UNIVERSITY OF
BATH

University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Graham Brown: The Limits of Linearity: A modest defence of the General Linear Model and of its critics

07 December 2014

The social world is not linear, but linear methods provide powerful and tractable forms of analysis for complex phenomena. Using linear forms of analysis on social phenomena requires simplifying assumptions and the analysis hence only holds insofar as those assumptions can be defended. In this, the social sciences differ little from natural sciences. The non-linear equations that describe the swing of a pendulum can be accurately approximated and solved under 'small angle' conditions.

Approximating a non-linear world with linear systems and linear forms of analysis is not unreasonable, but require careful attention to the limits of analysis and inference. In many areas of social science, linear analysis has become a form of Kuhnian normal science, in which anomalies and problems are 'explained away' (e.g. heteroscadisticity-robust regressions, see below) or, if still intractable, put into a pile entitled 'To Deal With – Later'. Occasionally, scholars such as Andrew Abbott (2001) or Bent Flyvbjerg(2001) have taken a look at this pile and suggested that it is high enough to warrant paradigmatic reconsideration, but such efforts have often met with robust and angry denouncement from the methodological mainstream (Laitin, 2003; Stolzenberg, 2003).

Abbott's critique of the GLM has itself been subject to severe criticism. Notably, Stolzenberg (2003) delivers a trenchant attack that derides Abbott's lack of statistical sophistication and his mischaracterisation (in Stolzenberg's view) of quantitative sociologists as subscribing to a 'wooden-headed' ontology. There are two separate issues at play here. The first is whether modern statistical methods are indeed linear in their modelling of reality; the second is whether a commitment to these methods entails a commitment to a more generalized linear view of social reality.

Stolzenberg and others are at pains to point out that modern statistical techniques are more than capable of handling non-linear relationships between variables; much of Stolzenberg's critique is dedicated to detailing different ways in which this can be dealt with statistically, and the subsequent decade has seen further development of these technique.   True, Stolzenberg admits, the linear model requires linearity in its *parameters* but asserts parenthetically that this is an 'entirely technical matter'. We will return, *passim*, to quite how parenthetical this observation should be.

The typical rejoinder to Abbott and Flyvbjerg's style of critique is that quantitative social scientists aren't nearly as naïve about statistical methods and their limitations as Abbott, Flyvbjerg and their ilk assert. This view is put strongly by Stolzenberg, and also by Laitin (2003: p. 163) who characterizes Flyvbjerg and other 'perestroikans' as exhibiting an 'abhorrence of all things mathematical… [that] reveals a fear of modern'.[1] Laitin, in turn, promotes a 'tripartite methodology', in which statistical analysis is complemented by formal modeling and 'narrative' techniques. While in Laitin's overall model, 'narrative' plays at best a supporting role – providing 'plausibility test' for formal models; identifying causal mechanisms; and, 'plant[ing] the seeds for future specification of variable' – he does acknowledge that in

---

[1] The term 'perestroikan' refers to the disciplinary dispute within American political science ignited by a widely circulated attack on a perceived methodological intolerance within the American Political Science Association signed pseudonymously by 'Mr Peterstroika'.

circumstances where statistical methods and formally modeling are unable to achieve significance, 'narrative would need to stand alone' (pp. 178-9).

If we accept Laitin's view then we have reason to be optimistic about the future of linear analysis as the sophistication of econometric techniques and the computer power to process these newer techniques progress rapidly. Two decades ago, Abbott acknowledged that a range of newer methods, notably event analysis and other epidemiological and demographic techniques, deal precisely with some of his concerns, although viewed this as an insufficient rejoinder to the underlying problem. Further statistical advances in the decades since Abbott published his critique have extended even further the sophistication of econometric methods, including multi-level models that allow for hierarchical nesting of observations with higher level fixed effects (Rabe-Hesketh and Skrondal, 2008) and advances to selection bias two-stage models to include interaction effects (Brown and Mergoupis, 2011). While these advances extend the range of contexts in which linear methods might be appropriately applied, however, they do not overcome the underlying problem identified by Abbott.

