Rethinking economics: Logical gaps – theory to empirical[1]

By Stuart Birks
Massey University
Palmerston North
New Zealand
k.s.birks@massey.ac.nz

Abstract

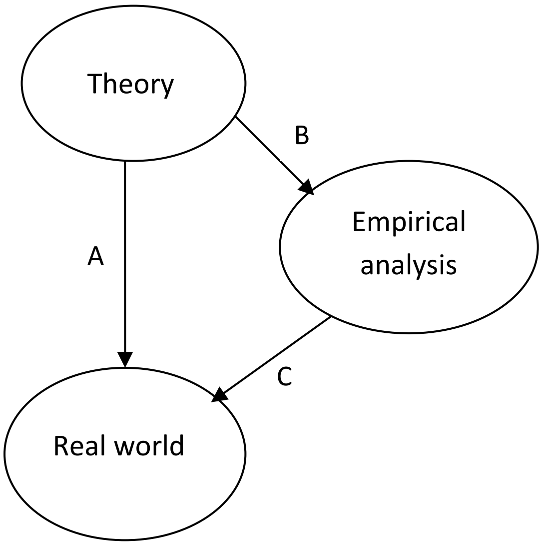
Economists have focused on the development of theory, with a heavy focus on internal logical consistency. There has also been much effort devoted to the development of suitable empirical techniques, in particular econometrics. Less attention has been given to the requirements for applying the latter, econometrics, to analyse relationships specified in the former, theory. This paper considers some aspects of this link. Two broad aspects are addressed, namely data issues and functional specification. Three data issues are discussed, those of aggregation, the use of proxy variables, and the nature and use of index variables. These aspects of functional relationships are then discussed. They are the use of control variables, the nature of causality, and the need for structural stability. The overall implications are that many problems are assumed away in empirical economic analysis. These can fundamentally distort the findings and hence provide misleading information for policy makers.

[1] Thanks are due to staff of the University of the West of England for helpful comments while visiting on sabbatical, with special thanks to Don Webber for his suggestions.

1. Introduction

Econometrics is currently the primary applied economics research technique taught to students to the extent that "research methods" courses are frequently simply econometrics courses with another name. Numerous textbooks for economists focus on econometrics, and economists' research techniques have generally focused heavily on quantitative methods of analysis. In any such analysis, there are steps to be taken before the quantitative analysis, and further steps once the quantitative results have been found. These aspects generally receive scant attention in the textbooks. However, they are important determinants of the practical value of the results of the analysis. Without due consideration the value of conclusions on practical application may be more rhetorical than actual, as in 'modes of argumentation' described by Dunn (2004). This paper draws on a three path structure as in Figure 1 and described in (Birks, 2012). That paper covered Path A, theory to the real world. This paper considers Path B the transition from theory to empirical formulation, and a third paper covers Path C, steps from empirical results to application to policy.

Figure 1: Logical errors, Types A, B and C



Consider the link between a theoretical formulation and an empirical formulation. Meghnad Desai (1981, p. 96) has made the point that the result of a statistical test of a theory depends on the validity of both the theory and the test. This paper considers whether formulations used for testing theories are likely to result in reliable tests. In particular, it focuses on i) data problems, and ii) problems of functional form.

This is the fifth of six Bristol Business School Economics Papers by Stuart Birks on Rethinking Economics. The full collection is:

- 1212 An economics angle on the law
- 1213 Rethinking economics: theory as rhetoric
- 1214 Rethinking economics: Downs with traction
- 1215 Rethinking economics: Economics as a toolkit
- 1216 Rethinking economics: Logical gaps – theory to empirical
- 1217 Rethinking economics: Logical gaps –empirical to the real world

Paper 1215 gives a general overview of the approach. Papers 1212 and 1214 illustrate the application. The approach includes three paths or types of potential error. Papers 1213, 1216 and 1217 cover paths A, B and C respectively.

2. Data problems

Where established data sources exist, there may be commonly accepted applications of those data. This can be dangerous in that, by habit or convention, definitional or measurement problems may be overlooked, or the suitability of data for particular uses may not be questioned. For numerous reasons, the variables used in theories may not match the data gathered.

Many economic theories relate to either short-run or long-run structures with a focus on equilibria or optima. The same data have to be used to estimate both short-run and long-run models[2], and the observations may not be of equilibria or optima.

Even if data are accurate, their use may be inappropriate. For example, health outcome measures such as QALYs[3] are commonly discounted in health care evaluations in the same way as is done for monetary measures, but it is not clear that the same approach can be taken or how it should be interpreted (Birks, 2006, 2009). Also, some index measures are ordinal in nature, but they are frequently used in regression analyses as if they were cardinal (Birks, 2007).

Two issues will be given particular attention here, aggregation and proxies. The first sub-section identifies spatial and temporal aggregation, using the latter to illustrate the possible problems. Patterns of aggregation influence the groupings used in an analysis and hence the results. As a general issue, this can shape perceptions as described for artificially engineered social groupings (Hargreaves-Heap & Varoufakis, 2002), or social justice (Tyler, 2000). Aggregation is more widespread and more of a problem for analysis than is commonly understood, however. This is followed by a subsection that considers what variables in an analysis might actually represent. Sometimes they are assumed to represent the given variable, while at other times they are taken to be acting as a proxy for something else. This possibility opens the door to several problems of interpretation.

