Inferences from regression analysis: are they valid?[1]

Steven J. Klees   [University of Maryland, USA]

The focus of this paper is regression analysis. Regression analysis forms the core for a family of techniques including path analysis, structural equation modelling, hierarchical linear modelling, and others. Regression analysis is perhaps the most-used quantitative method in the social sciences, most especially in economics and sociology but it has made inroads even in fields like anthropology and history. It forms the principal basis for determining the impact of social policies and, as such, has enormous influence on almost all public policy decisions.

This paper raises fundamental questions about the utility of regression analysis for causal inference. I argue that the conditions necessary for regression analysis to yield valid causal inferences are so far from ever being met or approximated that such inferences are never valid. This dismal conclusion follows clearly from examining these conditions in the context of three widely-studied examples of applied regression analysis: earnings functions, education production functions, and aggregate production functions. Since my field of specialization is the economics of education, I approach each of these examples from that perspective. Nonetheless, I argue that my conclusions are not particular to looking at the impact of education or to these three examples, but that the underlying problems exhibited therein generally hold to be true in making causal inferences from regression analyses about other variables and on other topics.

Overall argument

In some fields, regression analysis is used as an ad hoc empirical exercise for moving beyond simple correlations. Researchers are often interested in the impact of a particular independent variable on a particular dependent variable and use regression analysis as a way of controlling for a few covariates. Despite being common, in many fields such empirical fishing expeditions are frowned upon because the result of particular interest (the coefficient on the key independent variable under examination) will depend on which covariates are selected as "controls".

To the contrary, nowadays, most fields teach that one has to be serious about causal modeling in order to use regression analysis for causal inference. Causal models require certain conditions to hold for regression coefficients to be accurate and unbiased estimates of causal impact. While these conditions are often expressed as properties of regression residuals, they also may be expressed as three necessary conditions for the proper specification of a causal model examining a particular (or set of) dependent variable(s):

- All relevant variables are included in the model;

- All variables are measured properly; and
- The correct functional interrelationships of the variables are specified.

In order to achieve proper specification, one must have a very well elaborated theory that allows one to fulfill these conditions.[2] The fundamental problem with regression analyses is that we do not have sufficiently complete theories in any of our fields to properly specify causal models. Regression analysis application literatures therefore generally become discussions about the degree of misspecification and its consequences. Unfortunately, regression analysis theory is very unforgiving; with just one omitted variable, all regression coefficients may be biased to an unknown extent and in an unknown direction. While researchers sometimes use ad hoc reasoning to infer the direction of bias of particular omitted variables, they do so based on its potential correlation with a particular included independent variable of interest. However, this ad hoc reasoning is not valid. The direction of bias will depend on the intercorrelation of the omitted variable with *all* the included variables. Ad hoc reasoning does not offer a clue as to how biased included coefficients are.

More to the point, we are never talking about the simple case of a single omitted variable. We are faced with multiple failures of all three assumptions: many variables are always omitted, we have little idea of how to best measure the variables we are able to include; and we have hardly any idea of their functional form. This is best illustrated by looking at concrete examples of regression analyses literatures, as I do below.

Earnings functions

Earnings functions are used principally by economists and sociologists to investigate the determinants of earnings differences. It is probably one of the most-regressed topics of study and has been especially relevant to the economics of education as the source of rate of return to education estimates (Blaug, 1976; Psacharopoulos and Patrinos, 2004). I find earnings functions especially interesting because it is one of the few terrains where social scientists on the left and the right have competed, principally because of arguments about labor market segmentation. In economics, this was about a challenge to the neoclassical idea that there was just one big perfect labor market in which success was determined by your individual human capital characteristics. To the contrary, political economists and other critics of the neoclassical story saw an imperfect labor market with fractures (e.g., divisions into primary and secondary labor markets) and structures (e.g., large firms, unions, sexism, and racism) that greatly influenced whether an individual succeeded. In sociology, this was about a similar challenge to the idea generated from the dominant structural-functionalist theory and its derivative status attainment theory that, like economics, argued that individual success was determined chiefly by individual characteristics. To the contrary, critical sociologists, often sharing a conflict theory critique of structural functionalism, argued, like political economists, that success in the labor market was greatly determined by structural factors. Each side in this debate used regression analysis to "prove" their point of view (Klees and Milton, 1993).

