



Opinion

Volume 25 Issue 1 - August 2020
DOI: 10.19080/ARTOAJ.2020.25.556287

Agri Res & Tech: Open Access J

Copyright © All rights are reserved by Melkie Achenef Abate

Are Cross-National Empirical Studies of Sustainability, Agriculture and the Environment Cumulating Forward or Erring in an About Face?



Professor Edward L Kick*

Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC

Submission: August 13, 2020; **Published:** August 17, 2020

***Corresponding author:** Edward L kick, Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC

Opinion

I have had the privilege of being a senior (or “full”) professor since 1991, who has served as a department head at three different universities and edited, guest edited, or associate edited nine journals in sociology, agriculture, political and military sociology, statistics, food, and technology. I recently found that in the last two years I have been cited by about 1200 scholars in over 80 countries, in about 50 different subject areas. Broadened scholarly communication networks have given all of us a greatly expanded opportunity to consider the theories and methodologies adopted by authors from around the world and across a considerable number of disciplines. Regrettably, while we, as members of a curious time, have been given the potential opportunity for tremendous increases in output or scholarly “yield,” I am afraid the quality of theory and methodological applications have been waylaid, and cumulative knowledge in cross-national, quantitative research has been compromised. As an illustration, a lead corporation in the “citations business” notified me recently that there are over 7,000 published articles on the Kuznet’s curve, sustainability (including industrial agriculture) and environmental damage, written disproportionately by authors in the developing world examining their own countries. Recall the hypothesis: Economic growth creates ever more environmental damage through carbon production and agricultural toxins, among other related causes, until a nation becomes sufficiently wealthy that technology and environmental awareness turn the slope of the environmental damage curve around. I gave up after reading 10 sampled articles on the Kuznet’s curve, again drawn primarily from a sample of developing country authors. The obvious question is why all these researchers would be checking on the validity of this hypothesis for their own country and others like it, which are not wealthy now, nor will they likely be close to being “rich” in the foreseeable future. Perhaps the authors never read the original theory on this

dynamic, which requires the national attainment of substantial wealth, commonly measured by GNP/c.

A general goal of scientific inquiry is to accumulate knowledge for its own sake and for it to be applied to societal problematics so that the quality of global life may be sustained and improved upon. An initial step in the pursuit of scientific knowledge is to conceptualize the “order of things” (from Foucault). This requires care in the identification and definition of concepts such as “sustainability,” “ecological footprint” and the nature of each of their “causes.” Errors here are fundamental, and, unfortunately, frequent. I will not mention the specifics of the 10 articles I have just read. I will focus instead on a well-cited study by US authors (hereafter XYZ) that have a substantial and well-deserved readership, but one of their most respected articles reports some questionable procedures that serve as a counter example of the optimal conduct of inquiry. The article appeared in the *American Sociological Review* (ASR), and investigated the empirical veracity of different theorizations about “sustainability,” or perhaps the “ecological footprint.” It is very difficult to discern which of these concepts the authors intended to address. While the footprint and sustainability are very different concepts, they are used interchangeably throughout the XYZ article. Sustainability is accurately captured by the Brundtland Commission of the United Nations in 1987 (*Our Common Future*). There are three sustainability forms and only one of them directly reflects an environmental dynamic. Further discussion of the construction of the ecological footprint concept is provided in the pioneering work of Mathis Wackernagel and William Rees. Both of these conceptualizations were available long before the XYZ article was printed by the ASR. As an early step in the research process, exacting conceptualization is a requisite and it is hoped that researchers in developing and developed countries alike adopt this conventional procedure.

With regard to causes of the ecological footprint (and not, of course, social, economic, and environmental “sustainability”) one key causal force introduced in the ASR article is inspired by world-system theory (Wallerstein; see also Portes, Chase-Dunn and Snyder and Kick). The theory posits nations’ different positions in the world system determine variations in their internal attributes. For instance, national dominance in international trade, military power, cultural dominance, and political surveillance, the multiple networks as used by Snyder and Kick, have been shown to affect everything from national economic growth and inequality to food insecurity and environmental degradation. To my amazement, the XYZ paper chose “foreign debt” to conceptualize and measure world-system position, instead of the multiple international linkages measure empirically derived by Snyder and Kick, which is the industry standard. It has been cited and used frequently in this type of situation (Snyder and Kick; Kick et al.). Unsurprisingly, debt was found to have no impact whatsoever on the ecological footprint (sustainability?) variable by XYZ. However, the Snyder and Kick selection, or a more recent proxy of it, have been found to be causally associated with a variety of outcomes, including a number of environmental ones, such as the ecological footprint. XYZ recognize this frailty in the appropriate choice of the independent variable in their footnote 17, introducing Snyder and Kick’s alternative at some length, but they never explain WHY they chose the conceptually inferior “debt” variable in the first place. Again, the lesson is exacting conceptualization.

