ON SPATIAL HETEROGENEITY IN ENVIRONMENTAL COMPLIANCE COSTS

by

Randy A. Becker *
U.S. Bureau of the Census

CES 09-25 September, 2009

Abstract

This paper examines the extent of variation in regulatory stringency below the state level, using establishment-level data from the U.S. Census Bureau's Pollution Abatement Costs and Expenditures (PACE) survey to estimate a county-level index of environmental compliance costs (ECC). County-level variation is found to explain 11-18 times more of the variation in ECC than state-level variation alone, and the range of ECC within a state is often large. At least 34% of U.S. counties have ECC that are statistically different from their states'. Results suggest that important spatial variation is lost in state-level studies of environmental regulation.

Keywords: environmental costs, regulation, manufacturing, U.S. counties
JEL codes: Q52, R52, H73

* This paper has benefitted from the helpful comments of Ron Jarmin, Arik Levinson, Ron Shadbegian, Dan Weinberg, and the American Economic Association members of the Census Advisory Committee. Conference participants at the Third World Congress of Environmental and Resource Economists and the 2008 Comparative Analysis of Enterprise Data (CAED) conference also provided valuable feedback on earlier versions of this paper. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

1. Introduction

That the stringency of environmental regulation varies spatially in the United States hardly seems a noteworthy point anymore. It is rather well-established that states exercise discretion in their enforcement of federal environmental regulations, and states can of course adopt standards that are more stringent than those promulgated by the federal government. Over the past couple of decades, a number of different proxies that attempt to measure the extent of these regulatory differences between states have been constructed and subsequently used by researchers wishing to explore the impact of environmental regulations on industrial location and industrial activity (see Levinson 2001 and Brunnermeier and Levinson 2004 for reviews).

What is less well-studied and less appreciated is the degree of heterogeneity in regulatory stringency *below* the state level. Duffy-Deno (1992) uses the variation in pollution abatement expenditures across Standard Metropolitan Statistical Areas (SMSA) to examine the effects of environmental regulations on economic activity, but – with only 63 SMSAs – this analysis is not much richer than a state-level study and it obviously excludes a good deal of economic activity. Berman and Bui (2001) examine the impact on oil refineries of the uniquely stringent air quality regulations of the South Coast Air Basin (i.e., the Los Angeles area) versus those of the rest of California and the rest of the United States. Meanwhile, a growing number of studies have looked at the effects county non-attainment of the Clean Air Act's national ambient air quality standards (NAAQS) have had on manufacturing activity (e.g., McConnell and Schwab 1990, Henderson 1996, Kahn 1997, Becker and Henderson 2000, Greenstone 2002, List *et al.* 2003). While county-level NAAQS non-attainment status may be the best, most geographically-detailed measures of environmental regulation currently available, they cover only six air pollutants, and they are dichotomous (rather than continuous) in nature, thereby cloaking the true variation in regulatory intensity across counties, even within a state.

In this paper, I employ a unique database to measure and examine more fully the extent of variation in regulatory stringency below the state level. In particular, I use fourteen year's worth of establishment-level data from the U.S. Census Bureau's Pollution Abatement Costs and Expenditures (PACE) survey to estimate a county-level index of environmental compliance costs, as well as a comparable state-level index. Here, pollution abatement operating costs per unit of economic activity (output or employment) is modeled as a function of plants' industry, size, age, and year — factors known to determine both regulatory scrutiny and environmental expenditures — as well as plants' location. The resulting index captures extra-normal environmental costs at a detailed level of geography, due (if only in part) to additional environmental regulation faced by industry at the locale.

Results suggest that spatial heterogeneity in environmental compliance costs is real. County-level variation is found to explain 11-18 times more of the variation in environmental compliance costs than state-level variation alone, and the range of environmental compliance costs within a state is often large. I find that at least 34% of counties (containing 21% of U.S. manufacturing employment) have environmental compliance costs that are statistically different from their states'. Alternative specifications yield even more dramatic results. All told, there are only three states with counties with homogenous environmental compliance costs (in a statistical sense). These results suggest that important spatial variation is lost in state-level studies of environmental regulation.

The paper proceeds as follows. In the next section, I discuss the data and empirical specification used in estimating the county-level index of environmental compliance costs. Section 3 then documents the extent to which counties within a state are different from each other – and different from their state – in terms of their conditional environmental compliance costs. Section 4 discusses results using alternative specifications for the index, and Section 5 offers some concluding remarks, including a brief discussion of the potential uses for and the considerations surrounding a publicly-available county-level index of the sort introduced in this paper.

2. A County-level Index of Environmental Compliance Costs

The literature is full of studies that have used pollution abatement expenditure data from the PACE survey to measure geographic differences in the stringency of environmental regulations.[1] At the heart of each of these measures is an estimate of pollution abatement expenditures, divided by some measure of total manufacturing activity, such as gross state product, value added, or value of shipments. In recognition of the inherent variation in the pollution-intensiveness of industries, some measures attempt to adjust for a location's industrial composition (e.g., Bartik 1988, Levinson 1996, Gray 1997, Levinson 2001, Keller and Levinson 2002); others do not (e.g., Duffy-Deno 1992, Friedman *et al.* 1992, List and Kunce 2000, List and Co 2000). With the exception of Levinson (1996), all of these previous studies have used *published* PACE statistics, versus the underlying establishment-level microdata. And with the exception of Duffy-Deno (1992), who analyzed 63 SMSAs, the unit of geography is the state in all of these studies.

In this paper, I use the *establishment-level* data from the PACE survey.[2] For my purposes,

[1] The principal alternatives to such cost-based measures are various indexes and rankings produced by environmental organizations, which are often considered to be subjective in nature. See Levinson (2001) for a review and discussion.

[2] These survey data, as well as those from the Annual Survey of Manufactures and the Census of Manufactures, are confidential, collected and protected under Title 13 of the U.S. Code. Restricted access to these data can be arranged through the U.S. Census Bureau's Center for Economic Studies. See http://www.ces.census.gov/ for details.

these microdata have a few substantial advantages over the published PACE statistics that are commonly used. First and foremost, the location information associated with each establishment allows me to contemplate pollution abatement expenditure at the sub-state level. Second, the industrial classification of establishments in these data is the most detailed available, by *any* level of geography, which is extremely valuable in any effort to explain variation in pollution abatement expenditures. Finally, by merging these PACE microdata to information reported by these same establishments in the Annual Survey of Manufactures (ASM) and the Census of Manufactures (CM), I am uniquely able to control for the size and age of establishments — factors that have been shown to be important determinants of regulatory scrutiny and, hence, establishments' compliance expenditures (e.g., Becker and Henderson 2001; Becker 2005). By controlling for establishment size, I also control for economies of scale in abatement.

