real-world economics review, issue no. 62
subscribe for free

Rethinking economics: Logical gaps – empirical to the real world[1]

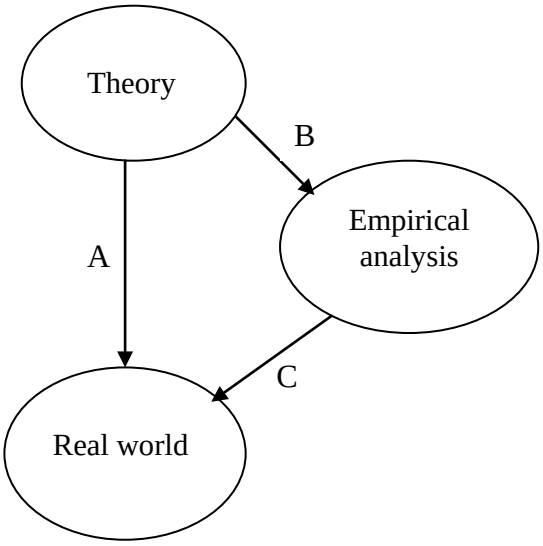
Stuart Birks [Massey University, New Zealand]

Abstract
One aspect of simplification that is apparent at the level of public debate is the way that policy conclusions are frequently drawn from limited statistical evidence. While economists and econometricians are generally cautious about specifying policy implications arising from econometric analyses, at the level of broader debate and media coverage there are fewer reservations. This paper explores some of the limitations and potential opportunities for policy-relevant findings from econometrics. In particular, it considers what can and cannot be deduced as a result of an explanatory variable being found to be statistically significant. In addition, it indicates what aspects to address or questions to raise if econometricians and economists are to extend this work to the point where it may be directly applicable in policy debate.

1. Introduction

This paper draws on a three path structure as in Figure 1 and described in (Birks, 2012c). That paper covered Path A, theory to the real world. A second paper (Birks, 2012b) considered Path B, the transition from theory to empirical formulation. This paper covers Path C, moving from empirical results to application to policy in the real world when research results may be used as a basis for policy.

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



There is often a tenuous relationship between research and policy, with research sometimes playing little or no part. There are several other components in the process of policy making, including political and media debate, response to pressure groups, and shaping or reacting to public opinion. Research is not necessarily directly focused on policy, and the approaches taken by researchers do not necessarily directly address policy questions. Nevertheless

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

researchers sometimes describe policy implications arising from their findings, research findings are sometimes used directly in policymaking and implementation, and findings can influence general understanding of issues. Consequently, the role of research in the policy process is worth exploring.[2]

There are several forces at work that result in a tendency to favour simplified views of issues.[3] They can be observed at each of the three "levels of discourse" described by Desai, namely theory, data analysis, and policy (Desai, 1981, p. 93). At the theoretical level, there is the value judgment associated with Occam's razor whereby simpler theories are preferred over more complex ones, *ceteris paribus*.[4] At the level of data analysis there are constraints of available data, limitations of techniques and problems with degrees of freedom.[5] At the policy level, public acceptance can be important, in which case a simple message is often required (Birks, 2012a). People do not have the time or motivation to understand in detail complex issues which may have little direct impact on them. Any influence is likely to arise through group action, where groups promote preferred agendas. This can be observed through the importance of interest groups and in the limited range of options presented for consideration.

One aspect of simplification that is apparent at the level of public debate is the way that policy conclusions are frequently drawn from limited statistical evidence. While economists and econometricians are generally cautious about specifying policy implications arising from econometric analyses, at the level of broader debate and media coverage there are fewer reservations. This paper explores some of the limitations and potential opportunities for policy-relevant findings from econometrics. In particular, it considers what can and cannot be deduced as a result of an explanatory variable being found to be statistically significant. In addition, it indicates what aspects to address or questions to raise if econometricians and economists are to extend this work to the point where it may be directly applicable in policy debate.

Section 2 briefly considers the link between statistical findings and policy recommendations by academic and public sector researchers and through media coverage of research. Section 3 outlines some basic statistical considerations, while section 4 considers issues associated with the step from statistically meaningful findings to more comprehensive policy analysis.

2. Using statistics for policy

One channel for statistical analysis to influence policy is through public presentation of research findings. This may affect general understanding of issues, shaping public opinion and influencing political priorities. This can happen even if the research was not intended for that purpose and if the results are misinterpreted at the public reporting stage. Where attention is created for political purposes the focus may be on a specific finding. In a small economy this can be associated with a visiting expert deliberately invited to promote a preferred perspective (as with agenda setting and framing, see Birks, 2012a). This is unlikely to result in high-level debate on alternative, possibly contradictory research findings.

[2] For a recent contribution, see Wolf (2007).
[3] See also Birks (2007)
[4] While theoretical simplicity is commonly lauded, more complex analyses with longer equations, more advanced mathematics and/or larger data bases are also associated high status.
[5] Concerns about techniques used by economists, especially econometrics, have been raised by various writers, including Swann (2006) and Thurow (1983). Alternative techniques are described in Swann (2006) and Allen (1978).

This section uses Desai's structure to briefly consider the research phase and then focus on the less commonly discussed but potentially very significant media phase. Specific analytic considerations are then discussed in more detail in section 3.

2.1 The research phase

McCloskey and Ziliak have identified problems in academic papers in the interpretation of statistical findings as being of significance for policy (McCloskey, 1998; Ziliak & McCloskey, 2004, 2008)[6]. One of their central points is that policy decisions should not be determined on the basis of statistical significance alone.[7]

Looking at recent issues of such economic journals as *Applied Economics*[8], *The Review of Economic Studies, Economic Record* and *Southern Economic Journal*, few articles actually refer to policy implications. This may reflect a difference in focus between academic economists and economists working as policy analysts.[9] Some of the discussion papers from the Reserve Bank of New Zealand use econometric models of the macroeconomy and relate the results to policy decisions.[10] The New Zealand Department of Labour research publications tend to rely heavily on more discursive forms of analysis with graphical representation of data.[11] The Ministry of Economic Development has papers outlining econometric analyses on microeconomic issues using disaggregated data.[12] The policy conclusions tend to be tentative, however, as in Maré and Timmins (2007, p. 53) on firm productivity. Given the range of industries and firms and the number of geographically related factors that can affect productivity, this study's findings are not entirely surprising. A strong statistical association would only arise if there is a fixed underlying structure that applies to highly heterogeneous units. This is unlikely to be the case. The same point could be made for many other studies using similar methodology.