But should we be so optimistic about the ability of ever-more advanced statistical methods to deal with an ever-wider range of contexts? It is notable that while Laitin acknowledges, at least in theory, that there are limits to the applicability of statistical methods, this is for him conceived of primarily in terms of *statistical significance* rather than epistemological appropriateness. He notes that the failure to pass statistical tests may often be attributed to improper specification through models including 'interactions, non-linearities, and so on'. Two points are worth making here. First, this is another instance of the approach – misguided in our view – that sees epistemological non-linearity as an appropriate response to ontological non-linearity. Secondly, his view of statistical significance as the ultimate arbiter of the applicability of statistical methods falls foul of Ziliak and McCloskey's (2008) trenchant critique of the 'cult of statistical significance'.

Even if linearly-sympathetic methodologists such as Laitin and others are careful to identify in principle the limitations of linear methods in social science, the accusation thrown by critics is not necessarily that these limitations are not recognized in principle, but that they are not recognized in practice: the industry of quantitative social science continues, they assert, in ignoring these limits and applying their methods without critical attention to their appropriateness to the issue at hand. In a contribution to the fiftieth anniversary edition of the highly ranked and highly quantitative *Journal of Peace Research*, Philip Schrodt (2014) lists a 'monocultural' obsession with linear methods in political science, describing a depressingly not-too-exaggerated *homo signficantus* who 'year after year, grinds out articles by downloading a dataset, knocks out a paper or two over the weekend by running a variety of specifications until – as will invariably occur – some modestly interesting set of significant coefficients is found'. Twenty years after Stolzenberg derided Abbott's supposed caricature of 'wooden-headed' allegiance to linear methods in social science, Schrodt sees little improvement.

Parenthetically, we could note that while Stolzenberg may be right that most practicing quantitative social scientists are not so simplistic about the ontology of social life as Abbott asserts, the percolation of their research and findings into policy and practice may often see such nuance and care stripped out in favour of simple (and simplistic) linear 'findings' that entail particular policy interventions. In the international development industry, for instance, Rihani (2002) observes a dominant linear paradigm among development practitioners, but one that is an 'assemblage of implicit assumptions that exert their influence in subtle ways' rather than an explicit ontological commitment.

At a more ontological level, Goldthorpe (2000) is concerned that this view mistakenly credits variable, or attributes, with causal efficacy. Actors and their (social) actions, Goldthorpe

insists, are the main agents and vehicles for causal effect through 'generative mechanisms'. A variable-based view of causality, from this perspective, does not provide us with any understanding of cause 'deeper' than the data themselves. Goldthorpe, of course, is a strong advocate of the use of regression analysis of large datasets in the social sciences; his concern is not to discredit quantitative social science but to put it on a surer ontological footing.

Goldthorpe's approach turns the central role of regularity in analysis on its head.

The Limits of Linearity in Practice

We can sum up the position we have outlined thus far as follows:

1. The social world is not linear but linear methods provide a powerful and appropriate approximation of non-linear reality under certain conditions;
2. Recent (and, by extension, future) statistical advances expand the range of contexts under which linear forms of analysis might be appropriate but do not (and cannot) negate the broader epistemological point;
3. Misguided caricatures on both parts aside, statements (1) and (2) are largely accepted in principle by both critics and defenders of the methodological mainstream;
4. What remains contentious is how far, in practice, mainstream quantitative social sciences do pay careful critical attention to the limits of linearity;
5. In part, the reason for (4) is because of the lack of methodological clarity over where the limits of linear analysis lie in social sciences.

The remainder of this paper is directed towards (5). We lay out three conditions that should be uncontroversial for limiting the contexts in which linear analysis may be appropriate and discuss ways in which we might pragmatically ascertain whether particular investigations conform to these conditions or not. We should note, however, that we do not see these three conditions as the end of the story; there may be other conditions for linear analysis of social reality that we have not addressed: our conditions are necessary for linear analysis, but not necessarily sufficient. Crucially, we want to argue that many of these techniques are available *within* the quantitative repertoire on methods and techniques, but that the methods available are often employed to 'explain way' non-linearity rather than to delimit linear analysis. Ours is hence not an attack on quantitative epistemology from 'outside', but a contribution to refining quantitative practice from the 'inside'.

The three conditions we identify are:

A. Homogeneity of effects, which can be broken down into homogeneity of entities and homogeneity of parameters;
B. Insignificantly interacting entities;
C. Stable and closed context.

