real-world economics review, issue no. 64
subscribe for free

Capturing causality in economics and the limits of statistical inference

Lars Pålsson Syll [Malmö University, Sweden]

> "Causal inference from observational data presents many difficulties, especially when underlying mechanisms are poorly understood. There is a natural desire to substitute intellectual capital for labor, and an equally natural preference for system and rigor over methods that seem more haphazard. These are possible explanations for the current popularity of statistical models.
>
> Indeed, far-reaching claims have been made for the superiority of a quantitative template that depends on modeling – by those who manage to ignore the far-reaching assumptions behind the models. However, the assumptions often turn out to be unsupported by the data. If so, the rigor of advanced quantitative methods is a matter of appearance rather than substance." David Freedman**:** *Statistical Models and Causal Inference*

Introduction

A few years ago, Armin Falk and James Heckman published an acclaimed article in the journal *Science*. The authors – both renowned economists – argued that both field experiments and laboratory experiments are basically facing the same problems in terms of generalizability and external validity – and that *a fortiori* it is impossible to say that one would be better than the other.

What is striking when reading both Falk & Heckman (2009) and advocates of field experiments – such as Levitt & List (2009) – is that field studies and experiments are both very similar to theoretical models. They all have the same basic problem – they are built on rather artificial conditions and have difficulties with the trade-off between internal and external validity. The more artificial conditions, the more internal validity – but also less external validity. The more we rig experiments/field studies/models to avoid confounding factors, the less the conditions are reminiscent of the real target system. To that extent, Falk & Heckman are probably right in their comments on the discussion of the field vs. experiments in terms of realism – the nodal issue is not about that, but basically about how economists using different isolation strategies in different "nomological machines" attempt to learn about causal relationships. By contrast with Falk and Heckman and advocates of field experiments, as Steven Levitt and John List, I doubt the generalizability of *both* research strategies, because the probability is high that causal mechanisms are different in different contexts and that lack of homogeneity/ stability/invariance does not give us warranted export licenses to the "real" societies or economies.

Experiments, field studies and the quest for external validity

If you mainly conceive of experiments or field studies as heuristic tools, the dividing line between, say, Falk & Heckman and Levitt & List, is probably difficult to perceive. But if we see experiments or field studies as theory tests or models that ultimately aspire to say something about the real target system, then the problem of external validity is central (and was for a long time also a key reason why behavioural economists had trouble getting their research results published).

Assume that you have examined how the work performance of Chinese workers, A, is affected by B ("treatment"). How can we extrapolate/generalize to new samples outside the original population (e.g. to the US)? How do we know that any replication attempt "succeeds"? How do we know when these replicated experimental results can be said to justify inferences made in samples from the original population? If, for example, P(A|B) is the conditional density function for the original sample, and we are interested in doing an extrapolative prediction of E [P(A|B)], how can we know that the new sample's density function P' is identical with the original? Unless we can give some really good argument for this being the case, inferences built on P(A|B) is not really saying anything on that of the target system's P'(A|B).

This is the heart of the matter. External validity/extrapolation/generalization is founded on the assumption that we can make inferences based on P(A|B) that is exportable to other populations for which P'(A|B) applies. Sure, if one can convincingly show that P and P' are similar enough, the problems are perhaps surmountable. But arbitrarily just introducing functional specification restrictions of the type invariance/stability/homogeneity, is, at least for an epistemological realist far from satisfactory. And often it is – unfortunately – arbitrary specifications of this kind that ultimately underpin neoclassical economists' models/experiments/field studies.

By this I do not mean to say that empirical methods *per se* are so problematic that they can never be used. On the contrary, I am basically – though not without reservations – in favour of the increased use of experiments and field studies within economics. Not least as an alternative to completely barren bridge-less axiomatic-deductive theory models. My criticism is more about aspiration levels and what we believe we can achieve with our mediational epistemological tools and methods in the social sciences.

Many experimentalists claim that it is easy to replicate experiments under different conditions and therefore *a fortiori* easy to test the robustness of experimental results. But is it really that easy? If in the example given above, we run a test and find that our predictions were not correct - what can we conclude? That B "works" in China but not in the US? Or that B "works" in a backward agrarian society, but not in a post-modern service society? That B "worked" in the field study conducted in year 2008, but not in year 2013? Population selection is almost never simple. Had the problem of external validity only been about inference from sample to population, this would be no critical problem. But the really interesting inferences are those we try to make from specific labs/experiments/fields to specific real world situations/institutions/structures that we are interested in understanding or explaining. And then the population problem is more difficult to tackle.