2.2 The media phase

Tentative conclusions are less commonly observed at Desai's third level, that of the media and policy discourse. Politicians are expected to appear clear and decisive, despite all the actual uncertainties surrounding policy issues. The public want information that has a clear point to make, or, for personal interest, that relates to a need to change behaviour, presented

---

[6] Ziliak and McCloskey (2008) has been critically reviewed (Spanos, 2008). However, the criticisms relate to proposed solutions to the problem. If anything, Spanos suggests that the problems themselves are more severe than suggested by Ziliak and McCloskey.

[7] This point has been made also in relation to policy in education:

> "Most research articles, after finding a set of things that is correlated with student performance, immediately go to a section on policy conclusions. The steps between the statistical analysis and the section on policy conclusions are seldom discussed." (Hanushek, 1997, p. 303)

[8] Based on their abstracts, econometrics was central to 10 of 11 articles in *Applied Economics* 39(21), December 2007. Of these, three drew some possible implications for decisions/policy.

[9] As a separate exercise, it may be interesting to see to what extent academic research impacts on policy decisions, and, if considered, whether the findings are correctly interpreted. Example 2 below is a case in point.

[10] http://www.rbnz.govt.nz/research/discusspapers/. Approximately 7 of the 15 papers in 2007 would fit into this category. The greater emphasis on policy implications from econometric analysis may be due to the Reserve Bank having defined objectives and a limited number of policy instruments. This to a degree constrains the range of analysis required, reducing the alternatives and associated costs and benefits to be considered.

[11] http://www.dol.govt.nz/publications-browse.asp?BrowseBy=Date&Year=2007, None of the 11 listed publications relied on econometric methods.

[12] http://www.med.govt.nz/templates/StandardSummary____22733.aspx. 3 of 5 papers for 2007 had significant econometric components and drew tentative policy-relevant conclusions.

without many complicating qualifications. Journalists may also lack the specialist knowledge required to handle complex issues, and they are constrained by the nature of their media to be concise and entertaining (Birks, 2008). Consequently, recommendations may be based on limited evidence and analysis, perhaps merely on a statistical association or ascribed to some designated "expert"[13]. To give three examples[14]:

Example 1: alcohol and brain damage

A *Dominion Post* article suggested that binge drinking 'damages brains' (Hill, 2007, 5 November). Arbias (Acquired Brain Injury Service) chief executive Sonia Burton suggested that "[e]ven so-called 'social drinking' could cause permanent brain damage". On the basis of this association, she called for an education programme and screening by health professionals that "should be as routine as a cholesterol check".

Example 2: Single parenthood and childhood risk

This is an example in a policy context where lack of statistical significance was used to draw policy conclusions. In paragraph 616 and footnote 299 of the Law Commission's *Preliminary Paper 47: Family Court Dispute Resolution* (Law Commission, 2002) there is reference to Fergusson (1998).[15] The paper is quoted in the footnote, "Collectively, the findings suggest that single parenthood, in the absence of social or family disadvantage, is not a factor that makes a major contribution to childhood risk".

This statement refers to a statistical finding on the significance of a variable. It is used to suggest that single parenthood may not be a concern as associated childhood problems are not observed when the study controls for certain factors.[16] The interpretation of this finding is a more complex matter. It must recognise the interconnectedness of many determining factors, such that the factors that are controlled for may be closely associated with single parenthood. It is therefore not realistic to simply treat single-parenthood as being independent of these determinants. This is made clear in the published study. Hence Fergusson states:

"The implications of these conclusions are clearly that social programmes and policies that are likely to be most effective in addressing the needs of at-risk families and their children are likely to involve multi-compartmental approaches that have sufficient breadth and flexibility to address the wide range of social, economic, family, individual and related factors that contribute to the development of childhood problems." (Fergusson, 1998, p. 172)

The Law Commission paper uses lack of statistical significance to contend that a factor is not important. The journal article presents its results with great care, but at the policy level it is selectively quoted to provide apparent support for a specific position. In fact, the impact of the factor may well be felt through other, related variables. This can happen due to more complex causal relationships, or because some variables (such as household income) are acting as a proxies for others.

[13] Note that one of Dunn's "modes of argumentation" in policy debate is "reasoning from authority" based on the achieved or ascribed status of the person presenting the information. (Dunn, 2004, p. 395)
[14] See also NZPA and Reuters (2007, 1 November) on obesity and cancer, (Palmer, 2007, 30 November) on job cancer risks, Perry (2008, 18 January) and Medical update (2002). It may not be coincidental that so many examples are health related. The media considers reader interest and this often requires a personal angle (Hamilton, 2004). The recommendations in the articles may shape perceptions, behaviour and policy.
[15] This example is discussed further in (Birks, 2002).
[16] Note the discussion of control variables in Birks (2012b).

Example 3: TV watching and attention problems

A research paper published in *Pediatrics* found a link between children's television watching and attention problems some years later (Landhuis, Poulton, Welch, & Hancox, 2007). On this basis, despite voicing reservations, the researchers recommended restricting children to no more than two hours watching per day. This example is discussed in more detail in section 0 below.

As a general point to draw from these examples, the information that is presented in reports of research contributes to the shaping of opinions and views on alternative issues and policies. At the very least, the news media do not always apply due caution in presenting these results. This is in part a consequence of inadequate specialist training and expertise. There may also be incentives to sensationalise.