A. Homogeneity of Effects

It is an obvious but important point that econometric methods require observations: rows of data. These observations are characterized by different levels of attributes (variables): the data columns. Observations might have other unobserved (and potentially unobservable) characteristics, which are captured under the individual error term. But beyond this, the entities that constitute our observations must be homogenous; they must be of the same 'type'. Put in Aristotlean terms, statistical observations may vary in their *attributes* but must be of the same *substance*. This assumption has been the subject of a range of criticisms from quantitatively- and qualitatively-inclined scholars alike. Andrew Abbott (1988) characterizes and rejects this as a view in which 'entities are fixed; attributes can change'. Entities, according to Abbott, can and do vary and mutate over time with important consequences.

When working with individual-level microdata, the assumption of entity homogeneity and stability may seem plausible enough; as long as we capture sufficient attributes – age, gender, education, and so forth – on the individual level, what remains is the human 'substance' that is indistinguishable across individuals. Statistically, there may be individual perturbations, but these are randomly distributed. But what if we are dealing with more complex entities like companies, or countries. Is it really plausible to assert that we can ever claim to characterize countries as of homogenous 'substance', no matter how much information we collect on their different 'attributes'? What are the limits of what we define as a country? We might take UN membership as a defining feature, but this rules out Taiwan, the Occupied Palestinian Territories, Western Sahara, Kosovo, to name but a few political entities with varying degrees of autonomous statehood. Moreover – and we return to this issue in section B below – the *reason* these entities are of disputed 'countryness' is itself heavily political; the status of Taiwan as a non-member of the UN is due to the contending geographical claims of Taiwan and the People's Republic of China and the relative geopolitical strength of the latter.

We might accept these problems but suggest that they are marginal; relative vagueness at the edges of our definition of entities does not affect the underlying relationships we seek to identify and test with linear methods. We might even test for them by running alternative models with different definitions and comparing our estimates. For instance, in considering the impact of ethnic polarization on the duration of civil wars, Montalvo and Reynal-Querol (2007) note the range of different definitions of civil war extant in the quantitative literature and test their results on two different definitions that are characterized by different fatality thresholds. In our example here, Montalvo and Reynal-Querol claim to find consistent results between the two definitions but, in doing so, they fall foul of Ziliak and McCloskey's critique of significance obsession. The two models do indeed show similar levels of statistical significance, but markedly different coefficient sizes: the coefficient on ethnic polarization in their most extensive model varies by a factor of four between the two definitions. Short of confirming the 'robustness' of their findings, we would suggest that this points towards the second part of the homogeneity problem: homogeneity of parameter effects.

In this respect, Brock and Durlauf (2001a; 2001b) provide a useful contribution towards thinking about the role of homogeneity in ascertaining the limits of linear methods. Their concerns are hard to ignore, coming from well-established mainstream economists, published under the auspices of the World Bank, and directed in part at the heart of development economics – the economic growth literature.

Fundamental to their critique are the notions of exchangeability and parameter heterogeneity. In relation to assumptions of homogeneity of entities in their particular domain of analysis – economic growth models - they assert that it is simply implausible to believe that the determinants of economic growth are identical across countries: 'Does it really make sense to believe that a change in the level of a civil liberties index has the same effect on growth in the United States as in the Russian Federation?' In the spirit outlined above, they do not deny that statistical advances – in this case, panel data fixed effects – mitigates some particular aspects, but such advances do not address the 'more general question'. Again, they do not want us to throw the inferential baby out with the statistical bathwater, but they believe that the assumption of parameter homogeneity underlying linear statistical methods 'is particularly inappropriate in studying complex heterogenous objects, such as countries'.

The assumption of parameter homogeneity is, for Brock and Durlauf, one manifestation of a broader problem of statistical inference – the assumption of exchangeability. Brock and Durlauf provide a formal definition and discussion of exchangeability in mathematical terms, but the idea is powerfully simple. The requirement of exchangeability is that once we have held for the explanatory factors we include in our analysis, 'no basis exists for distinguishing

the probabilities of various permutations of residual components in country-level growth'. If we have an extensive set of regressors that significantly explain economic growth, we should have no *a priori* reason for expecting the error on a particular observation to go one way or the other, to be small or large. Yet on any set of historical growth regressions, most economists would expect Japan to be a positive outlier. Japan is not exchangeable with other countries.