After exacting conceptualization, “measurement” or variable construction from concepts is the next critical step for cross-national researchers in this area. Yes, I know it is tedious to check each of the possible data values. The lesson for researchers is to make absolutely certain there are no out-of-range or anomalous scores in the data array of all the variables in the data. Ask, is every value in this range reasonable? Some will remember that a number of years ago a researcher presented blockbuster findings that started a theoretical sea change in an important area of sociological research. Fortunately for many of my colleagues, who felt the blockbuster findings had led to the premature end of their careers, one of my colleagues in graduate school at Indiana University, Dr. Glenn Firebaugh, then at Vanderbilt, was unable to replicate the blockbuster findings. However, when he purposively left missing value cases out of the analysis, the results changed dramatically. It seemed that Blockbuster’s program inadvertently kept the cases in the analysis even with missing values of -9. That was the reason Blockbuster’s findings were so anomalous! When reproduced with “missing cases” in the analysis, Firebaugh reproduced Blockbuster’ results, and when he took the missing cases and scores out of the data he found that Blockbuster’s results were grossly off the mark—a seemingly small but momentous mistake. Normal science could resume without fear dominant interpretations were inaccurate. It is just conjecture, but what would have happened if Firebaugh had not caught the mistake or had not attended a demanding graduate department that made all of us hand calculate solutions to factor and regression analyses?

In some of the papers I have read in cross-national research the sample is not large enough to generalize to the population. There are 195 countries in the world, and a sample of 40 of them (which was the case base in one of the papers I examined recently is not adequate to generalize to that population of 195. As an analogue, it is laughable that the quantitative scores my students assign to me every semester are actually a reflection of the population of students in my class. If your class is sized 25, a sample of 12 students (a fairly typical “response rate”) will not be large enough to generalize to the class population. In fact, you need all 25. Similarly for 50 students, where the population and sample size must be close in size for the mean scores given each item to be meaningful. In all likelihood your department Head will be using an unrepresentative sample size in reaching his or her conclusions about your teaching and that of your colleagues! Check to see—your teaching results may not be generalizable due to the sample size, and this may be true every semester you have been evaluated. For me that would be around 100 semesters. Apart from that, in cross-national research determine what population you are generalizing to—any “country” even if it is owned by another (such as Tasmania)? Is your cut-off a population size of 500,000 or more? Are you using “countries,” “nations,” “states,” or ‘nation-states” as both your population and sample? The lesson pertains to sample sizes and issues of conceptualization as well. Researchers often use these concepts interchangeably, much to the confusion of the readership.

I feel cheated as a reviewer or reader of a manuscript if the authors do not offer a correlation matrix showing bivariate statistical associations among all their dependent and independent variables, as well as the relevant summary statistics (means, standard variations among all variables and the residuals). Just as some newsreaders immediately go to the sports section, the front-page headlines, or the comics when they open a newspaper, I go to the correlations matrix, means and standard deviations, and then, the abstract in the manuscript. Any correlation of .80 or greater among the pool of regressors (independent variables) is suspect and immediately draws my attention. If several correlations of this magnitude are found, there is a probability of multicollinearity, which will affect the inferences reached in the article. In the broad scholarly area of national development, when so many indicators of development or modernization are candidates for use as independent (or dependent) variables, there is a good chance that any two variables a researcher picks will be highly correlated with one another. Different software programs offer different tests for multicollinearity, but you can be sure you have this problem if any of the beta coefficients (not Bs) in your results are greater than 1.00 or greater negative than -1.00. Researchers from the discipline of economics commonly report what I refer to as “impossible metrics” (“beta,” “standardized”) coefficients of 2 or 3. I encourage strict tests for muticollinearity in lieu of tests that deviate from the conventional, and are not thickly explained (e.g., XYZ). I admit sympathetic reviewers and editors may pass favorably on manuscripts that are riddled with errors for many reasons, including the “halo effect” of a theory, an author set,