More to the point here, is that there have literally been hundreds of earnings functions studies with each study using anywhere from somewhat to vastly different specifications. The three principal conditions necessary for the regression coefficients of an earnings function to be

---

[2] Statistical techniques for getting around these conditions are discussed briefly in conclusion., The essential problem with them is that they can't deal with rampant misspecification.

accurate estimators of true causal impact are very far from being fulfilled. First, all relevant variables that may affect earnings can never be included. Our theories literally posit dozens of variables, and which variables are included in a particular regression study is again idiosyncratic. Examples of variables that some researchers have considered relevant are: health status, years of schooling, quality of schooling, type of schooling, cognitive ability, race, ethnicity, religion, socioeconomic status, gender, immigration status, marital status, participation in a union, job search, occupation status and differentiation, labor market segment, firm and industry characteristics, and many more. Second, we do not know the "right" way to measure most of these variables. Measurement is ad hoc and varies from study to study. Third, the functional interrelationship between variables is not known. While it is common to use the natural logarithm of income as the dependent variable, even neoclassical economists admit the basis for doing so is very weak and, in actuality, what would be needed is to specify some unknown set of complex simultaneous equations filled with variables subject to complex interactions (Blaug, 1976; Klees and Milton, 1993).

The result of this state of affairs is endless misspecification – by necessity.[3] Each researcher has an almost infinite array of choices in how they specify the earnings function they estimate. Each regression study is never a replication but always different from others in many respects. The upshot is each regression study is idiosyncratic. Since it is relatively easy to get significant coefficients, especially with large data sets, everyone finds their particular variable of interest to be significant. When there is controversy, everyone finds empirical evidence to support their side of the debate. Every segmentation theorist finds labor market segments to be a significant factor in determining earnings and other labor market outcomes, yet no neoclassical economist or structural functionalist sociologist ever does.

With respect to education, most everyone finds some effect of education on earnings, reports it, and sometimes uses it to estimate a rate of return. But alternative specifications always yield different results, and so the estimates are notoriously unstable and inconsistent. Hanushek (1980, p. 240) argued that

> "…estimated rates of return for years of schooling particularly in regression estimates [on earnings], considering other individual differences appear very unstable: changes in sample, changes in time periods, changes in precise model specifications yield enormous changes in estimated rates of return."

The estimated impacts of education on earnings and associated rates of return are basically arbitrary, the result of ad hoc empiricism run rampant.

Education production functions

Another very common use of regression analysis is to estimate what are called education production or input-output functions (Levin, 1976; Hanushek, 1986). The dependent variable usually studied is a student's score on some achievement test. The three conditions for proper specification are again impossible to fulfill. First, the array of potential independent variables is huge, including, for example: socioeconomic status, gender, race, ethnicity, age,

---

[3] Imagine you collected data on the earnings of 100 people selected at random. Imagine delineating the literally dozens of factors that would explain why they had different earnings, only one factor being their educational differences. Despite sophisticated statistical techniques, it is simply impossible to accurately separate out the impact of education or any other variable from all other factors.

homework effort, computer use in the home, previous learning, ability, motivation, aspiration, peer characteristics, teacher degree level, teacher practices, teacher ability, teacher experience, class size, school climate, principal characteristics, curriculum policies, to name a few. Second, there is no agreement on how to measure most, if not all, of these variables. Third, again the possible functional interrelationships are innumerable. Contrary to the linear formulation usually run, recursive and simultaneous equation formulations with an array of interaction terms among the independent variables have been posited but little used (Levin, 1976; Hanushek, 1986).

Economists of education, sociologists of education, and other educational researchers have estimated hundreds of these functions. Again, with such an infinite array of specification choices, almost every study is unique and idiosyncratic. Hanushek (2004, 1986, 1979) has, over the long term, studied and summarized the results of such studies. Not surprisingly, he and others have found inconsistent results. However, he and the vast majority of quantitative researchers cling to the hope that improvements in models and data can eventually show some clear results. To the contrary, I see the complete indeterminacy of this form of research built into the very assumptions on which it is based.

A particularly destructive use of these functions is for so-called performance pay for teachers. The "value-added" to student achievement test score by individual teachers are ascertained through estimating an educational production function, usually using only a few control variables, with teacher effects determined by dummy variables or residuals (AERA, 2015). The problem, of course, is that with different control variables different teachers are ranked high or low and there is neither rhyme nor reason to choosing one specification over another. Yet around the U.S. teachers are being hired and fired based on these completely spurious results.