There is a clear Bayesian dimension to this logic: the inferences we draw from our growth regressions must be compared against our prior knowledge of, for instance, Japan as a country that has experienced exceptional growth or, as Brock and Durlauf also discuss in relation to Easterly and Levine's (1996) account of ethnic heterogeneity and growth retardation, much of Sub-Saharan Africa as likely negative outliers.

Brock and Durlauf do not see themselves as delivering a knock-out punch to linear analysis but rather as providing an avenue for new and fruitful research. Where the standard errors of our models conform to our prior expectation and hence violate the exchangeability requirement, this is not grounds to reject our inferences but is rather grounds for further enquiry. Hence, for instance, in replicating Easterly and Levine, Brock and Durlauf find that 'the operation of ethnic heterogeneity on growth is different in Africa, not just the levels of ethnic heterogeneity'. But this does not lead them to reject the finding, but to 'suggest a direction along which to extend their research. Our results illustrate how additional insights can be obtained by explicitly controlling for model uncertainty'. We can draw an instructive parallel here with the approach of Montalvo and Reynal-Querol above. Paying little attention to the differing coefficient sizes in their models, Montalvo and Reynal-Querol take the consistent statistical significance as evidence of the 'robustness' of their analysis and hence, implicitly, of the linearity of underlying effects across all cases. Looked at again through Brock and Durlauf's lenses, however, their findings suggest an alternative interpretation: that there *is* parameter heterogeneity; that by implication there *is not* a consistent linear effect across all cases; and that, therefore, more work remains to be done. Again, this is not throwing out the usefulness of linear analysis, but is showing how careful attention to the homogeneity of entities and parameters can both usefully delimit the extent of its applicability and point to new and interesting avenues of research.

Brock and Durlauf propose a methodology that allows us to explore the limits of linearity 'from within'. A second statistical phenomenon that helps us do this is heteroscedasticity. As we shall see, however, treatments of heteroscedasticity are more often than not geared toward 'explaining away' rather than embracing its potential to aid carefully delimited linear analysis.

Heteroscadisticity can be understood as a particular type of heterogeneity in parameter estimates, where the size of the individual error term is correlated with one or more predictor variable(see, e.g., Phillips and Xu, 2005). Typically, however, such attention is tuned towards finding for efficient estimators or 'correcting' for heteroscedastic variation in standard errors. While this may be appropriate in certain contexts, in other contexts the presence of heteroscedasticity may be taken precisely as an indicator of the limits of linearity.

Consider a standard textbook example of heteroscedasticity: the relationship between income and consumption. At lower levels of income, the relationship between income and consumption is typically quite close (standard errors are small) as people on low income have little disposable income to save (so consumption does not fall far below income) and little access to credit (so consumption does not exceed income substantially). At higher levels of income, however, the standard errors are much high – individuals with higher income exhibit much greater variation in their saving and borrowing behaviour.

In the standard econometric manuals (see, e.g., Wooldridge, 2013: Ch.8), heteroscedasticity is treated as a 'technical' problem that can (and should) be 'corrected' for either through

manipulation of the data (through, for instance, using a log transformation) or, if that fails, using heteroscedasticity-robust statistics. Because, from this perspective, the issue at stake is the distribution of standard errors rather than the coefficient, 'correcting' for heteroscedasticity does not change the point prediction for an observation with an unknown left-hand side, but it does affect the confidence interval. In our textbook example, correcting for heteroscedasticity would generate the same predicted consumption based on an individual's income, but would have a wider confidence interval (potentially beyond the normal range of statistical significance).

Is this reasonable? On the one hand, it might be argued that heteroscedasticity-robust standard errors give us a more plausible prediction (or, to be more precise, predicted range) of consumption at higher levels of income than would be implied by OLS estimation. But by the same token, it is giving us a *less* precise prediction of consumption at lower levels of income than we might reasonably expect given the data. If we were to partition the data – albeit at some arbitrary point – we could plausibly find a strong, significant and homoscedastic relationship between income and consumption in the lower partition and a weak or even statistically absent relationship in the upper partition.