or both (Merton). However, to aid cumulative science I would strongly suggest cross-national quantitative researchers report: 1. means, 2. standard deviations (realizing their importance in ratio form in the error term), 3. correlation matrices, 4. unstandardized and standardized coefficients, 5. the countries used in the analysis (and why they were retained), 6. the countries dismissed from analysis (and why they were dismissed), and 7. recognize the problems with regressors that are redundant with one another. When multicollinearity is present, coefficients can swing wildly, are very sensitive to small changes in the model, and one may not be able to trust the p-values used to identify independent variables that are statistically significant. As just one example, you may find a beta coefficient that is moderately high in magnitude but is not statistically significant, or variables that seemingly have switched signs compared with the results in the correlation matrix, or redundant variables that appear to push each other in different directions (one toward stronger and positive, and the other toward the negative). These signals become rather transparent in data researchers have worked with for 40 years, but even then the strict usage of multicollinearity tests in the software program are recommended.

I have focused on XYZ in part because they note, in reference to the standard test for multicollinearity (the VIF), that the VIF score they achieved of 8.55 suggests to them “problems with multicollinearity are not dramatic.” I have seen blogs that mention 10 as the absolute cutoff point. On the other hand I have seen even more statistics textbooks or journal articles that argue for a value of 2 or 4 (Hair et al). For the sake of advancing cumulative research I suggest in well-established areas such as sustainability or the ecological footprint that, all else equal, we adopt more rigorous standards. In completely unexplored areas we should accept relatively higher VIF levels and less demanding levels of statistical significance for the testing of coefficients. There has been a notable change in this research area in the last two decades or more of reporting B (unstandardized) coefficients instead of beta (standardized) coefficients and also the deletion of correlation matrices. There may be an excellent reason for this such as greater interest in the application of results to public policy, and less interest in adjudicating competing theories. However, in the absence of betas, descriptive statistics and correlation coefficients, it becomes more difficult without built-in tests in the software to detect problems with multicollinearity. The VIF remains, but it is apparent that test levels are unacceptably controversial. For the sake, once again, of cumulative knowledge,

I think it best that we report all the statistics mentioned above, and in their conventional as well as contemporary forms. In fact, for the sake of positive movement in science, I request that all Editors require the routine publication of both Bs and betas as well as means, standard deviations, and correlations for all cross-national multiple regression studies. If Editors themselves do not begin requiring these in the immediate future, I hope the norms of science empower reviewers and researchers to do so. When we know why the cumulative body of research on one question is huge but inconclusive, we had better begin considering the methodological choices of the body of researchers in the area.

It would be possible to continue itemizing a number of other concerns in cross-national analyses. Some suggested topics include heteroscedasticity (residuals vary with the magnitudes of regressors), time-series (evaluation of X and Y over time, with other regressors placed in the error term), identifying and analyzing outlying cases (cases lying far off the best fitted line that, if not eliminated, will pull the line in their direction), proper identification of models, use of network analyses to identify a unit's position among other units, skewed distributions, log transformations, autocorrelation, dummy and proxy variables, type I and type II errors, Poisson regression and count data, instrumental variables in reciprocal causation, and so on.

Here we have just discussed some of the more pernicious conditions that cause research to stumble, often fall backwards, and fail to cumulate knowledge in both the short and longer-runs. These are not the more dramatic events Kuhn has treated so famously in his discussion of anomalies, revolutionary science, and paradigmatic changes. Unfortunately, the errors considered here, if they are repeated into the future, may instead culminate in another 7,000 studies of the same phenomenon, which fail despite their abundance to converge on the same empirical understandings.

Some Reasons Why Cross-National Empirical Studies of Sustainability, Agriculture and the Environment May Not be Cumulating Forward, But Instead Erring in an About Face *

*I wish to thank Drs. Thomas Burns, Jeffrey Kentor, and Andrew Jorgenson for past discussions of the issues appearing in this manuscript. They are not in any way responsible for the content of the current manuscript. This work was supported by the USDA National Institute of Food and Agriculture, Hatch projects 1002045.



This work is licensed under Creative Commons Attribution 4.0 License
DOI: [10.19080/ARTOAJ.2020.25.556287](https://doi.org/10.19080/ARTOAJ.2020.25.556287)

**Your next submission with Juniper Publishers
will reach you the below assets**

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
(Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
<https://juniperpublishers.com/online-submission.php>