studies

The Official Journal of the Southern Regional Science Association

How (not) to Estimate GDP at the Sub-State Level: The Usefulness of the Economic Census*

José G. Caraballo-Cueto

Department of Business Administration,University of Puerto Rico at Cayey, PR

Abstract: An accurate estimation of GDP for sub-state divisions is needed for studies on the overall size and growth of smaller economies, assessment of the industrial composition by subdivisions, and the impact of economic policies, among other reasons. In this paper, I review two popular methods to infer GDP in counties, both of which rely heavily on payroll, without controlling for labor productivity differences between subdivisions. By considering labor productivity heterogeneity across counties, I propose a new methodology that appears to provide a better map of the GDP of counties. For forecasting, my approach uses payroll statistics, but to a lesser extent. The new method is tested with data from Puerto Rico, a suitable jurisdiction given its significant heterogeneity between counties. I am able to map changes in economic geography associated with changes in the state's economic structure. Like many other jurisdictions, Puerto Rico has undergone a deindustrialization process that ended up concentrating GDP in one geographical cluster.

Keywords: county GDP, economic geography, regional development

JEL Codes: R1, R3, O2

1. INTRODUCTION

The relevance of sub-state level data to evaluate trends in regional development has been stated in the related literature, from Andrews (1954) to Shaffer et al. (2004). The reasons to use reliable estimates of GDP or GNI at the sub-state level are many: to study the overall size and growth of smaller economies, to assess industrial composition by counties and the impact of economic policies, to inform planners about industries that have a better effect on GDP, to manage development grants and allocate industrial incentives, to support regional developers and planners in economic strategies for underdeveloped areas, and to allow analysts to circumscribe an area for individual analysis, among other things. The business community needs an accurate measure of GDP to plan the location of retail stores and to find appropriate clusters (Heizer et al., 2004). The Bureau of Labor Statistics uses sub-state GDP to estimate the unemployment rate at the sub-state level.

However, the current estimates of GDP at the sub-state level are based on wage bill or employment levels, by making relatively strong assumptions, as discussed below. Here, I do not discuss the methodologies that use employment levels to infer regional GDP because their sub-

* Acknowledgements: I have benefited from the suggestions of the three excellent anonymous reviewers and the editor, to whom I am deeply grateful. I am also grateful to Stephanie Rodríguez-Galí for her research assistance, to Yoel Velázquez, Mario Medina, and to participants in the working paper seminar at the University of Puerto Rico for their helpful comments. The usual disclaimer applies.

The author is an assistant professor and the director of the Census Information Center in the University of Puerto Rico at Cayey, PR. E-mail: jose.caraballo8@upr.edu.

© Southern Regional Science Association 2017.
ISSN 1553-0892, 0048-749X (online)
www.srsa.org/rrs

optimality has already been discussed elsewhere (Lahr, 2001). The first methodology discussed here assumes that worker productivity levels (sales divided by employment; henceforth, productivity) at the state level are the same at each sub-state level (assumption of spatially-invariant technology). The second method assumes that the productivity levels in a given industry in a given subdivision are equal to the productivity level of the same industry at the state level. As shown below, the productivity level differences are relevant, not necessarily to explain the source of growth between counties, but to avoid overestimating or underestimating the GDP estimates in a given county. For instance, if a dollar paid in payroll is related to a GDP of $200 in county X and to a GDP of $100 in county Y, then factor payments of $2 in each county would point–under current methodologies–to a similar GDP in each county, while county Y actually produces half of the GDP of county X.

Is there a better procedure to estimate GDP at the county level with public data? I contribute to this literature by improving the GDP estimates at the county level. My innovation is to use value added and sales data from the economic census to anchor the estimation in a given year. Because an economic census is published every five years, I also show a simple and effective way to extrapolate GNI or GDP in a given county for the years preceding the next economic census. Thus, this benchmarking method does not assume productivity levels to be the same at any level. This is an important consideration because, in jurisdictions such as Puerto Rico where there is a relatively large and diverse manufacturing sector in some counties but not in others, assuming that productivity is equal to the state level would lead to suboptimal estimations. Due to these considerations, Puerto Rico is an interesting case study to test the proposed methodologies.

I argue that a reasonable method should map qualitative observations on the state, should be consistent with the trends observed in other data sources, and should be parsimonious (relatively simple to calculate) with the level of information made available to the public (private estimates are not generalizable and are not revised by statistical agencies), as these estimates must be done on a regular basis. In that sense, the methodology discussed by Lahr (2001) to estimate the GDP of subdivisions can be useful and can be combined with my method, but it is not feasible in this case study.[1]

The proposed methodology provides a better map of the qualitative observations on economic geography than the current two methods, shows parsimony, and is internally consistent with the trends observed in other data sources. This method accounts for productivity level differences or production shares that are lost in industry aggregation.

In the next section, I discuss two of the current methodologies, explain the proposed new method, and provide descriptive statistics. In Section 3, results from the different methods are presented and I discuss how the structural change that Puerto Rico has experienced has changed the economic development in non-metropolitan counties. Section 5 concludes.

[1]In Puerto Rico, the last input-output table published corresponds to 2002 and there is no aggregate data for an industry mix below the 4-digit NAICS level thereafter.

© Southern Regional Science Association 2017.

2. METHODOLOGIES AND DATA

2.1 Data

In this section, I briefly describe the methodology that the Census Bureau disclosed about the Economic Census, as shown on their website. The Economic Census gathers information every five years from all private companies that are contained in the government records. It has several benefits. For example, most of the sectors are not subject to sampling errors. In addition, the Census contains information on sales, payroll, and employment by establishment size for different industries and it reports sales and employment by county. The Census Bureau cannot share this information with the Internal Revenue Service or related departments. The Economic Census is based on employer establishments, so companies must file a separate report for each location or establishment. If a company operates in multiple markets (e.g., manufacturing and retailing), a separate report for each activity is requested. Title 13 of the Code of Laws of the United States requires companies to answer the questionnaires.[2]

Like any other instrument, economic censuses are also subject to non-sampling errors, but they are not measured directly. The Census Bureau states on its website that the Bureau, "employs quality control procedures in all phases of the collection, processing, and tabulation of the data to minimize the effects of non-sampling error." For sample-based estimates, measures of the sampling variability are provided only for the construction sector.

Other payroll data used in this research comes from the Quarterly Census of Employment and Wages (QCEW). Data regarding GDP and GNI comes from the Planning Board (in the case of Puerto Rico) and the BEA (in the case of Illinois).

2.2 Methodologies

The GDP can be measured by the expenditure approach, by adding the value added activities of each sector, or by the income approach. We cannot apply any of these approaches directly to counties, because most countries (including the U.S.) do not have these data disaggregated at the county level. Thus, we have to estimate the economic activity of counties based on the available public data.

The two current methodologies for estimating GDP at the county level have two similarities: both are based on wage data and both are based on certain assumptions about productivity levels. These methods are discussed in sections2.2.1 and 2.2.2. In section 2.2.3, Ipresent my methodology.

2.2.1 Unweighted Payroll Ratio

The first method can be named the "unweighted payroll ratio" because the estimated GDP of county *i* with respect to the state is estimated by dividing the total wages. In particular,

$$Y_i = X_{pr} * \left(\frac{W_i}{W_{pr}}\right) \tag{1}$$

where Y_i is the estimated GDP at county i, X_{pr} is the measured GDP in national accounts (in the case of Puerto Rico, we used the GNI as computed by the Planning Board in Puerto Rico, for reasons that will become clear below), W is the wage bill, and the pr indicates Puerto Rico. The advantage of this method is its computational ease. Further, it only requires data from the

[2] See http://uscode.house.gov/ for more information.

© Southern Regional Science Association 2017.

QCEW. The main disadvantage is that this method is based on the idea that the productivity levels in county i map the productivity of the whole territory, in this case, Puerto Rico. This can be problematic, because a rural county with a low share of the highly productive sectors (e.g., manufacturing), such as Villalba, has a very different productivity level than the rest of the country.

The opposite would apply to a county in the Metropolitan area, such as Carolina. For instance, according to the Economic Census of Puerto Rico, the average county had a ratio of sales to wage bill of 5.0 in 2012, but there was a relatively high standard deviation of 4.7. In Carolina, the ratio of sales to wage bill went up to 15 and in Villalba, it went down to 1.1. In most of the U.S., there is a similar situation. For instance, in the 2012 Economic Census of Illinois, a state with a relatively large mining sector, Douglas County had a ratio of sales to wage bill of 17, but Cook County had a ratio of 47. Labor productivity differences may be the explanatory factor behind these differences in the sales to wage ratio. With such a high dispersion, the assumption that the productivity in county i is the same as that at the state level is probably too strong.

