

Empirics and the public good

University of Wollongong Annual Economic and Social Policy Lecture 2026

Michael Brennan | 17 September 2026

Let's cut to the chase. No speech about economics and the social sciences can ignore our context: the world is a troubling place.

We are living through a period of heightened anxiety – rapid change and the equally rapid disintegration of some of old verities: Faith in democracy, a world order built on peaceful coexistence and mutual gains from trade. A rough consensus about the proper role of governments, markets and communities.

In their place: hot wars, a fraying trading order, democracies fragmenting from within, a climate transition whose costs keep colliding with the costs of climate change itself, and fiscal models built for an era of ever-expanding services now creaking under their own weight.

There is a desensitising element to all this. A lack of bandwidth. A temptation for the citizen to turn away or turn inward.

But what about the social scientist? What are we to do?

For us, the inward turn would be to forget the world, retreat into the workshop, and focus our energies on the things we can know about and control.

Is that our lot? Or is to get out there and wrestle with the big, ugly questions that come at us, not necessarily of our choosing?

Just to pose the question is to answer it. What is the point of the social sciences, if not to be useful at times like this?

After all, for most of us, the taxpayer has made a substantial investment in our knowledge and skill set. Are they not entitled to a return?

Moreover, surely we want to be useful. The biggest problem for many of us is working out how.

My contention is that being useful, as social scientists, has to mean engaging fully with the world around us, but doing so in a very specific way. A way that fits with our distinctive contribution.

We always have to remember that the social sciences differ from physics or chemistry. It is chastening to reflect on the extent of progress in those hard sciences over decades – the medical cures, the engineering feats, the moon landings.

On the face of it, progress in the social sciences looks and feels a bit different to the technological breakthroughs that came via the physical sciences.

We shouldn't be defeatist about that. What all science shares is a rigorous approach to identifying *what is not true*. Ideally, we can shrink the remaining 'credible worlds' down to a smaller set. In the physical sciences, this happens with great precision.

The economist's set of credible worlds does not narrow as much. We can't really say that we have found the precise policy technology that solves poverty, chronic disease prevention or even regional economic development.

But still, we know more than we once did. The things we know more about mostly concern frameworks – public goods, rational expectations, game theory or difference in differences analysis.

And these frameworks are invaluable in that core scientific task of seeing and explaining what isn't true.

Given the challenge of pin-pointing narrow technological solutions, one could even argue that the comparative advantage of economics is that task of exposing fallacy and contradiction. Of rebutting bad policy ideas, especially when those ideas are paraded as though they involve no cost or any painful trade-off.

This is one source of great value in economic reasoning.

We differ somewhat from the physical sciences partly because humans are infinitely more complex in their behaviour than are molecules or bacteria.

But there is another thing. Even with all the computational power in the world, we would come back to the basic problem that – in economics – we cannot make inter-personal comparisons of utility.

That is, we have no way to scientifically weigh the interests of one person against those of another when they are in conflict.

Genuine Pareto improvements are vanishingly rare in public policy. They are common in everyday life: they happen all the time in market exchange. But the issues that fall to public policy to deal with almost always involve some gains and losses across individuals, on which the economist is not qualified to arbitrate.

Those are value judgments – only resolved by other means; best resolved at the ballot box.

Not because politics is especially good at reconciling competing worldviews, but because it's a peaceful way of accommodating conflicts of interest that are, at root, irreconcilable.

Economics is a tool for social explanation. Not for social optimisation or control.

Despite the language economists use – optimise, maximise, solve – our job is generally not to calculate or pinpoint the 'right' choice. It is to offer a rigorous framework for thinking through the issues: the trade-offs, the unintended consequences, the way individual actions respond to incentives and aggregate into social outcomes, and how policy, institutions or the rules of the game might change all of that.

And so, our role is mainly to educate, not to advocate.

We are more like cartographers than captains. We can draw the map of the terrain – helping people see a version of order from fragmented experience – but we don't choose the destination, and we have no special authority to steer the ship.

That is because when economists enter the public debate, their primary audience is actually not the government, nor should it be. The main audience is the public.

If we are true to our mission and rise to the occasion in the face of what seem like unprecedentedly complex and grave threats, then it is because we help the community, through a higher quality public debate, to make sense of the choices.

We should never be glib about policy. It is almost always hard. And the community gets benefit from a candid explanation as to why that is: the trade-offs and competing priorities inherent in any policy choice.

Rather than lecturing governments, it often means defending them. Governments face hard choices every day – the economist has to explain why services can't expand without some rise in taxes, why rent caps have unintended consequences, why tariffs get paid by domestic consumers rather than foreign producers. Nothing in public policy is easy; if it were, it would have been done already.