Aggregate production functions

While many economics of education studies have looked at the impact of education inputs on student achievement and others have focused on the connection between education and earnings, as a proxy for productivity, some studies have tried to look more directly at the connection of education and productivity by looking at the effect of education on economic growth, as measured by GNP. Indeed, some of the earliest work on human capital examined the correlation between levels of education or school enrolments in a country and its GNP (Bowman, 1966; Blaug, 1970). However, correlation is not causation, and these studies were quickly dismissed as neither controlling for other differences between countries nor demonstrating which was cause and which was effect (Blaug, 1970).

The most significant early, and still widely quoted, work that tried to take a more sophisticated look at the connection between education and GNP was by Edward Denison (1961, 1967). Denison focused on a particular form of what economists call an "aggregate production function". Like an earnings function tries to look at all the variables that might affect earnings, production functions look more directly at all the variables that might affect production output in a particular industry. An "aggregate" production function, as the name implies, looks at the effect of inputs on total production output, that is, GNP. This approach, *in theory*, could get around the need to assume earnings reflect productivity by directly looking at the impact of education on output. However, Denison's famous work did *not* do this. Instead of estimating an aggregate production function, it *assumed* one of a particular form and then used

education's association with earnings as the evidence of education's impact on GNP, thus offering nothing different than the results yielded by the problematic education-earnings connection discussed above. Blaug (1970, p. 100) dismissed all this early research: "In short, we learn from international comparisons [of education and GNP]…that we do not learn from international comparisons."

Attempting to connect education directly to GNP generally fell out of favor until the late 1980s and 1990s when a few works in the area of what was called "new growth theory" signaled a broader vision of education's contribution (Romer,1986; Lucas, 1996 Psacharopoulos and Patrinos, 2004).[4] This vision is theoretically interesting in that education is seen not just as contributing to worker productivity but as enhancing growth through a variety of mechanisms and externalities. However, empirically these new directions have proven extremely difficult to model mathematically. Almost every researcher who attempts to estimate these connections therefore uses a different model and the results are, as one would expect, typically idiosyncratic, unstable, and inconsistent (Psacharopoulos and Patrinos, 2004; Stevens and Weale, 2004). In 1970, Blaug said the "Mecca of the economics of education lies elsewhere" (p. 100), and I think that holds true today, for reasons similar to the ones I discussed for education and earnings and for educational inputs on outputs.

As I said, the results of the empirical research estimating the impacts above have been idiosyncratic, unstable, and inconsistent. The same is true for the impact of education on GNP for similar reasons. First, there is no agreement on how to measure the stock or flow of human capital in a country. Various proxies have been used but, as Psacharopoulos and Patrinos (2004, pp. 13-14) admit, such measurement may be the "weakest point" of these studies: "Such data have serious intertemporal and inter-country comparability problems, and there are data gaps often filled with constructed data based on interpolations and extrapolations."

Second, more to my general point, as Psacharopoulos and Patrinos (2004, p. 15) also admit: "Countries also differ in many other aspects than those measured by physical and human capital stock…" that can affect GNP. Estimates of aggregate production functions have literally used dozens of different variables as inputs, such as climate, latitude, access to waterways, transportation infrastructure, technological development, investment climate, cultural and political differences, fiscal and monetary policy, etc. (Stevens and Weale, 2004; Hulten and Issakson, 2007; Hulten, 2009).[5] Empirical studies idiosyncratically choose some of these input variables, from those available in the data set being used, and always omit many others. As Psacharopoulos and Patrinos (2004, p. 15) again admit: "These omitted variables can lead to margins of error of hundreds of per cent in accounting for differences in the economic growth path between countries."

Third, it is widely recognized by economists that the linear functional form so commonly used in regression analyses studies is not applicable to aggregate production functions. However, there is considerable debate over what functional form to use and different functional forms yield different estimates of the impact of education (and of all other inputs) on GNP (Stevens and Weale, 2004). There is even a respected school of economics that says that there is no theoretical basis for even believing that an aggregate production function actually exists. Each

---

[4] Sociologists periodically study the determinants of GNP but usually without any pretense of theoretical justification for an aggregate production function (e.g., Kentor, 1998).
[5] Tan (2014, p. 426) argues that "social, political, institutional, and cultural factors need to be considered when the impact of education" on economic growth is estimated.

good and service may have a "production function," meaning some mathematical regularity in how inputs like land, labor, capital, and technology combine to produce televisions, yachts, insurance policies, hamburgers, etc. However, since there is no physical process by which aggregate GNP is produced, nor, from this perspective, is there some way to aggregate and measure physical capital, trying to specify an aggregate production function is seen as nonsensical (Cohen and Harcourt, 2003). Guerrien and Gun (2015, p. 100) note that Paul Samuelson, Nobel laureate in economics, pointed out that aggregate production functions wrongly offer "a statistical test of an accounting identity (which is by definition always true)". They argue (p. 99) for the need "to convince everyone to definitively abandon the aggregated [sic] production functions, both in theory and practice" (also see Felibe and McCombie, 2013).