2.1. Aggregation

Aggregation is well recognized in terms of macroeconomic variables, and we might also acknowledge different levels of aggregation in industry or employment classifications. Aggregation reduces the number of variables, and hence the number of relationships between variables, to be considered. If aggregates are to provide valid simplifications, then it is necessary to have homogeneity of the components of an aggregate in terms of their relationships with the other variables under consideration (except in the presumably unusual situation of constant composition of the aggregate as its size changes). Homogeneity here is not an all-encompassing requirement, but is application-specific. Homogeneity is required only with respect to the functional relationships under consideration. Even then, heterogeneity may not distort if all changes on the value of the aggregate result from changes in the number, but not the composition, of a “representative unit”. A further requirement for simplicity of analysis would be heterogeneity between aggregates, as otherwise the number of variables could be reduced by combining further.

Aggregates are far more common than is commonly stated. One distinction between macroeconomics and microeconomics is that the former is based on aggregate variables. However, even in microeconomics there is aggregation over suppliers of goods and heterogeneous units of factors of production such as labour.

[2] The distinction in economics, being based on the assumption of values of certain elements being fixed or variable, is conceptual, not temporal. If, at the microeconomic level, there are short-run and long-run responses to a change in circumstances, and if the short run is not a fixed length of time for all situations, then according to economic theory there will presumably be complex and variable patterns of lagged effects from any change.

[3] Quality-adjusted life years.

There are additional common aggregations that are almost universal and generally unstated. Think of aggregation i) over time, which has implications for the timing of impacts of one variable on another, and ii) over space, as with disregard for distance, required assumptions about market boundaries, or cross-country regressions which disregard differences in country size and hence weights by country and within-country averaging. The problems arising from spatial aggregation are more complex than those arising from aggregation over time. This is because time is uni-dimensional whereas space is in two (or three) dimensions, and because spatially aggregated units can be of widely varying size. In addition, while time intervals such as months may vary in terms of say the number of days or working days, the variation is far less than that of spatial characteristics. It is unrealistic to assume an isotropic surface.

The example here focuses on the issue of aggregation over time. While the economy operates in continuous time, the data are aggregated into discrete time intervals, as with daily, monthly, quarterly or annual data. Analysis then considers simultaneous, lead, or lagged relationships with these aggregate data.

Consider quarterly money supply and price level data. A simultaneous relationship running from money to prices could be interpreted as a change in money supply in this quarter having an effect on the price level also in this quarter. The aggregation issue results in an assumption that a money supply change will have the same effect on this quarter's price level whether it occurs on the first or last day of the quarter. Similarly, it is assumed that an effect on the next quarter's price level will be quite different if a money supply change occurs on the last day of this quarter rather than the first day of the next.

Aggregation also influences the interpretation of results. Consider a distributed lag model with a series of estimated coefficients. The coefficient of the one period lag on quarterly data should not be interpreted as the impact to be expected 4-6 months after a money supply change. For a change on the last day of the previous quarter, the next quarter reflects the impact in the first three months following the change. To treat the lag coefficients as the effects to be expected in the current, next and subsequent quarters is to assume that all the money supply changes occur on the first day of the quarter. Note that this issue arises because the actual current and lagged impacts depend on the timing of an event and the lengths of the time periods over which data are aggregated. Longer intervals may ease the problem, but they would not generally eliminate it and they come at the cost of a reduced number of observations and less explanation due to averaging within observations.

Even if there is in reality a fixed temporal relationship between cause and effect, aggregation problems can result in highly uncertain and imprecise estimates of the relationship. To illustrate this, consider the following simple problem.

First, two series, X and Y, are generated by the following steps (the data are in the Appendix):

1. Let X be the value of a throw of a die. Take 50 observations, 1-50.
2. Toss a coin 50 times, one for each X observation.
3. If it comes up heads, the X value counts towards its matching Y.
4. If it comes up tails, the X value counts towards the next period Y.

One interpretation (A) of this structure is that there is one occurrence of X each period, either at the start or the end of the period. There is a deterministic relationship between X and Y such that $Y_t = X_{t-½}$. If X happens at the start of a period, the full impact on Y is felt in that period. If it happens at the end of a period, the full impact is felt in the next period. Which of these occurs is determined by the throw of a die. As an alternative interpretation (B), it could be considered as a model with variable lags, having a 50 percent chance that the effect is felt in either of the current or the next period. In either event both X and Y actually occur at a particular instant, not spread over time. This differs from the more complicated effect of a change such as a tax cut, where the effect on disposable income (X) and any resulting impact on consumption (Y), say, can take some time.

In generating Y, it happened that 23 of the 50 X observations resulted in lagged impacts on Y. Pairwise correlations were as follows: (X_t, Y_t) = 0.565516; (X_{t-1}, Y_t) = 0.06752; (X_t, X_{t-1}) = -0.01865.