Here I use the establishment-level data from the PACE surveys of 1979-1982, 1984-1986, and 1988-1994.[3] As in most of the aforementioned studies, I will employ data on total pollution abatement operating costs (*PAOC*), which includes salaries & wages, parts & materials, fuel & electricity, capital depreciation, contract work, equipment leasing, and other operating costs associated with a plant's abatement of its air and water pollution as well as its solid waste in that calendar year. To this I merge data on these establishments from the ASM or CM, including employment, value of shipments, four-digit SIC industry, county, and plant vintage (as measured by an establishment's first appearance in the Census of Manufactures). After restricting the sample to cases that had linkable PACE and ASM/CM records in a given year, and after

[3] The collection of these data began with 1973. The establishment-level data from 1974-1978 have only recently been uncovered and are not employed here. Establishment-level data from 1973 and 1983 are still missing. A survey for reference year 1987 was not conducted. The PACE survey was suspended over the past decade and a half (1995-1998, 2000-2004, 2006-present), though it is hoped that annual collection will soon resume. The usefulness of the data from the 1999 PACE survey is rather limited (see Becker and Shadbegian 2005). The recently released data for reference year 2005 are also not employed here.

eliminating inactive establishments, plants in Alaska and Hawaii, and those with missing or incomplete data on critical items, there are 200,532 establishment-years of observations for my empirical work. This rather sizable sample contains approximately 49,000 unique manufacturing plants, encompassing virtually all four-digit SIC manufacturing industries and located in 2,514 different U.S. counties.

In harmony with the previous literature, the basis for my index is an establishment's PAOC intensity — that is, its pollution abatement operating costs per unit of economic activity. In this paper, I mainly use as the denominator a plant's output – namely, its value of shipments (*VS*). Previous studies (cited above) have also used gross product, value added, or value of shipments. Later in the paper, I discuss some results from alternate specifications that instead use plant employment (*EMP*) as their denominator.

The degree of regulatory scrutiny faced by a manufacturing plant – and hence its PAOC intensity – is most certainly dependent on the industry it is in (its inherent pollution-intensiveness), as well as the year and its size, and it has been shown that the *combination* of these three factors can affect regulatory intensity (Becker and Henderson 2000). Accordingly, I model PAOC intensity as a function of an industry-year-size quartile effect. In lieu of an overwhelming number of dummy variables (of which there would be over 21,000), I employ the data at hand to compute an estimate of the expected PAOC intensity for each industry-year-size quartile class. In particular, for an establishment in industry n', year t', and size quartile q', the relevant PAOC intensity is assumed to be[4]

$$\underset{j\in\{n=n',t=t',q=q'\}}{median}(PAOC_j / VS_j). \tag{1}$$

Here, an establishment's size quartile is determined by its position in the employment-weighted

[4] Alternatives to this include the PAOC intensity of the mean establishment in the set $\{n', t', q'\}$ or the weighted mean. These will be explored below.

plant employment distribution for its industry in that period.[5]

In principle, a plant's age category could be a fourth dimension used in computing an establishment's expected PAOC intensity in equation (1), but this would significantly increase the number of applicable cells and severely reduce the average number of observations per cell. Instead, establishment age is controlled for by a separate series of plant vintage indicators (**V_k**) based on an establishment's first appearance in the CM, with $k \in \{1963, 1967, 1972, 1977, 1982, 1987, 1992, 1997\}$. It is therefore assumed that plant vintage has equivalent effects on regulation and environmental compliance costs across industry, year, and size classes.[6]

My county-level index of environmental compliance costs is the vector of ϕ_m parameters from the following regression equation:

$$\ln\left(PAOC_i/VS_i\right) = \alpha + \beta \cdot \ln\left(\underset{j \in \{n',t',q'\}}{median}\left(PAOC_j/VS_j\right)\right) + \sum_{k \in K}\left(\gamma_k \cdot V_k\right) + \sum_{m \in M}\left(\phi_m \cdot C_m\right) + \varepsilon_i \qquad (2)$$

where observation *i* is an establishment in industry *n´*, year *t´*, size quartile *q´*, and *j* also indexes establishments in the sample. *K* is the set of possible first CM appearances, less one omitted possibility (1963). *M* is the set of U.S. counties, *m* indexes those counties, and C_m is one in a series of county indicator variables, less one omitted county (Washington DC). ε_i is an error term.

A comparable state-level index can be estimated from a version of equation (2) with $\sum_{s \in S}\left(\phi_s \cdot C_s\right)$ in place of $\sum_{m \in M}\left(\phi_m \cdot C_m\right)$, where *S* is the set of U.S. states, *s* indexes those states, and

[5] That is, the inter-quartile cutoffs are chosen so that each quartile contains one-fourth of the industry's employment, rather than one-fourth of its establishments. I conjecture that this grouping more closely approximates the manner in which regulators prioritize their scrutiny of plants within an industry. Because the PACE and ASM sample larger establishments more heavily, the distribution across these size quartiles in this sample is more uniform than it might be, with 18.8%, 26.9%, 29.8%, and 24.5% in the quartiles with the largest to smallest plants, respectively. Data on an industry's plant employment distribution is taken from the prior or contemporaneous CM.
[6] Demonstrating whether this is indeed the case is beyond the scope of this current paper. It should be noted that *any* use of plant vintage here is more than has been (or could be) done by any previous study.

C_s is one in a series of state indicator variables, less one omitted state (again, Washington DC). Note that since the same omitted category is employed in both specifications (i.e., establishments in the "county" and "state" of Washington DC that were in existence as early as the 1963 Census of Manufactures), the county- and state-level indexes are identically scaled and directly comparable.

Since the value of the dependent variable is bounded from below for a significant number of observations, the parameters of equation (2) are estimated via a Tobit specification.[7] The parameter α is the estimated constant, representing the omitted group (establishments in Washington DC that were in existence as early as the 1963 Census of Manufactures). In estimation, β is restricted to be equal to one, forcing the notion that an establishment is *expected* to have PAOC intensity equivalent to the estimate for its industry-year-size class, as specified in equation (1). Deviations from this are explained by differences in plant vintage, as captured by the γ_k parameters,[8] and by differences between counties, as measured by the estimated ϕ_m parameters — the county-level index. The index, assumed here to be time invariant, reveals any extra-normal environmental compliance costs, due to above- or below-normal environmental regulation faced by manufacturers at the county level. The index also includes potential geographic differences in prices related to pollution abatement, such as the salaries of environmental workers, cost of low-sulfur coal, price of electricity, fees for solid waste hauling and disposal, and so forth. Exploring the variation in this index is the subject of the next section.

[7] Establishments are asked to report their expenditures in *thousands* of dollars. Therefore, with rounding, a response of zero reflects expenditures of less than \$500. The `cnreg` (censored normal regression) command in Stata is a generalization of the standard Tobit procedure that allows the censoring point to vary by observation. In this case, left-censoring occurs at $ln(0.5/VS_i)$ for about 18% of the observations in this sample.

[8] Regressions show that, other things being equal, establishments of older vintages have higher PAOC intensity.

3. On Spatial Heterogeneity in Environmental Compliance Costs

I begin by noting that, according to the respective pseudo-R^2 statistics, the county dummy variables in equation (2) explain over 18 times more of the variation in excess PAOC intensity than a version of equation (2) with state dummy variables in their place, relative to a model with no geography variables at all. The R^2 statistics from the analogous OLS regressions tell a similar story. Here, the county dummy variables explain over 13 times more of the variation in excess PAOC intensity than do state dummy variables. Even the adjusted-R^2 statistics suggest that county variation explains 11 times more of the variation in excess PAOC intensity than state variation, relative to a model without any geographic controls.

