Why inferential statistics are inappropriate for development studies and how the same data can be better used

Ballinger, Clint

6 January 2011

Why inferential statistics are inappropriate for development studies and how the same data can be better used

Clint Ballinger
Working copy, February 2011
clintballinger@cantab.net
Comments welcome

The purpose of this paper is twofold:

1) to highlight the widely ignored but fundamental problem of 'superpopulations' for the use of inferential statistics in development studies. We do not to dwell on this problem however as it has been sufficiently discussed in older papers by statisticians that social scientists have nevertheless long chosen to ignore; the interested reader can turn to those for greater detail.

2) to show that descriptive statistics both avoid the problem of superpopulations and can be a powerful tool when used correctly. A few examples are provided.

The paper ends with considerations of some reasons we think are behind the adherence to methods that are known to be inapplicable to many of the types of questions asked in development studies yet still widely practiced.

I The problem of 'superpopulations'

There is of course an enormous body of literature on problems with statistical approaches to the social sciences.[1] There is one problem, however, that is both more fundamental than many other problems and particularly salient for development studies and related fields. It nevertheless receives a curiously small amount of attention. This is the problem of superpopulations.

---

[1] Much of the best criticism of frequentist statistics comes from Bayesians and vice versa. Western and Jackman do a wonderful job of explaining why frequentist statistics suffer from apparent populations. However, they are less convincing that Bayesian methods—which are also inferential statistics—avoid the same problems. See Firebaugh 1995; see Efron 1986 for problems with Bayesian statistics more generally; see also footnote 12 in this paper.

Inferential statistics are based on taking a random sample from a larger population (e.g., survey data in the social sciences, samples from a factory for quality control in manufacturing and so on) and attempting to draw conclusions about a) the larger population from that data and b) the probability that the relations between measured variables are consistent or are artifacts of the sampling procedure.

However, in political science, economics, development studies and related fields the data often represents as complete an amount of data as can be measured from the real world (an 'apparent population'). It is not the result of a random sampling from a larger population.[2] Nevertheless, social scientists treat such data as the result of random sampling.

Because there is no source of further cases a fiction is propagated—the data is treated *as if* it were from a larger population, a 'superpopulation' where repeated realizations of the data are imagined. Imagine there *could be* more worlds with more cases and the problem is fixed.

An obvious possible objection is that this assumption makes a metaphysical commitment to indeterminism, i.e. there must be the possibility of different outcomes in the new cases or else further draws from it would not lead to any 'new' data. (Western and Jackman 1994, Berk et. al. 1995)

Crucially, however, this is not the fundamental problem. The real problem is that *even the assumption of indeterminism does not make the superpopulation fix viable.* Why? Imagine a superpopulation from which cases with different data could be drawn from. What 'draw' from this imaginary superpopulation does the real-world set of cases we have in hand represent? This is simply an unanswerable question. The current set of cases *could* be representative of the superpopulation, and it *could* be an extremely unrepresentative sample, a one in a million chance selection from it. There

---

[2] The incompleteness of the types of datasets often used in development and related studies cannot be used to defend them as some sort of 'sample'; it only represents a rather severe form of exclusion bias, hardly a plus for reliable inference.

As example, Goodchild considers this type of problem in the context of a study using data from U.S. counties and notes: 'First, one might limit analysis to a truly random sample of the roughly 3100 U.S. counties, limiting the number and spacing them far enough apart so that the assumption of independence is tenable. But this strategy requires rejecting the vast majority of the available data, widening the confidence limits of conclusions, and possibly missing some important patterns and effects.' (Goodchild 2009, 412). The same problem applies to international and other similar datasets.

is simply no way to know, not even in theory. 'Take the example of a confidence interval for a mean where we can conclude that under repeated realizations (which are acknowledged to be impossible), the interval would cover the true mean 90% of the time. *We have no way of knowing whether the current interval is one of the fortunate 90% and no possibility of further replications.*' (Western and Jackman 1994, 414, emphasis added). The metaphysical appeal to indeterminism and imaginary superpopulations (itself a questionable metaphysically-based 'fix') is simply not a fix at all for inferential statistics when dealing with the type of data used in development studies.

As Western and Jackman emphasize, 'It is simply not relevant for the problem at hand to think of observations as draws from a random process when further realizations are impossible in practice and lack meaning even as abstract propositions. In short, frequentist inference answers a question that comparative researchers are not typically asking' (Western and Jackman 1994, 413). The kinds of questions many social science researchers ask involve variation between cases each representing highly complex and unique spatial concatenations of political, economic, geographic, demographic, institutional, and cultural factors and the interdependencies between these. The problem is not one of statistics that need to be fixed. Rather, it is a problem of the misapplication of inferential statistics to non-inferential situations.