The point here, necessarily dusted with cautionary scare quotes, is that 'standard' procedures for 'correcting' for heteroscedasticity appear to be primarily motivated by a desire to 'deal' with what is perceived as a 'technical problem' – the parallel with Stolzenberg's language is telling. Our assertion is that heteroscedasticity may, rather, tell us something more fundamental about our data and the limits of linear inference. The former behaviour is consistent with an epistemology-dependent ontology; the latter position more critical. Put in more prosaic language, heteroscedasticity in the education-income relationship tells us that there is something fundamentally different going on in the determinants of income at higher levels of education than at lower levels. Why, then, would we want to 'correct' for this just to maintain the illusion of a consistent linear relationship across the entire domain?

B. Insignificantly-Interacting Entities

Let us return to our basic description of linear analysis: a set of homogenous entities (observations) characterized by different attributes and consistent causal relationships between dependent and independent attributes. One of the fundamental assumptions that undergird linear statistical techniques is the independence of observations. One implication of this is that the relationship between independent and dependent variables acts separately but consistently on the different observations and that effect on one observation is not affected by the effect on another observation. Variable might interact; observations do not.

Clearly, at the most basic level, this is not a great description of the social world. Individuals, companies, countries do not operate in a vacuum, isolated from each other. Moreover, we might note that this assumption sits at odds with much formal modeling in economics and, increasingly, other social sciences that seek to model the behaviour of interacting agents relative to their position within a particular system. This interaction might be direct – as in the classic principal-agent dilemma – or indirect – as in the account of urban segregation dynamics in Schelling's model.

But we might reasonably assert that at the macroscopic level, these influences counteract each other randomly, allowing us to identify clear signals with linear analysis; this would be the line taken by Goldthorpe, for instance. But how far can we push this assumption?

Let us return to the education-income relationship. Across an entire economy, we might reasonably assert such a simplifying assumption. But within a particular economy …

There is a clear link here with the issue of homogeneity of entities. When we cannot assume homogenous entities – where, in Brock and Durlauf's terminology, we lack 'exchangeability' –

this is often precisely because the actions of one or more entities are likely to have a significant effect on other entities. Consider, for instance, the extensive quantitative literature on the determinants of democratization (e.g. Hegre et al. 2001). Using historical datasets, this literature has sought to identify the geopolitical and economic characteristics of countries that tend to lead to democratization. But these countries do not interact. Is this plausible?

The Arab Spring saw a wave of regime change – some permanent, some reversed, some wholly unfinished – across the Middle East and North Africa. Statistically, many of these countries were negative outliers – they were ripe for democratization. In this sense, the Arab Spring could be seen as conforming to a linear account of democratization. But the timing of the revolutions can hardly be taken as independent or coincidental. Likewise, the fall of the Berlin Wall had clear and direct demonstrative effects on other authoritarian regimes in Eastern Europe. Moreover, all of these transitions were enabled by Gorbachev's repudiation of the Brezhnev Doctrine as part of the partial democratization of the Soviet Union under his leadership. Indeed, the wider literature on democratization is concerned with the 'waves' with which democratization has occurred (see, e.g. Huntington 1986) – phenomena that cannot be explained with simple linear models.

As above, we can note that there are plenty of statistical techniques that allow us to 'correct' for this to some extent, notably spatial autoregression techniques to hold for geographical 'neighbourhood effects' (see, e.g. Brinks and Coppedge 2006; Ward and Gleditsch 2008) and for more abstract 'spatial' relationships such as trade connections (see Beck, Gleditsch, and Beardsley 2006). But while these might account to some degree for spatial and temporal 'clustering', they retain the assumption that interactions between individual observations are smoothed out in a linear way.

Consider, for instance, the effect of 'special cases' in democratization processes. Spatial clustering may appear to account for the wave of democratization in Eastern Europe at the end of the Cold War period, but these were predicated, quite specifically, on the actions of the Soviet Union *qua* the Soviet Union, not as an indistinguishable (but powerful) 'neighbour'. Thinking counterfactually, had Russia taken a path towards more consolidated and competitive democracy in the post-Soviet era than it did, it seems implausible to assert that this would not have had dramatic effects on the other post-Soviet states. But could one make the same claim about Belarus? Again, one might seek to hold for this by adjusting autoregressive geographical weights to hold for some measure of regional power or 'hegemony'. But this increasingly smacks not so much of 'theory-saving' as 'method-saving'. The point here is Brock and Durlauf redux: in accounting for democratization in Eastern Europe, the Soviet Union lacks 'exchangeability' with other observations, even in geographically-weighted analysis. The implications of violation of this condition, however, are far stronger when we consider interacting agents than when we consider parameter homogeneity on its own. Brock and Durlauf show how we can begin to delimit an appropriate realm for linear analysis by excluding (or partitioning) cases where residuals are systematically and *ex ante* predictably extreme. But they if we have reason to believe that the interaction of those 'unexchangeable' cases with the remaining cases affects outcomes in the remaining cases, exclusion or partitioning may no longer be sufficient.