In addition to those just discussed, Table 1 also presents the R^2 statistics from three "placebo" specifications that address the possibility that any large set of "random" dummy variables might also explain variation in PAOC intensity. In the Placebo #1 specification, 25 "random" dummy variables are included, based on the first letter of the name of the county in which the establishment is located. In the Placebo #2 [#3] specification, 243 [2,403] "random" dummy variables are included, based on the first letter of the name of the county in which the establishment is located <u>and</u> the last digit [last two digits] of the establishment's total employment. The results suggest that these "random" dummies indeed add explanatory power, relative to the baseline model with no geography variables. For example, in the case of Placebo #3, the 2,403 dummy variables add 2.9% to the pseudo-R^2. However this is far less than the 10.4% that the 2,513 county dummy variables add. These results clearly demonstrate that (a) geography – whether state or county – explains some portion of environmental compliance costs, and (b) collectively, counties have substantially more explanatory power than do states, on the matter of environmental compliance costs.

Table 2 begins to illustrate the degree of heterogeneity in environmental compliance costs *within* each of the 48 states in scope to this analysis. In particular, I present two statistics here: The first is the *range* of the index values in the state – i.e., the difference in the index values between the counties with the maximum and minimum index value. The second is the *mean absolute deviation (MAD)* of the index values in the state – i.e., the average deviation of the county indexes from the average of the indexes. While computing these two statistics, it was necessary for me to ignore the index values of 571 counties, for confidentiality reasons.[9] The measures reported in Table 2 therefore may understate the true degree of heterogeneity observed in the state.[10] Nonetheless, the values presented in Table 2 are fairly correlated with their *true* values (calculated without these suppressions): For *Range* the pairwise and Spearman's rank correlations are 0.6441 and 0.6818, respectively, and for *MAD* those two correlations are 0.5423 and 0.5774, respectively. A state's rank (highest value = 1) is also shown in Table 2 for each of these two measures.

We see that the state with the highest *Range* is Oklahoma, where the difference in the maximum and minimum county-level index value is almost 8 points. In terms of *MAD*, the state with the highest value is New Mexico, where the average county index is 1.3 points from the mean index in the state. In this measure, Oklahoma ranks second. Meanwhile, Delaware ranks lowest in terms of both *Range* and *MAD*, with Connecticut, Rhode Island, Utah, Massachusetts, New Jersey, and New Hampshire also exhibiting relatively low levels of heterogeneity.

While a full exploration of the state-level determinants of this heterogeneity is beyond the scope of this current paper, casual observation suggests that the states with the lowest *Range* and *MAD* (and therefore the *least* heterogeneity among their counties) tend to be northeastern states,

---

[9] This is in addition to the 600-plus counties that we do not observe at all in this database.
[10] Indeed, 35 of the 48 *true* minima were suppressed, as were 26 of the 48 *true* maxima.

and tend have the smallest land area, highest population density, and the smallest number of counties. To examine whether there are such relationships, Table 3 presents the pairwise correlation coefficients between *MAD*, *Range*, land area, population density, and number of counties. The table also contains correlations with a Herfindahl-Hirschman Index (HHI) of county population (since population concentrated in relatively few counties may concentrate political power and lead to homogeneity) and with the states' environmental compliance cost index. The table confirms casual observation. Heterogeneity is indeed positively correlated with states' land area and number of counties, and negatively correlated with states' population density, HHI, and environmental compliance cost index value.

Since these state characteristics are also usually significantly correlated with each other, a simple OLS regression is used to examine their independent impacts on heterogeneity. Table 4 reveals that, controlling for the other state characteristics, population density is the only statistically significant determinant of *MAD*. Exactly why heterogeneity tends to be highest in the least densely populated states – and whether this is perhaps picking up other omitted variables – is worth future investigation. Meanwhile, of the set of state characteristics examined here, the number of counties is the only statistically significant determinant of *Range*. The likely explanation here is that, controlling for land area, states with a large number of counties tend to have fewer plants per county, leading to more imprecisely measured index values, including ones that fall toward and become the minimum and maximum in the state. We will see this in graphical form shortly.

Returning to Table 2, among large manufacturing states (in terms of employment), Texas exhibits the largest degree of heterogeneity, measured both by *Range* (6.5) and *MAD* (0.65). Michigan also has a relatively high heterogeneity, with a *Range* and *MAD* of 5.1 and 0.57, respectively. Meanwhile, California (the largest manufacturing state by any measure) and New York exhibit much less heterogeneity, among large manufacturing states, with Ohio, Illinois, and Pennsylvania falling somewhere in the middle.

To help further illustrate heterogeneity, Figure 1 plots the county- and state-level index values for Texas, Michigan, and California, respectively. One particularly nice feature of these graphs is their depiction of the confidence intervals around the state and county point estimates.[11] This makes obvious the fact that many of the counties toward the extrema are clearly statistically different from their respective states, in terms of the environmental compliance costs faced by their manufacturing establishments. What is less obvious here is whether the many counties with confidence intervals that overlap with their state's are in fact statistically different from their state.

To test for this statistical difference, for county m' in state s', note that the variance of the difference between the estimated county and state index values is given by

$$\mathrm{var}(\hat{\phi}_{m'} - \hat{\phi}_{s'}) = \mathrm{var}(\hat{\phi}_{m'}) + \mathrm{var}(\hat{\phi}_{s'}) - 2 \cdot \mathrm{cov}(\hat{\phi}_{s'}, \hat{\phi}_{m'}) \tag{3}$$

where $\mathrm{var}(\hat{\phi})$ is the square of the estimated standard error (*se*) associated with the respective index value ($\hat{\phi}$). Here, $\hat{\phi}_{m'}$ and $\hat{\phi}_{s'}$ are equivalent regression coefficients from two different models, estimated on the same sample, where one model contains an additional set of explanatory variables (i.e., indicators for all the other counties in the state). In cases such as this, Clogg *et al.* (1995) – as further refined in their reply to Allison (1995) – demonstrate that (3) becomes

$$\mathrm{var}(\hat{\phi}_{m'} - \hat{\phi}_{s'}) = \mathrm{var}(\hat{\phi}_{m'}) + \mathrm{var}(\hat{\phi}_{s'}) - 2 \cdot \mathrm{var}(\hat{\phi}_{s'}) \cdot (\hat{\sigma}_C^2 / \hat{\sigma}_S^2) \tag{4}$$

[11] Here and throughout, I employ robust standard errors that allow for the potential correlation of within-plant observations (i.e., between repeated observations of the same plant).

where $\hat{\sigma}_C^2$ and $\hat{\sigma}_S^2$ are the estimated sum of squared errors from the county- and state-based regression models, respectively. The corresponding 90% confidence interval is therefore

$$(\hat{\phi}_{m'} - \hat{\phi}_{s'}) \pm 1.645 \cdot \sqrt{se_{m'}^2 + se_{s'}^2 - 2 \cdot se_{s'}^2 \cdot (\hat{\sigma}_C^2 / \hat{\sigma}_S^2)} \quad (5)$$

which is easily computed from standard regression output.