Figure 1   Visualizing the 'Superpopulation' Assumption

Manufacturing Globes in Italy. The assumption that we can treat the world as part of a 'superpopulation' is something like the idea that endless worlds roll off an assembly line, usually with the additional assumption of indeterminism - that each world is different from the others.[3]

Image by Sandro Michahelles in Povoledo 2007

Inferential statistics do not and cannot do what they are called on to do in development studies. This is all summarized and shown clearly in Freedman 1991, Western and Jackman 1994 and related statistical literature. We briefly hypothesize as to why a method that has been shown to be so substantially in error nonetheless continues to be used in the postscript. Rather than reargue the case, it is more useful to

---

[3] As with Czech history, so too the world: 'If Czech history could be repeated, we should of course find it desirable to test the other possibility each time and compare the results. Without such an experiment, all considerations of this kind remain a game of hypotheses' (Milan Kundera 1984, 223). Economist Georgescu-Roegen makes a similar observation: 'We may recall in this connection the elementary remark of C.S. Peirce that universes are not as common as peanuts. Because there is only one Western civilization, the question of whether its historical development merely follows a trajectory determined entirely by the initial condition or whether it represents a hysteretic process can be settled neither by an effective experiment nor by the analysis of observed data. *Unfortunately, the answer to this sort of question has an incalculable bearing upon our policy recommendations, especially those with a long-run target—such as the policies of economic development'* (Georgescu-Roegen, 1966, 65, emphasis added). Likewise E.O. Wilson observes that 'if ten thousand humanoid histories could be traced on ten thousand Earthlike planets, and from a comparative study of those histories empirical tests and principles evolved, historiography—the explanation of historical trends—would already be a natural science' (Wilson 1998, 9).

consider what can be done positively. An immense amount of interesting data has been developed for comparative social studies for statistical analysis. How can it be used? Shalev 1998 and 2007 shows that descriptive statistics can be used not only in place of the inferential statistical techniques commonly utilized, but that they are actually superior in expressing, supporting, rejecting, analyzing and forming new theoretical arguments.

II Why Descriptive Statistics are Useful

One way to use data developed for statistical manipulation is visually. Simple maps of data can do much that inferential statistics are imagined to do but more clearly. This is an increasingly common observation in recent years with an extensive literature so we do not discuss visual methods further.[4] Descriptive statistics themselves are a visual method to some degree, spatially arranging data in a way that helps show the relationship between various kinds of data. One of their advantages over mapping and other complex visual methods is that descriptive statistics have

---

[4] In the social science literature maps and other visual tools are seldom relied on to express or test theories (Orford et. al. 1999; surveying more recent social science literature makes clear that this trend has improved at best only slightly). Indeed remarkably, *even in geography* the use of maps has become rare (Martin 2000). Yet maps are able to convey a great deal of information with many of the benefits concerning descriptive statistics (maintaining case identity, spatial juxtaposition) in a manner that is fundamentally compatible with human cognitive skills.

A well known demonstration of the superiority of simple visualization of data versus statistical tables is provided in Tufte (1997, 38-54) concerning the 1986 space shuttle Challenger disaster. Data expressed with visual methods, ranging from tabular analysis to maps and other visualization methods allow researchers to effectively express subtle and complex meanings or *gestalts* (Cleveland 1994, 224) with the same data that might otherwise be used in regressions and other statistical methods. Indeed, one could argue that the form of presenting data in statistical arguments is highly maladapted for many of the purposes in the social sciences. As Haining notes, 'Cleveland (1993) argues that the human eye and brain are good at detecting patterns and relationships from visual evidence when that visual evidence is presented and the visual displays engineered in ways that are grounded in perceptual and cognitive psychology' (Haining 2003, 192). Proper use of descriptive statistics and maps, especially when combined, can be used to convey 'statistical reasoning' without many of the problems (Change of Support Problem [COSP], the Moveable Area Unit Problem [MAUP] and so on) associated with the type of data used in development studies.

virtually no special technological requirements. This simplicity has the additional benefit that assumptions and manipulations are more transparent to end users than those used in frequentist and Bayesian statistics. There are other significant advantages of descriptive statistics beyond simply being more accessible, however, as we hope to demonstrate.

Perhaps because they are 'old fashioned' and 'low-tech' compared to *either* visual methods or statistical analysis, descriptive statistics are often overlooked as important tools. Shalev (1998, 2007) argues, and we fully agree, that they are superior in many cases to other methods.

Descriptive statistics – a natural solution to the superpopulation problem (and many others)

It is interesting that when the problem of superpopulations is discussed, descriptive statistics are often cited as the natural or obvious solution. For example, Berk et. al. observe that when data are comprised by the whole population 'the data may be treated as a (fixed) population…Descriptive summaries of the population can be used' (Berk et. al. 1995, 425). Similarly Michael Goodchild, a leading expert in spatial statistics, more recently notes:

> the world of controlled experiments is itself the exception, being limited to those situations in which a researcher is able to identify a population, select cases from it randomly and independently, and then make inferences about the population. This ideal situation occurs so rarely in geographic research that one wonders why the associated assumptions have persisted, and indeed dominated, for so long. Why are we not content simply to describe specific parts of the heterogeneous world that we see around us [i.e. descriptive statistics], using the rigorous methods

of science, and why do we insist instead that inferences be made about some poorly conceived and nonexistent hypothetical world?   (Goodchild 2009, 414)[5]

In economics, discussing Hoover's assertion that 'Econometrics is not about measuring covering laws' (Hoover 2001, 173) Kittel argues that 'accepting this rescue of the econometric methodology drowns the very ambition of macroquantitative research.' If econometrics is not about searching for covering laws 'then the approach gives up the aim of testing theories assumed to be independent of space and time; *it becomes a highly sophisticated variant of descriptive statistics*' (Kittel 2006, 11, emphasis added). Although Kittel says this in *defense* of theory testing 'independent of space and time', we believe he has inadvertently 'hit the nail on the head': What is needed to answer the kinds of questions comparativists ask, especially the spatial questions asked about uneven development, are better and clearer (generally *less* 'sophisticated') descriptive statistics. Simple methods of presenting the right data can not only replace inappropriately used inferential statistics. They are *superior* to them for the types of questions asked in political economy and comparative development type studies.