Randomization – in search of a gold standard for evidence-based theories

Evidence-based theories and policies are highly valued nowadays. Randomization is supposed to best control for bias from unknown confounders. The received opinion is that evidence based on randomized experiments therefore is the best. More and more economists have also lately come to advocate randomization as the principal method for ensuring being able to make valid causal inferences.

Renowned econometrician Ed Leamer (2010) has responded to these allegations, maintaining that randomization is not sufficient, and that the hopes of a better empirical and quantitative macroeconomics are to a large extent illusory. Randomization promises more than it can deliver, basically because it requires assumptions that in practice are not possible to maintain:

> "We economists trudge relentlessly toward Asymptopia, where data are unlimited and estimates are consistent, where the laws of large numbers apply perfectly and where the full intricacies of the economy are completely revealed. But it's a frustrating journey, since, no matter how far we travel, Asymptopia remains infinitely far away. Worst of all, when we feel pumped up with our progress, a tectonic shift can occur, like the Panic of 2008, making it seem as though our long journey has left us disappointingly close to the State of Complete Ignorance whence we began.
>
> The pointlessness of much of our daily activity makes us receptive when the Priests of our tribe ring the bells and announce a shortened path to Asymptopia… We may listen, but we don't hear, when the Priests warn that the new direction is only for those with Faith, those with complete belief in the Assumptions of the Path. It often takes years down the Path, but sooner or later, someone articulates the concerns that gnaw away in each of us and asks if the Assumptions are valid… Small seeds of doubt in each of us inevitably turn to despair and we abandon that direction and seek another…
>
> Ignorance is a formidable foe, and to have hope of even modest victories, we economists need to use every resource and every weapon we can muster, including thought experiments (theory), and the analysis of data from nonexperiments, accidental experiments, and designed experiments. We should be celebrating the small genuine victories of the economists who use their tools most effectively, and we should dial back our adoration of those who can carry the biggest and brightest and least-understood weapons. We would benefit from some serious humility, and from burning our 'Mission Accomplished' banners. It's never gonna happen.
>
> Part of the problem is that we data analysts want it all automated. We want an answer at the push of a button on a keyboard… Faced with the choice between thinking long and hard versus pushing the button, the single button is winning by a very large margin.
>
> Let's not add a 'randomization' button to our intellectual keyboards, to be pushed without hard reflection and thought."

Especially when it comes to questions of causality, randomization is nowadays considered some kind of “gold standard”. But just as econometrics, randomization is basically a deductive method. Given the assumptions (such as manipulability, transitivity, Reichenbach probability principles, separability, additivity, linearity etc) these methods deliver deductive inferences. The problem, of course, is that we will never completely know when the assumptions are right. As Nancy Cartwright (2007) formulates it:

> “We experiment on a population of individuals each of whom we take to be described (or ‘governed’) by the same *fixed causal structure* (albeit unknown) and *fixed probability measure* (albeit unknown). Our deductive conclusions depend on that very causal structure and probability. How do we know what individuals beyond those in our experiment this applies to?... The [randomized experiment], with its vaunted rigor, takes us only a very small part of the way we need to go for practical knowledge. This is what disposes me to warn about the vanity of rigor in [randomized experiments].”

Although randomization may contribute to controlling for confounding, it does not guarantee it, since genuine ramdomness presupposes infinite experimentation and we know all real experimentation is finite. And even if randomization may help to establish average causal effects, it says nothing of individual effects unless homogeneity is added to the list of assumptions.

Real target systems are seldom epistemically isomorphic to our axiomatic-deductive models/systems, and even if they were, we still have to argue for the external validity of the conclusions reached from within these epistemically convenient models/systems. Causal evidence generated by randomization procedures may be valid in “closed” models, but what we usually are interested in is causal evidence in the real target system we happen to live in.
So, when does a conclusion established in population X hold for target population Y? Usually only under very restrictive conditions! As Nancy Cartwright (2011) – succinctly summarizing the value of randomization - writes:

> “But recall the logic of randomized control trials… They are ideal for supporting ‘it-works-somewhere’ claims. But they are in no way ideal for other purposes; in particular they provide no better bases for extrapolating or generalising than knowledge that the treatment caused the outcome in any other individuals in any other circumstances… And where no capacity claims obtain, there is seldom warrant for assuming that a treatment that works somewhere will work anywhere else. (The exception is where there is warrant to believe that the study population is a representative sample of the target population – and cases like this are hard to come by.)”