The distortions may be widespread. Quite apart from statistical estimation and functional form problems, the information deduced from these findings may be flawed. This raises a fundamental question, how should we present findings to generate more effective interpretation, especially for policy formation?[17] In addition, given the answer to this first question, what additional questions should be asked to more effectively address the requirements for good policy decisions?

3. Consideration of the problems

A paper at a health economics conference in Auckland in November 2005 illustrates a common problem with the use of statistical results for policy purposes. To give fictitious data, imagine a prevalence of 18 percent for some negative health measure for low income groups, compared to 16 percent for high income groups. This indicates a possible relationship between income and the prevalence of this problem. Does this justify policies to improve the income of low income groups? Quite aside from causality and the issue of the costs of the problem and the costs of alleviating the problem, income may be the wrong measure to look at. Changing income may not be effective, and even if it is, the best that could be achieved is a two percent improvement for those on low income, and no gain for those on high income. This might be considered fairly minimal in terms of addressing the problem, even though it is the type of policy inference commonly made.[18]

The following discussion will be based on a simple regression equation as it provides a useful structure for explanation. Consider a basic single equation multiple regression model where Y is a target variable of policy interest and X_1 can be affected by policy:

$$Y = b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n + u$$

[17] For a discussion of the use of econometrics in law, including reservations and qualifications, see Harkrider (2005).
[18] Stringer's justification for action research is based on the limited value of studies seeking generalised patterns such as these (see also section 3.3 of Birks, 2012b).

Statistical analysis can give results such as a finding based on whether or not X_1 is statistically significant as a determinant of Y.[19] With a superficial assessment, it might be concluded that:

- If it is not significant, there is no relationship, so the variable can be ignored.
- If it is significant, there is a relationship, so there can be a policy recommendation to change X_1.

In other words, there is heavy emphasis on the statistical significance of the estimate of b_1. Such reasoning is flawed. Statistical significance cannot be interpreted as answering all the questions required for deciding on policy intervention. Even if the relationship is one between a policy variable and a target variable, many aspects remain to be considered. For policy, it is important to know the magnitudes of impact, the variability of impact, the costs and possible side-effects of intervention, and, ideally, alternative policy options should also be considered. The first two of these are basic but often overlooked. They are briefly discussed here. A more fundamental issue relating to statistical hypothesis testing is then considered, followed by an issue of option identification from statistical results. The second two points are addressed in section 00.

i) Magnitude of policy impact – if X is changed, how much change is there in Y?

Harkrider gives a good legal example of this point when he distinguishes between "practical significance" and statistical significance:

"Practical significance means that the magnitude of the effect being studied is not de minimis – it is sufficiently important substantively for the court to be concerned. For example, econometric evidence in the context of a publishing merger may reveal that titles published by new entrants are .0001 percent less profitable than titles published by existing entrants. That result may be statistically significant, but not substantively important." (Harkrider, 2005, p. 15)

Similarly for policy, it could be asked whether the relationship between the variables and the available options for change in X_1 result in realistic and effective policy options. In addition to required magnitudes of change and costs of change, the apparent answer could be a result of the representation of the issue, as in the formulation of the equations that are estimated. These problems have been discussed in relation to Path B in Birks (2012b).

ii) Variability of policy impact

Often relatively little attention is given to the overall R-squared for an equation. Sometimes a relationship may be only poorly specified by the equation. A statistically significant explanatory variable may then be a small factor in the overall determination of the value of the dependent variable.

Also, even though the significance of a coefficient is commonly discussed (as with the t-test results), this may not be carried over to consider the possible variability of response to a policy of changing X_1. The estimated coefficient may be significantly different from zero, but

[19] There is scope to debate the criteria for determining whether results are "statistically significant". This discussion takes statistical findings as given, looking at the subsequent stage of interpretation of results for policy purposes.

the true value may still be quite different from the estimated value, and the effect of a change in X_1 on Y may also be variable across individual cases.

3.1 McCloskey and Ziliak on interpretation of statistical significance

McCloskey and Ziliak have repeatedly identified problems in academic papers in the way that policy inferences are drawn from statistical findings. These are described in detail in Ziliak and McCloskey (2008), which includes a quote from Thomas Schelling on the back cover:

"McCloskey and Ziliak have been pushing this very elementary, very correct, very important argument through several articles over several years and for reasons I cannot fathom it is still resisted."

3.1.1 A digression on rhetoric

The reason for the resistance may be found by noting the focus of logic on proof, and rhetoric on persuasion. McCloskey's *The rhetoric of economics* is now in its second edition (McCloskey, 1998). It may have proved its point (the arguments are logical), but it has not persuaded many economists (the rhetoric is weak).

In economic theory, little attention is generally paid to processes and persuasion. This is perhaps inevitable, given the focus on static analysis and assumptions of exogenous preferences and rationality. However, persuasion may be important in terms of both our understanding of economic phenomena and the development of economics as a discipline.

There is other literature that incorporates concepts such as traction, agenda setting, and framing. This shows clearly that it is not enough simply to present a correct argument. One reference that addresses these issues in a political context is Cobb and Ross (1997). The title, *Cultural strategies of agenda denial: Avoidance, attack, and redefinition*, suggests that there are reasons why people with a heavy investment in established positions may be unwilling to change.[20] [21]

---

[20] Consider the discussion on accepting or rejecting a theory in Birks (2012c).