At the e61 Institute, we often talk about our role in highlighting drivers and trade-offs. That is, the underlying causes of an observed problem, and the pros and cons of different responses.

We have put out short explainers on tobacco excise: noting the vexed dilemma that addressing the black market probably requires a substantial cut in tobacco excise. As a country, by raising the excise we got ourselves on the wrong side of the so-called Laffer Curve (the relationship between tax rates and tax revenue, which is usually positive but

can become negative when high taxes lead to an activity being reduced or the taxes being avoided).

But the solution of an excise cut is also vexed. First, it's a big win for big tobacco. Many people (including public health advocates) will be unhappy with that. Second, we can't be sure that cutting the excise now will recover all the lost revenue. Once the black market steps in, it's hard to win back market share.

So, it's hard. Harder than it looks.

We have done similar explainers on resource taxation. Not to advocate a single, optimal approach; but more to highlight that there is no free lunch. If we want resource projects, then we have to try and tax the economic rents from them, but not the normal return; Nor can we just impose a levy on an arbitrary slice of revenue.

But to tax rents, we need to take account of project risk. And to do that, government either has to explicitly share risk with private investors in the construction phase or provide some allowance for risk in calculating the rate of return above which rent taxes apply.

There are alternative designs to the current Petroleum Resource Rent Tax, but no simple answers that offer uncomplicated rivers of gold.

In these and other examples, we try to bring quantitative rigour to the task of explaining, which is where economists and other social scientists are on strongest ground. What this implies is that if we are to meet the challenges of the moment, we need a rare combination of ambition and humility.

More ambition to focus on the biggest social problems of our time. But more humility about what we know and the limits of our competence in advocating solutions, as distinct from our important educative role.

The rise of empirical research

Today I want to explore that conjunction of ambition and humility in a particular context: namely how economists do empirical work.

Over the last two decades, economics has placed far more weight on empirical research.

To be sure, data and quantification have always mattered – for testing theoretical models and estimating the parameters that flesh them out.

But in recent years this has gone to a new level.

First there has been the significant increase in the amount of data. In Australia, this includes the creation of large, linked administrative data sets by the Australian Bureau of Statistics.

Second there have been significant advances in empirical technique. It's often called the credibility revolution, because it puts data and empirics at the fore, testing theoretical predictions against real-world evidence.

It involves a more reduced-form, more atheoretical approach: usually with a sharp focus on causal inference – working out whether a genuine causal relationship exists 'in the data,' not just a correlation.

There is still a role for theory, but it becomes more parsimonious: it's mainly to generate a hypothesis about the causal chain that we could be observing, and how we might get to an estimate of the causal effect.

Theory operates to discipline the identification strategy.

A central concept is that of the Average Treatment Effect – the impact that some event or policy has on an outcome, relative to an unobserved counterfactual (that is, what would have happened to the same individual in the absence of the event or policy).

In a perfect world, to estimate that treatment effect credibly, we typically need something that mimics random assignment: a natural experiment, a policy quirk.

Basically, some source of exogenous variation that allows us to test what happens under one scenario against what happens under an alternative one.

Obviously, this is not original to economics. The principle of randomisation started in agricultural research. It is the stock in trade of medical trials. Other approaches, like regression discontinuity (comparing what happens just below an arbitrary cut off with what happens just above it) was originally used in the education literature to test the effect of scholarships.

But economics has embraced these quantitative techniques along with explicit claims of a paradigm shift (hence the 'credibility revolution').

e61's work makes use of these core empirical techniques.

We have used the 'event study' approach to examine what happens when workers in coal-fired power stations lose their jobs.

Because of the longitudinal nature of the administrative data, and because the data is linked across a large number of variables, it is possible to identify the average effects on those displaced workers in terms of lost income, the length of time it takes to find alternative employment, the wages earned in the next job and the likelihood of having to move to another region.

Moreover, these effects can be compared to those who lose their jobs in other sectors. Other work we have done on wage-scarring shows that the effect of job loss on future earnings is highest in industries like manufacturing, lower in services and lowest in public sector roles.

We have tested the impact of the Baby Bonus on fertility rates. Using seventeen years of linked administrative birth records, e61 researchers compared births before and after the introduction of the policy, net of the underlying downward drift in the birth rate.

We found the Bonus lifted monthly births by 6.5 per cent – and, by following mothers through to 2022, we found evidence that this was a genuine, permanent increase in completed family size, not just couples moving forward births they would have had anyway.

That distinction, between shifting timing and actually changing outcomes, is what a good identification strategy is built to catch, and which a raw correlation would not accurately reveal.

So, these empirical approaches are doing real work in improving the evidence base for policy.

In neither of these examples did e61 express a value judgment on policy. Regarding the Baby Bonus, we identified that the policy had an effect on fertility.