Moreover, even in the absence of 'special cases' that interact significantly with other observations and skew their behaviour accordingly, there are other conditions under which interacting agents can affect the linearity of outcomes in ways that are not so identifiable *ex ante*. Evolutionary modeling and complexity literature points to the importance of runaway 'feedback loops' that can produce dramatic macroscopic effects from the interactions of apparently 'unspecial' cases. The rise of ethnic Chinese wealth in much of Southeast Asia and ethnic Lebanese wealth in much of West Africa are examples of such processes. Ex ante, there would have been no particular reason to pick these groups out as separate from other groups

in their likelihood to gain the dominant positions in the local economy that they have. It was the particular nature of the interactions *between* Chinese merchants in Southeast Asia that enabled them to gain the economic advantage that they did (see, e.g. Tagliacozzo and Chang 2011).

We have seen, then, that the assumptions of linear analysis are problematic in domains where we can reasonably expect interactions between observations to have significant impacts upon individual outcomes. Where these interactions are random noise – in the sense that they counteract each other at the macroscopic level – linear analysis of the system may still be appropriate and useful. But we have identified two conditions where this seems more problematic: where we have good reason to believe that 'special cases' determine the outcome of other cases in a non-symmetrical way; and, where feedback loops among 'non-special cases' may generate runaway macroscopic effects. The former of these is clearly linked to the first of our conditions, that of homogeneity of parameter effects – special cases violate this. The latter of these links to our third condition, that of a stable and consistent context. It is to this that we now turn.

C. Stable and Closed Context

[incomplete]

REFERENCES

Abbott, A. (1988).'Transcending General Linear Reality' *Sociological Theory,* 6 (2): 169-186.

Abbott, A. (2001). *Time Matters: On Theory and Method.* Chicago: University of Chicago Press.

Brown, G. K. and T. Mergoupis (2011).'Treatment interactions with non-experimental data in Stata' *Stata Journal,* 11 (4): 545-555.

Flyvbjerg, B. (2001). *Making Social Science Matter: Why Social Inquiry Fails and How It Can Succeed Again.* Cambridge: Cambridge University Press.

Goldthorpe, J. H. (2000). *On Sociology: Numbers, Narratives, and the Integration of Research and Theory.* Oxford: Oxford University Press.

Laitin, D. D. (2003).'The Perestroikan challenge to social science' *Politics and Society,* 31 (1): 163-184.

Phillips, P. C. B. and K.-L. Xu (2005).'Inference in autoregression under heteroskedasticity' *Journal of Time Series Analysis,* 27 (2): 289-308.

Rabe-Hesketh, S. and A. Skrondal (2008). *Multilevel and Longitudinal Modeling Using Stata.* College Station, TX: Stata Press.

Rihani, S. (2002). *Complex Systems Theory and Development Practice.* London: Zed Books.

Schrodt, P. A. (2014).'Seven deadly sins of contemporary quantitative political analysis' *Journal of Peace Research,* 51 (2): 287-300.

Stolzenberg, R. M. (2003).'Book Review: Time Matters: On Theory and Method. By Andrew Abbott' *Sociological Methods and Research,* 31 (3): 420-427.

Stolzenberg, R. M. (2003).'Book review: Time Matters: On Theory and Method. By Andrew Abbott' *Sociological Methods & Research,* 31 (3): 420-427.

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach (5th edition).* South Western College.

Ziliak, S. T. and D. N. McCloskey (2008). *The Cult of Statistical Signfiicance: How the Standard Error Costs Us Jobs, Justice, and Lives.* Ann Arbor: University of Michigan Press.