2.2.2 Weighed Payroll Ratio

Another approach to estimate GDP in a given county is obtained by following the methodology of Kendrick and Jaycox (1965),

$$Y_i = \sum_{n=1}^{N} \left(\frac{W_{n,i}}{W_{n,pr}}\right) * X_{n,pr} \tag{2}$$

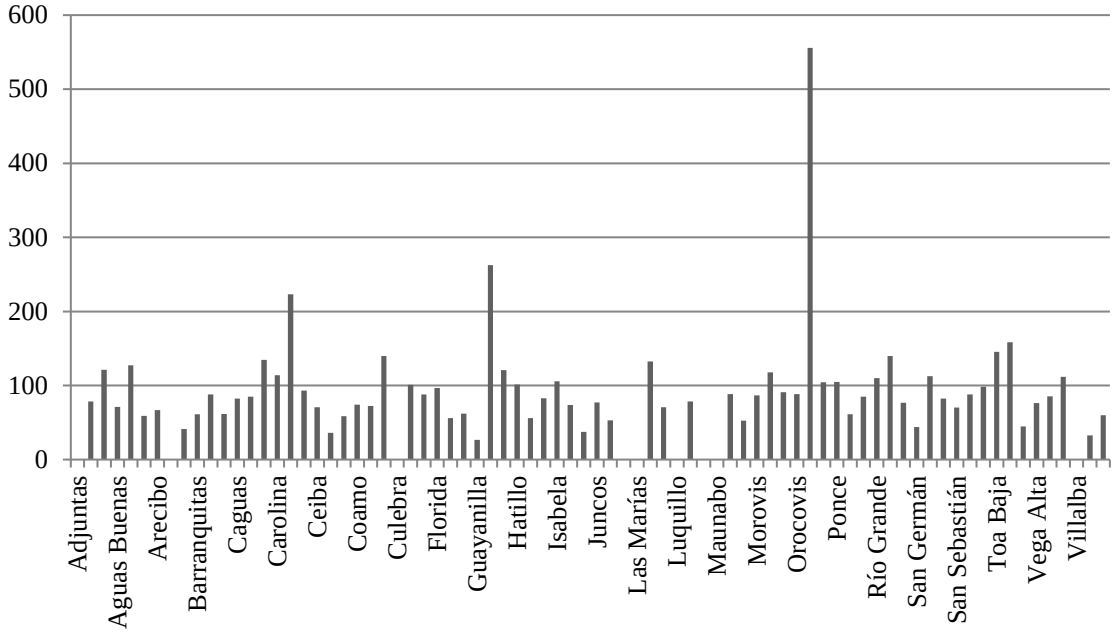
where Y_i is the estimated GDP at county i, X is the measured GDP in national accounts, W is the wage bill, the pr indicates Puerto Rico, and the n indicates an industry. In other words, the largest difference between Equations (1) and (2) is that, in this method, the total wage bill is weighted by the wage bill in every industry. A similar approach to Equation (2) was developed by Niemi (1972). This method is based on the premise that the worker productivity in a given industry at the county level matches the productivity at the state level in that industry. However, the premise that worker productivity at the county level in a given industry matches its productivity counterpart at the state level can also be problematic if there is a large diversity in the same industry across counties. For instance, based on the Economic Census, the worker productivity in the manufacturing sector in 2012 in Barceloneta (Northern region) was $4,236, while in Adjuntas (Center-West region) it was $42, and in Puerto Rico as a whole it was $914.

Drilling down manufacturing to four digits would not necessarily reveal major differences. In the Illinois 2012 Economic Census, Printing had a labor productivity of $97 in Jefferson County and $223 in Kane County. In addition, there is data for only 25 percent of all four-digit sectors at the county level, vis-à-vis 55 percent of the two-digit NAICS sectors that are available at the county level. In the case of the Metropolitan Statistical Areas or MSAs (core urban areas with a population of 50,000 or more, plus adjacent territories), 44 percent of all four-digit sectors are available at the county level, but this is still well below the 75 percent reached with the two-digit sectors. Thus, in the case of the U.S., the MSAs are recommended (as done by the BEA) and, in most cases, I concentrate this analysis on the two-digit sectors, given data availability.[3]

[3]If greater efforts are made in the future to collect U.S. sales data at the county level, our method can be applied to the county level in every state in the U.S.

© Southern Regional Science Association 2017.

Figure 1: Labor Productivity Levels in Construction, 2007



Note: Labor productivity is equal to value added in construction divided by number of workers.
Source: Author's calculation based on Economic Census (2007).

A similar variance is observed in the rest of the sectors. For instance, based on the 2007 Economic Census of Puerto Rico, worker productivity in retail sales was $200 for Puerto Rico as a whole, but in counties such as Guaynabo, it can be as high as $290, and in counties such as Arroyo it can be as low as $109. In Figure 1 we observe how the productivity in construction also differs significantly across counties. Similar outcomes are found if we move down to the four-digit sectors. In 2012, counties in Illinois, such as Douglas, had a labor productivity of $57 in Accounting and $50 in Computer Systems, but other counties, such as DuPage, had $114 in Accounting and $202 in Computer Systems.

2.2.3 A Proposal

Is there a better approach? In the following, I propose a new and relatively efficient method to anchor the GDP of county *i* in Puerto Rico,

$$Y_i = X_{pr} * \sum_{n=1}^{N} \left(\frac{C_n}{C_{pr}}\right) \tag{3}$$

where Y is the estimated GDP at county i, X is the measured GDP in national accounts, pr indicates Puerto Rico, and C is sales obtained from the economic census. In Equation (3), I estimate the GDP of county i by adding the value added of manufacturing and construction with the total sales in the rest of the two-digit NAICS sectors described in Table 1. This total sales is then divided by the sum of all of two-digit NAICS sectors available at the state level. In the economic census, "value added" is only measured for manufacturing and construction.[4]

[4]In the census data, value added differs a bit from value added in the national accounts, because the Census Bureau used the accounting book value of depreciation, not the economic value and cost of purchased services, but the differences are relatively small.

© Southern Regional Science Association 2017.

Table 1: Sales Share in Economic Census and Aggregate Accounts, 2012

Sectors	Sales Share in Economic Census	Sector Share in Aggregate Accounts
Agriculture, Forestry, Fishing and Hunting	0.05%	0.8%
Mining, Quarrying, and Oil and Gas Extraction	0.1%	0.04%
Construction	2.2%	1.3%
Manufacturing	43.8%	46.3%
Retail Trade	14.3%	4.7%
Transportation and Warehousing	1.1%	0.9%
Information	7.1%	2.4%
Finance and Insurance	8.4%	5.1%
Real Estate and Rental and Leasing	0.9%	15.2%
Professional, Scientific and Technical Services	1.8%	1.6%
Management of Companies and Enterprises	0.48%	0.07%
Administrative and Support and Waste Management and Remediation Services	1.7%	1.8%
Educational Services	0.3%	0.7%
Health Care and Social Assistance	3.5%	3.3%
Arts, Entertainment, and Recreation	0.1%	0.1%
Accommodation and Food Services	2.4%	1.8%
Other Services, except Public Administration	0.4%	0.4%

Source: Author's calculation based on Economic Census (2012) and Planning Board (2016).

However, in 2012, the average county share of manufacturing and construction represent more than 46 percent of the total GDP, as shown in Table 1. Furthermore, more than 80 percent of the products sold in Puerto Rico are imported (Curet-Cuevas, 2003), agriculture representedless than 1 percent of the total GDP in the average county, and I did not include sales from wholesale trade. Thus, I do not have major concerns that a double counting of the final sales in this approach would distinctively contaminate the computed GDP of county i. As there are data limitations at the county level in the U.S., our approach can be applied to MSAs because most of the sectoral data are available (75 percent, as noted before).

Note in Table 1 that the sales shares in the economic census in 2012 closely matched the sectorial shares in Puerto Rico's GDP, except for Retail Trade, Information, and Real Estate. Although care should be taken when independently analyzing these three sectors at the county level, the service sectors showed greater complexities when measured and the rest of the sectors do not show great divergences from the national accounts. Below, I map the economic geography obtained when I use an industry weighed version of Equation (3) such as,

$$Y_i = \sum_{n=1}^{N} \left(\frac{C_{n,i}}{C_{n,pr}} \right) * X_{n,pr} \tag{4}$$

© Southern Regional Science Association 2017.

where Y_i is the estimated GDP at county i, X is the measured GDP in national accounts, pr indicates Puerto Rico, the C indicates sales obtained from the economic census, and n indicates industry. Equation (4) can be an even better approach for subdivisions in the U.S. because state GDP is measured by adding value added from different industries.[5] In this paper, I focus on Equation (3) since the output of manufacturing and retail trade in Puerto Rico, and possibly in Delaware, can be contaminated by income transferred from other jurisdictions through the transfer pricing mechanisms of multinational companies (see Grubert and Slemrod, 1998; Federal Reserve Bank of New York, 2012).[6] However, I show below that most of the counties have a close correlation with both Equations (3) and (4), except for San Juan.