A regression, $Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \varepsilon$, was estimated. Despite interpretation A being based on a deterministic relationship between X and Y, the specification of time periods resulted in an R^2 of only 0.3259. A standard interpretation might be that variations in X explain only 33 percent of variations in Y. It could then be suggested that other variables and random error are responsible for the remaining variation. In other words, X is a poor predictor of Y.

Coefficients for the βs, and standard errors, are given in Table 1. Note that the coefficient on the lagged variable is small and not statistically significant. It would be concluded that there is no lagged effect with this particular sample, despite 23 of the 50 observations actually having a lagged impact. With the estimate of β_1 being significant and near to unity, it would be concluded that the magnitude of the effect is as in the underlying model, but that all the impact occurs in the same period as X.

Table 1: Lagged impact regression results

	β_0	β_1	β_2
Coefficient	-0.29101	0.961213	0.135912
Standard error	0.982915	0.205265	0.210735

This example illustrates two points. First, there are conventional ways in which econometric results can be interpreted. Such interpretations are consistent with the results and so may reflect the underlying situation. However, other interpretations may also be possible. Consequently we may be overstating the strength of findings if we only follow the conventions. Second, one way to identify possible alternatives is to use data for which the underlying structure is known. Clearly, the results are then consistent with that particular structure.

The problem is not one of misspecification through an incorrect functional form (it is linear), nor one of an omitted variable[4]. Rather, it arises from incorrect aggregation over time. It shows the potential for econometrics to give false negatives simply because of aggregation within discrete time periods. Models with quarterly data do not merely require that the effects be felt within a quarter. They actually require the effects to be felt instantaneously if all the effects are to be observed within the same quarter as the cause. Lagged impacts inevitably result in impacts being spread inconsistently over time periods, resulting in imprecise estimates of the underlying relationships.

This presents an additional problem for economists. Not only are economic effects not instantaneous, but also microeconomic theory suggests distinct short-run and long-run adjustments so the effect of a change in a variable is likely to have a complex temporal structure. Are lags likely to be of fixed length? Consider a few economic examples. There can be lags between an event occurring and a decision being made on a response. Petrol prices change, and people decide to get more economical cars. Some new equipment might be needed, but a decision has to be made at an infrequent committee meeting. There can then be lags between the decision and resulting action. How long does it take to sell a house, or find a new job, or to get a tradesperson to do some work? There can be a lot of variation in the time between an event that can cause the initiation of an action and the end result.

In summary, time, and aggregation over time, are not trivial matters for economists. Given the potential problems as illustrated by the numerical example here, care must be taken when interpreting estimates of lag structures and their significance or otherwise. There are alternative

[4] Omitted variable bias is a paradoxical problem for economists. Theories about the economy are simplified representations of the real world. When testing or applying any theory, the aim is to include only the few most important variables, to the exclusion of others.

interpretations of the same results, and the conventionally accepted conclusions may not be universally valid. Cross-section data do not eliminate these problems. Rather, they are likely to exacerbate them. Given the importance of population groupings in much policy analysis, this raises serious questions about interpretation of evidence. If the limitations are not recognised, then the accuracy of the evidence is overstated and any resulting deliberation based on those findings may be flawed.

2.2. Proxies

Statistical models only "explain" in terms of finding statistical associations between strings of numbers. The results depend on the numbers alone, with no regard to the specific variables underlying the numbers. Two distinct variables with identical data series would give identical statistical results. The identified variable may be acting as a proxy for one or more other variables which are, individually or collectively, correlated with it. The interpretation of statistical results as referring to specific variables depends heavily on this issue of proxies, especially in situations where related variables are "controlled for".

Sometimes an index is used as a variable in regression so as to reduce the number of variables being considered. This is based on the concept of several variables moving together, so that a variable can be constructed that will represent them all and pick up their combined effects. An extreme approach would involve using a single indicator variable instead of a composite index. In other words, a single variable can be used to pick up the effects of several variables that are individually or in some linear combination correlated with that variable.

If a variable can be used that way, can it ever be said with any certainty that any explanatory variable in a regression is not serving that function? In other words, whenever an equation is estimated, the results may represent the impact not of the individual explanatory variables included in the equation, but of all the correlated variables and combinations of variables for which the included variables can act as a proxy.

This point can be illustrated by the use of a variable, "mother's education". Several documents present results about the effects of education levels of mothers along with consequential policy recommendations. For example, from the World Bank Living Standards Measurement Study, Working Paper No.123, *The Demand for Medical Care: Evidence from Urban Areas in Bolivia*:

> "Our results also show that income and education are also important determinants of demand for medical care. For children, mother's education is far more influential than father's." (Ii, 1995, p. ix)

Similar reasoning is found in Working Paper No.128, *How Does Schooling of Mothers Improve Child Health? Evidence from Morocco* (Glewwe, 1997, p. viii) and Working Paper No.132, *The Role of the Private Sector in Education in Vietnam: Evidence from the Vietnam Living Standards Survey* (Glewwe & Patrinos, 1998, p. ix). Elsewhere numerous examples can be found, including: an OECD report (Morrisson, 2002, p. 30); a UK Treasury document (Doney & Wroe, 2006, p. 5); and a Joseph Roundtree Foundation document (Machin & McNally, 2006, p. 13).