Table 5 summarizes the results of this statistical testing.[12] Overall, I find that 855 (34.0%) of the 2,513 counties in these 48 states have an environmental compliance cost index statistically different from the index of their respective state. These 855 counties contained 21% of U.S. manufacturing employment and 21% of U.S. manufacturing establishments in 2002.[13] I find that 546 (21.7%) of the 2,513 counties have an environmental compliance cost index statistically *higher* than the index of their respective state, while 309 (12.3%) have an index statistically *lower* than the index of their state. These two groups contained 10.6% and 10.4% of U.S. manufacturing employment, respectively. Table 6 lists the largest of these counties (ranked by their 2002 manufacturing employment) and the direction of their difference vis-à-vis their state. Note that this list includes major counties in New York City, Chicago, San Francisco, Detroit, Dallas, and other large cities.

Table 7 shows the states with the highest and lowest percentage of their counties that are statistically different from the state. We see that Massachusetts, Rhode Island, and Delaware exhibit no heterogeneity whatsoever (from a statistical standpoint). At the other extreme, nearly two-thirds of Nebraska's counties are different from their state-level index, followed closely by Montana. Earlier we noted that Oklahoma, New Mexico, and Texas were among those with the

[12] Counties in which a standard error could not be estimated are assumed to be *not* statistically different than their state. A small number of underlying observations appears to be the primary reason why standard errors could not be estimated. There are 111 such counties, containing just 0.3% of U.S. manufacturing employment in 2002.
[13] Note that these states also have about 600 counties not in my sample and therefore without an index value. These counties accounted for just 0.5% of the manufacturing employment in these states.

largest *Range* and *MAD*, and all three appear here in the top 10, which is not necessarily surprising.

In Figure 2, I present county maps for five large manufacturing states that figure prominently in Tables 2 and 6: Texas, Michigan, California, New York, and New Jersey. Here, counties are grouped [and shaded] by whether their index value is statistically smaller than the state's index [light grey], statistically indistinguishable [medium grey], or statistically larger than the state's index [black]. Counties with no data or with an index value suppressed for confidentiality reasons are also depicted [white].

The maps of Figure 2 allow for some casual observation. One potentially interesting question is whether there is any clustering of "high" and "low" index values within a state. This could arise for a number of reasons — e.g., adjacent counties may share the same set of state regulators and/or nearby counties may regulate themselves similarly, to avoid inter-jurisdictional competition. It has also been shown that environmental regulation may be more lax where exposure to emissions is more likely to fall outside the state, such as in border counties, and particularly those on a state's eastern edge (Helland and Whitford 2003).

In Texas, there doesn't appear to be any clustering of counties with high index values, or of counties with low index values – a few Dallas-Fort Worth counties being an exception. In Michigan, the counties with high index values are chiefly non-metropolitan counties, and a cluster of two counties with low index values appear in Detroit. In California, there is a cluster of high index counties east of San Francisco, and a cluster of low index counties just south of San Francisco. In New York, a couple of low index counties appear in New York City, while many of New York's other large cities (Buffalo, Albany/Troy, Syracuse) have high index values. Moreover, 6 of the 14 Finger Lake counties have high index values and none have a low index value. Finally, New Jersey's three low index counties are all clustered in the New York City area. Future analyses could explore whether clustering (to the extent it appears here) is an actual phenomenon or occurs merely by chance.

4. Results from Alternative Specifications

Here, I briefly explore alternate specifications for the index. In particular, I examine two choices. One is the option of using the mean or weighted mean in computing expected PAOC intensity, instead of median, in equations (1) and (2). There seems no compelling reason to prefer one over the other two. The other choice is the option of using plant employment (*EMP*) in the denominator of PAOC intensity (and expected PAOC intensity), instead of value of shipments (*VS*). While the use of *VS* has precedence in the literature, the $PAOC_i/EMP_i$ ratio might be said to encapsulate a regulator's implicit choice between environmental protection and jobs. Environmental regulation may impact *EMP* more than *VS*.

Comparing the county-level indexes using the median, mean, and weighted mean, I find the pairwise correlations are never less than 0.989 and the Spearman's rank correlations are never less than 0.928. Indexes using the mean and weighted mean are the most closely correlated, with Spearman's rank correlations of 0.988 and 0.997, for the *VS*- and *EMP*-based indexes, respectively. Comparing the various *VS*- and *EMP*-based indexes, pairwise correlations range from 0.984 to 0.994, and the Spearman's rank correlations range from 0.881 to 0.948.

Clearly, the index is very robust to the choices of median, mean, weighted mean, *VS*, and *EMP*, but the correlations are not perfect. To further examine the similarities and differences between these alternate indexes, I calculate the number of counties that have an index value that is statistically different (higher and lower) from the index value of the respective state. Table 8 presents the results. Several points are worth making. First, the particular index used and discussed in the previous sections of this paper (i.e., using *VS* and median; labeled specification #1) is the most "conservative" of the six, in terms of the number of counties that are statistically different from their state (855). This was a deliberate choice on my part. Second, mean and weighted mean yield very similar numbers, while the median-based indexes yield fewer *total* counties that are statistically different, and in particular fewer counties that are statistically *higher* that their state index. Finally, the *VS*-based indexes yield fewer *total* counties that are statistically different, fewer counties that are statistically *lower*, but *more* counties that are statistically higher.

Therefore, choosing *EMP* rather than *VS*, and choosing (weighted) mean rather than median, makes at least some difference. To illustrate the matter more concretely, I redo the analyses of Tables 5 and 6, using the *EMP*-based and weighted mean version of the index (specification #6). Results are summarized and compared in Table 9. We see that specification #6 has 43 additional counties that are statistically different from their state, relative to specification #1. These 898 counties contained nearly 26% of U.S. manufacturing employment in 2002, compared with 21% for the 855 counties in specification #1. The two specifications have 700 counties in common, though 23 differ in their direction relative to the state index. Relative to specification #1, 198 new counties appear in specification #6 and 155 disappear. Considering just the largest manufacturing counties, as in Table 6, several are statistically different from their state in specification #6 but not in specification #1. These include major counties in New York City (lower than state), Houston (higher than state), Dallas (lower than state), Philadelphia (lower than state), Miami (lower than state), and so forth. Other counties that appear in Table 6 (specification #1) are not statistically different from their state in specification

#6, including Santa Clara County, CA and Macomb County, MI.

These exercises suggest that even though the indexes are very highly correlated with each other, they do tell somewhat different stories in terms of the number of counties different than their state, which counties differ from their state, and even whether a county is statistically higher or lower than its state's index.

5. Concluding Remarks

The results in this paper suggest that spatial heterogeneity in environmental compliance costs is real. County-level variation is found to explain 11-18 times more of the variation in environmental compliance costs than state-level variation alone. And the range of environmental compliance costs within a state is often large. Using the most "conservative" of the alternative specifications, I find that 34% of counties (containing 21% of U.S. manufacturing employment) have environmental compliance costs that are statistically different from their states' – including many large manufacturing counties. Less conservative but equally defensible specifications yield even more dramatic results. All told, there are only three states with counties with homogenous environmental compliance costs (in a statistical sense). This paper's results suggest that important spatial variation is lost in state-level studies of environmental regulation.