[21] This point is frequently made and illustrated. To quote Simon, "Legitimacy may sometimes be achieved (and even attention secured) by the usual credentials of science…[b]ut many an impeccable report is ignored, and many a report without proper credentials gains a high place on the agenda" (Simon, 1971, p. 50). Note also:

> "Science can destroy religion by ignoring it as well as by disproving its tenets. No one ever demonstrated, so far as I am aware, the nonexistence of Zeus or Thor, but they have few followers now." (Clarke, 1953, p. 21)

And, "[M]ost papers are never read at all. No matter what a paper did to the former literature, if no one else does anything else with it, then it is as if it had never existed at all." (Latour, 1987, p. 40)

Reasoned assessment may prove more complex than this, as indicated by Pope John Paul II:

> "[T]here are in the life of a human being many more truths which are simply believed than truths which are acquired by way of personal verification...This means that the human being—the one who seeks the truth—is also *the one who lives by belief.*" (Pope John Paul II, 1998, para.31)

Political motivation on gender issues is suggested by Margaret Mayman when she said: "Stereotypical notions of femininity have tended to obscure women's violence or alternatively, it has been actively denied by feminists, both first and second-wave, because it complicates the interpretation of women's 'innocence' and non-culpability in intimate violence" (Mayman, 2003). Durie on ethnicity also cautioned that it is, "important that the researcher should not to be captured by current ideologies that manicure a perception of the past to suit a current purpose" (Durie, 1999). This may also help to explain why the figures from one New Zealand study on the economic costs of family violence (Snively, 1994) continue to be used despite the study having serious flaws (Birks, 2000), whereas a PhD thesis identifying gender biases in sentencing (Jeffries, 2001) has had little impact.

In relation to challenges to statistical significance, economists may still be involved in avoidance. Cobb and Ross talk of "identification groups", people who raise an issue in the first place, and "attention groups" who then promote the issues more widely (Cobb & Ross, 1997, p. 7). Without the latter, the issue will not get off the ground.[22]

3.1.2 Interpretation of statistical significance

As Schelling states, McCloskey and Ziliak's points are elementary. Two of their central points are outlined here. They can both be simply illustrated.

On the first point, the impact of sample size on statistical results, consider the gender pay gap:

1. With earnings data for one man and one woman, nothing can be said about the significance of any difference between them as nothing is known about the distribution of male and female earnings. More than one observation for each is required.

2. With a larger sample, assumptions can be made and tests undertaken for a difference in average incomes.

3. At the other extreme, if observations are available for every man and every woman in the population, the average male and female earnings can be calculated precisely. The estimate equals the true population value, the variance of the estimate is therefore zero. A difference as low as 1c is therefore statistically significant.

Hence, a finding that a gender pay gap does or does not exist depends on the sample size. However, this has nothing to do with significance for policy.

To state this point more generally, statistical results depend on the underlying situation and the test that is applied, where the test is sensitive to sample size, N. N is generally not related to the underlying situation. Consequently, policy decisions should not be determined on the basis of statistical significance alone.

There is a second criticism of the interpretation of statistical significance. It involves a problem with the conventional interpretation of null hypothesis significance tests. It has been illustrated through a class of examples that have been presented in several places (such as Cohen, 1994, pp. 998-999; Taleb, 2005, pp. 206-207). The examples have tended to take the following form. There is a test for some illness that picks up say 95 per cent of true cases. Someone gets a positive result. What is the likelihood that the person has the illness? Many assume that the answer is 95 per cent, but they are wrong. While only 5 per cent of those

[22] Cobb and Ross's more detailed formulation (identification groups, the attentive public and attention groups, and the mass public) has parallels with an earlier description by Bryce of:

> "…three classes of person who have to do with the making of public opinion…the men who seriously occupy themselves with public affairs…those who, though comparatively passive, take an interest in politics…[and] all that large residue of the citizens which is indifferent to public affairs, reading little and thinking less about them." (Bryce, 1929, pp. 176-177)

The need for a small pressure group to gain wider support is also stated in, "...this kind of group is wholly dependent on the socialization of conflict...this is a trigger organisation which may start a chain reaction..." (Schattschneider, 1960, pp. 47-48)

without the illness are expected to have positive test results, this could be a large proportion of results if many of those tested are well.

The explanation involves computing tables, and/or equations of conditional probabilities. These present the underlying logic, but have done little to change behaviour. Apparently, logical arguments are not necessarily very persuasive. It may be helpful to illustrate the point in the example by taking an extreme case. This reduces the detail required.

Consider a society that has such advanced technology that all coins are so well made that none of them are ever biased. Someone tosses a coin 6 times. Whatever side came up on the first toss is repeated for the next 5 tosses. The chance of such a result with an unbiased coin is $(½)^5$ or about 3%. An analyst is likely to reason that the chance of this occurring with an unbiased coin is so low that the null hypothesis of unbiasedness would be rejected, concluding that there is a high likelihood that the coin is biased. For this example, **this would be wrong every single time** that the reasoning is followed. It is known that false positives can occur, but, it cannot be determined if a positive result is a false or a true positive (e.g. biased coins in this example) without using additional information.

There is a difference between a statement that the outcome is unlikely if the coin is unbiased and a statement that, given the outcome has been observed, the coin is likely to be biased. The reasoning uses the former to claim the latter. Ziliak and McCloskey refer to this as the "fallacy of the transposed conditional" (Ziliak & McCloskey, 2008, p. 17).[23]

David Hendry also demonstrates the importance of an awareness of the fallacy, "If, say, 1000 possibly lagged, non-linear functions of a set of candidate exogenous variables in a model with many breaks are checked for relevance at a significance level of 0.1%, and all are indeed irrelevant, then on average **one** will be retained adventitiously..." (Hendry, 2009, p. 41)

It also appears in Gorard, Prandy and Roberts (2002, p. 11), who then describe the "prosecutor fallacy", giving an example of a fingerprint or DNA test:

"Prosecutors tend to use the probability of such a match (e.g. 1 in 10,000) as though it were the reverse of a probability of guilt (9,999 in 10,000). However, they have to argue also that there is no human error in the matching process, that the match signifies presence of the suspect at the crime scene, that presence at the scene necessarily entails guilt, and so on.