Given these fundamental problems with fulfilling the conditions for regression analysis to yield accurate estimates of causal impact (discussed earlier), it is no wonder that consistent results of the impact of education on GNP are not found. Reviews of this literature report a bewildering array of idiosyncratic methodological choices resulting in a bewildering array of different results (Stevens and Weale, 2004). Patrinos and Psacharopoulos (2004, p. 15) quote Temple and Voth (1998, p. 1359): "[A]ttempting to impose the framework of an aggregate production function is almost certainly the wrong approach for many developing countries." I would say that this is the wrong approach for any country.[6]

It should be noted that almost all these studies only offer some measure of the quantity of education, not its quality. In a widely quoted recent study, Hanushek and Woessmann (2008) try to remedy this by adding country average PISA test scores as a proxy for the quality of education in a country, concluding that a one standard deviation difference in test scores yields a 2 percentage point higher growth rate of GNP/capita. In the light of the foregoing problems, I find this claim completely unreasonable and its uncritical reception due to ignorance of the fundamental problems with human capital theory and empirics discussed in this paper (also see Klees, forthcoming). Hanushek and Woessmann's measures of the quantity and quality of education, choice of other inputs to control for, and choice of functional form are all idiosyncratic.[7] They are only one of literally thousands of reasonable alternative specifications of an aggregate production function. Different specifications will yield different results.[8]

Discussion

While I have approached the examples above as an economist most interested in the impact of education, the problems are identical in looking at the impact of any of the other myriad

---

[6] Levine and Renelt (1992, p. 942) discuss the "variety of economic policy, political, and institutional factors" that need to be accounted for in GNP regressions: "Given that over 50 variables have been found to be significantly correlated with growth in at least one regression, readers may be uncertain as to the confidence they should place in the findings of any one study…We find that only a few findings can withstand slight alterations in the list of explanatory variables." I would add that real world alterations are far from slight. Also see Sala-i-Martin's (1997) article on the subject, entitled "I just Ran Two Million Regressions."

[7] Hanushek and Woessmann (2008) control for only two of the literally dozens of variables that they could have included, and these two variables reflected neoliberal development ideology (openness to trade and security of property rights). Different controls would, of course, yield very different estimates of the impact of the "quality" of education.

[8] Hanushek and Woessmann (2015) then take these invalid estimates and make country-by-country projections decades in the future of what GNP would be if PISA results improved. Their results have no validity, depending on literally hundreds of assumptions and completely tenuous causal linkages.

independent variables in these equations. Moreover, as far as I can see, the impossibility of proper specification is true generally in regression analyses across the social sciences, whether we are looking at the factors affecting occupational status, voting behavior, etc. The problem is that as implied by the three conditions for regression analyses to yield accurate, unbiased estimates, you need to investigate a phenomenon that has underlying mathematical regularities – and, moreover, you need to know what they are. Neither seems true. I have no reason to believe that the way in which multiple factors affect earnings, student achievement, and GNP have some underlying mathematical regularity across individuals or countries. More likely, each individual or country has a different function, and one that changes over time. Even if there was some constancy, the processes are so complex that we have no idea of what the function looks like.

Researchers recognize that they do not know the true function and seem to treat, usually implicitly, their results as a good-enough approximation. But there is no basis for the belief that the results of what is run in practice is anything close to the underlying phenomenon, even if there is an underlying phenomenon. This just seems to be wishful thinking. Most regression analysis research doesn't even pay lip service to theoretical regularities. But you can't just regress anything you want and expect the results to approximate reality. And even when researchers take somewhat seriously the need to have an underlying theoretical framework – as they have, at least to some extent, in the examples of studies of earnings, educational achievement, and GNP that I have used to illustrate my argument – they are so far from the conditions necessary for proper specification that one can have no confidence in the validity of the results.