The good, the bad and the distracting

The key question is: taking my earlier claim – that the social sciences should aim for relevance on the big issues, combined with humility about optimal policy – then what should we make of the credibility revolution?

Is this a step forward?

Has it gotten economics closer to that ideal? Has it pushed it further away, or had no effect?

Predictably, the answer to that is not simple.

The answer is never simple. My contention is that the effect has been mixed.

Much of it is good. But there are some downsides. And in some cases, I will argue, that it has been a bit of a distraction from the main game.

The good

I will start with the good.

The good news is that the heightened emphasis on empirical research has meant that Micro-economics has become much more grounded in the real world. The oft-heard traditional critique of economics – too abstract, too theoretical – is a much harder case to make today.

If the subject matter is fertility policy or job losses in traditional industries, then we are fairly anchored in the real world. At e61 we have also done work on zoning restrictions, NDIS, entrepreneurship and teacher shortages for example.

And the canonical empirical studies associated with the credibility revolution involve testing the effect of minimum wages, class sizes and charter schools to name just a few practical, real world policy interventions.

There are also cases where quantification is important in and of itself.

Our work on stamp duty is a great example.

Every economist will tell you that stamp duty is a terrible tax. Why? Because it is a transaction-based tax which therefore discourages mobility. In the Australian context, there are multiple models that will tell you it is our worst, least efficient tax.

But those models are essentially theory-driven. What we have lacked in Australia is any reliable estimates of what has actually occurred when stamp duty rates have changed. Has increased stamp duty actually led to fewer transactions, against a credible counter-factual?

And how big is the effect, really?

We looked at a natural experiment that arose when Queensland raised the stamp duty rate on owner-occupier housing in 2011, relative to the rate on investment properties.

We were able to compare the relative impact on owner-occupier transactions relative to turnover in the investor market – where the rate did not change – and to compare those results in individual localities in Queensland with statistically similar localities in the states that left stamp duty unchanged, such as New South Wales.

That latter comparison helped to ensure a credible counter-factual and to control for some of the things that might have been affecting the national housing market over the same period.

In the end, our study suggested that each one-percentage-point rise cut home purchase volumes by 7.2 per cent, with a similarly sized effect on the rate at which people simply changed address. This is an example of the reduced form, data-first approach. Not derived from a model where assumptions about behaviour are fed in at the start – but read straight off what people actually did when policy changed.

So, the empirical work supports the theory. And it gives us a sense of magnitude. In this case, the magnitude was perhaps surprisingly high.

Quantification clearly fits within the appropriate role of the economist as already discussed: to put the facts on the table, not to pinpoint the right policy response.

Then there is a third benefit of the empirical turn in economics. By being much more disciplined about causation, economists have gotten better at busting another kind of popular myth.

In a way, economists have always been avid myth-busters. We are natural contrarians. So, it's a particularly satisfying part of the job. I have already mentioned the ways economists use their conceptual toolkit to show counter-intuitive things about rent controls or tariffs.

The empiricists are equally adept at debunking another specific set of observational fallacies: namely all those correlations that masquerade as causation.

A dog-walking acquaintance of mine recently told me that dog owners live longer than non-dog owners. I have no idea of the source of that claim. I know the claim was causal, because my acquaintance highlighted the channel: walking the dog each morning is good for your mental health.

That implies a causal connection: if you get a dog, you are likely to live longer.

But the empiricist will tell us that any estimate of that effect is bound to be biased.

Dog owners clearly differ from non-dog owners in ways that could be correlated with life expectancy but which are not caused by dogs – they probably walk more (which influenced their decision to get a dog), they're more likely to be partnered, to have kids and to own a home.

Those are all confounders – things that are correlated with dog ownership and life expectancy but don't operate via any causal link between the former and the latter.

We could control for those things statistically. But the deeper problem is that dog ownership might be correlated with any number of unobservable attributes which also affect life expectancy independently.

There is no way to randomise dog ownership, so it's very hard to make a credible causal claim.

This is perhaps a trivial example. But I think it is fair to say that the focus on empirics has made economists much more alive to these rhetorical leaps from observed correlation to implicit causal claims.

So that is another positive of the empirical turn.

With all these good things, what's not to like? As they say in the classics, what can possibly go wrong?

The bad

The less positive side is that, just as the empirical turn opened up new scope for knowledge, it has also opened up new scope for that well known malady of economists: intellectual over-confidence.

The hubris is partly how much we can really 'know' from estimating a single reduced-form empirical relationship.

It is the problem of external validity. Will a result observed in a particular time and place, with a specific treated population, apply more generally? If a family violence program works in one community, how confident can we be that it will work the same way in a different one?

In medical science, we have reasonable grounds for confidence that a drug trial – if the treatment and control groups are large enough and properly randomised – will give a result of fairly general applicability. Biology doesn't vary all that much from one patient to the next.