All these studies are making claims that could be used to support policy decisions. However, a Google Scholar search on ["mother's education" +proxy] provided over 2,500 results. From *Demography*:

> "[W]e argue that maternal education may be a proxy for the socioeconomic status of the household as well as for characteristics of the community of residence. Hence, we consider the possibility that the observed correlation between maternal education and various markers of child health may be spurious...we demonstrate that controls for a few socioeconomic variables and for community of residence substantially attenuate the maternal education/child health link. Based on these results, we argue that the relationship between maternal education and child health is considerably weaker than is commonly believed." (S. Desai & Alva, 1998, p. 71)

As far as the computations are concerned, any variable that is included in a regression is considered as a string of numbers. Whatever the numbers actually represent is irrelevant for those processes, but it is essential for interpretation of the results. However, there is good reason to expect a variable to pick up the effects of other (at least numerically) related variables, which allows for the use of proxy variables and can also yield spurious results. As they rely solely on the numbers, there is nothing in the statistical techniques to allow an analyst to deduce whether the results represent the effects of the variable, of some other variable(s) for which it is a proxy, or pure chance. It would seem that some policy recommendations are claiming to be based on a stronger evidence base than is justified.

2.3. More on indices

Indices were briefly mentioned at the start of the above subsection on proxies. More can be said about indices in general. Consider first the construction of a price index, say the consumer price index. The process could be set out as three steps:

1. Select the items
2. Choose quantities for the items
3. Add prices

Steps one and two are guided by the requirement that the items and their weights reflect a representative consumption bundle. While this would not match the consumption of every (or perhaps any) consumer, and actual consumption can change over time for a variety of reasons (such as changes in relative prices, tastes, or available products), there are criteria on which to base the decisions and against which they can be assessed. The third step involves identification of prices. These may vary over time and outlet, and consumers may shop around, so the selected prices may not match those actually paid. However, there can still be a clearly specified process to follow.

There has been a proliferation of indices used by economists in recent years. To name a few, there are the Healthy Housing Index; Financial Liberalisation Index; Index of Globalisation; Control of Corruption Index; Corruption Perceptions Index; Health Utility Index; Human Development Index; Australian Unity Wellbeing Index; and Economic Living Standards Index. For each of these, some equivalent of the above three steps must be followed. To give them in a general form, the steps are:

1. Select items
2. Choose weights
3. Add values

The process is described for the Economic Living Standards Index in Birks (2007).

1. Items for inclusion are answers to: seven questions on ownership of a telephone, washing machine, etc. (yes, or no because did not want, could not afford, or other reason); seven activities such as visit to hairdresser, friends for dinner wish similar response options; eight economising measures ("Spent less time on hobbies than you would like to help keep down costs", no, a little, a lot); three self-assessment wellbeing questions on four or five point scales. There is no clear reason for the number and choice of these items.
2. The questions were all given equal weight.
3. Values were then based on these responses, coded up (0 or 1 for some questions, 0 to 4 in integers of others, etc.) with no clear reason for these values. With the same orderings, responses to questions could have been given values 0 or 50, or 0, 1, 4, 10, and 30, say.

Given the changes to values on individual questions, or altered weightings, the ordering of index measures could be quite different. Even if the ordering were maintained, the actual index values could follow a very different pattern (exponential compared to linear, for example). This is a major concern where index measures are used as variables in regressions. As was shown in Birks (2007), changing the scaling while maintaining the ordering can result in a change not only in magnitude, but also in the sign of a regression coefficient. With no basis for preferring one scaling over another, the sign cannot be determined with any accuracy. Policy directions are recommended on the basis of signs of coefficients. Hence a positive relationship can result in a suggestion to increase X so as to bring about an increase in Y. Consequently the potential to change the sign can be very important. Of course, this problem is not generally recognised with only one representation of the index being used. It is a general problem when ordinal measures are treated as if they were cardinal, and many indices are, at best, ordinal.

3. Functional forms

Theory generally does not give the specific form of relationship to be estimated.[5] This is particularly so with static models, as timing and lagged effects are not considered. One simple, but powerful description of the functional form problem was presented in a seminar at Massey University several years ago by a visiting geography professor. He used the concepts of an input wave and an output wave. In a single equation model, an independent variable provides an input wave, a particular pattern of impact over time. There is then a resulting effect on the dependent variable, which can be described as an output wave, a particular pattern of effect over time. In a model, the pattern and timing of impact are fixed by the functional form. With a linear relationship, the output wave matches the pattern of the input wave. All the regression does is estimate sign and the magnitude of the effect. This is highly restrictive, but it is a starting assumption in most estimated relationships. It is not the only strong restriction imposed by linear models, another being additive separability. Two additional concerns, controlling for other factors and causality, are described here.