The United States' states have long been used as a laboratory to explore social and economic phenomena, including the impact of environmental regulations on industrial location and industrial activity. An index of the sort the introduced in this paper could potentially improve such regulatory analyses by expanding the laboratory to include U.S. counties. With such an index, researchers could (re-)explore the effects of environmental regulation on industrial location, employment, output, investment (including foreign direct investment), industrial emissions, ambient pollution levels and so forth *at the county level*. An obvious advantage of such an index over the occasionally-used county-level NAAQS non-attainment status is that it encompasses more than just six air pollutants and is continuous (rather than dichotomous or categorical) in nature.

Future research will be directed toward producing a publicly-available county-level index of the sort introduced in this paper. There are a number of issues to consider. There is the matter of which of the six specifications presented here is the most ideal, or whether there is another specification that might be more preferred. Also, the index could (potentially) be improved in at least two ways. First, a separate index could be produced for each of the three pollution media: air, water, and solid waste. Second, an *annual* county-level index could be estimated.[14] I further discuss this below. Finally, I'll note that confidentiality requirements may limit the ability to release multiple versions of the index, or to make subsequent revisions to the specifications. Therefore, particularly careful forethought is needed in producing and releasing such an index.

Finally, regarding an *annual* county-level index, while this is certainly conceptually appealing – since environmental regulations may strengthen (or weaken) over time within counties, and at varying rates between counties – a county-year index comes at some cost. Given the observations in the sample, and given confidentiality restrictions, an index value could only be published for about 14,000 county-years, representing just over 1,500 counties. Besides the absence of about 1,600 counties altogether, index values would be absent for some years for many counties. Of the 1,500 counties with at least one publishable index value, the average county would have 9.2 years' worth of values, one-quarter of counties would have only 1-4

[14] Related to this, seven additional years of PACE microdata can now be utilized, in addition the 14 years use here in this paper. These additional years include 1974-1978, 1999, and 2005.

years' worth, and only about 600 would have index values for all 14 years.[15] A more serious issue is the precision of the estimated county-year index values, since the average publishable index value would have only 13 observations underlying it, in contrast to the 102 observations underlying the average (time-invariant) county-level index discussed in this paper. Whether the increased richness is worth the sacrifice in quality is worth serious consideration.

[15] These 600 counties do contain approximately 75% of U.S. manufacturing employment however.

References

Allison, Paul D. "The Impact of Random Predictors on Comparisons of Coefficients between Models: Comment on Clogg, Petkova, and Haritou," *American Journal of Sociology*, 100(5), 1294-1312, March 1995.

Bartik, Timothy J. "The Effects of Environmental Regulation on Business Location in the United States," *Growth & Change*, 19(3), 22-44, Summer 1988.

Becker, Randy A. "Air Pollution Abatement Costs under the Clean Air Act: Evidence from the PACE Survey," *Journal of Environmental Economics and Management*, 50(1), 144-169, July 2005.

Becker, Randy and Vernon Henderson. "Effects of Air Quality Regulations on Polluting Industries," *Journal of Political Economy*, 108(2), 379-421, April 2000.

Becker, Randy A. and J. Vernon Henderson. "Costs of Air Quality Regulation," in Carlo Carraro and Gilbert E. Metcalf (eds.), *Behavioral and Distributional Effects of Environmental Policy*, Chicago: The University of Chicago Press, 2001.

Becker, Randy A. and Ronald J. Shadbegian. "A Change of PACE: Comparing the 1994 and 1999 Pollution Abatement Costs and Expenditures Surveys," *Journal of Economic and Social Measurement*, 30(1), 63-95, 2005.

Berman, Eli and Linda T. M. Bui. "Environmental Regulation and Productivity: Evidence from Oil Refineries," *The Review of Economics and Statistics*, 83(3), 498-510, August 2001.

Brunnermeier, Smita B. and Arik Levinson. "Examining the Evidence on Environmental Regulations and Industry Location," *Journal of Environment and Development*, 13(1), 6-41, March 2004.

Clogg, Clifford C., Eva Petkova, and Adamantios Haritou. "Statistical Methods for Comparing Regression Coefficients between Models," *American Journal of Sociology*, 100(5), 1261-1293, March 1995.

Duffy-Deno, Kevin T. "Pollution Abatement Expenditures and Regional Manufacturing Activity," *Journal of Regional Science*, 32(4), 419-436, November 1992.

Friedman, Joseph, and Daniel A. Gerlowski, and Johnathan Silberman. "What Attracts Foreign Multinational Corporations? Evidence from Branch Plant Location in the United States," *Journal of Regional Science*, 32(4), 403-418, November 1992.

Gray, Wayne B. "Manufacturing Plant Location: Does State Pollution Regulation Matter?" *NBER Working Paper Series*, #5880, January 1997.

Greenstone, Michael. “The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures,” *Journal of Political Economy*, 110(6), 1175-1219, December 2002.

Helland, Eric and Andrew B. Whitford. “Pollution Incidence and Political Jurisdiction: Evidence from the TRI,” *Journal of Environmental Economics and Management*, 46(3), 403-424, November 2003.

Henderson, J. Vernon. “Effects of Air Quality Regulation,” *American Economic Review*, 86(4), 780-813, September 1996.

Kahn, Matthew E. “Particulate Pollution Trends in the United States,” *Regional Science and Urban Economics*, 27(1), 87-107, February 1997.

Keller, Wolfgang and Arik Levinson. “Pollution Abatement Costs and Foreign Direct Investment Inflows to U.S. States,” *The Review of Economics and Statistics*, 84(4), 691-703, November 2002.

Levinson, Arik. “Environmental Regulations and Manufacturers’ Location Choices: Evidence from the Census of Manufactures,” *Journal of Public Economics*, 62(1), 5-29, October 1996.

Levinson, Arik. “An Industry-Adjusted Index of State Environmental Compliance Costs,” in Carlo Carraro and Gilbert E. Metcalf (eds.), *Behavioral and Distributional Effects of Environmental Policy*, Chicago: The University of Chicago Press, 2001.

List, John A. and Catherine Y. Co. “The Effects of Environmental Regulations on Foreign Direct Investment,” *Journal of Environmental Economics and Management*, 40(1), 1-20, July 2000.

List, John A. and Mitch Kunce. “Environmental Protection and Economic Growth: What Do the Residuals Tell Us?” *Land Economics*, 76(2), 267-282, May 2000.

List, John A., Daniel L. Millimet, Per G. Fredriksson, and W. Warren McHone. “Effects of Environmental Regulations on Manufacturing Plant Births: Evidence from a Propensity Score Matching Estimator,” *The Review of Economics and Statistics*, 85(4), 944-952, November 2003.

McConnell, Virginia D. and Robert M. Schwab. “The Impact of Environmental Regulation on Industry Location Decisions: The Motor Vehicle Industry,” *Land Economics*, 66(1), 67-81, February 1990.

U.S. Bureau of the Census. *Pollution Abatement Costs and Expenditures, 19__*. Washington, DC: U.S. Government Printing Office, various years.