Moreover, what researchers do in practice invalidates their results even further. In theory, when using regression analysis, you are supposed to start with a complete model specification, and then take your data and estimate it, a one-shot deal. Given the indeterminacy of model specification, no one does that in practice. In his now classic article, "Let's Take the Con Out of Econometrics," Leamer (1983, p. 36) describes regression analysis in the real world and its consequences:

> "The econometric art as it is practiced at the computer … involves fitting many, perhaps thousands, of statistical models….This searching for a model is often well-intentioned, but there can be no doubt that such a specification search invalidates the traditional theories of inference. The concepts of unbiasedness, consistency, efficiency, maximum likelihood estimation, in fact, all the concepts of traditional theory utterly lose their meaning by the time an applied researcher pulls from the bramble of computer output the one thorn of a model he likes best, the one he chooses to portray as a rose."

The practical question to me then becomes whether we have learned anything from all this research? Most quantitative researchers would say they have, but I believe that such learning, if examined, would turn out to be from a subset of studies done from a perspective with which the researcher agreed. As Leamer (1983, p. 37) put it: "Hardly anyone takes data analyses seriously. Or, perhaps more accurately, hardly anyone takes anyone else's data analyses seriously" (also see Leamer, 2010). Hardly anyone ever uses anyone else's specification without "improving" on it, arguing explicitly or implicitly that the previous study was incorrect.

These remarks do not imply that, at least within paradigms, there is no cumulative learning from one another's arguments. Such learning does take place. However, the argument here suggests that regression-based causal inference is simply an excuse for theorizing but does not provide any valid evidence for it. There's an old saw in economics: "If you torture the data long enough, nature will confess." In reality, nature never confesses. Studies from the three examples I have chosen have commanded the attention of educators and policymakers for over 50 years, yet, in reality, I believe that this approach has no validity, providing no reliable, or even approximate, information to help a sensible allocation of societal resources.

Econometricians and other regression analysts do recognize that there are many sources for biases of regression coefficients. They spend a lot of time on ways to correct for things like sample selection bias and measurement error – without much success unless you are willing to make some heroic assumptions. But these problems are minor compared to rampant misspecification. Regression analysts have tried to deal with one misspecification problem – that of omitted variables – through the use of instrumental variables (IVs). But this generally requires accurate measurement of included variables and correct specification of functional form, none of which is ever true. Instrumental variable techniques give different results depending on the IV chosen, as well as have other problems (Heckman and Urzua, 2009; Leamer, 2010). Again, these and other techniques (regression discontinuity, differences-in-differences) require heroic assumptions to deal with any aspect of misspecification (Angrist and Pischke, 2009).[9]

I believe that, unfortunately, regression analysis methodology is a dead end, no better than alchemy and phrenology, and someday people will look back in wonder at how so many intelligent people could convince themselves otherwise. This is not a problem that better modeling, techniques, and data can fix.[10]

Alternatives

I do not see the essence of the problem as quantification. Nor do I think it futile to try to look for causes and consequences of our practices and policies. Quantifying social phenomena clearly has its limits and, at best, yields approximations (Samoff, 1991). But cross-tabulations and correlations are useful to suggest interrelationships. As is well-known, however, any associations found may be spurious or have a myriad of alternative explanations. For example, crosstabs may reveal that, on average, women earn $.75 compared to $1.00 earned by men. We can unpack this some by looking at women and men working full-time, where perhaps the data show a comparison of $.80 to the $1.00. We can further look at college-educated women working full-time compared to men in similar circumstances, perhaps giving us a comparison of $.90 to the $1.00. Crosstabs can give even finer comparisons. These comparisons, despite limitations, offer real, descriptive, face valid data. Unfortunately, social sciences' hope that we can control simultaneously for a range of factors

---

[9] At best, deviations from proper specification have to be very well-behaved for corrective actions to work – and they never are. Another approach for dealing with misspecification is sensitivity analysis but even the extreme form of this, advocated by Leamer (1983, 2010), that would be needed to account for the extremely wide variations in specification, are unlikely to uncover any robust regularities, as is evident from most empirical literatures (also see McAleer, Pagan, and Volker, 1985, and Saltelli, 2008).

[10] I come to the opposite conclusion of Hendry's (1980) article, "Econometrics – Alchemy or Science?" It is interesting to note that Hendry's litany of Keynes' critique of statistical methods parallels my own. As Hendry says: "Keynes came close to asserting that no economic theory is ever testable" (p. 396) (also see Pratten, 2005).

like education, labor force attachment, discrimination, and others is simply more wishful thinking.