---

[5] Spatial autocorrelation and superpopulations are both related to the statistical attempt to overcome the same problem - there is one earth and we can neither imagine new ones nor unproblematically divide the earth up to create more cases. For example, as noted in footnote 2, Goodchild considers the problem in U.S. counties, which applies equally to other extensive datasets of the type used in development studies. 'Counties are largely arbitrarily defined reporting zones, with boundaries that cut across the grain of landscape organization, and it is impossible to argue convincingly that two counties have been drawn independently from some larger population, especially when the counties are adjacent' (Goodchild 2009, 412). Goodchild says there are three options:

1) sample from the data – but in the types of studies like development studies sampling is counterproductive (see footnote 2);

2) try spatial statistics, although these, despite the best efforts of Goodchild and many others, remain plagued with numerous unresolved problems;

3) finally 'one might reject inference altogether, arguing that what U.S. counties reveal is true only of U.S. counties, and that generalization to some hypothetical and undefined population has no merit. *This last option is the most problematic, because it flies in the face of the inferential tradition, despite the straightforward and compelling arguments that support it*.' (Goodchild 2009, 412, emphasis added)

Some examples

> 'There are no important findings in sociology or criminology that are so complex that they cannot be communicated with graphs and tables that are intelligible to intelligent laymen and policy makers.' (Goertzel 2002, 16)

We present here two examples of approaching social science questions using descriptive statistics from Shalev 1998. They are slightly older, but they are used to illustrate, with as brief examples as possible, the ways in which descriptive statistics are useful. This author has used similar approaches that provide more recent, extended examples in other development related fields (Ballinger 2011a, 2011b, 2011c, 2011d, 2011e). For what this author views as exemplary use of descriptive statistics see work by economic historian Paul Bairoch on complex relations and controversial theories concerning population, development, imperialism and related topics (Bairoch 1985 and Bairoch 1993). Goertzel 2002 and Goertzel and Goertzel 2008 also present similar examples.

In a widely cited paper, Bo Rothstein (1990) studies the cross-country variation in union membership in 18 OECD countries, and why it is particularly high (over 70%) in only a few of these. He focuses on the 'Ghent system' where unions are responsible for administering unemployment insurance, which has the effect that in periods of economic crisis union membership is unlikely to decrease and may even increase. In addition to the presence/absence of the Ghent system, however, are other possible causes of high union membership, such as high left party participation in government and total potential union membership (the absolute number of employed and unemployed wage earners).

Rothstein uses multiple regression to attempt to discern which of these factors are important, and to what degree. Shalev slightly modifies[6] and replicates the regressions, with the following result (coefficients are standardized *betas*):

Union Density= .47(Ghent) + .28(Left Government) - .34(Log of Potential Membership)

---

[6] Including dropping Iceland from the analysis, giving 17 cases.

The coefficients are significant at standard levels (with Left Government only marginally so) and the adjusted R-squared is .73. The coefficient for the Ghent variable shows a substantial net difference between Ghent and non-Ghent systems of 27 percentage points.

Although this regression seemingly supports Rothstein's hypothesis, there are several problems, especially concerning the common assumption that the explanatory variables are exerting independent rather than interdependent effects. This problem is compounded because, with only 17 cases, it would be impossible to add even a few other factors into the model, which limits further use of controls or modeling of interaction effects. Shalev proposes that a multiway flowchart showing different unionization values with countries clustered by cross tabulation based on categorical versions of Rothstein's measures on other causal variables both more clearly supports Rothstein's argument as well as points out possible further avenues of research (Table 1, next page).

Table 1 divides the group of countries into small and large, and then further subdivides these by the strength of Left parties and into Ghent and non-Ghent systems. Beside each case is the percentage of union membership. If one looks at the Ghent cases, it is clear that they have a significantly higher percentage of union membership. One also notices other relationships straight away, that, for example, only small countries with medium to strong left parties have the Ghent system.