TABLE 1
Explanatory Power of Geography Variables†

Specification	Number of dummy variables	Pseudo-R^2 from Tobit	R^2 from OLS	Adjusted-R^2 from OLS
Baseline model (no geography variables)	0	0.1062	0.3771	0.3771
State effects	48	0.1068	0.3793	0.3791
County effects	2,513	0.1172	0.4065	0.3990
Placebo #1	25	0.1063	0.3776	0.3775
Placebo #2	243	0.1065	0.3782	0.3774
Placebo #3	2,403	0.1093	0.3855	0.3781

† Placebo #1 consists of the baseline model (with no geography variables), plus dummy variables based on the first letter of the name of the county in which the establishment is located. Placebo #2 consists of the baseline model, plus dummy variables based on the first letter of the name of the county in which the establishment is located and the last digit of the establishment's total employment. Placebo #3 consists of the baseline model, plus dummy variables based on the first letter of the name of the county in which the establishment is located and the last two digits of the establishment's total employment.

TABLE 2
Measures of Within-State Heterogeneity in Environmental Compliance Costs

State	Range	(Rank)	Mean absolute deviation	(Rank)
Alabama	4.563	(9)	0.619	(10)
Arizona	1.545	(40)	0.419	(32)
Arkansas	3.552	(21)	0.482	(26)
California	3.010	(29)	0.388	(37)
Colorado	3.523	(22)	0.413	(34)
Connecticut	0.723	(45)	0.137	(47)
Delaware	0.129	(48)	0.049	(48)
Florida	4.250	(15)	0.511	(21)
Georgia	3.822	(19)	0.535	(18)
Idaho	2.384	(34)	0.522	(19)
Illinois	3.965	(18)	0.469	(27)
Indiana	3.255	(26)	0.438	(29)
Iowa	4.575	(8)	0.577	(12)
Kansas	2.913	(30)	0.549	(17)
Kentucky	3.439	(24)	0.508	(24)
Louisiana	4.410	(11)	0.444	(28)
Maine	2.444	(33)	0.438	(30)
Maryland	1.876	(37)	0.337	(40)
Massachusetts	0.981	(44)	0.150	(46)
Michigan	5.123	(6)	0.573	(13)
Minnesota	3.450	(23)	0.623	(9)
Mississippi	4.014	(17)	0.510	(23)
Missouri	3.701	(20)	0.629	(6)
Montana	3.090	(28)	0.715	(4)
Nebraska	2.518	(32)	0.568	(15)
Nevada	1.218	(43)	0.390	(36)
New Hampshire	1.361	(42)	0.303	(42)
New Jersey	1.478	(41)	0.279	(43)
New Mexico	5.692	(5)	1.292	(1)
New York	3.137	(27)	0.373	(39)
North Carolina	5.118	(7)	0.510	(22)
North Dakota	1.715	(38)	0.386	(38)
Ohio	4.276	(14)	0.413	(33)
Oklahoma	7.836	(1)	0.908	(2)
Oregon	4.332	(12)	0.629	(8)
Pennsylvania	4.147	(16)	0.434	(31)
Rhode Island	0.573	(47)	0.175	(44)
South Carolina	2.166	(36)	0.336	(41)
South Dakota	2.533	(31)	0.558	(16)
Tennessee	6.450	(4)	0.521	(20)
Texas	6.549	(3)	0.647	(5)
Utah	0.631	(46)	0.158	(45)
Vermont	2.363	(35)	0.570	(14)
Virginia	7.666	(2)	0.617	(11)
Washington	4.429	(10)	0.629	(7)
West Virginia	4.303	(13)	0.742	(3)
Wisconsin	3.397	(25)	0.497	(25)
Wyoming	1.685	(39)	0.402	(35)

TABLE 3
Correlation Coefficients between State-level Characteristics†

	Mean absolute deviation	Range	Land area	Population density	Number of counties	HHI of county population	State-level ECC index
Mean absolute deviation	+1.000						
Range	+0.727*	+1.000					
Land area	+0.569*	+0.473*	+1.000				
Population density	–0.473*	–0.068	–0.666*	+1.000			
Number of counties	+0.519*	+0.724*	+0.692*	–0.209	+1.000		
HHI of county population	–0.445*	–0.557*	–0.433*	+0.173	–0.750*	+1.000	
State-level ECC index	–0.398*	–0.243*	–0.687*	+0.644*	–0.457*	+0.246*	+1.000

† Mean absolute deviation and range are taken from Table 2. Land area (in square miles), population density (in 2000 population per square mile), and the number of counties are in natural logs. The HHI of county population is computed using 1980 population. Statistical significance at the 10% level is indicated by an asterisk.

TABLE 4
Impact of State-level Characteristics on the Degree of Heterogeneity within States[†]

	Mean absolute deviation	Range
Land area	+0.022	+0.019
	(0.047)	(0.368)
Population density	–0.058*	+0.013
	(0.031)	(0.244)
Number of counties	+0.069	+1.293**
	(0.058)	(0.454)
HHI of county population	–0.223	–0.644
	(0.362)	(2.810)
State-level environmental compliance cost index	+0.013	+1.145
	(0.248)	(1.926)
R-squared	0.4204	0.5380
Number of observations	48	48

[†] Mean absolute deviation and range are taken from Table 2. Land area (in square miles), population density (in 2000 population per square mile), and the number of counties are in natural logs. The HHI of county population is computed using 1980 population. Statistical significance at the 10% and 5% level are indicated by single and double asterisks, respectively.

TABLE 5
Differences between County- and State-level Indexes of Environmental Compliance Costs

	Number	Percent of U.S. manufacturing employment in 2002	Percent of U.S. manufacturing establishments in 2002
Counties that are significantly **higher** than their state at the 90% level	546	10.6%	9.7%
Counties that are significantly **lower** than their state at the 90% level	309	10.4%	11.3%
Counties that are **not** significantly different from their state at the 90% level	1,658	78.5%	77.9%
Counties that are not in the sample	597	0.5%	1.2%

TABLE 6
Largest Manufacturing Counties that are Statistically Different from Their State

County	Manufacturing employment in 2002	Index relative to state's index
Santa Clara County, CA	160,734	Lower
Dallas County, TX	132,968	Lower
Cuyahoga County, OH	91,803	Lower
Alameda County, CA	88,262	Higher
Macomb County, MI	75,040	Lower
Oakland County, MI	73,500	Lower
DuPage County, IL	66,165	Lower
Hamilton County, OH	60,975	Lower
Erie County, NY	56,473	Higher
New York County, NY	47,838	Lower
Bergen County, NJ	47,625	Lower
Queens County, NY	38,889	Lower
Catawba County, NC	36,991	Lower
Washington County, OR	30,203	Lower
Pima County, AZ	29,755	Higher
Essex County, NJ	29,080	Lower
Allen County, IN	28,905	Higher
Onondaga County, NY	27,482	Higher
San Mateo County, CA	26,402	Lower
Brown County, WI	24,263	Lower

TABLE 7
States with the Highest and Lowest Percentage of Their Counties Statistically Different from the State

1.	Nebraska	64.4%
2.	Montana	61.9%
3.	New Mexico	50.0%
4.	Missouri	45.5%
	Oklahoma	45.5%
6.	West Virginia	44.4%
	Nevada	44.4%
8.	Vermont	41.2%
9.	Kansas	41.0%
10.	Texas	40.7%
	⋮	⋮
39.	California	24.1%
40.	Ohio	22.7%
41.	Oregon	21.2%
42.	Maryland	13.0%
43.	Connecticut	12.5%
44.	South Carolina	11.1%
45.	New Hampshire	10.0%
48.	Delaware	0.0%
	Rhode Island	0.0%
	Massachusetts	0.0%