[23] Consider the following table where A-D are the four possible outcomes of a hypothesis test:

	Null Hypothesis True	Null Hypothesis False
Reject Null Hypothesis	A (false positive)	B
Accept Null Hypothesis	C	D (false negative)

The 5 percent significance level means that 19xP(A|Ho) = P(C|Ho). In other words, there is a 5 percent probability of rejecting the null hypothesis, given that the first **column** applies. It is then often assumed to mean that there is a 5 percent chance of being wrong, when the null hypothesis is rejected. This is relating A to B and is conditional on being in the first **row**. The reject criterion is unlikely to be met if the null hypothesis is true (column 1). If the reject criterion is met (row 1), it is then commonly assumed that the null hypothesis is unlikely to be true. The condition has been transposed from the column to the row. This assumption is wrong because it considers no information about the situation when the null hypothesis is false. It is based solely on probabilities assuming the null hypothesis to be true. Hence the fallacy. Note also that Type I errors (A) and Type II errors (D) are both conditional in their respective columns. When an analyst has a test result the concern is for the row-conditional probability. See also Schmidt (1996) for misinformation from significance tests in a paper written for psychologists.

Above all, they have to demonstrate that the number of potential suspects is so small that a 1 in 10,000 chance is the equivalent of 'beyond reasonable doubt'." (Gorard, et al., 2002, p. 12)

A legal example is also given by Volokh. He describes how a claim that few women make false rape claims can be misinterpreted as meaning that claims of rape are unlikely to be false (Volokh, 2005).[24] In an example based on case law, Robertson points out that it is wrong to use the point that "36 per cent of adult survivors of abuse suffered from PTSD", to assert that the presence of PTSD is evidence of this abuse (Robertson, 2003). As Robertson states, "Its value as evidence cannot be assessed without a figure for similarly placed non-abused people..."

3.2 Interpretation: change X_1 or change b_1?

There is a fundamental point that is often overlooked. Even when the policy options under consideration are restricted to the relationship between X_1 and Y, the outcome depends on both the value of X_1 and the relationship between X_1 and Y. Researchers tend to pick one of these, most commonly a change in X. For example, more education is statistically associated with higher earnings, and so a recommendation aimed at increased earnings could be to provide more education (a change in X). For some variables, such an option is not available. Consider a statistical relationship between gender and earnings. As a general rule, a person's gender cannot be changed, so a policy recommendation might be for a change in the relationship between gender and earnings through regulation or market intervention such as affirmative action on pay and/or employment. These amount to policy changes to alter b_1 rather than X_1. For many policy questions, both X_1 and b_1 may be variable, so both options should be available for consideration.

A common economics textbook illustration of this point can be seen with the treatment of externalities. Consider a market for a product with external costs of production. The standard treatment involves the addition of a "social cost" curve which comprises marginal private cost plus marginal external cost. A tax can be imposed to move the supply curve in recognition of this external cost (Doyle, 2005, p. 148; Stiglitz, 1993, p. 180). Some texts describe such an equilibrium point as the social optimum or the efficient point (Gwartney, Stroup, & Sobel, 2000, p. 128; Mankiw, 2007, p. 206; McTaggart, Findlay, & Parkin, 2003, p. 353; Sloman & Norris, 2008, p. 162). The assumption for this latter claim to be true is that there is a fixed relationship between the marginal cost of the externality and the output of the good.[25] An alternative, if the option is available, would be to target the externality directly. This would acknowledge the possibility of varying the external cost at any given level of output, which is analogous to a variation of b_1 (Mankiw, 2007, p. 217; McTaggart, et al., 2003, p. 352; Stiglitz, 1993, p. 589). Even if an optimal reduction in external costs is achieved, the outcome may not be optimal overall, given that losers are not compensated.

The approach of targeting the externality directly can be taken further. In the supply and demand diagram, the externality is measured not in terms of the volume, but in terms of the value of the externality associated with an additional unit of output. Policies that target the externality directly and vary the volume of the externality assume a fixed value (cost) per unit of externality. Instead, it may be possible to alter this value. Coase (1960) gives the example where people who are affected by an externality could move away so as to avoid the effects,

[24] Douglas Adams, in *The hitchhiker's guide to the galaxy* used the same flawed reasoning in his "proof" of the non-existence of God (Adams, Moore, Jones, Jones, & Moore, 1981, p. 60).
[25] Even that is not enough as there is no compensation paid to the losers.

thereby reducing the costs of the externality. In other words, a reduction in an external cost can be achieved through altering output, altering the production process, or altering the behaviour of those affected by the externality. In general, there may be many options available to alter the relationship between a variable, X_1 and another variable, Y.

There is an additional stage that could be added to Coase's assessment. His starting point is where the activities are in place. The courts then allocate property rights, after which the parties can then negotiate a mutually beneficial agreement.[26] As this discussion illustrates, an issue can be considered in increasing detail by allowing changes in additional variables. Instead of starting with activities in place as in Coase's farming example, consider a set of rules, or possible rules, on allocation of property rights and people considering strategies on location of an activity. The rules open up the possibility of game playing, or one party threatening or actually imposing costs on another. Coase writes of a cattle raiser expanding his activity when he would have to compensate a neighbouring crop farmer for any damage caused. He rightly suggests that this would not result in the neighbour expanding production to benefit from the compensation (assuming the crop farmer is a price taker). However, consider two cattle raisers as neighbours. If a crop farmer where to consider buying the land from one of the cattle raisers, this would impose costs on the remaining cattle raiser not directly involved in the purchase transaction. Consequently, existing rules can be used strategically (just as they can in sport, as with forcing or conceding penalties). In other words, economic analysis can become more complex, but also possibly, more realistic, through increased relaxation of ceteris paribus conditions. The role of law in Coase's example should be noted. A further step could relate to the determination of the rules. Beyond this, there are additional aspects to consider, including process and development over time with path-dependent changes in institutions and expectations, phenomena discussed further below. This illustrates the limited extent of the explanation contained in the initial supply and demand depiction of the effects of externalities. It also indicates that estimated models are bounded in their assumptions as to what can be changed, thereby limiting the range of policy options considered.