Shalev argues that this 'unsophisticated' method of presenting the data 'vindicates Rothstein's thesis much more effectively than his regression. It makes clear precisely the things he wants to demonstrate: that the Ghent effect is large and not spurious, and that it comes into play in the countries where other conditions are broadly favorable to unions' (Shalev 1998, 13).
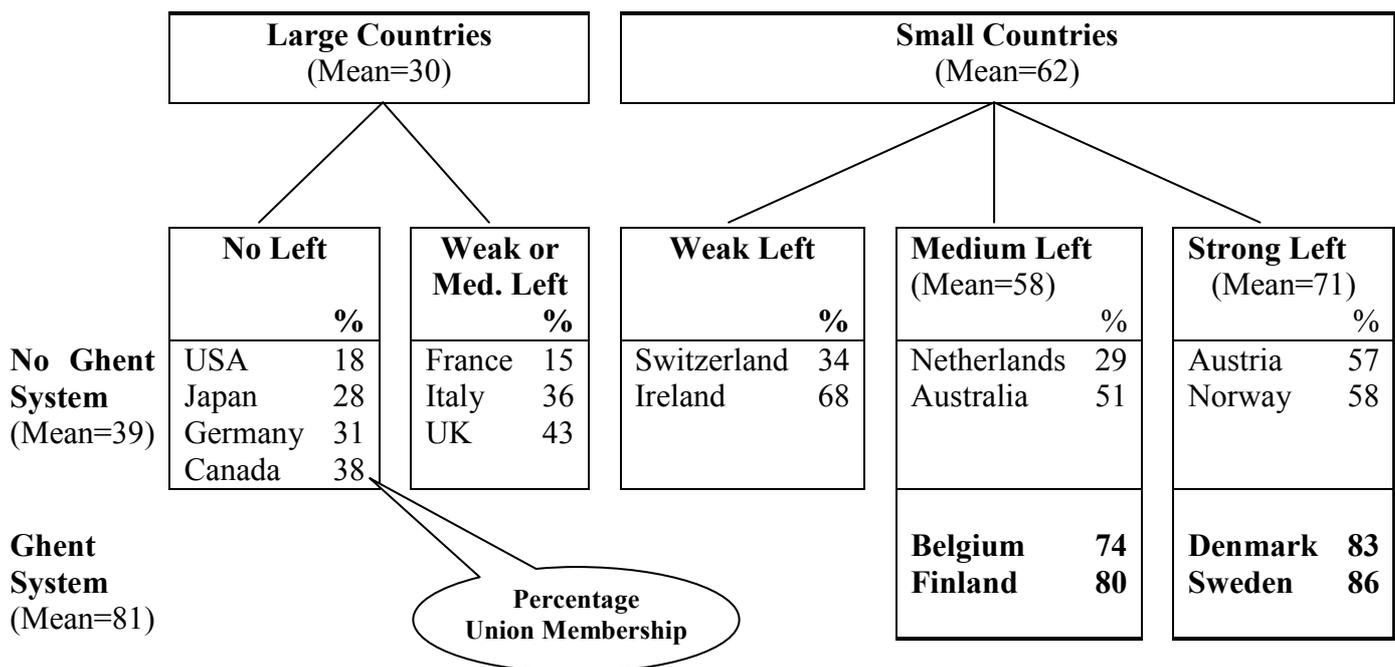
	Large Countries (Mean=30)		Small Countries (Mean=62)		
	No Left %	**Weak or Med. Left** %	**Weak Left** %	**Medium Left** (Mean=58) %	**Strong Left** (Mean=71) %
No Ghent System (Mean=39)	USA 18 Japan 28 Germany 31 Canada 38	France 15 Italy 36 UK 43	Switzerland 34 Ireland 68	Netherlands 29 Australia 51	Austria 57 Norway 58
Ghent System (Mean=81)		Percentage Union Membership		Belgium 74 Finland 80	Denmark 83 Sweden 86

Table 1 Reanalysis of Rothstein (1990) by Shalev (1998)

Equally important, this method points 'the interested researcher to the most fertile questions for selective case comparisons that might help nail down how important Ghent really is. In particular, it must be questioned whether the Ghent system alone can explain the very large differences in density between the members of two otherwise well-matched pairs of countries: Belgium and the Netherlands, and Sweden and Norway' (Shalev 1998, 13). Other factors seem likely to explain these particular cases, such as the fact that Belgian unions are present in the workplace, whereas in the Netherlands they are not. The difference between Norway and Sweden may be attributable to differences in sectors in the countries, such as differences in gender in the workplace, private sector trade and services, and percentage of white collar workers. Two other anomalies the chart suggests deserve further research are the wide gap between the United States and Canada, and the unexpectedly high union membership in Ireland.[7]

Crucial for better analysis of development data, however, is that this method makes clear where further research might be pursued, what factors drive what particular clusters of cases, and as in the pairs 'Belgium-Netherlands' and 'Sweden-

---

[7] All of these observations are by Shalev.

Norway', what additional factors might be confounding the results. Few or none of these insights are obtained from Rothstein's original multiple regression analysis. The more technically sophisticated method of multiple regression is *less* useful than Shalev's tabular flowchart using the same data for supporting Rothstein's hypothesis and pointing out where other variables are important.

Descriptive statistics make conceptual errors more obvious and analysis more clear – e.g., missing cases and spatial autocorrelation

One of the problems with statistical methods noticeable in the Rothstein example above is that they usually hide the cases behind them. This makes it difficult to see important aspects of the information, such as when cases or variables should be included that are not, the existence of spatial autocorrelation in data, and anomalies that might fruitfully lead to more refined or new theories.

For example, Ballinger 2011c analyzes Acemoglu, Johnson and Robinson's (2002) argument that the material well-being of countries shows a 'reversal of fortune' between 1500 and today, and shows how tabular analysis of their regression data makes more evident that they left out African countries and how seriously this skews their results. The inclusion of these would completely eliminate the statistical basis of their argument, adding sixteen cases, all strongly counter to their hypothesis, to an original dataset of 41 cases.