3.1. Controlling for other factors

A Google Scholar search for “vector of control variables” conducted on 27 February 2009 produced “about 4,140” results. Repeated on 24 October 2012 the number had grown to “about 8,780”. It has become common practice to convert basic models with a few variables into ostensibly more complex and realistic models simply by adding “control variables”. It is then claimed that the effects of those variables have been taken into account, with the results for the variables under investigation being those observed having made full allowance for the other effects.

Without further explanation of the functional form, this is pure rhetoric. Consider what it means. It has to be assumed that, for a unit change in a control variable, when the dependent variable is:

- A number - the control variable has a fixed numerical effect;
- A log - the control variable has a fixed proportionate effect;
- A total value (such as GDP) - the control variable has a fixed total effect;
- A per capita value - the control variable has a fixed per capita effect;
- A nominal value - the control variable has a fixed nominal effect;
- A real value - the control variable has a fixed real effect;
- A first difference - the control variable has a fixed effect on the first difference.

Should it be simply assumed that, whichever of these is used, it results in the relationship being correctly specified? As alternative specifications are chosen for the dependent variable, can it be expected that the functional form for the control variable will always be correct? These would appear to be highly implausible.

[5] Even when it does, the approach has been criticised. For example, Mende (2005) suggests that researchers are too quick to move into estimation and application of quantitative techniques, thereby possibly missing numerous other possible perspectives. Similarly, Lindblom (1990) and Dunn see much standard social science policy research as being locked in to established frameworks and “little more than a ratification of conformity” (Dunn, 1997, p. 277). For a broader criticism of the requirement of a stable underlying structure, “The most telling point against [econometrics is that] significant invariant event regularities...have yet to be uncovered in economics” (Lawson, 1997, p. 70), and Carr (2008, p. 171) talks of history being concerned with processes of change (i.e. changing structures) while economists “take cover” in econometrics.

As an alternative explanation, consider standard approaches to the use of data in different currencies (adjusting for exchange rate differences by converting to a common currency) or different price levels (adjusting for inflation by converting from nominal to real values). Instead of these adjustments, would it be considered acceptable to add an exchange rate or a price index as an additional variable in a linear regression? The implication would be that the effect of say a price level change is independent of the magnitudes of all other variables. If such an approach is unacceptable, why should it be assumed sufficient when controlling for any other influences?

Interpretation of results is also problematic. Control variables may sometimes be considered as a means of adjusting (in a restricted way) for aggregation problems in the data. For examples, dummy variables can do this by allowing for some difference in overall affect according to category. In other cases, their use lends itself to an interpretation such as, "if policies were introduced **to successfully eliminate the effects of the control variables**, then the relationship between the other variables may be as estimated". An additional rejoinder is then needed, "however, the policies do not have to be introduced", which bears a close similarity to the reasoning underpinning the Kaldor-Hicks compensation principle.

Similar issues have been identified in psychology, especially building on the work of Urie Brofenbrenner, who recognized the impact of wider social factors on individual behaviour. As one illustration, Brofenbrenner challenged the idea that the effects of certain factors can be taken out so as to analyse interactions between others:

> "[Brofenbrenner] went on to explain that it made no sense at all to control for ethnicity, social class, or household composition in an attempt to isolate "pure" process. No processes occur outside of a context. And if we want to understand context, we need to take it into account, not pretend to control it away." (Steinberg, Darling, & Fletcher, 1995, p. 424)

This suggests a major problem with much of the quantitative work in economics. Unstated and untested assumptions are being made right at the specification stage. This will generally have an impact on the results, but not in any predictable or meaningful way. Specification error is a concern raised in econometrics textbooks, but not commonly in relation to control variables.

3.2. Causality

There are well-recognised causality issues, such as:

- The distinction between correlation and causation and whether a statistical relationship is actually causal at all;
- If it is causal, whether it is a relationship between two identified variables or involving some additional variable(s);
- The direction of causality between variables (one way from A to B or B to A) or two way);
- The timing of the impact.

As indicated in several examples above, these matters may be glossed over in many studies.[6] Economic analysis relies heavily on econometrics to estimate relationships between variables. The results may then be used to influence policy decisions. The recommendations are based on simplified assumptions about causality. The existence of more complex relationships is indicated by some recommendations that do not seem to fit very well within this framework. In Israel, a

[6] Granger causality uses statistical tests based on a presumption of *post hoc ergo propter hoc*. Some treat this as a test of causality, but it is really a case of Schopenhauer's second stratagem for winning an argument, "the homonymy", where "[the] trick is to extend a proposition to something which has little or nothing in common with the matter in question but the similarity of the word; then to refute it triumphantly, and so claim credit for having refuted the original statement" (Schopenhauer, c1851, 3.II). The problem can be seen by consideration of an increase in retail spending followed by Christmas. The former is a response to the latter, despite it occurring first. (This highlights an additional problem for economists, in that behaviour is based on perceptions of the environment. These may differ in substance and timing from events in the environment themselves, however most data series are attempts to measure real rather than perceived phenomena.)

comprehensive stabilization programme was effective using a range of policies working together whereas traditional approaches relying on fewer policy instruments had failed (Cukierman, 1988, p. 48). A combination of measures was required to achieve a health objective in African countries (James et al., 2006, p. 148). In youth smoking prevention, Ross et al. (2006) describe a range of policies used to change the whole environment, considering that their success depended on the multidimensional approach. An OECD Policy Brief mentioned above, (Morrisson, 2002), refers to mothers' education and poverty. While it mentions the statistical link between mother's education and child health, and talks of a relationship between mother's education and the demand for education for children, it also talks of "clusters of factors" and "combinations of disadvantages".