Equation (3) represents the method to benchmark the county GDP estimation in the year when the economic census is published (the last one was 2012), but I propose the following interpolation between the censuses:

$$\frac{Y_i}{W_i} = w_i \rightarrow Y_{i,t+1} = w_i * W_{i,t+1} * \hat{e}_{pr,t} \tag{5}$$

where $\hat{e}_{pr,t}$ is the labor productivity growth in Puerto Rico in period t, as reported by the statistical agency (in the case of Puerto Rico, the Planning Board calculates productivity as equal to real GNP divided by employment). In other words, for the year after the publication of the economic census, we adjusted the sales to wage bill ratio obtained from the economic census by multiplying it by the wage bill from the QCEW in period $t + 1$ and by the productivity growth at the state level.

The overall growth term in labor productivity in Equation (5) should have a small impact and should not affect estimation in a given county, because in most of the subdivisions, productivity is relatively stable. Even in Grundy, a county with a relatively large mining sector, labor productivity, in nominal terms, was $383 in 2012 and $353 in 2007, while labor productivity in Illinois was $109 and $124, respectively. In Bloomington, an MSA in Illinois, the overall nominal labor productivity was $147 in 2012 and $151 in 2007. However, in states with relatively high productivity swings between periods, the productivity growth term in Equation (5) can be dropped to forecast GDP in their subdivisions.[7]

Perhaps a better procedure to verify the forecasting accuracy of Equation (5) is to compare the county estimates of 2007 and 2012 obtained from Equations (1), (2), and (5) with the actual county shares obtained from the 2007 and 2012 published economic censuses. The following Theil t-statistic is used to compare the accuracy of the three methods:

[5]https://www.bea.gov/regional/pdf/gsp/GDPState.pdf Here, BEA confirms that GDP by state differs from GDP for the nation for two minor reasons: in revision schedules and in that GDP by state excludes compensation of federal personnel stationed abroad and government consumption of fixed capital for structures and equipment located abroad. In addition, BEA states in this report: "In general, there are two procedures for estimating GDP by state and its components, one uses state-level Census Bureau value-added data for the goods-producing industries to estimate GDP by state for these industries, and the other utilizes Census Bureau receipts and payroll data, or company financial data to estimate Gross Operating Surplus for the services-producing industries" (p. 2).

[6]Gross National Product (GNP) is generally used in Puerto Rico in preference to GDP because much of the GDP is distorted by the transfer pricing of multinational companies. Thus, for the first three equations we used GNP and for equation (4) we used GDP. Delaware is also found to receive income from other jurisdiction through accounting practices (Bruce et al., 2007).

[7]This treatment is similar to the BEA approach, wherein no productivity adjustments are made to the earnings bill in their estimation.

© Southern Regional Science Association 2017.

$$T = \frac{\left[\frac{1}{i}\sum_{i=1}^{K}(C_i - P_i)^2\right]^{0.5}}{\left[\frac{1}{i}\sum_{i=1}^{K} {C_i}^2\right]^{0.5} + \left[\frac{1}{i}\sum_{i=1}^{K} {P_i}^2\right]^{0.5}} \tag{6}$$

This statistic was originally developed by Theil (1958) and has been used extensively in the literature, including Greene (2003), to measure the forecasting accuracy of a given method. Thus, the lower the t-statistic, the higher the accuracy of a given estimation.

Because there are few possibilities to cross validate a national accounting system at the county level, I cross validate the estimates with the overall trends across the Puerto Rican economy. For instance, counties with the largest presence of manufacturing or with significant economic activity in many sectors should appear as the most dynamic sectors. Thus, a methodology that ranks a county with low manufacturing or low service output as a leading county should be considered suboptimal.

Thereby, I take advantage of the estimates to evaluate whether the deindustrialization in Puerto Rico motivated a higher GDP concentration or clustering in a few counties. This final analysis would help us to understand the heavy traffic, among other planning issues, occurring every working day on every highway that heads towards the Metropolitan area, which could be reducing labor productivity (by losing time driving), decreasing leisure time, increasing economic inequality between counties, and causing even more pollution in the Metropolitan area (from the additional cars). To this end, we compare the clusters obtained using the Anselin Local Moran's I (Anselin, 1995) from both the unweighted payroll method and our method. The Anselin Local Moran's I identifies the areas of clustering by location and by values similar in magnitude, and is given by,

$$I_i = \frac{Y_i - \bar{Y}}{S_i^2}\sum_{j=1, j\neq i}^{N} w_{i,j}(Y_i - \bar{Y}) \tag{7}$$

where w is the spatial weight between i and j, and

$$S_i^2 = \frac{\sum_{j=1, j\neq i}^{N} w_{i,j}}{N-1} - \bar{Y}^2. \tag{8}$$

A positive and statistically significant I would indicate that the county is surrounded by counties with similar values, thus creating a cluster.

We also apply the Getis-Ord G_i* analysis (Mitchell, 2005), better known as the Hot Spot Analysis, which is given by

$$G_i^* = \frac{\sum_{j=1}^{N} w_{i,j}Y_j - \bar{Y}\sum_{j=1}^{N} w_{i,j}}{S\sqrt{\frac{\left[N\sum_{j=1}^{N} w_{i,j}^2 - \left(\sum_{j=1}^{N} w_{i,j}\right)^2\right]}{N-1}}}, S = \sqrt{\frac{\sum_{j=1}^{N} Y_j^2}{N} - (\bar{Y})^2}. \tag{9}$$

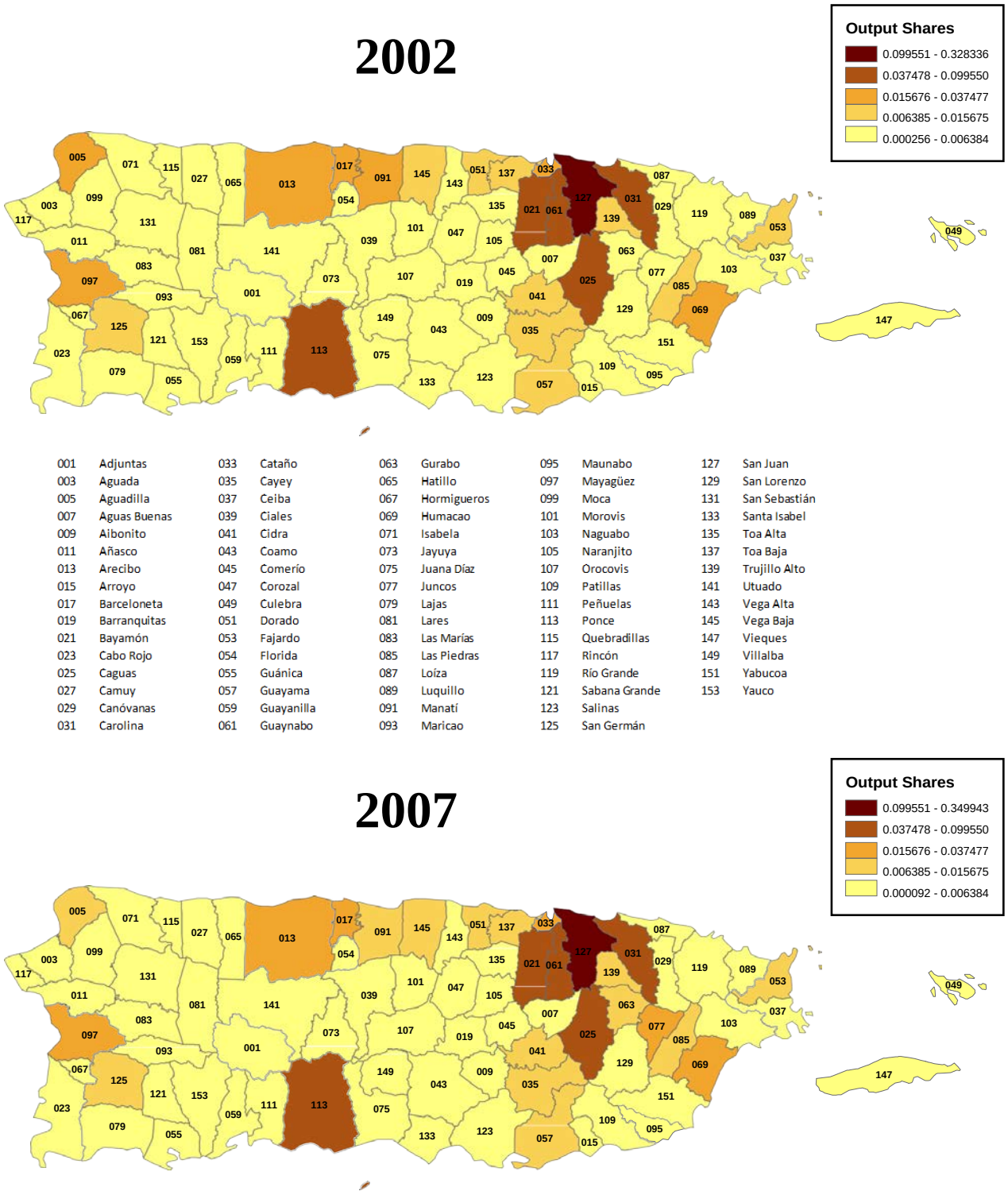
A Hot Spot would be a county that has a high $G^*$and is surrounded by other features with high values as well. A Cold Spot is a county with a negative and statistically significant G^* that is surrounded by similar counties with low values.