Ideally controlled experiments (the benchmark even for natural and quasi experiments) tell us with certainty what causes what effects – but only given the right closures. Making appropriate extrapolations from (ideal, accidental, natural or quasi) experiments to different settings, populations or target systems, is not easy. “It works there” is no evidence for “it will work here.” Causes deduced in an experimental setting still have to show that they come with an export-warrant to the target population/system. The causal background assumptions made have to be justified, and without licenses to export, the value of “rigorous” and “precise” methods is despairingly small.

Here I think Leamer's button metaphor is appropriate. Many advocates of randomization want to have deductively automated answers to fundamental causal questions. But to apply "thin" methods we have to have "thick" background knowledge of what's going on in the real world, and not in (ideally controlled) experiments. Conclusions can only be as certain as their premises – and that also goes for methods based on randomized experiments.

An interesting example that illustrates some of the problems with randomization – spillovers and the bridging of the micro-macro gap – was recently presented in an article by Pieter Gautier *et al.* (2012):

> "In new research, we study a Danish job search assistance programme which, according to a randomised experiment, leads to large positive effects on exit rates to work... We show, however, that because of spillover effects, a large-scale implementation will only marginally reduce unemployment without increasing welfare...
>
> The empirical results suggest that considering both negative and positive spillover effects is important when evaluating the job search assistance programme. The Danish programme essentially increases the job search effort of participants by requiring them to make more job applications. The effect on vacancy supply is modest, so when participants send out more applications, this reduces the probability that a specific job application gets selected.
>
> It is often argued that randomised experiments are the golden standard for such evaluations. However, it is well know that a randomised experiment only provides a policy-relevant treatment effect when there are no spillovers between individuals. In the study discussed above, we have shown that spillovers can be substantial. Despite the success of a small-scale implementation of the programme at the micro level, we find it to be ineffective at the macro level. The results of our study are no exception."

So this example does pretty well explain one reason for randomized controlled trials not at all being the "gold standard" that it has lately often been portrayed as. Randomized controlled trials usually do not provide evidence that their results are exportable to other target systems. The almost religious belief with which its propagators portray it, cannot hide the fact that randomized controlled trials cannot be taken for granted to give *generalizable* results. That something works somewhere is no warranty for it to work for us or even that it works *generally*.

Econometrics and the difficult art of making it count

In an article that attracted much attention, renowned econometrician and Nobel laureate James Heckman (2005) writes (emphasis added):

> "A model is a set of possible counterfactual worlds constructed under some rules. The rules may be laws of physics, the consequences of utility maximization, or the rules governing social interactions ... *A model is in the mind. As a consequence, causality is in the mind*."

Even though this is a standard view among econometricians, it is – at least from a realist point of view – rather untenable. The reason we as scientists are interested in causality is that it's a part of the way the world works. We *represent* the workings of causality in the real world by means of models, but that doesn't mean that causality isn't a fact pertaining to relations and structures that exist in the real world. If it was only "in the mind," most of us couldn't care less.
The reason behind Heckman's and most other econometricians' nominalist-positivist view of science and models, is the belief that science can only deal with observable regularity patterns of a more or less lawlike kind. Only data matters, and trying to (ontologically) go beyond observed data in search of the underlying real factors and relations that generate the data is not admissable. All has to take place in the econometric mind's model since the real factors and relations according to the econometric methodology are beyond reach since they allegedly are both unobservable and immeasurable. This also means that instead of treating the model-based findings as interesting *clues* for digging deeper into real structures and mechanisms, they are treated as the *end points* of the investigation. Or as Asad Zaman (2012) puts it:

> "Instead of taking it as a first step, as a clue to explore, conventional econometric methodology terminates at the discovery of a good fit... Conventional econometric methodology is a failure because it is merely an attempt to find patterns in the data, without any tools to assess whether or not the given pattern reflects some real forces which shape the data."