A second example used by Shalev, in addition to expressing the original authors' argument more clearly than the statistical method the author used, also demonstrates the second and third points above concerning spatial autocorrelation and anomalies. It is based on a study of the common idea in political economy that independent central banks have a strong favorable effect on economic performance. Hall and Franzese (1998) argue that while it is true that independent central banks are anti-inflationary, they can have negative effects on unemployment; specifically, in economies with uncoordinated wage bargaining, central banks might fail to respond to employment signals, resulting in higher unemployment. Using data for 18 OECD countries over the period 1955-1990, they support their argument first in a simple tabular format (without case identity), and then with the application of multiple regression with controls for economic, political and institutional variables. Although Shalev notes that their regression analysis is of very high quality, it is 'heroic' to think that 18

observations can statistically support an argument based on 9 independent variables (including interaction terms).[8]

---

[8] Hall and Franzese use pooling of data to increase their sample size. This method is not useful for the long-term questions of recent development studies where it is precisely the influence *on* cross-sectional variation *from* longitudinal variation that is of interest. Even in short-term studies Shalev and others argue that pooling as a method of increasing sample size has serious problems, especially the assumption that cross-sectional and longitudinal variation are comparable. As Shalev concludes, the 'most likely consequence of a turn to pooling is to muddy the causal waters still further' (Shalev 1998, 17).

Table 2 shows the data from Hall and Franzese divided by five categories on the coordination measure and three categories on the central bank independence measure. The upper table uses numerical data expressing unemployment percentages. The lower table shows the corresponding real cases.

Central Bank Independence	Coordination of Wage Bargaining					Mean
	0	0.25	0.5	0.75	1	
Low (<.4)		1.3 5.6			2.0	2.7
Medium	4.9 8.1	4.1	4.9	3.3		**4.5**
High (>.6)	6.1			0.9 3.1	2.2	3.7
Mean	6.3	3.8	4.9	2.8	2.0	3.9

Percentage of Unemployment

Central Bank Independence	Coordination of Wage Bargaining					N
	0	0.25	0.5	0.75	1	
Low (<.4)		ITA NZL			NOR SWE	4
Medium	IRE UKG	AUS FRA	BEL NET	DEN FIN JAP		9
High (>.6)	**CAN USA**			**SWI GER**	**OST**	5
N	4	4	2	5	3	

(Corresponding real cases)

Table 2 Reanalysis of Hall and Franzese (1998) by Shalev (1998)

One weakness in the Hall and Franzese argument the above chart makes clear is that it is in the group of countries with *medium* central bank independence that unemployment is highest. However, this might be explained by the controls Hall and Franzese use in their regression (although again, the meaningfulness of these are questionable in such a small sample size). The hypothesis that in countries with strongly independent central banks coordinated wage bargaining reduces unemployment is supported by the unemployment difference between Austria, Germany and Switzerland versus uncoordinated North American countries. However, Shalev argues that Hall and Franzese need to persuade us that 'it is indeed

coordination, rather than factors like economic openness or the political representation of labor, which explains this specific gap' (Shalev 1998, 16).

Crucially, the phrase 'North American' alerts Shalev to the clusters of Canada/United States, Belgium/Netherlands, Norway/Sweden etc. and the possibility that spatial autocorrelation might be a problem; 'each one of these sub-groups of countries constitutes a cultural, political and economic region with numerous internal similarities but which is worlds apart from the other. Regression's assumption of the independence of characteristics that cluster so strongly appears particularly misplaced in such a context' (Shalev 1998, 16).

Shalev does not use the phrases 'spatial autocorrelation' or 'Galton's problem', (perhaps indicative of the degree to which these problems are often ignored in comparative political economy[9]). The important point, however, is the ease with which spatial autocorrelation is noticed and how descriptive statistics with case identity maintained allows for data gathered for multiple regression analysis to be analyzed in a way that is sensitive to this problem, as well as to the problems of a small number of cases and the interaction of factors.[10]

Keeping cases visible also helps in recognizing patterns and anomalies that statistical methods obscure, an important consideration given the importance anomalies can play in theory generation (Köbben 1967, Rogowski 1995). For example, why is Japan clustered with Scandinavian countries, or the antipodean countries with France and Italy? These types of anomalies may reveal interesting hitherto unsuspected relations between factors.

---

[9] Spatial autocorrelation has often received very little attention in research areas (besides anthropology where 'Galton's problem' is often discussed) such as comparative political economy. Ross and Homer (1976) was one exception; Bollen et. al. (1993) in a survey of 108 large-scale comparative sociology research books and articles found that only five percent considered the problem of the non-independence of cases. In economics more recently 'there is actually very little discussion of spatial autocorrelation in most standard econometrics texts.' (Darmofal 2006, 3).

[10] Of course there are tests (Moran's I, Geary's C and others) for spatial autocorrelation in data. But with real-world data, what is the point of losing the real cases, with the rich bundle of real-world context and information each conveys, for a single numerical measure such as provided by statistical tests of autocorrelation? As Ballinger 2011a shows, even with larger datasets, in development type studies autocorrelation is easy to see, and can be examined in more nuanced and useful ways than numerical measures.

Overall, descriptive statistics do not commit fundamental statistical errors as inferential statistics do. They actually work *better* than inferential statistics, more clearly expressing and testing the types of causal relations made in comparative political economic and development research.

Conclusion

The generally ignored problem of 'super' (sometimes also called 'apparent') populations means many of the questions asked in development studies cannot be answered using inferential statistics. There are no 'fixes'—even the superpopulation 'fix' that is turned to when the problem is pointed out can not rectify what is a fundamental misapplication of statistics.