4. There are standard policy questions not covered by the econometrics

The examples in subsection 2.2 above indicate that policy conclusions may be drawn or behaviour changes suggested on the basis of statistically significant relationships between variables. This can lead to poor decisions as there are additional aspects that must be considered for a proper assessment. To illustrate, Example 3 is discussed here in more detail. While laws are unlikely to be drafted on the basis of this particular analysis, it illustrates the sort of thinking that may be used to justify policy interventions, and in some instances these involve legislative measures or decisions.

4.1.1 An example – TV watching and attention problems

In September 2007 there was media coverage of a study on childhood television viewing and attention problems (Landhuis, et al., 2007). It serves as a useful illustration of the potential problems that can arise if policy recommendations are made on statistical association alone.

[26] Coase recognised the significance of transaction costs and suggested that this might limit the negotiation. Consequently he considered it important for efficient activity that the courts make the right allocation in the first place. This point has often been missed, as he indicated in his Nobel Prize lecture (Coase, 1991).

One report in *The Press* (Hann, 2007) included the sort of information contained in a media release by the researchers (Hancox, 2007), together with further information from one of the researchers and a personal angle from a Christchurch mother. The main finding of the study was that "children who watched at least two hours of television a day were more likely to have short attention spans, and have difficulty concentrating on tasks".

Hann quoted Hancox, "Although teachers and parents have been concerned that television may be shortening the attention span of children, this is the first time that watching television has been linked to attention problems in adolescence". To put this in other words, people had suspected a causal relationship, but until now there had not even been any observed statistical relationship. Readers might be excused for thinking that a causal relationship had been found, although that is not what was said. The published study says, "As with any observational study, we were unable to prove that childhood television causes attention problems in adolescence". It also presents possible alternative explanations for the observed relationship, but reasonably suggests that there may be some causal link, and that some limiting of viewing may be prudent for heavy viewers. The study includes a recommendation, "It, therefore, seems prudent to observe the recommendation of the American Academy of Pediatrics to limit children's television viewing to a maximum of 2 hours per day" (Landhuis, et al., 2007, p. 536).[27]

There are several additional questions that could have been asked. On the statistical findings, it was found that childhood television viewing was associated with adolescent attention problems with a standardised regression coefficient of 0.12 and p of 0.0001. When adolescent television viewing was added to the equation, the coefficient fell to 0.06 and p rose to 0.0515, with results for adolescent television viewing being 0.16 and $p < 0.0001$ (Landhuis, et al., 2007, p. 534). If television viewing hours when young are correlated with viewing hours when older as this suggests, care should be taken in concluding that younger viewing causes problems later. It may not be possible to separately identify the effects of earlier viewing as suggested.

Questions could also be asked on the interpretation of the results in terms of recommended actions. Should the matter be a concern? What are "attention problems"? Are they really problems, and how serious are they? How many children have these problems, and what is the actual difference associated with extra hours of television viewing? What magnitude of benefits might be expected from reducing younger children's viewing? If viewing is reduced, what would the affected children be doing otherwise (do the average results apply to all)? If there are benefits from improved attention, what other ways might there be to bring about this change? Might any of these alternatives be easier to achieve or more effective? Are there benefits from television watching that might counterbalance the costs? It would appear that there are a number of additional questions that should be considered before deciding on policy responses.

4.1.2 Policy questions

As indicated by the example in 0, not only are there statistical issues to consider when drawing policy conclusions, but there are also a number of specific policy questions to ask. An

[27] Figure 1 of the study (Landhuis, et al., 2007, p. 535) indicated fewer attention problems among those watching for 1-2 hours per day compared to those watching less than 1 hour per day, which suggests first that the relationship may be non-linear, and second that increased viewing may be beneficial for low watchers (if the relationship is causal).

"ideal" economic approach to policy decisions (assuming perfect information and zero costs of analysis) involves identifying all the available policy options, determining their effects, valuing them to calculate costs and benefits, and then applying a decision rule to select the best option. A statistically significant relationship in a regression equation tells nothing about alternative options. Nor does it address the question of costs and benefits. All it demonstrates is that it **may** be possible to alter the value of Y by changing the value of X_1. Outstanding questions[28] include:

a. Can you change X?
b. How can X be changed?
c. At what cost?
d. How much control is there over this change (how precise are the changes in X)?
e. How variable are the effects on Y?
f. What lags are there?
g. What is the **value** of the resulting change in Y (what is the benefit, does it outweigh the cost)?[29]
h. Are there any distributional effects (gainers, losers)?
i. Are there any side-effects?
j. Are there other policy options available (including changing the relationship)?

In summary, it is important to consider the ability to change the target variable, and the costs and benefits of such a change, along with those of alternative policy options to address the same problem. This information is not provided through a t-test.

While this point may be readily understood, there is an additional dimension to consider. It illustrates a group having its own perspectives, techniques and conventions. The above questions do not fit many of the conventions and standard dimensions for critical assessment of econometric analyses. The same could apply to other disciplines and to professions. Each may have its own perspectives, techniques and conventions, and these may not be regularly re-evaluated. Group cultures and group beliefs that may not match those of other groups, and which may not stand up to careful scrutiny, are to be expected within disciplines, professions, and political and social groups. Collier presents a telling comment which may demonstrate a key the institutional barrier to broader criticism:

"Particular Institutions and false beliefs about them may be in a *functional* relation, such that the false beliefs serve to preserve the institutions that they are about...to propound the truth is not just to criticise, but to undermine the institution." (Collier, 1994, p. 172)

5. Conclusions

This paper has shown that there is a large rhetorical component in the application of results from econometric models to real world issues.