3. RESULTS

In Figure 2, I observe the economic geography obtained from the unweighted method. In all the maps, I used the Jenks natural breaks classification method to cluster the counties into different categories, a method that seeks to reduce the variance within categories and to

© Southern Regional Science Association 2017.

Figure 2: Output Shares based on the Unweighted Method, 2002 and 2007



Source: Author's calculation based on Economic Censuses of corresponding years.

© Southern Regional Science Association 2017.

maximize the variance between categories (Jenks, 1967). I used the same thresholds in the following maps that estimate GNI at the county level to allow a comparison between estimation methods, but the last map uses its own natural break, as it is estimating GDP at the county level.[8]

Manatí and Humacao are ranked in the third bracket for 2002, suggesting that these counties produced more than Cidra or Guayama, which were countries with higher manufacturing activity. In particular, Cidra's value added in manufacturing was triple that of Manatí in 2002. Guayama's manufacturing value added represented $5.5 million in 2002 vis-à-vis $466,000 in Humacao.

There appear to be even more incongruities when we observe the time-series dimension. Since 1996, Puerto Rico has undergone a deindustrialization process, in part thanks to the end of the U.S. incentives that provided tax cuts to manufacturing companies established in Puerto Rico (Collins et al., 2006). No vigorous economic strategy emerged afterwards that could counteract this reduction in economic activity, especially in the country-side where many of these companies were operating. Thus, a method that estimates GDP in counties should show a progressive reduction in GDP outside of the Metropolitan area. But in Figure 2, we see that the first estimation method shows a very similar economic geography, which could indicate stabilization. Furthermore, new production clusters appear to emerge in 2007: Gurabo was producing at a level similar to Cidra and Juncos was suddenly one of the most economically active municipalities, surpassing Guayama.[9]

Similar inconsistencies were shown when the payroll weighed method is applied, shown in Figure 3. For instance, Cataño was in the top tier of counties in 2002 in terms of economic activity. On the other hand, Barceloneta, the county with the largest manufacturing value added in Puerto Rico, representing 20% of the total value added in Puerto Rico's manufacturing, was in the lowest tier of GDP shares, with less than 1% (suggesting that Barceloneta produced less than 1% of Puerto Rico's total GDP). In terms of trends, this second methodology also showed counties such as Isabela and Juana Díaz, which, in 2002, were in the lowest tier of GDP shares, and in 2007, were some of the most dynamic municipalities. In fact, in 2007, Isabela and Juana Díaz were in the same bracket as Carolina, a county in the Metropolitan area where most of the economic activity took place. Thus, instead of observing a progressive dismantling of the economic activity in these rural counties, the first two methodologies appear to show relatively more economic activity in rural counties and higher equality of GDP across the counties.

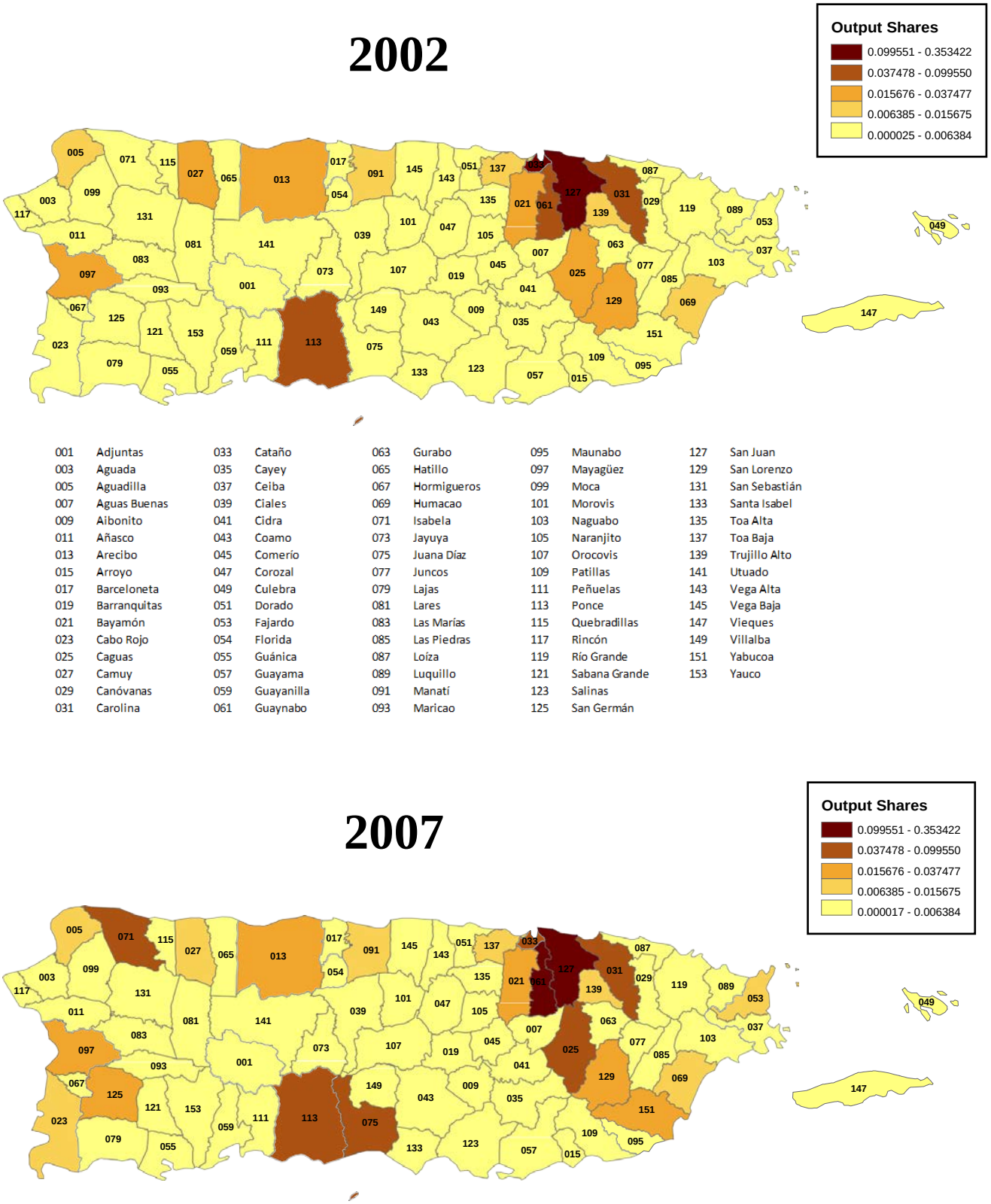
Similarly, these payroll methods confront challenges in the U.S. as well. For instance, according to the BEA, the GDP of the Bloomington, Illinois MSA was estimated at $10.5 billion in 2012. This GDP estimation of Bloomington could be potentially affected by the relatively large labor productivity differences of the service sector in Bloomington vis-à-vis Illinois. In particular, Bloomington had a labor productivity of $98 in Professional Services (NAICS 54), while Illinois had $193 in 2012. If I use the proposed method described in Equation (3) to estimate Bloomington's GDP, it would show that the GDP in Bloomington was $6,659 million.

[8]To assess the time-series dimension, there should be an analysis with our clustering maps, i.e. the last ones.

[9]Juncos was one of the counties that received a severe hit from this deindustrialization process with the disappearance of many factories.

© Southern Regional Science Association 2017.