The problem is that the causal relations underlying such associations are so complex and so irregular that the mechanical process of regression analysis has no hope of unpacking them. One hope for quantitative researchers who recognize the problems I have discussed is the use of experimentation – with the preferred terminology these days being randomized controlled trials (RCTs). RCTs supposedly get around the issues faced by regression analysis through the use of careful physical, experimental controls instead of statistical ones. The idea is that doing so will let one look at the effect of an individual factor, such as whether a student attended a particular reading program. In order to do this, one randomly assigns students to an experimental group and control group, which, in theory, will allow for firm attribution of cause and effect. Having done this, one hopes that the difference in achievement between the groups is a result of being in the reading program. Unfortunately, it may or may not be. You still have the problem that the social and pedagogical processes are so complex, with so many aspects for which to account, that, along some relevant dimensions, the control and experimental group will not be similar. That is, if you look closely at all potentially relevant factors, control groups almost always turn out systematically different from the experimental group, and the result is we no longer have the ability to make clear inferences. Instead, we need to use some form of statistical analysis to control for differences between the two groups. However, the application of statistical controls becomes an ad hoc exercise, even worse than the causal modeling regression approach. In the latter, at least there is a pretence of developing a complete model of potentially intervening variables whereas with the former a few covariates are selected rather arbitrarily as controls. In the end, one does not know whether to attribute achievement differences to the reading program or other factors (Leamer, 2010).

If we are interested in looking at quantitative data, I am afraid we are mostly stuck with arguing from cross-tabulations and correlations. This is a dismal prospect for most quantitative researchers who have spent years becoming virtuosos at data analysis and see the implications of my argument as essentially abandoning the research enterprise. Fortunately, for many of us, the research enterprise is alive and well, with a myriad of more qualitative alternative methodologies with which to investigate our educational and social world.

When I went to graduate school, introduction to research methods courses often focused on regression analysis or on Campbell and Stanley's (1963) examination of the design and analysis of quantitative experimental and quasi-experimental studies. This is still true today within certain fields and university departments. However, the past 30 years has seen a blossoming of alternative approaches to research methods, especially in education, but in other fields as well. Education has been at the forefront of such changes, largely, in my view, because many of the changes were generated within the field of program evaluation which grew, in large part, from the educational evaluations that were mandated by the U.S. Congress in the 1960s and 1970s. Many of those involved in evaluation fieldwork simply found that the quantitative approach to research and evaluation could not capture the experience of the programs they were studying and drew upon other traditions, such as in sociology or anthropology, or invented new approaches. In subsequent years, these forays yielded a wide array of alternative methods for research and evaluation (Mertens, 2015).

For a number of years, I have been fortunate to teach our department's Introduction to Research Methods course. While any grouping of methods is somewhat arbitrary and their labeling always problematic, the course is divided in three, focusing in turn on quantitative/positivist methods, qualitative/ interpretive methods, and critical/transformative methods. There is a large literature on the qualitative/quantitative debate. Some argue too much has been made of it, while others, whom I agree with, argue that there are fundamental theoretical differences in outlook that need to be considered (Smith and Hesushius 1986; Mertens, 2015). Regardless, it is clear that there are lots of qualitative alternatives to quantitative experimental and regression analysis approaches, including: case study, ethnography, grounded theory, phenomenology, narrative, and oral history, to name a few.

Additional methodological alternatives are offered by critical/transformative perspectives which come out of the array of theories in the social sciences and applied fields such as radical political economy, critical sociology, feminisms, queer theory, and others focused on issues of marginalization (Klees, 2008). These perspectives generally criticize the fundamental lack of objectivity of positivist/quantitative research and qualitative/interpretive research, arguing that there is no neutral research, and that too often such studies are done in support of dominant interests. Critical/transformative research takes an explicit position to work in the interests of marginalized people. This includes research under the labels of participatory, action, feminist, indigenous, critical, critical ethnography, and critical race (Denzin et al. 2007; Mertens 2015).[11]

Proponents of quantitative research recognize that some of these alternative methods exist but usually, at best, relegate them to the realm of generating ideas, not to the scientific process of building knowledge of the social world. To the contrary, many proponents of alternative methods argue that they are as or more valid, reliable, and generalizable than quantitative.[12] For example, Miles and Huberman (1994, p. 434) go so far as to argue;

> "Qualitative studies…are especially well-suited to finding causal relations; they look directly and longitudinally at the local processes underlying a temporal series of events and states, showing how these led to specific outcomes, and ruling out rival hypotheses. In effect we get inside the black box; we can understand not just that a particular thing happened, but how and why it happened."