The common feature of these examples is that individual policies might be ineffective. For successful outcomes, a package of policies is required. This is consistent with an INUS approach to causality, whereby outcomes require the coincident occurrence of a specific combination of factors, and there may be several alternative combinations that can achieve the same outcome (Addison, Burton, & Torrance, 1984).[7] Policies in isolation may be ineffective, but in the right combination they can work, and there may be several alternative approaches to choose from. To rephrase this, a policy might be effective in some situations but ineffective in others as a combination of (policy and other) circumstances are required.[8]

This does not fit well within an econometric modelling approach because econometric models have difficulty including conditional multivariate components. Even in the relatively simple situation of binary variables (a policy is "on" or "off") a complex mix of dummy variables and interactive terms would be required. Take a simplified version of the first example, consider a dummy variable, D_1, equal to 1 when there is a tight monetary policy, another, D_2, for tight fiscal policy, and D_3 for a prices and incomes policy. It would then be necessary to include a variable $D_1xD_2xD_3$.

If INUS conditions are included, there may be several possible combinations to consider. If so, for the structure to be correctly specified it would be necessary to include all the combinations that occurred in the sample period, set out as combinations rather than just individual variables. This raises an additional problem. If the combinations of policies that can give the desired outcome are not already known, are there ways that they can be identified and/or tests undertaken to find them? This is important in a general policy dimension. A basic point is that it may be desirable to consider broad-based, multi-dimensional, multifaceted policy. Instead, regression approaches tend to be based on the assumption that there are individual policy variables with a fixed marginal impact on the target variables.

3.3. Structural stability

Alan Greenspan's statement to the House Oversight and Government Affairs Committee on 23 October 2008 included the following:

> "In recent decades, a vast risk management and pricing system has evolved, combining the best insights of mathematicians and finance experts supported by major advances in computer and communications technology. A Nobel Prize was awarded for the discovery of the pricing model that underpins much of the advance in derivates markets. This modern risk management

[7] An INUS condition is **I**nsufficient on its own, but is a **N**ecessary component of a set of conditions that are **U**nnecessary but **S**ufficient for an outcome to arise. For example, there are several reasons why a house might burn down (cooking fire, electrical fault, candle), but each reason is contingent on other factors (proximity of flammable material, lack of oversight, no smoke alarm, etc.). A similar point is made by David Byrne and Charles Ragin using the term "qualitative comparative analysis". In an interview, Byrne defined this as, **"**a method which is 'set theoretic' i.e. it understands causal relations in the social world in terms of relationships in combination – sets, rather than the unique contribution of single variables**"** (Byrne, 2009).

[8] This point is central to Rodrik's case for country-specific policies, "…straightforward borrowing (or rejection) of policies without a full understanding of the context that enabled them to be successful (or led them to be failures) is a recipe for disaster" (Rodrik, 2007, pp. 4-5).

paradigm held sway for decades. The whole intellectual edifice, however, collapsed in the summer of last year because the data inputted into the risk management models generally covered only the past two decades, a period of euphoria. Had instead the models been fitted more appropriately to historic periods of stress, capital requirements would have been much higher and the financial world would be in far better shape today, in my judgment." (Greenspan, 2008)

This indicates a problem that can arise in understanding situations that may occur infrequently, as with some shocks[9], or long cycles or structural changes, when people are drawing on experience only of the recent past.

A broader point is that generations each have their own experiences of the world. An older generation may make decisions in reaction to their experience of earlier issues or problems. A younger generation, observing these decisions, will not have experienced the context on which they are based. Therefore, their interpretation of their current experiences can differ from that of the older generation. A certain behaviour (or policy) could be seen as "the way things are done", as with consumption of fast foods. Alternatively, it could be rejected as misguided or outdated behaviour of older people. This could possibly help to explain differing attitudes to debt of those who lived through the depression of the 1930s, those born after the Second World War, and those born in the 1980s.

A further potential problem that Greenspan did not mention is that, depending on the complexity of the estimation of risk of default on mortgages, the calculation may be giving the average risk of default, whereas the important measure, at least in an expanding market, would be the marginal case. As more loans are given to people within a category, risks may be rising as the marginal and average are not the same. A similar point can be observed in an example by Musgrave and Musgrave. They present results on returns to education, after which they state, "...since there is considerable slack in student input (only about 50 percent of students complete high school and only one-third of high school graduates proceed through college), it appears that considerable scope for increased investment in higher education still exists" (Musgrave & Musgrave, 1973, p. 196). However, the statistical result is the average return, and additional (marginal) students may have other characteristics and outcomes (even if only through changing pay by qualification as numbers of qualified change).