TABLE 8
Comparing Alternative Specifications

Specification	Measure of PAOC intensity	Measure of expected PAOC intensity	Number of counties that are statistically different from their state		
			Total	Lower	Higher
#1	PAOC/VS	median (PAOC/VS)	855	309	546
#2	PAOC/VS	mean (PAOC/VS)	874	307	567
#3	PAOC/VS	weighted mean (PAOC/VS)	872	309	563
#4	PAOC/EMP	median (PAOC/EMP)	866	339	527
#5	PAOC/EMP	mean (PAOC/EMP)	894	342	552
#6	PAOC/EMP	weighted mean (PAOC/EMP)	898	345	553

TABLE 9
Major Differences between Two Alternate Specifications

	#1	#6
Measure of PAOC intensity	PAOC/VS	PAOC/EMP
Measure of expected PAOC intensity	median (PAOC/VS)	weighted mean (PAOC/EMP)
Number of counties that are statistically different from their state	855	898
Percent of U.S. manufacturing employment in 2002	21.0%	25.7%
Number of counties in common	700	700
Same direction of difference	677	677
Reversal of difference	23	23
Number of counties unique to specification	155	198
Higher than state index	99	113
Lower than state index	56	85
Largest manufacturing counties unique to specification (manufacturing employment / index relative to state's index)	Santa Clara County, CA (160,734 / Lower)	Harris County, TX (147,339 / Higher)
	Macomb County, MI (75,040 / Lower)	Tarrant County, TX (83,010 / Lower)
	Pima County, AZ (29,755 / Higher)	San Bernardino County, CA (66,352 / Higher)
	Brown County, WI (24,263 / Lower)	Marion County, IN (57,373 / Lower)
		St. Louis County, MO (53,255 / Lower)
		Dade County, FL (50,568 / Lower)
		Middlesex County, NJ (43,299 / Higher)
		Philadelphia County, PA (36,411 / Lower)
		Multnomah County, OR (36,058 / Higher)
		Kings County, NY (35,137 / Lower)
		Nassau County, NY (32,981 / Lower)
		Broward County, FL (31,731 / Lower)
		Duval County, FL (27,296 / Higher)
		Passaic County, NJ (26,910 / Lower)

FIGURE 1
Heterogeneity in Environmental Compliance Costs: Three Examples

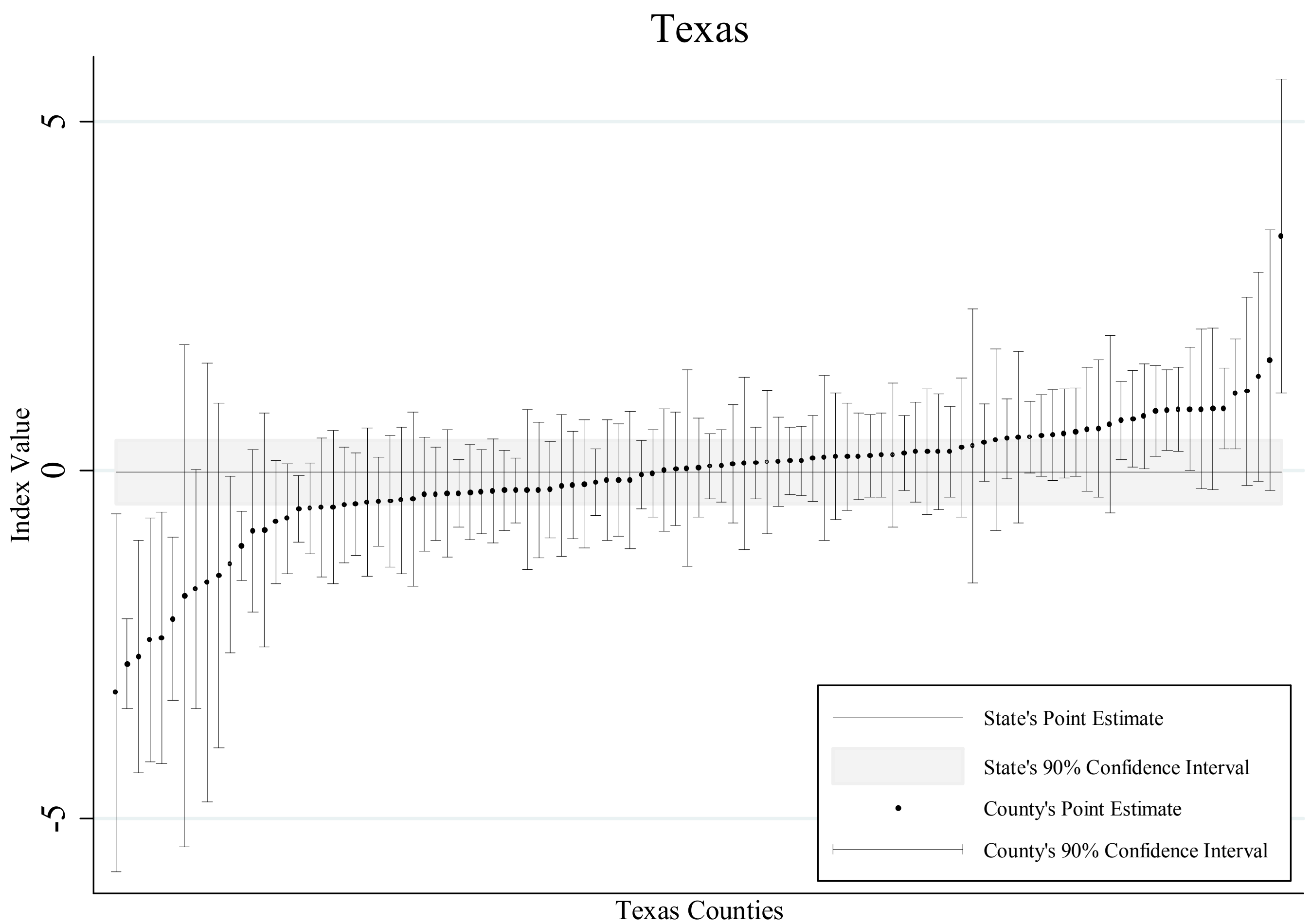

FIGURE 1 (cont.)