Similarly, strong arguments are made for the transferability and generalizability of qualitative and critical research (Donmoyer 1990; Mertens, 2015).

Conclusions

For a number of years, a sociologist of education colleague and I taught a regression lab course. We used a good national data set and had the students spend the semester running alternative specifications of education production functions. Each class, groups came in and explained their specifications and their results. As expected, different specifications of

[11] These classifications are not clear and neat. For example, you can do grounded theory or feminist research from positivist, interpretive, or critical perspectives. To make things more confusing still, critical paradigm researchers may avail themselves of any method, including quantitative ones.
[12] Given the problems with the positivist paradigm, there is considerable literature examining alternative criteria for good research for the interpretive and critical paradigms (Mertens, 2015).

included variables, decisions on how to measure variables, and functional forms yielded substantially different results. Each group was asked to explain their results as if they had written them up for a journal article. My colleague used to always comment on a group's explanation of their results with: "That makes sense." And it always did. As do the articles in the literature I reviewed above. We can always make sense of our results and always do. When running regressions, we stop making the many adjustments to our regressions – that always must be made – when we get results that make sense to us. However, taking these literatures as a whole, they simply result in divergent findings, all based on reasonable – at least to some – alternative specifications of their regression equation models.

In conclusion, I wish to say I'd like to be wrong about my argument in this paper. It would be useful if the emperor's not-so-new clothes were more than the nakedness researchers seem to avoid looking at too closely. Unfortunately, theory and practice seem to strongly indicate otherwise. The theoretical conditions for regression analysis to "work" are never close to being met. And, in practice, regression analysis applications seem to result in interminable debates because specifications are so loose that researchers seem to be able to use this family of techniques to "prove" almost anything they want. Nonetheless, while we cannot find the simple cause-effect regularities that regression analysts would like to uncover, at the very least there are still many alternatives methods for investigating our educational and social world.

References

AERA. 2015. "AERA Statement on the Use of Value-Added Modles (VAM) for the Evaluation of Educators and Educator Preparation Programs." *Educational Researcher*, 44 (8): 448-452.

Angrist, Joshua D. and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Guide.* Princeton: Princeton University Press.

Blaug, Mark. 1970. *An introduction to the economics of education.* London: Penguin.

Blaug, Mark. 1976. "The Empirical Status of Human Capital Theory: A Slightly Jaundiced Survey." *Journal of Economic Literature.* 14 (September): 827-55.

Bowman, Mary Jean. 1966. (Spring) "The Human Investment Revolution in Economic Thought." *Sociology of Education.* 39 (2): 111-37.

Cohen, A. J. and G. C. Harcourt. 2003. "Retrospectives: Whatever Happened to the Cambridge Capital Theory Controversies?" *Journal of Economic Perspectives.* 17 (1): 199–214.

Denison, Edward F. 1961. *The Sources of Economic Growth in the United States and the Alternatives Before Us.* NY: Committee for Economic Development.

Denison, Edward F. 1967. *Why Growth Rates Differ?* Washington: DC: Brookings.

Denzin, Norman, Lincoln, Yvonna, and Smith, Linda. 2007. *Handbook of Critical and Indigenous Methodologies.* Thousand Oaks, Ca.: Sage.

Donmoyer, Robert. 1990. "Generalizability and the Single-Case Study." In W. Eisner and A. Peshkin, eds., *Qualitative Inquiry in Education.* NY: Teachers College Press.

Felipe, Jesus and John McCombie. 2013. *The Aggregate Production Function and the Measurement of Technical Change: Not Even Wrong.* Cheltenham, UK: Edward Elgar.

Guerrien, Bernard and Ozgur Gun. 2015. "Putting an End to the Aggregate Function of Production." *Real-World Economics Review*, 73: 99-109.

Hanushek, Eric. 1979. "Conceptual and Empirical Issues in the Estimation of Educational Production Functions." *Journal of Human Resources,* 14 (3): 351-88.

Hanushek, Eric. 1980. "Alternative Models of Earnings determination and Labor Market Structures." *Journal of Human Resources,* 16: 238-59.

Hanushek, Eric. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature*, 24 (3): 1141-1177.

Hanushek, Eric. 2004. "What If There Are No 'Best Practices'?" *Scottish Journal of Political Economy*, 51 (2):156-72.

Hanushek, Eric and Ludger Woessmann. 2008. "The role of cognitive skills in economic development." *Journal of Economic Literature*. 46 (3): 607-668.