The average versus marginal issue has similarities with that of outliers. Should such observations be discarded as aberrations which distort statistical results, or are they valuable sources of information. A classic example of the value of outliers is that of the Broad Street pump (Hempel, 2007). In 1854 a cholera outbreak produced numerous cases clustered around a pump in Broad Street, London. Dr John Snow noticed this when he plotted the deaths on a map. The connection with the pump was demonstrated more conclusively by the lack of deaths in a factory close by which did not use that water, and the death of a woman some distance away who was regularly supplied with water from the pump.[10]

The value of reliance on a common structure has also been questioned, as in cases where only low R^2s are produced. Action research, often applied to education issues, aims to bring change through consideration of issues at a micro level. This is because the individual influences may be more important than the more generalised ones found through quantitative research. Egon Guba, in a foreword to Stringer's book on action research, writes:

> "We have witnessed, over the past half-century or so, determined efforts to find general solutions to social problems...The cost to national economies has been prodigious, and there is precious little to show for it...It ought to be apparent by now that generalized, one-size-fits-all solutions do not work...Without intimate knowledge of local context, one cannot hope to devise

[9] Taleb refers to these as "black swans" (Taleb, 2005, 2007). Winston (2005) distinguishes between perspective and analytic time, pointing out the problem for economics of unique events.
[10] The existence of relevant factors besides proximity to the pump suggests the existence of INUS conditions.

solutions to local problems. *All* problems are de facto local; inquiry must be decentralized to the local context.” (Stringer, 2007, p. ix)

Parallels might be considered with the law, giving generalised solutions, and its implementation, through which individual circumstances may be considered. If the specific are more important than the general, there is limited scope for laws themselves to address the problems. There are also lessons for the value of much economics and econometrics research in that this commonly seeks to obtain generaliseable findings.

There is an additional consideration also. Of all the available information, only a small component is in the form of large, quantitative data series. This is particularly true with the growth of new information technologies, providing large quantities of readily accessible textual and other data along with means to quickly search and analyse them. Consequently, other approaches to research and the use of alternative types of data may give additional valuable insights into phenomena of interest to economists.

4. Conclusions

Data problems and limited functional forms can place significant constraints on empirical analyses. This is not entirely surprising for two reasons. First, econometrics is constrained by the restrictive nature of relationships it is used to estimate, with many potential alternatives assumed away at the initial specification stage. Second, econometric results focus on the average, or the common features, which may tell only a small part of the whole story. Nevertheless, as Dunn has suggested, as a mode of argumentation, these techniques may be persuasive. This may result in decisions being based on misinformation and undue confidence in quantitative research results.

References

Addison, J. T., Burton, J., & Torrance, T. S. (1984). Causation, Social Science and Sir John Hicks. *Oxford Economic Papers, 36*(1), 1-11.

Birks, S. (2006). Discounting Health Effects in Pharmacoeconomic Evaluations: Current Controversies, *PharmacoEconomics* (Vol. 24, pp. 1273-1274): ADIS International Limited.

Birks, S. (2007). A Word of Caution on Indices. *SSRN eLibrary*. Retrieved from http://ssrn.com/paper=1080328

Birks, S. (2009). Some problems/assumptions in pharmacoeconomic analysis In R. J. G. Arnold (Ed.), *Pharmacoeconomics: From theory to practice* (pp. 133-148). Boca Raton, Fl: CRC Press.

Birks, S. (2012). Rethinking economics: theory as rhetoric. *Bristol Business School Economics Papers Series, 12*(13). Retrieved from http://www2.uwe.ac.uk/faculties/BBS/BUS/Research/economics2012/1213.pdf

Byrne, D. (2009). Interview with David Byrne (interviewer Brian Castellani). Retrieved 26 October 2012, from http://sacswebsite.blogspot.co.uk/2009/05/interview-with-david-byrne.html

Carr, E. H. (2008). *What is history?* Harmondsworth: Penguin.

Cukierman, A. (1988). The End of the High Israeli Inflation: An Experiment in Heterodox Stabilisation. In M. Bruno, G. D. Tella, R. Dornbusch & S. Fischer (Eds.), *Inflation stabilization: the experience of Israel, Argentina, Brazil, Bolivia, and Mexico* (pp. 48-94). Cambridge, Mass.: MIT Press.

Desai, M. (1981). *Testing monetarism*. London: Pinter.

Desai, S., & Alva, S. (1998). Maternal Education and Child Health: Is There a Strong Causal Relationship? *Demography, 35*(1), 71-81.

Doney, M., & Wroe, M. (2006). Keeping our promises: Delivering education for all Available from http://www.dfid.gov.uk/Documents/publications/education-delivering-promises.pdf

Dunn, W. N. (1997). Probing the Boundaries of Ignorance in Policy Analysis. *American Behavioral Scientist, 40*(3), 277-298.