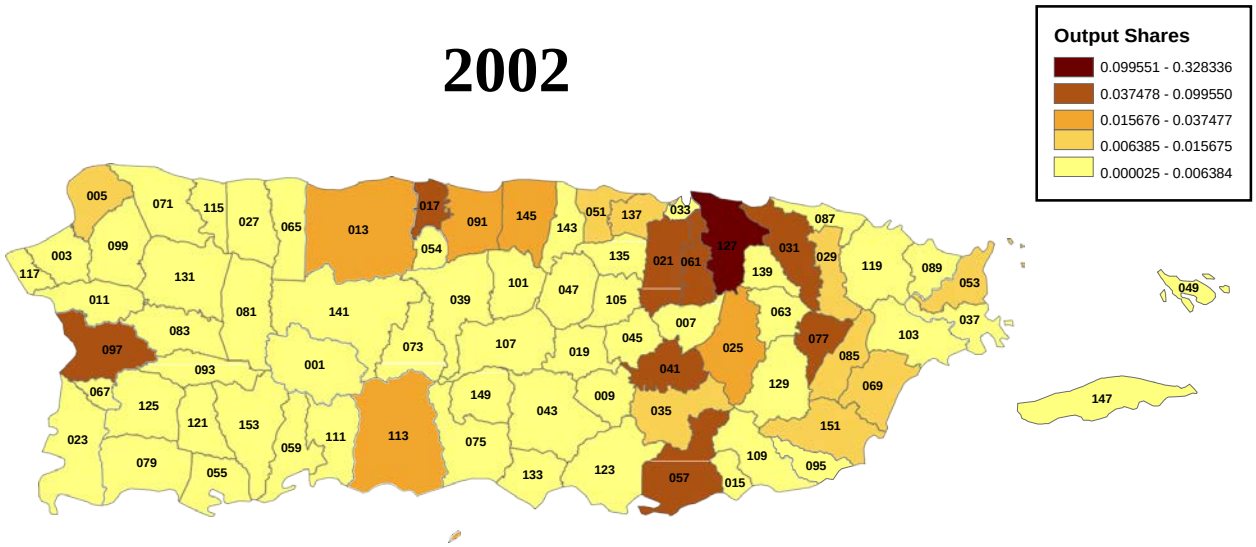
Figure 3: Output Shares based on the Weighed Method, 2002 and 2007



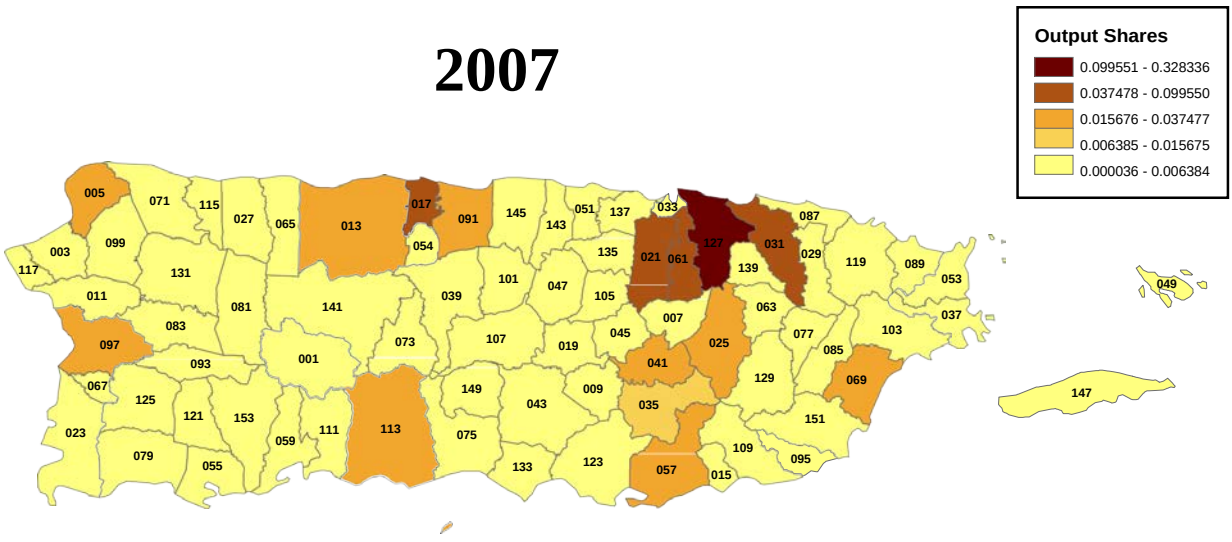
Source: Author's calculation based on Economic Censuses of corresponding years.

© Southern Regional Science Association 2017.

Figure 4: Output Shares using Economic Census, 2002-2007



001	Adjuntas	033	Cataño	063	Gurabo	095	Maunabo	127	San Juan
003	Aguada	035	Cayey	065	Hatillo	097	Mayagüez	129	San Lorenzo
005	Aguadilla	037	Ceiba	067	Hormigueros	099	Moca	131	San Sebastián
007	Aguas Buenas	039	Ciales	069	Humacao	101	Morovis	133	Santa Isabel
009	Aibonito	041	Cidra	071	Isabela	103	Naguabo	135	Toa Alta
011	Añasco	043	Coamo	073	Jayuya	105	Naranjito	137	Toa Baja
013	Arecibo	045	Comerío	075	Juana Díaz	107	Orocovis	139	Trujillo Alto
015	Arroyo	047	Corozal	077	Juncos	109	Patillas	141	Utuado
017	Barceloneta	049	Culebra	079	Lajas	111	Peñuelas	143	Vega Alta
019	Barranquitas	051	Dorado	081	Lares	113	Ponce	145	Vega Baja
021	Bayamón	053	Fajardo	083	Las Marías	115	Quebradillas	147	Vieques
023	Cabo Rojo	054	Florida	085	Las Piedras	117	Rincón	149	Villalba
025	Caguas	055	Guánica	087	Loíza	119	Río Grande	151	Yabucoa
027	Camuy	057	Guayama	089	Luquillo	121	Sabana Grande	153	Yauco
029	Canóvanas	059	Guayanilla	091	Manatí	123	Salinas		
031	Carolina	061	Guaynabo	093	Maricao	125	San Germán		



Source: Author's calculation based on Economic Censuses of corresponding years.

© Southern Regional Science Association 2017.

Given these incongruities, is there a better approach? In Figure 4, I show the estimation using Equation (3). Generally, this is correlated with the previous ones, but with important improvements. No inconsistencies are found between counties, as in the previous methods. For instance, Barceloneta (9 percent), Mayaguez (4 percent), and Cidra (3.6percent) are the greatest clusters of production outside the Metropolitan area.

This method shows that four counties in the Metropolitan area produce close to 55 percent of Puerto Rico's GDP. This can explain the heavy traffic congestion every morning on every road heading to the Metropolitan area. According to the Puerto Rican Community Survey 2009-2014, on average 29 percent of the workers that lived in Utuado took more than 45 minutes to travel to work each day, a proportion that inched up to 38 percent in Rio Grande and 31 percent in Toa Baja. Similar congestion has been observed in other counties in other jurisdictions such as New York City and Mexico City, situations that may decrease worker productivity and may decrease the motivation to work (Weisbrod et al., 2003).

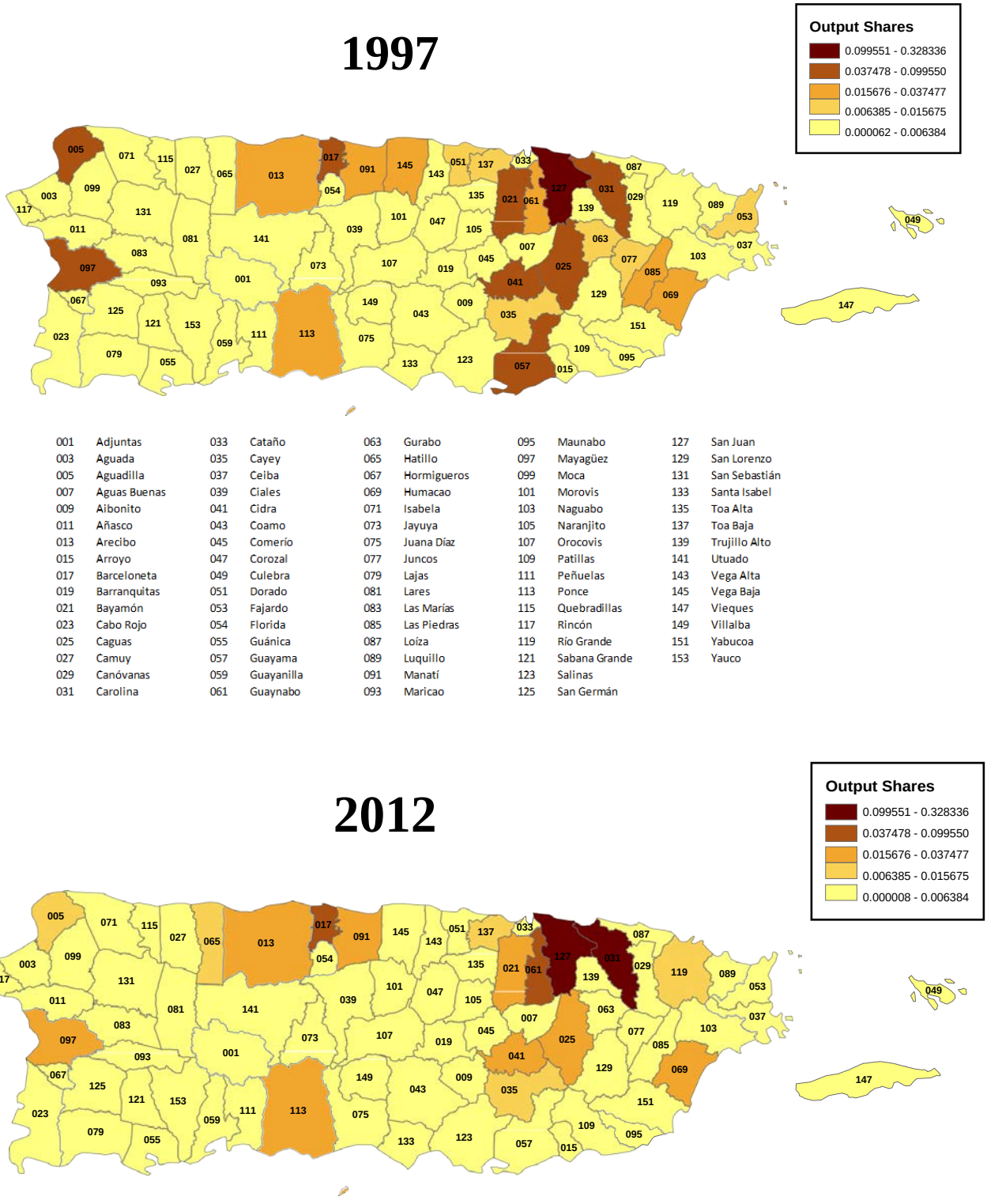
The productive dismantling of the country-side is also easier to capture under myproposal. Manufacturing represented almost 58 percent of total GDP in Puerto Rico in 2007, while in 2012, it represented 46 percent, a clear indication of the power that deindustrialization had after the elimination of section 936 of the federal tax code, which provided federal tax incentives for U.S. manufacturing companies operating inPuerto Rico. However, this deindustrialization was not distributed homogeneously across the island. In Figure 4, we see how, from 2002 to 2007, the economic geography became more unequal in terms of production. This is even more evident in Figure 5.