David Freedman (2010) raises a similar critique:

> "In my view, regression models are not a particularly good way of doing empirical work in the social sciences today, because the technique depends on knowledge that we do not have. Investigators who use the technique are not paying adequate attention to the connection – if any – between the models and the phenomena they are studying. Their conclusions may be valid for the computer code they have created, but the claims are hard to transfer from that microcosm to the larger world.
>
> Given the limits to present knowledge, I doubt that models can be rescued by technical fixes. Arguments about the theoretical merit of regression or the asymptotic behavior of specification tests for picking one version of a model over another seem like the arguments about how to build desalination plants with cold fusion and the energy source. The concept may be admirable, the technical details may be fascinating, but thirsty people should look elsewhere."

Most advocates of econometrics and regression analysis want to have deductively automated answers to fundamental causal questions. Econometricians think – as David Hendry expressed it in *Econometrics – alchemy or science?* (1993) – they "have found their Philosophers' Stone; it is called regression analysis and is used for transforming data into "significant' results!" But as David Freedman (2010) poignantly notes – "Taking assumptions for granted is what makes statistical techniques into philosophers' stones." To apply "thin" methods we have to have "thick" background knowledge of what is going on in the real world, and not in idealized models. Conclusions can only be as certain as their premises – and that also applies to the quest for causality in econometrics and regression analysis.

Without requirements of depth, explanations most often do not have practical significance. Only if we search for and find fundamental structural causes, can we hopefully also take effective measures to remedy problems like e.g. mass unemployment, poverty, discrimination and underdevelopment. A social science must try to establish what relations exist between different phenomena and the systematic forces that operate within the different realms of reality. If econometrics is to progress, it has to abandon its outdated nominalist-positivist view of science and the belief that science can only deal with observable regularity patterns of a more or less law-like kind. Scientific theories ought to do more than just describe event-regularities and patterns – they also have to analyze and describe the mechanisms, structures, and processes that give birth to these patterns and eventual regularities.

Modern econometrics is fundamentally based on assuming – usually without any explicit justification – that we can gain causal knowledge by considering independent variables that may have an impact on the *variation* of a dependent variable. This is however, far from self-evident. Often the *fundamental* causes are *constant* forces that are not amenable to the kind of analysis econometrics supplies us with. Or as Stanley Lieberson (1985) has it:

> "One can always say whether, in a given empirical context, a given variable or theory accounts for more variation than another. But it is almost certain that the variation observed is not universal over time and place. Hence the use of such a criterion first requires a conclusion about the variation over time and place in the dependent variable. If such an analysis is not forthcoming, the theoretical conclusion is undermined by the absence of information...
>
> Moreover, it is questionable whether one can draw much of a conclusion about causal forces from simple analysis of the observed variation... To wit, it is vital that one have an understanding, or at least a working hypothesis, about what is causing the event *per se*; variation in the magnitude of the event will not provide the answer to that question."

Our admiration for technical virtuosity should not blind us to the fact that we have to have a more cautious attitude towards probabilistic inference of causality in economic contexts. Science should help us penetrate to "the true process of causation lying behind current events" and disclose "the causal forces behind the apparent facts" (Keynes (1971-89), **XVII**). We *should* look out for causal relations, but econometrics can never be more than a starting point in that endeavour, since econometric (statistical) explanations are not explanations in terms of mechanisms, powers, capacities or causes. Firmly stuck in an empiricist tradition, econometrics is basically concerned with the *measurable* aspects of reality. But there is always the possibility that there are other variables – of vital importance and, although perhaps unobservable and non-additive, not necessarily epistemologically inaccessible – that were not considered for the model. Those who *were* can hence never be *guaranteed* to be more than potential causes and not real causes.

A rigorous application of econometric methods in economics really presupposes that the phenomena of our real world economies are ruled by stable causal relations between variables. A perusal of the leading econometric journals shows that most econometricians still concentrate on fixed parameter models and that parameter values estimated in specific spatio-temporal contexts are simply *assumed* to be exportable to totally different contexts. To warrant this assumption one, however, has to convincingly establish that the targeted acting causes are stable and invariant, so that they maintain their parametric status after the bridging. The endemic lack of predictive success of the econometric project indicates that this hope of finding fixed parameters is a hope for which there really is no other ground than hope itself.