[28] While these questions are raised in relation to econometric studies, they apply to all policy options where one (policy) variable is altered so as to bring about a change in another (target) variable.
[29] This is one of the key concerns raised in Ziliak and McCloskey (2008). Note that effects on Y, and the policy significance of the resulting Y, may not always be continuous. This can cause particular problems, especially where variability of outcome assumes particular significance. Consider the difference, for a non-swimmer standing in a tank of water, between a situation where the water level is exactly at shoulder height and one where the water level is, on average, at shoulder height. Econometric estimation gives average impacts only.

At the policy stage, type C errors may arise for those who go straight from quantitative analysis to policy recommendations. This is due in part to problems in the nature of statistical tests and the interpretation of the results. The criterion commonly used in statistical tests has been questioned, especially in terms of misinterpretation of results due to the fallacy of the transposed conditional. A bigger difficulty is the limited value of econometric results for addressing many of the questions which economists might wish to ask when considering choice of policy options. The results from regression models provide estimates of specified relationships between the chosen variables. This provides part of the information required for a subset of policy options. It is part of the information because it does not consider aspects such as the changes that can be achieved, the cost and value of those changes, and alternative policy options. It is a subset of policy options because the focus is on changing variables, not relationships, to achieve the outcomes, and the search is for common determinants that might be widely effective, while policies more closely focused on individual circumstances are not considered. Economic analysis is commonly tightly structured within a theoretical framework. While that can be very helpful, it means that certain aspects are excluded or assumed away.

While many analysts may present their findings carefully, there is also the danger that others, including the media and the public, will draw false inferences from the results. They could also be misused in a political environment, as suggested by Dunn's argumentation by method (See section 2 of Birks, 2012a). Consideration of these factors and attention to the additional issues could increase the value of econometric analyses by placing them in a wider analytical context. Recognitions of the issues also results in an additional range of research questions, types of data to consider, and associated research methods, that economists could profitably consider.

References

Adams, D., Moore, S., Jones, S., Jones, P., & Moore, R. (1981). *The hitch-hiker's guide to the galaxy*: Music for Pleasure.

Allen, T. H. (1978). *New methods in social science research: policy sciences and futures research*. New York: Praeger.

Birks, S. (2000). Comments on the economic cost of family violence. In S. Birks & G. Buurman (Eds.), *Research for policy: informing or misleading* (pp. 17-25). Palmerston North, New Zealand: Centre for Public Policy Evaluation, Massey University.

Birks, S. (2002). Submission to the Law Commission on Preliminary Paper 47: Family Court Dispute Resolution. Retrieved 7 January, 2008, from http://econ.massey.ac.nz/cppe/issues/LCPP47sub.htm

Birks, S. (2007). To KISS or not to KISS. *Asymmetric Information,* (28, March), 11-12. Retrieved from http://www.nzae.org.nz/news/newsletters/AINo28-March2007.pdf

Birks, S. (2008). The Media and an Informed Electorate - An Economist's Perspective. *Power & Place: Refereed Proceedings of the Australian & New Zealand Communication Association Conference, Wellington, July 9-11*. Retrieved from http://www.massey.ac.nz/?aaac11238s

Birks, S. (2012a). Rethinking economics: Downs with traction. *Bristol Business School Economics Papers Series, 12*(14). Retrieved from http://www2.uwe.ac.uk/faculties/BBS/BUS/Research/economics2012/1214.pdf

Birks, S. (2012b). Rethinking economics: Logical gaps - theory to empirical. *Bristol Business School Economics Papers Series, 12*(16). Retrieved from http://www2.uwe.ac.uk/faculties/BBS/BUS/Research/economics2012/1216.pdf

Birks, S. (2012c). 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

Bryce, J. (1929). *Modern democracies* (Vol. 1). London: Macmillan.

Clarke, A. C. (1953). *Childhood's end*. New York: Harcourt, Brace and World.

Coase, R. H. (1960). The Problem of Social Cost. *Journal of Law and Economics, 3*, 1-44.

Coase, R. H. (1991). The institutional structure of production - Nobel Prize lecture. Retrieved 3 March 2012, from http://nobelprize.org/nobel_prizes/economics/laureates/1991/coase-lecture.html

Cobb, R. W., & Ross, M. H. (Eds.). (1997). *Cultural strategies of agenda denial: Avoidance, attack, and redefinition*. Lawrence: University Press of Kansas.

Cohen, J. (1994). The Earth Is Round (p < .05). *American Psychologist, 49*(12), 997-1003.

Collier, A. (1994). *Critical realism: an introduction to Roy Bhaskar's philosophy*. London: Verso.

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

Doyle, E. (2005). *The economic system*. Chichester: Wiley.

Dunn, W. N. (2004). *Public policy analysis: an introduction* (3rd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Durie, E. T. (1999). Ethics and values. *Indigenous peoples and the law,* (November). Retrieved from http://www.kennett.co.nz/law/indigenous/1999/39.html

Fergusson, D. M. (1998). The Christchurch Health and Development Study: an overview and some key findings. *Social Policy Journal of New Zealand,* (10), 154-176.

Gorard, S., Prandy, K., & Roberts, K. (2002). An introduction to the simple role of numbers in social science research (ESRC Occasional Paper Series, Paper 53) Available from http://www.tlrp.org/rcbn/capacity/Papers/roleofnumbers.pdf

Gwartney, J. D., Stroup, R. L., & Sobel, R. S. (2000). *Economics: Private and public choice* (9 ed.). Fort Worth: Dryden.

Hamilton, J. T. (2004). *All the news that's fit to sell: how the market transforms information into news*. Princeton, N.J.: Princeton University Press.

Hancox, R. J. (2007). Childhood TV viewing linked to teenage attention problems, Media release, 4 September, Dunedin School of Medicine, Department of Preventive and Social Medicine. Retrieved from http://dunedinstudy.otago.ac.nz/Publications%20and%20Documents/Summary%20and%20Media%20Sheets/TV%20attention%20FINAL.pdf

Hann, A. (2007, 5 September). TV linked to teen attention problems. *The Press*. Retrieved from http://www.stuff.co.nz/4190202a7144.html

Hanushek, E. A. (1997). Outcomes, Incentives, and Beliefs: Reflections on Analysis of the Economics of Schools. *Educational Evaluation and Policy Analysis, 19*(4), 301-308.