However, descriptive statistics can be utilized in far more sophisticated ways than is generally realized. Combining the methods such as those Shalev demonstrates for reanalyzing the arguments by Rothstein and by Hall and Franzese with other forms of visualization of data (footnote 4) is most appropriate for the study of uneven development. Crucially, besides fitting the type of problem better than methods based on inferential statistics this method allows for the immense archival and creative effort put into constructing relevant datasets to be leveraged for further analysis rather

than rejecting data driven approaches entirely as has happened at times in some fields.[11]

Postscript – On the persistence of unsuitable statistical methods

The problem of superpopulations is fundamental. There is no correction for what is at bottom not a shortcoming of statistical methods but a misapplication of them. Because the problem is not fixable with 'more of the same'—debates and more papers on method, it represents a threat to established practice (and practitioners). Simply pointing it out and moving on are not an option, it can only be either a) ignored completely or b) visual and descriptive statistics turned to. There is no point in paraphrasing what has already been said clearly:

---

[11] Perhaps due in part to the observation that statistical approaches were not working well (as well as due to other intellectual trends of the time, such as viewing comparative methods as wrong because of cultural relativism, and anti-positivism and anti-empiricism), in some fields not just statistics but its *data* has been rejected. Human geography provides an example. There has been widespread objection to comparative approaches based on empirical data except at very small 'intensive' scales (case studies and usually small-scale survey data). 'Analysis of large data sets has become totally passé, the object of suspicion or even derision as "empiricist" ' (Hamnett 2003, 2). Markusen notes the 'disdain for empirical work in the tradition of verification in much recent geographical writing' (Markusen 1999, 881) and Martin and Sunley that 'There is almost a disdain for in-depth empirical analysis (let alone hypothesis testing, which is dismissed as being identified with positivistic research and criticized for failing to reveal the detailed processes at work), whereas intensive, interview-based research is celebrated for its "explanatory depth." ' (Martin and Sunley 2001, 154). Reasons include objections to definitions and operationalization of culture and institutions, and a more general rejection of 'cross-cultural' comparisons by both anthropologists and geographers based on Boasian 'historical particularist' assumptions and traditions. '[I]n their righteous indignation at the frequently both grandiose and incorrect formulations of the evolutionists, the cultural relativists virtually cast out the comparative method…each society became unique. The comparison of different societies, therefore, made no sense' (Udy 1973, 255). This is unfortunate, as Graeber notes concerning anthropology (2004, 96): 'While anthropologists are, effectively, sitting on a vast archive of human experience, of social and political experiments no one else really knows about, that very body of comparative ethnography is seen as something shameful…it is treated not as the common heritage of humankind, but as our dirty little secret.' Descriptive statistics allow the 'vast archives of human experience' developed partly for statistical manipulation to be utilized in a better way.

The scientific community does not have good procedures for acknowledging the failure of a widely used research method. Methods that are entrenched in graduate programs at leading universities and published in prestigious journals tend to be perpetuated.

In 1991, David Freedman, a distinguished sociologist at the University of California at Berkeley and the author of textbooks on quantitative research methods, shook the foundations of regression modeling when he frankly stated "I do not think that regression can carry much of the burden in a causal argument. Nor do regression equations, by themselves, give much help in controlling for confounding variables" (Freedman, 1991: 292). Freedman's article provoked a number of strong reactions. Richard Berk (1991: 315) observed that Freedman's argument "will be very difficult for most quantitative sociologists to accept. It goes to the heart of their empirical enterprise and in so doing, puts entire professional careers in jeopardy." (Goertzel 2002).

This applies equally or more so to the improper use of inferential statistics in development studies and related fields that use similar data.

A less directly important reason descriptive statistics are not more widely used is that analysis of descriptive statistics is demanding, in many ways *more* so than complex inferential statistics, and 'riskier' as well as they demand clearly articulated (and thus more easily refuted) relations between factors, i.e., causal arguments or explanations. As Freedman notes, when the assumptions of statistical analysis are not met in the ways discussed here, 'Then formal statistical procedures are irrelevant, and we are reduced (or should be) to old fashioned thinking.' (Freedman 1991, 305). Western and Jackman note that when all the data is collected, rather than turn to an imagined superpopulation as justification for inferential statistics, one could simply eschew statistical inference as unnecessary and that 'this position is certainly valid, but it still commits the researcher to a substantive theory of how the data were generated…(Western and Jackman 1994, 413). To us this is a very *good* thing to commit researchers to doing. Rather than being hidden behind a veil of statistics, the clear causal language demanded by descriptive statistics commits researchers more clearly to useful arguments; furthermore, in their clarity these can more easily be supported or refuted.

Clear causal assertions are in turn attacked from two sides—as 'scientistic' or positivist on the one hand, and as not *sufficiently* 'scientific' or rigorous on the other by those who nevertheless rely on what are in truth causal arguments hidden behind the veil of statistics.