Hatillo and Rio Grande used to produce 0.5 percent of Puerto Rico's GDP in 1997, which increased to 0.7 percent and 0.9 percent, respectively, in 2012. The relevance of manufacturing in Hatillo's GDP increased from 5 percent in 2002 to 15 percent in 2012. In Rio Grande, the growth was mostly due to an increase in retail trade. Aside from these two small increases, the rest of the counties that had a relatively vibrant local economy were worse in 2012, reducing working opportunities for their populations and exacerbating problems with traffic and access to income. Some of the counties most affected by deindustrialization were Vega Baja, Dorado, Guayama, Juncos, Fajardo, Gurabo, and Las Piedras. Mayaguez, Cidra, Caguas, Aguadilla, and Bayamon were other municipalities with declining GDP.

This higher concentration in the Metropolitan area is verified by spatial statistical analysis, such as the one shown in Figure 6. The map shows the Anselin Local Moran's I, which highlights four types of clusters: a cluster of high values (HH), a cluster of low values (LL), an outlier in which a high value is surround primarily by low values (HL), and an outlier in which a low value is surrounded primarily by high values (LH). The progressive concentration of GDP in a few counties was clear in every period from 1997 to 2012. In 1997, there were four clusters of production: one in the North-Western region, one in the Southern area, one in the Northern region, and a relatively high focus in the Metropolitan area. In 2012, all of the relatively small centers of production in the rest of the territory disappeared: Only the Metropolitan area had a high and statistically significant niche of production.

© Southern Regional Science Association 2017.

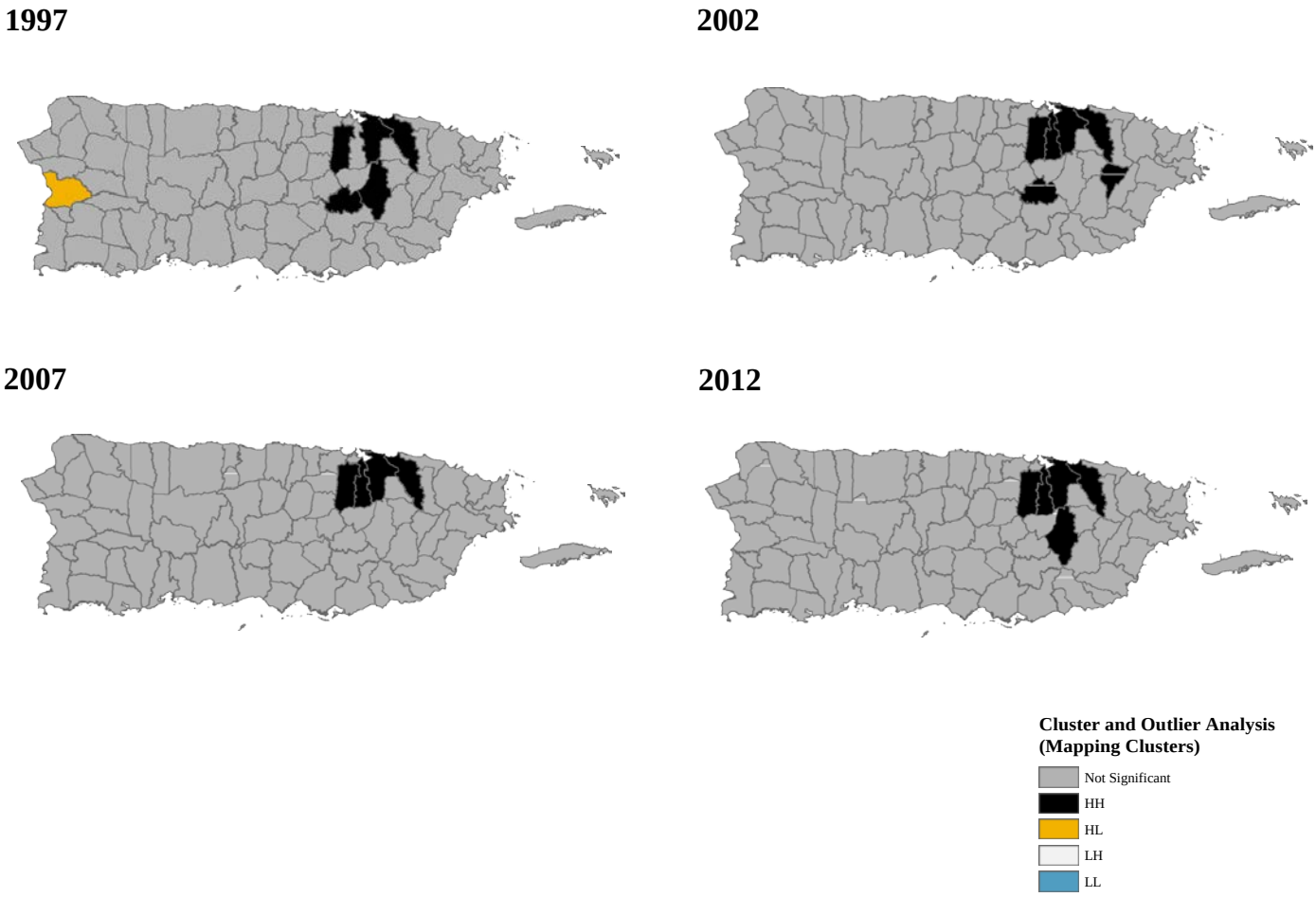
Figure 5: Impact of Deindustrialization, 1997-2012



Source: Author's calculation based on Economic Censuses of corresponding years.

© Southern Regional Science Association 2017.