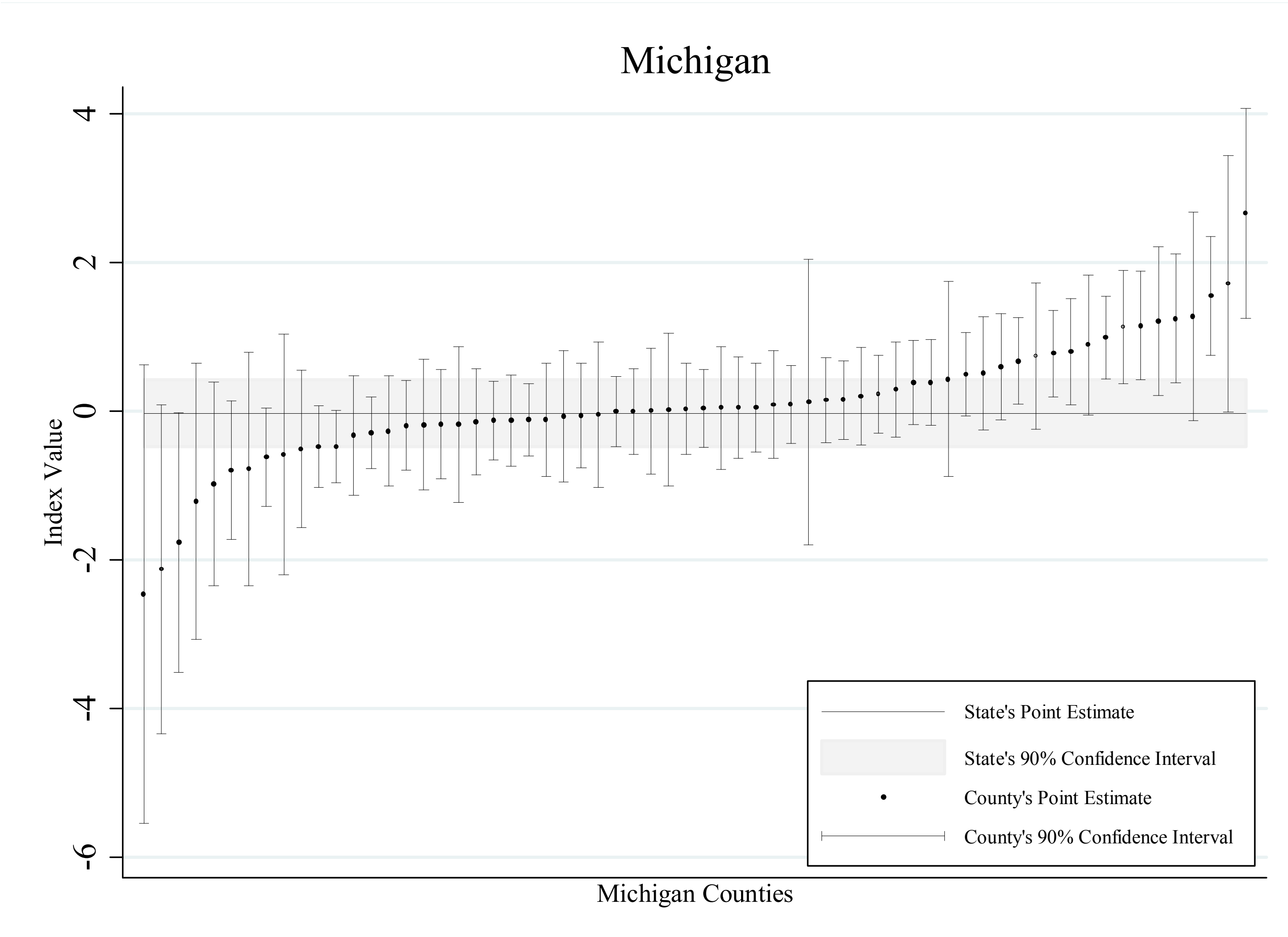

FIGURE 1 (cont.)

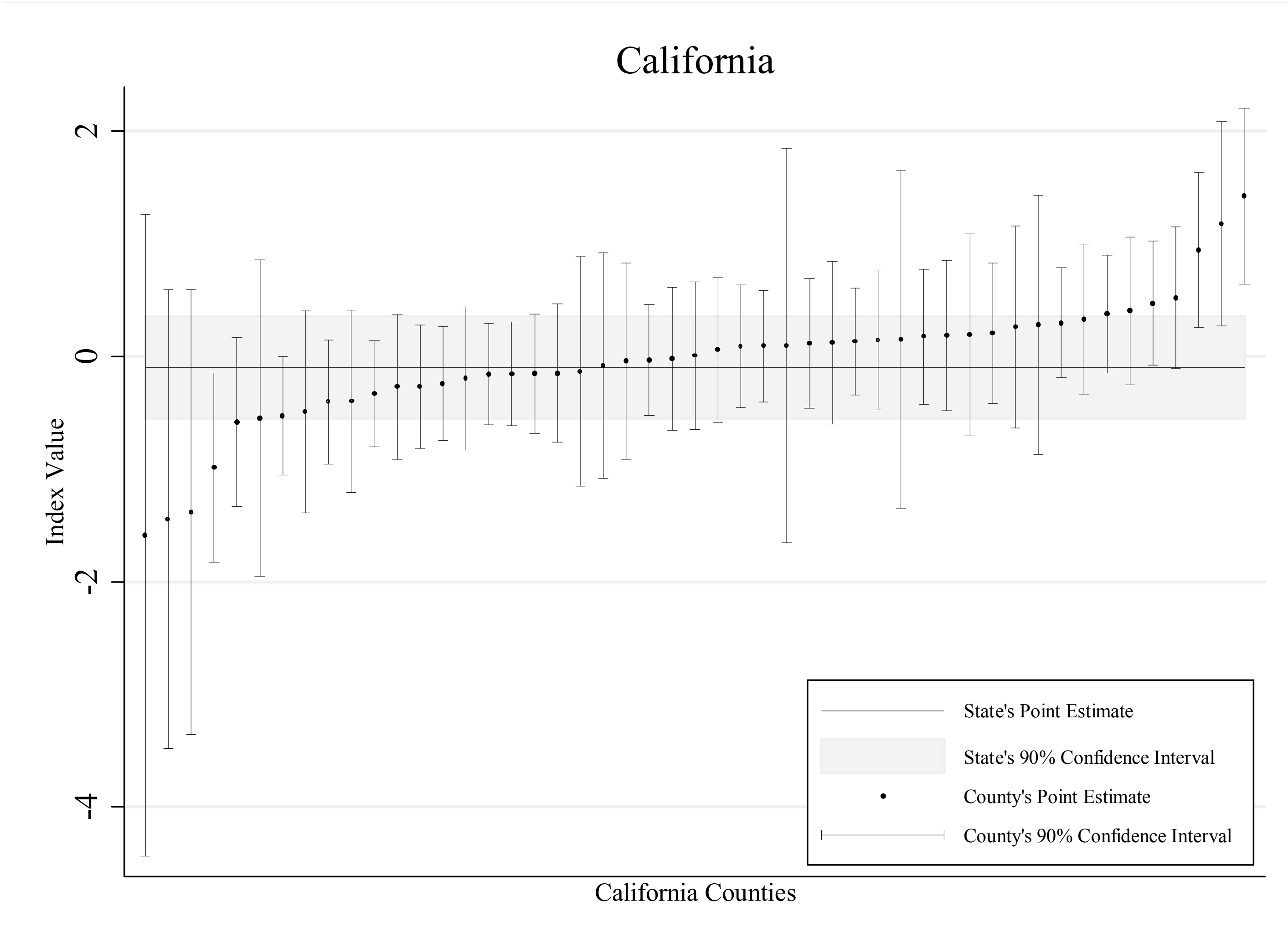

FIGURE 2
Counties that are Statistically Different from Their State