As Pearl (2003) notes, there is strong resistance to causal (vis-à-vis 'associational' or 'correlational') arguments among those favoring statistical approaches. This is due to fundamental differences between correlational versus causal assumptions, two of which are:

1. 'associational assumptions can be expressed in the familiar language of probability calculus, and thus assume an aura of scholarship and scientific respectability. Causal assumptions, as we have seen before, are deprived of that honor, and thus become immediately suspect of informal, anecdotal or metaphysical thinking.'[12]

2. Causal arguments are resisted due to 'their intimidating clarity. Assumptions about abstract properties of density functions or about conditional independencies among variables are, cognitively speaking, rather opaque, hence they tend to be forgiven, rather than debated. In contrast, assumptions about how variables cause one another are shockingly transparent, and tend therefore to invite counter-arguments and counter-hypotheses' (Pearl 2003: 288).[13]

Among economists perhaps the strangest aspect of the prevalent insistence on mathematical and econometric modeling as opposed to simpler causal language is that

---

[12] '[T]his difference becomes illuminated among Bayesians, who are accustomed to accepting untested, judgmental assumptions, and should therefore invite causal assumptions with open arms— they don't. A Bayesian is prepared to accept an expert's judgment, however esoteric and untestable, so long as the judgment is presented as a probability expression. *Bayesians turn apprehensive when that same judgment is cast in plain causal English…*' (Pearl 2003, 288, emphasis added).

[13] Pearl concludes that it is ironic that it is this intimidating clarity 'that often deters researchers from articulating assumptions in causal vocabulary. Indeed, since the bulk of scientific knowledge is organized in causal schema, scientists are incredibly creative in constructing competing alternatives to any causal hypothesis, however plausible. Statistical hypotheses in contrast, having been several levels removed from our store of knowledge, are relatively protected from such challenges, and offer therefore a safer ride toward the conclusion.' (Pearl 2003, 288-289)

the former have perhaps not been near as important to the development of economic theory as is commonly perceived. Discursive arguments rather than mathematical models were the basis for such important concepts as Coase's theory of the firm, Mundell's optimal currency area (OCA) theory, and the consideration of the market for 'lemons' in used-automobile markets which was the basis for the theory of imperfect competition by George Akerlof, all of which led to Nobel prizes in economics.

Economist Bryan Caplan lists ten of the most influential ideas of mainstream academic economics since 1949:

1. Human capital theory
2. Rational expectations macroeconomics
3. The random walk view of financial markets
4. Signaling models
5. Public choice theory
6. Natural rate models of unemployment
7. Time consistency
8. The prisoners' dilemma, coordination games, and hawk-dove games
9. The Ricardian equivalence argument for debt-neutrality
10. Contestable markets

Almost none of these ideas originated with mathematical models, but instead through observation of the real world and/or descriptive statistics, intuition, and discursive arguments. 'Out of the whole list, there are few plausible cases where mathematics was more than an afterthought: maybe idea #2, and possibly #3. Even there, intuition, not math, probably played the leading role.' (Caplan 2003)

Caplan continues: 'The contributions of econometrics to economics are similarly meager—particularly because econometrics has "crowded out" traditional qualitative economic history…When simple econometrics failed to yield universal agreement among informed economists, this merely provided the impetus for econometric theorists to supply increasingly complex estimators and other tools. Truly, this is a case of looking for car keys underneath the streetlight because it is brighter there. (Caplan 2003)

Bibliography

Bairoch, Paul. 1985/1988. *Cities and Economic Development.* Translated by Christopher Braider. Chicago: University of Chicago Press.

Bairoch, Paul. 1993. *Economics and World History: Myths and Paradoxes*. London: Harvester Wheatsheaf.

Ballinger, Clint. 2008. Initial Conditions as Exogenous Factors in Spatial Explanation. Ph.D. Dissertation, Department of Geography, University of Cambridge. Available online at:
http://www.scribd.com/full/46617327?access_key=key-1vbhnscdlxd5fdu49s74

Ballinger, Clint. 2011a. Comparative Economics in a World Divided: Spatial Autocorrelation and World Regions.
http://philosophyofscience.webstarts.com/working_papers.html

Ballinger, Clint. 2011b. Acemoglu, Johnson, and Robinson: Natural Experiments or Geographic Theories of Development?
http://philosophyofscience.webstarts.com/working_papers.html

Ballinger, Clint. 2011c. Two Fatal Flaws in the 'Reversal of Fortune' (Acemoglu, Johnson, and Robinson 2002) Argument.
http://philosophyofscience.webstarts.com/working_papers.html

Ballinger, Clint. 2011d. Mercantilism and Uneven Development.
http://philosophyofscience.webstarts.com/working_papers.html

Ballinger, Clint. 2011e. World Regions, International Development and Cluster Analysis.
http://philosophyofscience.webstarts.com/working_papers.html

Berk, Richard A., Bruce Western and Robert E. Weiss. 1995. Statistical Inference for Apparent Populations. *Sociological Methodology* 25: 421-458.

Berk, Richard A., Bruce Western and Robert E. Weiss. 1995. Reply to Bollen, Firebaugh, and Rubin. *Sociological Methodology* 25: 481-485.

Berk, Richard A. and David Freedman. 2001. 'Statistical Assumptions as Empirical Commitments'. UCLA Department of Statistics Papers, Paper 2001080101.

Bertin, J. 1983. *Semiology of Graphics*. Madison: University of Wisconsin Press.

Bollen, Kenneth A, Barbara Entwisle and Arthur S Alderson. 1993. Macrocomparative research methods. *Annual Review of Sociology* 19: 321-51.