Figure 6: Anselin Local Moran's I using Economic Census, 1997-2012



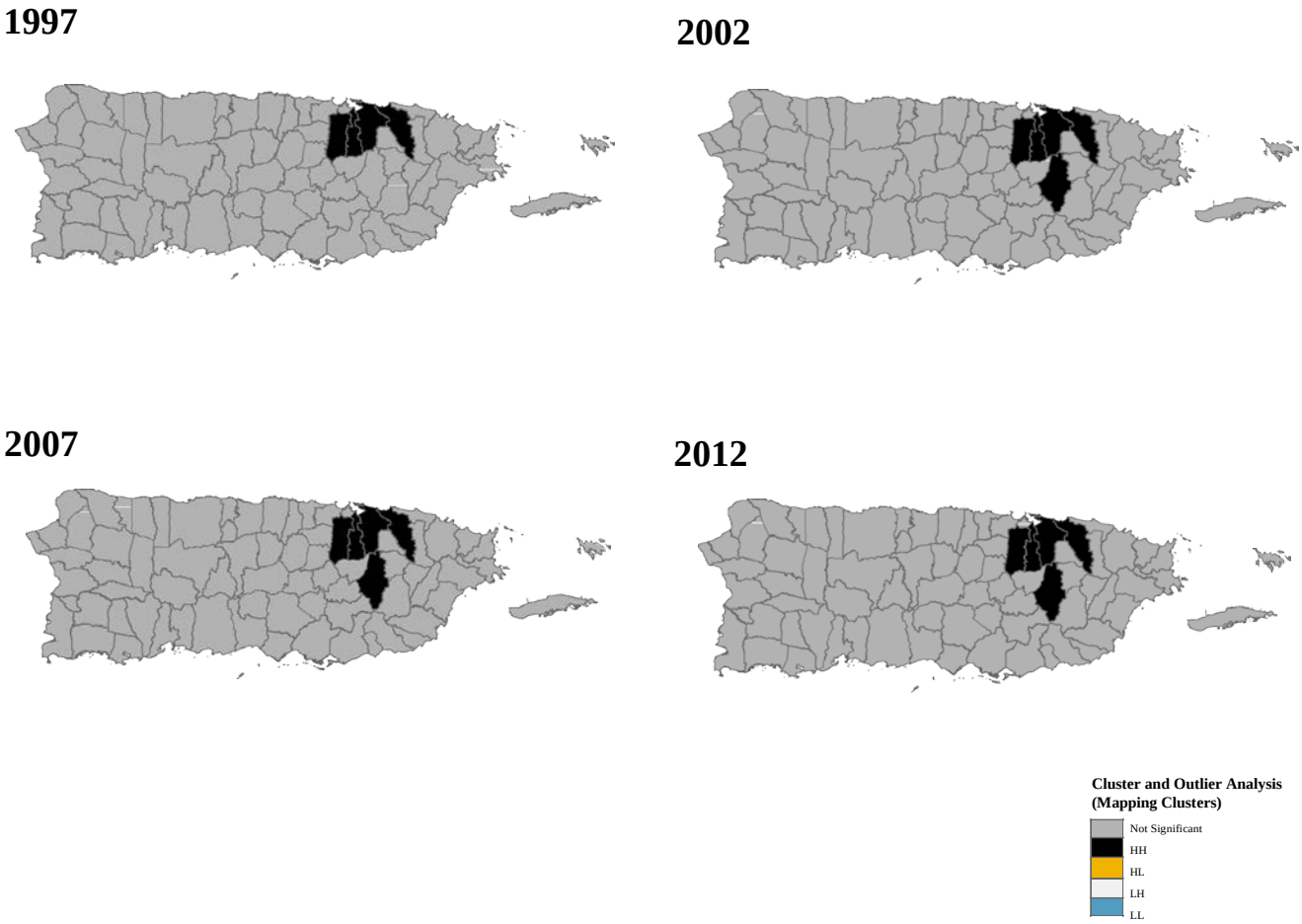
Notes: I use Fixed Distance Band as the conceptualization of spatial relationships. Euclidean distance method was used. No standardization or thresholds were applied.
Source: Author's calculation based on Economic Censuses of corresponding years.

However, if I use the unweighted payroll method to analyze spatial clustering, I will not be able to find a distinctive pattern from 1997 to 2012 except for the new cluster in Caguas (central region), as shown in Figure 7. In fact, the clustering mapped is completely equal from 2002 to 2012. This would stress the sub-optimality of this method: it does not appear to be cross-validated with changes in economic geography observed in other data sources.

These findings are corroborated by the use of another method of spatial analysis, the Hot Spot approach shown in Figure 8. The G_i^* score is a Z score. Thus, counties with a yellow color were not statistically significant clusters. In this analysis, the statistically significant small centers of production are represented as "cold spots" (blue color). Note the progressive dismantling of GDP clusters that occurred from 1997 to 2012. That is, although in every period, the Metropolitan area led the economic activity of the Island, in 2007 and 2012, the small clusters observed in the Western area disappeared.

© Southern Regional Science Association 2017.

Figure 7: Anselin Local Moran's I using unweighted Payroll Method, 1997-2012



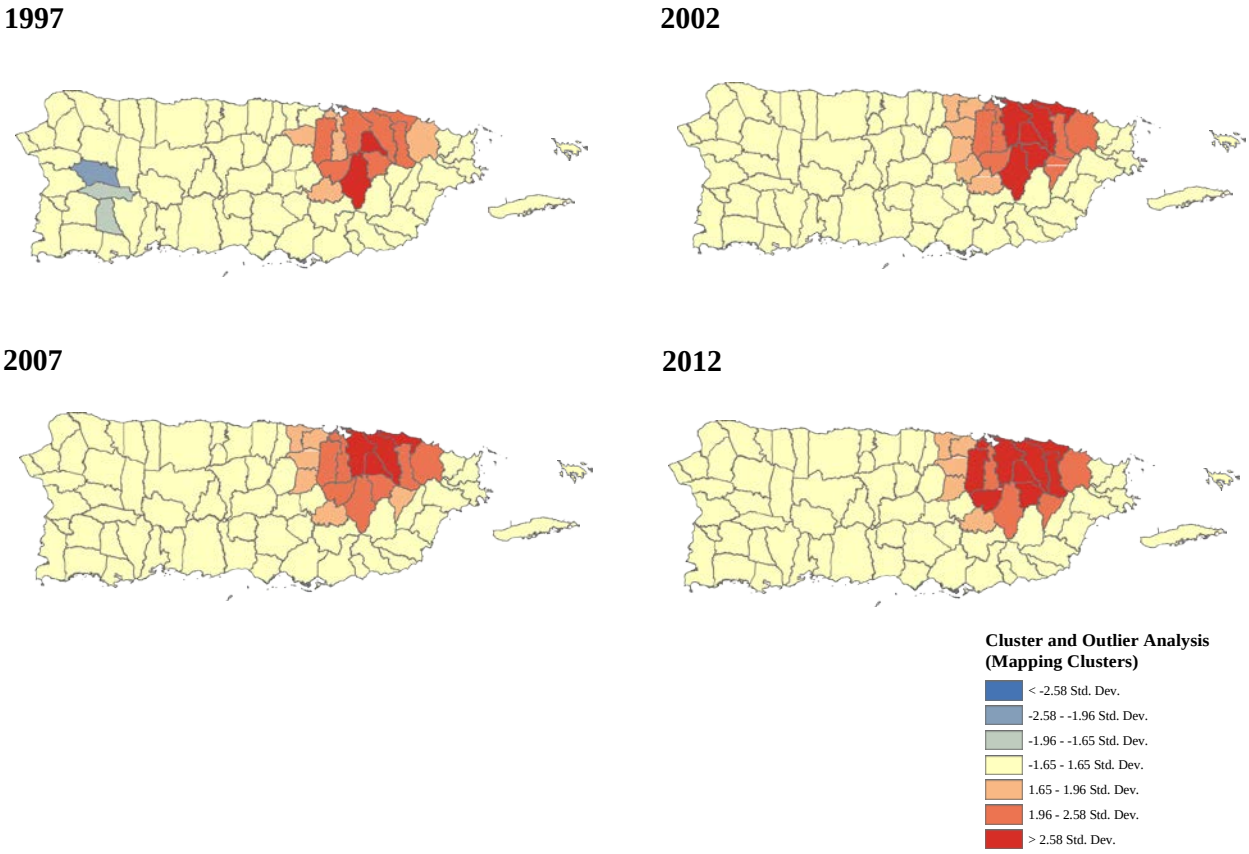
Notes: I use Fixed Distance Band as the conceptualization of spatial relationships. Euclidean distance method was used. No standardization or thresholds were applied.
Source: Author's calculation based on Economic Censuses of corresponding years.

Are we able to forecast the GDP in the counties between economic censuses with this method? With Equation (5), I can complement the information in the economic census with the information provided by the payroll data in the QCEW. This combined method, in fact, provides a better prediction of the GDP between the economic censuses than both the unweighted and weighed payroll methods. Using Equation (6) as my barometer of accuracy, Table 2 illustrates the precision of each method in forecasting the GDP of county *i* in 2007 and 2012 with respect to the GDP obtained from Equation (3) using the economic censuses of 2007 and 2012.

Although every method makes use of the payroll data from the QCEW in one way or another, my proposal appears to be superior to the other methods in predicting GDP from the next economic census. After comparing the three methodologies for year 2007, I find that the weighed payroll method illustrated in Equation (2) has the lowest fitness with respect to the GDP obtained from the economic census of 2007. The unweighted method appears to be more closely connected, but in both periods, 2007 and 2012, Equation (5) was the estimation that more closely matched the GDP obtained in Equation (3). This is an important finding, given the few lags associated with the publication of the economic census: With Equation (5), I am able to have an

© Southern Regional Science Association 2017.

Figure 8: Hot Spot Analysis, 1997-2012



Notes: I use Fixed Distance Band as the conceptualization of spatial relationships. Euclidean distance method was used. No standardization or thresholds were applied.
Source: Author's calculation based on Economic Censuses of corresponding years.

accurate approximation of GDP in county *i* every year, before the publication of an economic census.