Keynes's (1951(1926)) critique of econometrics and inferential statistics was based on the view that real world social systems are not governed by stable causal mechanisms or capacities:

> "The atomic hypothesis which has worked so splendidly in Physics breaks down in Psychics. We are faced at every turn with the problems of Organic Unity, of Discreteness, of Discontinuity – the whole is not equal to the sum of the parts, comparisons of quantity fails us, small changes produce large effects, the assumptions of a uniform and homogeneous continuum are not satisfied. Thus the results of Mathematical Psychics turn out to be derivative, not fundamental, indexes, not measurements, first approximations at the best; and fallible indexes, dubious approximations at that, with much doubt added as to what, if anything, they are indexes or approximations of."

The kinds of laws and relations that econometrics has established, are laws and relations about entities in models that presuppose causal mechanisms being atomistic and additive. When causal mechanisms operate in real world social target systems they only do it in ever-changing and unstable combinations where the whole is more than a mechanical sum of parts. If economic regularities obtain they do it (as a rule) only because we engineered them for that purpose. Outside man-made "nomological machines" they are rare, or even non-existent. Unfortunately that also makes most of the achievements of econometrics – as most of contemporary endeavours of economic theoretical modeling – rather useless.

Conclusion

Statistics and econometrics should not – as already Keynes (1973(1921)) argued – primarily be seen as means of inferring causality from observational data, but rather as descriptions of patterns of associations and correlations that we may use as *suggestions* of possible causal relations.

Causality in social sciences – and economics – can never solely be a question of statistical inference. Causality entails more than predictability, and to really in depth explain social phenomena require theory. Analysis of variation – the foundation of all econometrics – can never in itself reveal *how* these variations are brought about. First when we are able to tie actions, processes or structures to the statistical relations detected, can we say that we are getting at relevant explanations of causation. Too much in love with axiomatic-deductive modeling, neoclassical economists especially tend to forget that accounting for causation – *how* causes bring about their effects – demands deep subject-matter knowledge and acquaintance with the intricate fabrics and contexts.

References

Cartwright, Nancy (2007): “Are RCT's the Gold Standard?” *Biosocieties*, **2**, 11-20.

– (2011): “Will this Policy Work for You? Predicting Effectiveness Better: How Philosophy Helps,” *Presidential Address, PSA 2010*.

Falk, Armin & Heckman, James (2009), “Lab Experiments Are a Major Source of Knowledge in the Social Sciences,” *Science* 23 October.

Freedman, David (2010), *Statistical Models and Causal Inference,* Cambridge: Cambridge University Press.

Gautier, Pieter *et al*. (2012), “Estimating equilibrium effects of job search assistance,” CEPR Discussion Papers 9066, C.E.P.R. Discussion Papers.

Heckman, James (2005), “The Scientific Model of Causality,” Sociological Methodology*, **35**, 1–97.*

Hendry, David F. (1993), *Econometrics: alchemy or science?* Oxford: Blackwell.

Keynes, John Maynard (1951 (1926)), *Essays in Biography.* London: Rupert Hart-Davis.

– (1971-89), *The Collected Writings of John Maynard Keynes*, vol. I-XXX. D E Moggridge & E A G Robinson (eds), London: Macmillan.

– (1973 (1921)), *A Treatise on Probability*. Volume VIII of *The Collected Writings of John Maynard Keynes*, London: Macmillan.

Leamer, Edward (2010), “Tantalus on the Road to Asymptopia,” *Journal of Economic Perspectives,* **24***,* 31–46.

Lieberson, Stanley (1985), *Making it count: the improvement of social research and theory*, Berkeley: University of California Press.

Levitt, Steven & List, John (2009), “Field experiments in economics: The past, the present, and the future,” *European Economic Review* **53**, 1-18.

Zaman, Asad (2012), “Methodological Mistakes and Econometric Consequences,” *International Econometric Review*, **4**, 99-122.

Author contact: Lars Pålsson Syll lars.palsson-syll@mah.se

SUGGESTED CITATION:
Lars Pålsson Syll, “Capturing causality in economics and the limits of statistical inference”, *real-world economics review*, issue no. 64, 2 July 2013, pp. 81-89, http://www.paecon.net/PAEReview/issue64/Syll64.pdf