Dunn, W. N. (2004). *Public policy analysis: an introduction* (3rd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
Glewwe, P. (1997). How Does Schooling of Mothers Improve Child Health? Evidence from Morocco Available from http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2000/02/24/000009265_3971031092438/Rendered/PDF/multi_page.pdf
Glewwe, P., & Patrinos, H. A. (1998). The Role of the Private Sector in Education in Vietnam: Evidence from the Vietnam Living Standards Survey Available from http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/1998/03/01/000009265_3980624090958/Rendered/PDF/multi_page.pdf
Greenspan, A. (2008). Testimony to the Committee of Government Oversight and Reform, October 23. Retrieved 24 October 2008, from http://oversight.house.gov/documents/20081023100438.pdf
Hargreaves-Heap, S., & Varoufakis, Y. (2002). Some Experimental Evidence on the Evolution of Discrimination, Co-operation and Perceptions of Fairness. *Economic Journal, 112*(481), 679-703.
Hempel, S. (2007). *The strange case of the Broad Street pump: John Snow and the mystery of cholera*. Berkeley: University of California Press.
Ii, M. (1995). The demand for medical care: evidence from urban areas in Bolivia Available from http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/1996/02/01/000009265_3961214182415/Rendered/PDF/multi0page.pdf
James, C. D., Hanson, K., McPake, B., Balabanova, D., Gwatkin, D., Hopwood, I., et al. (2006). To Retain or Remove User Fees?: Reflections on the Current Debate in Low- and Middle-Income Countries. *Applied Health Economics and Health Policy, 5*, 137-153.
Lawson, T. (1997). *Economics and reality*. London: Routledge.
Lindblom, C. E. (1990). *Inquiry and change: the troubled attempt to understand and shape society*. New Haven, Conn.: Yale University Press.
Machin, S., & McNally, S. (2006). Education and child poverty: A literature review Available from http://www.jrf.org.uk/sites/files/jrf/9781859354773.pdf
Mende, J. (2005). The poverty of empiricism. *Informing Science Journal, 8*, 189-210.
Morrisson, C. (2002). Policy Brief No.19: Health, Education and Poverty Reduction Available from http://www.oecd.org/dataoecd/59/57/1851366.pdf
Musgrave, R. A., & Musgrave, P. B. (1973). *Public finance in theory and practice* (2 ed.). New York: McGraw-Hill.
Rodrik, D. (2007). *One economics, many recipes: globalization, institutions, and economic growth*. Princeton: Princeton University Press.
Ross, H., Powell, L. M., Bauer, J. E., Levy, D. T., Peck, R. M., & Lee, H.-R. (2006). Community-Based Youth Tobacco Control Interventions: Cost Effectiveness of the Full Court Press Project. *Applied Health Economics and Health Policy, 5*, 167-176.
Schopenhauer, A. (c1851). The art of controversy Available from http://etext.library.adelaide.edu.au/s/schopenhauer/arthur/controversy/
Steinberg, L., Darling, N. E., & Fletcher, A. C. (1995). Authoritative Parenting and Adolescent Adjustment: An Ecological Journey. In P. Moen, G. H. Elder, K. Luscher & U. Bronfenbrenner (Eds.), *Examining lives in context: perspectives on the ecology of human development* (pp. 423-466). Washington, DC: American Psychological Association.
Stringer, E. T. (2007). *Action research* (3rd ed.). Los Angeles: Sage Publications.
Taleb, N. N. (2005). *Fooled by randomness: the hidden role of chance in life and in the markets* (2nd ed.). New York: Random House Trade Paperbacks.
Taleb, N. N. (2007). *The black swan: the impact of the highly improbable*. New York: Random House.

Tyler, T. R. (2000). Social Justice: Outcome and Procedure. *International Journal of Psychology, 35*(2), 117-125.

Winston, G. C. (2005). Three problems with the treatment of time in economics: perspectiveness, repetitiveness, and time units. In G. C. Winston & R. F. Teichgraeber (Eds.), *The boundaries of economics* (pp. 30-52). Cambridge: Cambridge University Press.

Appendix

Data for regression in Section 2.1

Obs. no.	Y	X	Obs. Cont.	Y cont.	X cont.
1	3	3	26	5	4
2	0	2	27	0	3
3	2	5	28	6	3
4	5	1	29	0	1
5	1	3	30	6	5
6	9	6	31	6	6
7	0	6	32	1	1
8	7	1	33	0	3
9	1	1	34	3	5
10	5	5	35	5	2
11	2	2	36	2	3
12	2	2	37	8	5
13	0	4	38	0	4
14	10	6	39	9	5
15	0	1	40	3	3
16	7	6	41	0	1
17	0	6	42	5	4
18	9	3	43	1	1
19	0	1	44	0	1
20	7	6	45	1	1
21	3	3	46	1	1
22	0	1	47	5	4
23	3	2	48	1	1
24	1	1	49	0	3
25	0	1	50	9	6

X_1 was 3 and heads, so Y_1 equals 3
X_5 was 3 and tails, so contributes towards Y_6. X_6 was 6 and heads, so also contributes to Y_6, hence Y_6 equals 9.