Lastly, in Figure 9, I illustrate the map obtained when the proposal is weighed by industry, as shown in Equation (4). The largest difference between Estimations (3) and (4) is in the counties with a relatively high service sector, such as San Juan, because in the 2012

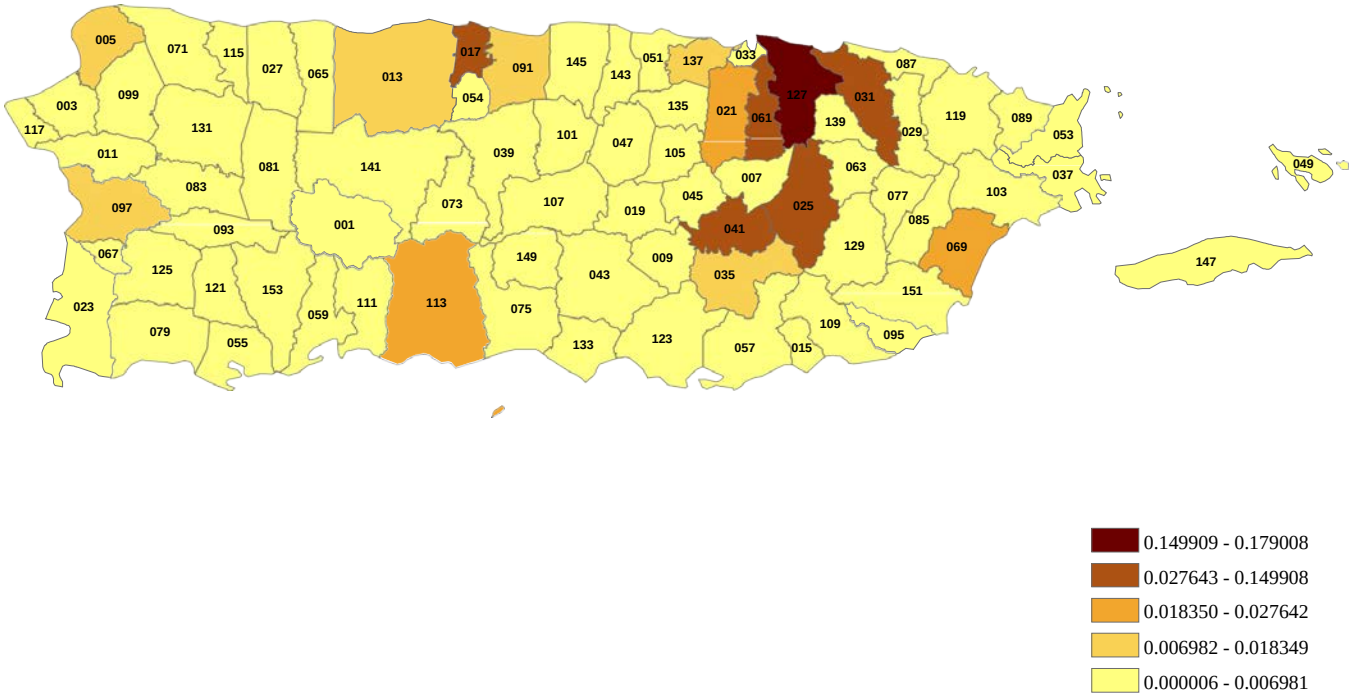
Table 2: Theil T-Statistic Scores of Estimation Methods, 2007-2012

	Estimated Year	
Method	2007	2012
Our proposal	0.174	0.108
Unweighted payroll	0.183	0.154
Weighted payroll	0.292	

Note: Because of data limitations in some sectors, I calculate Theil T for weighted payroll method only for 2007.
Source: Author's calculation based on Economic Censuses and QCEW data of corresponding years.

© Southern Regional Science Association 2017.

Figure 9: Estimated Output Weighed by Industries, 2012



Note: Its own natural breaks were used in this case because the GNI differs from GDP significantly in Puerto Rico.

Economic Census, San Juan had 60 percent of its GDP from Information, Finance and Insurance, but in national accounts, those sectors had very low weights (2.4 percent and 5.1 percent, respectively). Manufacturing represented 1.5 percent of the total GDP in San Juan in 2012, while Puerto Rico's manufacturing output has a relatively high weight of 46 percent, perhaps because of transfer pricing and other factors, as explained above. As a result, if I use Equation (3), San Juan produced 32 percent of the total GDP in Puerto Rico in 2012, but if I use Equation (4), its GDP share shrinks to 18 percent. Other counties that had a lower GDP with Equation (4) are Hatillo and Rio Grande, whose retail trade represented, in 2012, 64 percent and 81 percent of their respective GDP, while retail trade only represented 4.7 percent in Puerto Rico's GDP. However, in general, both methods have a high correlation: Excluding San Juan, the estimations obtained from Equations (3) and (4) had a Theil t-statistics of 0.10.

4. CONCLUSIONS

Although GDP is not a measure of well-being, it is an important indicator of economic development, and thus, can shed light on many related factors, such as regional economic trends. Thus, searching for an accurate measurement of GDP at the county level is a rewarding task for researchers in development, planning, geography, business, and urban policy. Universities can use these data for their strategic plans. For instance, in Figure 5, we observe that in counties where almost all of the campuses of the University of Puerto Rico are established (San Juan,

© Southern Regional Science Association 2017.

Bayamon, Carolina, Cayey, Humacao, Ponce, Mayaguez, Aguadilla, and Arecibo), there is relatively higher economic activity, even after deindustrialization.

However, currently, it is not possible to measure GDP at the county level, because most of the data needed are only made public at the state level. Thus, interested parties must rely on estimation methods. I have found that the current estimation methods are not optimal and can provide misleading results. One issue with the current methods is that they depend on strong assumptions about productivity levels across counties. I developed a simple method that does not require those assumptions because it is based on sales data from the economic censuses and we showed the relative superiority of our proposal. I also developed a method to estimate the GDP in a given county between censuses with respect to the state and showed that it can better predict the outcomes obtained from the economic census.

This approach can be applied to other jurisdictions that have economic census data. Also, the GDP obtained can be used as an independent variable in a spatial analysis to explain different topics, such as poverty in the counties, because the closer a person lives to a place with relatively high economic activity, the higher his/her chances of earning a sufficient level of income, and the lower his/her probability of suffering poverty. Our GDP estimation can also be used as the dependent variable in a spatial analysis that seeks to explain differences in developmental levels or regional trends. However, those types of analyses exceed the scope of this paper, and should be developed further elsewhere.

REFERENCES

Andrews, Richard B. (1954) “Mechanics of the Urban Economic Base: The Problem of Base Measurement,” *Land Economics*, 30, 52–60.

Anselin, Luc. (1995)“Local Indicators of Spatial Association – LISA,” *Geographical Analysis*, 27, 93–115.

Bruce, Donald, John Deskins, and William F. Fox. (2007) “On the Extent, Growth, and Efficiency Consequences of State Business Tax Planning,” *Taxing Corporate Income in the 21st Century*, 226, 253–54.

Collins, Susan M., Barry P. Bosworth, and Miguel A. Soto-Class. (2006)*Restoring Growth in Puerto Rico: Overview and Policy Options*.Brookings Institution Press: Washington, DC.

Curet-Cuevas, Eliezer. (2003) *Economía Política de Puerto Rico: 1950 a 2000*.M.A.C. Editions: San Juan, Puerto Rico.

Federal Reserve Bank of New York. (2012) “Report on the Competitiveness of Puerto Rico’s Economy.”Last accessed January 2016 athttp://www.newyorkfed.org/regional/puertorico/report.pdf.

Greene, William H. (2003) *Econometric Analysis*. Pearson Education:Upper Saddle River, NJ.

Grubert, Harry, and Joel Slemrod. (1998) “The Effect of Taxes on Investment and Income Shifting to Puerto Rico,” *Review of Economics and Statistics*, 80, 365–373.

Heizer, Jesse H., Barry Render, and Howard J. Weiss. (2004)*Operations Management*, 8th ed.Prentice Hall: Upper Saddle River, NJ.

© Southern Regional Science Association 2017.

Jenks, George F. (1967) "The Data Model Concept in Statistical Mapping," *International Yearbook of Cartography*, 7, 186–190.

Kendrick, JohnW., and C. Milton Jaycox. (1965) "The Concept and Estimation of Gross State Product," *Southern Economic Journal*, 32, 153–168.

Lahr, Michael L. (2001) "Reconciling Domestication Techniques: The Notion of Re-exports and Some Comments on Regional Accounting," *Economic Systems Research*, 13, 165–179.

Mitchell, Andy. (2005) *The ESRI Guide to GIS Analysis*, 2nd ed. ESRI Press: Redlands, CA.

Niemi, A.W. (1972) "A Reexamination of the Kendrick-Jaycox Method of Estimating Gross State Product," *Review of Regional Studies*, 2, 14–22.

Shaffer, Ron, Steve Deller, and Dave Marcouiller. (2004)*Community Economics: Linking Theory and Practice*, 2nd ed. Blackwell Publishing: Oxford.

Theil, Henry. (1958) *Economic Forecasts and Policy*. North-Holland: Amsterdam.

Weisbrod, Glen, Don Vary, and GeorgeTreyz. (2003) "Measuring Economic Costs of Urban Traffic Congestion to Business," *Transportation Research Record: Journal of the Transportation Research Board*, 1839, 98–106.

© Southern Regional Science Association 2017.