Harkrider, J. (Ed.). (2005). *Econometrics: Legal, Practical, and Technical Issues*: American Bar Association.

Hendry, D. F. (2009). The Methodology of Empirical Econometric Modelling: Applied Econometrics Through the Looking-Glass. In T. C. Mills & K. Patterson (Eds.), *Palgrave Handbook of Econometrics* (Vol. 2: Applied Econometrics, Chapter 1, pp. 3-67). Basingstoke: Palgrave Macmillan.

Hill, R. (2007, 5 November). Binge drinking 'damages brains'. *The Dominion Post*. Retrieved from http://www.stuff.co.nz/4261574a20475.html

Jeffries, S. (2001). *Gender Judgments: An Investigation of Gender Differentiation in Sentencing and Remand in New Zealand*. Christchurch: University of Canterbury.

Landhuis, C. E., Poulton, R., Welch, D., & Hancox, R. J. (2007). Does Childhood Television Viewing Lead to Attention Problems in Adolescence? Results From a Prospective Longitudinal Study. *Pediatrics, 120*(3), 532-537.

Latour, B. (1987). *Science in action: how to follow scientists and engineers through society*. Cambridge, Mass.: Harvard University Press.

Law Commission. (2002). Preliminary Paper 47: Family Court Dispute Resolution Available from http://www.lawcom.govt.nz/UploadFiles/Publications/Publication_87_207_PP47.pdf

Mankiw, N. G. (2007). *Essentials of Economics* (4 ed.). Mason: Thomson.

Maré, D. C., & Timmins, J. (2007). Geographic Concentration and Firm Productivity Available from http://www.med.govt.nz/upload/51016/07-01.pdf

Mayman, M. (2003). Raised voices: ending intimate violence: Lecture two - heroes of their own lives: women and children empowered (2 December). Retrieved from http://satrs.standrews.org.nz/mmiv2of2.doc

McCloskey, D. N. (1998). *The rhetoric of economics* (2nd ed.). Madison, Wis.: University of Wisconsin Press.

McTaggart, D., Findlay, C., & Parkin, M. (2003). *Economics* (4 ed.). Frenchs Forest: Pearson Education.
Medical update. (2002, July). Smoke screen. *Reader's Digest (New Zealand Edition), 161,* 19.

NZPA and Reuters. (2007, 1 November). Obesity, inactivity 'strongly linked to cancer'. *Stuff,* (1 November). Retrieved from http://www.stuff.co.nz/4257962a20475.html

Palmer, R. (2007, 30 November). Cancer study uncovers job risks. *The Dominion Post*. Retrieved from http://www.stuff.co.nz/4297221a19716.html

Perry, M. (2008, 18 January). Solariums found to raise skin cancer risk in young. *New Zealand Herald*. Retrieved from http://www.nzherald.co.nz/section/story.cfm?c_id=204&objectid=10487507

Pope John Paul II. (1998). Encyclical letter fides et ratio of the Supreme Pontiff John Paul II to the Bishops of the Catholic Church on the relationship between faith and reason (September 14). Retrieved from http://www.vatican.va/holy_father/john_paul_ii/encyclicals/documents/hf_jp-ii_enc_15101998_fides-et-ratio_en.html

Robertson, B. (2003). Evidence. *New Zealand Law Journal,* (July), 252.

Schattschneider, E. E. (1960). *The semisovereign people; a realist's view of democracy in America*. New York: Holt, Rinehart and Winston.

Schmidt, F. L. (1996). Statistical Significance Testing and Cumulative Knowledge in Psychology: Implications for Training of Researchers. *Psychological Methods, 1*(2), 115-129.

Simon, H. A. (1971). Designing organisations for an information-rich world. In M. Greenberger (Ed.), *Computers, communications, and the public interest* (pp. 38-52). Baltimore: The Johns Hopkins Press.

Sloman, J., & Norris, K. (2008). *Principles of economics* (2nd ed.). Frenchs Forest, N.S.W.: Pearson Education Australia.

Snively, S. L. (1994). *The New Zealand economic cost of family violence*. Wellington, N.Z.: Family Violence Unit, Dept. of Social Welfare.

Spanos, A. (2008). Review of Stephen T. Ziliak and Deirdre N. McCloskey's The cult of statistical significance: how the standard error costs us jobs, justice, and lives. *Erasmus Journal for Philosophy and Economics, 1*(1), 154-164.

Stiglitz, J. E. (1993). *Economics*. New York: Norton.

Swann, G. M. P. (2006). *Putting econometrics in its place: a new direction in applied economics*. Cheltenham, UK: Edward Elgar.

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.

Thurow, L. C. (1983). *Dangerous currents: the state of economics*. Oxford: Oxford University Press.

Volokh, E. (2005, 15 April). Research on false rape reports. Retrieved 15 December, 2008, from http://volokh.com/posts/1113597035.shtml

Wolf, A. (2007). Misleading statistical studies. *Policy Quarterly, 3*(4), 22-29.

Ziliak, S. T., & McCloskey, D. N. (2004). Size matters: the standard error of regressions in the American Economic Review. *Journal of Socio-Economics, 33*(5), 527-546.

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

Author contact: K.S.Birks@massey.ac.nz

SUGGESTED CITATION:
Stuart Birks, "Rethinking economics: Logical gaps – empirical to the real world", *real-world economics review*, issue no. 62, 15 December 2012, pp. 51-67, http://www.paecon.net/PAEReview/issue62/Birks62.pdf

You may post and read comments on this paper at http://rwer.wordpress.com/2012/12/14/rwer-issue-62/