Caplan, Bryan. 2003. 'Why I am not an Austrian Economist'. Unpublished manuscript, Department of Economics, George Mason University. Accessed January 2011.
http://www.econfaculty/gmu.edu/bcaplan/whyaust.htm

Cleveland, William S. 1993. *Visualizing Data*., Summit, NJ: Hobart Press.

Cleveland, William S. 1994. *The Elements of Graphing Data*. Summit, NJ: Hobart Press.

Darmofal, David. 2006. 'Spatial Econometrics and Political Science'. Paper presented at the Annual Meeting of the Southern Political Science Association, Atlanta Georgia, January 5-7, 2006.

Efron, B. 1986. Why isn't everyone a Bayesian? *The American Statistician* 40: 1-11.

Firebaugh, Glenn. 1995. Will Bayesian Inference Help? A Skeptical View. *Sociological Methodology* 25: 469-472.

Freedman, David. A. 1991a. Statistical Models and Shoe Leather. *Sociological Methodology* 21: 291-313.

Freedman, David A. 1991b. A Rejoinder to Berk, Blalock, and Mason. *Sociological Methodology* 21: 353-358.

Georgescu-Roegen, Nicholas. 1966. *Analytical Economics: Issues and Problems*. Cambridge, Mass: Harvard University Press.

Goertzel, Ted. 2002. 'Econometric Modeling as Junk Science' unpublished manuscript, Department of Sociology, Rutgers University. (A simplified version without graphs is published as 'Myths of Murder and Multiple Regression', *The Skeptical Inquirer*. 26(1): 19-23; the graphs, however, are important.)

Goertzel, Ted and Benjamin Goertzel, 2008. Capital Punishment and Homicide Rates: Sociological Realities and Econometric Distortions. *Critical Sociology* 34(2). 239-254.

Goodchild, Michael F. 2009. What Problem? Spatial autocorrelation and geographic information science. *Geographical Analysis* 41(4): 411-417.

Graeber, David. 2004. *Fragments of an Anarchist Anthropology*. Chicago: Prickly Paradigm Press.

Hall, Peter A. and Robert J. Franzese. 1998. Mixed Signals: Central Bank Independence, Coordinated Wage Bargaining, and European Monetary Union. *International Organization* 52: 505-535.

Haining, Robert. 2003. *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press.

Hamnett, Chris. 2003. Contemporary Human Geography: Fiddling While Rome Burns? *Geoforum* 34: 1-3.

Hoover, Kevin D. 2001. *Causality in Macroeconomics*. Cambridge: Cambridge University Press.

Kittel, Bernhard. 2006. A Crazy Methodology? On the Limits of Macro-Quantitative Social Science Research. *International Sociology* 21: 647-677.

Köbben, A. J. F. 1967. Why Exceptions? The Logic of Cross-Cultural Analysis. *Current Anthropology* 8(1-2): 3-34.

Kundera, Milan. 1984. *The Unbearable Lightness of Being*. London: Faber and Faber.

Markusen, Anne. 1999. Fuzzy Concepts, Scanty Evidence, Policy Distance: The Case for Rigour and Policy Relevance in Critical Regional Studies. *Regional Studies* 33: 869-884.

Martin, R. L. 2000. Editorial: In memory of maps. *Transactions of the Institute of British Geographers* 25: 3-5.

Martin, Ron and Peter Sunley. 2001. Rethinking the 'Economic' in Economic Geography: Broadening Our Vision or Losing Our Focus? *Antipode* 33: 149-161.

Pearl, Judea. 2003. Statistics and Causal Inference: A Review. *Sociedad de Estadística e Investigación Operativa (Test)* 12(2): 281–345.

Orford, Scott, Richard Harris, and Daniel Dorling. 1999. Geography: Information Visualization in the Social Sciences, A State-of-the-Art Review. *Social Science Computer Review* 17: 289-304.

Povoledo, Elisabetta. 2007. In Italy, Creating Worlds Takes Precision, Yes, and Politics. *The New York Times*. August 15.

Rogowski, Ronald. 1995. The Role of Theory and Anomaly in Social Science Inference. *The American Political Science Review* 89:467-470.

Rothstein, Bo. 1990. 'Labour Market Institutions and Working Class Strength', in *Structuring Politics: Historical Institutionalism in Comparative Analysis,* Steinmo, S., K. Thelen and F. Longsteth, eds. Cambridge: Cambridge University Press.

Ross, Mark Howard and Elizabeth Homer. 1976. Galton's Problem in Cross-National Research. *World Politics* 29(1): 1-28.

Shalev, Michael. 1998. 'Limits of and Alternatives to Multiple Regression in Macro-Comparative Research', presented at the 2nd conference on The Welfare State at the Crossroads, Stockholm, June 12-14.

Shalev, Michael. 2007. Limits and Alternatives to Multiple Regression in Comparative Research. *Comparative Social Research*. 24: 267-314.

Tufte, Edward R. 1997. *Visual Explanations: Images and Quantities, Evidence and Narrative.* Cheshire, Connecticut: Graphics Press.

Udy, Stanley H, Jr. 1973. Cross-Cultural Analysis: Methods and Scope. *Annual Review of Anthropology* 2: 253-270.

Western, Bruce and Simon Jackman. 1994. Bayesian Inference for Comparative Research. *The American Political Science Review* 88(2): 412-423.

Wilson, Edward O. 1998. *Consilience: The Unity of Knowledge.* London: Little, Brown and Company.