Hanushek, Eric and Ludger Woessmann. 2015. *Universal Basic Skills: What Countries Stand to Gain*. Paris: OECD.

Heckman, J. J. and S. Urzua. 2009. "Comparing IV with Structural Models:What Simple IV Can and Cannot Identify." National Bureau of Economic Working Paper 14706, Washington, DC.

Hendry, David. 1980. "Econometrics: Alchemy or Science?" *Economica*, 47 (188): 387-406.

Hulten, Charles R. 2009. "Growth Accounting," National Bureau of Economic Research Working Paper 15341, Washington, DC.

Hulten, Charles R. and Anders Isaksson. 2007. "Why Development Levels Differ: The Sources of Differential Economic Growth in a Panel of High and Low Income Countries," National Bureau of Economic Research Working Paper 13469 (October) Washington, DC.

Kentor, Jeffrey. 1998. "The Long-Term Effects of Foreign Investment Dependence on Economic Growth, 1940-1990." *American Journal of Sociology*, 103 (4): 1024-1046.

Klees, Steven. 2008. "Reflections on Theory, Method, and Practice in Comparative and International Education." *Comparative Education Review,* 52 (3) (August): 301-28.

Klees, Steven. Forthcoming. Human Capital Theory and Rates of Return: Brilliant Ideas or Ideological Dead Ends." *Comparative Education Review*.

Klees, Steven and Milton Sande. 1993. "Inferences From Regression Analysis: The Case of Earnings Functions." Paper presented at the Comparative and International Education Society annual meeting, Kingston, Jamaica, March 11-19, 1993.

Leamer, Edward E. 1983. "Let's Take the Con Out of Econometrics." *American Economic Review* 73 (March): 31-43.

Leamer, Edward E. 2010. "Tantalus on the Road to Asymptopia." *Journal of Economic Perspectives*, 24 (2): 31-46.

Levin, Henry M. 1976. "Concepts of Economic Efficiency and Educational Production." In *Education as an Industry*, ed. J.T. Froomkin. Cambridge, MA: Ballinger.

Levine, Ross and David Renelt. 1992. "A Sensitivity Analysis of Cross-Country Growth Regressions." *American Economic Review*, 82 (4): 942-963.

Lucas, Robert. 1996. "On the mechanics of economic development." In G. Grossman (Ed.), *Economic growth: Theory and evidence*, Vol. 1, pp. 284-323. Cheltenham, UK: Elgar.

McAleer, Michael, Adrain R. Pagan, and Paul A. Volker. 1985. "What Will Take the Con Out of Econometrics." *American Economic Review*, 75 (3): 293-307.

Mertens, Donna. 2015. *Research Methods in Education and Psychology: Integrating Diversity with Quantitative and Qualitative Approaches* (4th ed.). Thousand Oaks, CA: Sage.

Miles, M., and Huberman, A. 1994. *Qualitative Data Analysis* (2nd ed.). Newbury Park, CA: Sage.

Pratten, Stephen. 2005. "Economics as Progress: The LSE Approach to Econometric Modelling and Critical Realism as Programmes for Research." *Cambridge Journal of Economics*, 29 (2): 179-205.

Psacharopoulos, George and Harry A. Patrinos. 2004. "Human Capital and Rates of Return." In *International Handbook on the Economics of Education*, eds. Geraint Johnes and Jill Johnes. Northhampton, MA: Edward Elgar.

Romer, P. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy*. 94: 1002-37.

Sala-i-Martin, Xavier X. 1997. "I Just Ran Two Million Regressions." *American Economic Review*, 87 (2) 178-183.

Saltelli, Andrea et al. 2008. *Global Sensitivity Analysis: The Primer*. Chichester, UK: Wiley.

Samoff, Joel. 1991. "The Façade of Precision in Education Data and Statistics: A Troubling Example from Tanzania," *Journal of Modern African Studies*, 29 (4): 669-689.

Smith, John, and Heshusius, Louis. 1986. "Closing Down the Conversation: The End of the Quantitative-Qualitative Debate among Educational Inquirers." *Educational Researcher,* 15 (1): 4-13.

Stevens, Philip and Martin Weale. 2004. "Education and Economic Growth." In *International Handbook on the Economics of Education*, eds. Geraint Johnes and Jill Johnes. Northhampton, MA: Edward Elgar.

Tan, Emrullah. 2014. (September) "Human Capital Theory: A Holistic Criticism." *Review of Educational Research*. 84 (3): 411-45.

Author contact: sklees@umd.edu

_____

You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-74/