In economics, we rarely have that luxury. Randomised control trials are expensive and hard to administer faithfully at scale. What we usually have is quasi-experimental evidence: situations in which some external event has given us something that looks a little like randomisation.

But even when the internal logic of a study is airtight – even when we're confident A really did cause B, in that place, at that time (what we call internal validity) – we've said comparatively little about whether the same mechanism holds anywhere else.

Part of the challenge is that internal and external validity can often pull in different directions: the more tightly a study isolates its effect, the narrower the population and setting to which that effect can be said to apply with any confidence.

When economists have compared the estimated effect sizes across many rigorous evaluations of the same type of intervention – the same kind of cash-transfer program, say, run in different places, they have found the results vary almost as much from one context to the next as they do across completely different kinds of interventions altogether.

Context, it turns out, often does as much work as the treatment itself.

Yet results are often reported in ways that don't fully reflect the uncertainty attached to them.

Too often we hear policy advisers saying things like ‘the evidence says...’ or ‘we know that...’ When in fact what we know, or what the evidence tells us, is much more uncertain, much more contingent than that language suggests.

That’s particularly stark in cases where economists overcome the lack of true randomisation via an instrumental variable (IV).

The IV approach has had some great successes in delivering causal estimates – a Local Average Treatment Effect (LATE) in cases where there are common causes of the treatment and the effect which would otherwise bias the estimate of treatment effect.

Using draft lotteries, assignment of defendants to judges, birth dates in relation to years of schooling, are just some of the innovations under the IV approach.

But there are reasons to believe the results may not always travel well.

What is often forgotten is that the IV estimator delivers a LATE specifically for the ‘complier’ population – those who, if they received the treatment, did so specifically because of the instrument; and, if they did not receive it, would have if the instrument was applied to them. It does not describe the treatment effect for everyone who received, or did not receive, the treatment.

So, the estimated effect is specific to the instrument used. A different instrument, isolating a different set of compliers, can produce a different answer to what looks like the same question.

But that subtlety is often lost in how results get described or used. Hence the tendency for economists to speak more confidently about a particular study’s findings than the design actually supports.

More broadly, there is arguably an irreducible endogeneity to any policy choice. We are naturally interested in what happened when Queensland reduced stamp duty. Or when a locality relaxed some regulatory restriction.

But why did they do it in the first place? Was there pent-up demand that the stamp duty or the regulation was holding back?

That underlying driver could mean the effect will be different in another state or locality.

Even if the policy change was forced by some exogenous shock (a treaty obligation, say, rather than a considered local decision), the measured effect still a LATE for the complier population– roughly, what happened to the people and places that received the treatment because of the specific instrument, against a counterfactual. There is no guarantee that a different place, weighing the *voluntary* adoption of the same policy (for whatever reasons), would experience the same treatment effect.

This is the frustration of piecemeal reduced-form work: it is still an open question whether we are building a solid structure of knowledge, brick by brick, from the results of individual empirical studies – or just accumulating a pile of loosely related facts, each one true only in its own time and place.

There are other ways that causal estimates can be over-interpreted.

Having found a statistically significant effect, it is often tempting to leap to a firm policy conclusion. But the estimated effect is almost never the end of the story.

I mentioned e61's work on the Baby Bonus and on stamp duty. In neither case did we make a definitive finding as to whether the policy was worthwhile.

Was the Baby Bonus value for money? It depends partly on the cost relative to the fertility effect. Much of the money spent is simply a transfer to parents of newborns. Whether that transfer is the right policy is a value judgment.

Stamp duty might prevent mobility, but that does not amount to a call for its abolition. To make that call, you would have to say something about how the revenue would be replaced. Almost certainly there would be winners and losers.

Economists can identify and quantify those effects but cannot adjudicate between individual winners and losers.

What this highlights is that, even in these impressively empirical times, we cannot dispense with theory.

Theory gives us more than an identification strategy. It is the vital element that allows us to generalise. Whether we acknowledge it or not, there is always a model - explicit or implicit – that motivates and sits behind reduced form empirical work.

The broader relevance of a causal estimate operates via that model. It's the model, possibly implicit, that links the individual empirical finding to counterfactual in broader contexts.

Theory is also indispensable if we are to say anything about welfare implications. And for all the problems of making inter-personal utility comparisons and solving for optimal policies, economists (again) play a vital role via welfare analysis in showing what is not true. (Or indeed showing what would need to be true in order for a particular policy to be optimal).

The distracting

There's a bigger question here: has the credibility revolution become a distraction from the main game of economics?

The drive for clean identification pulls us toward settings with proxies for randomisation – where choice has effectively been engineered away – enrolment lotteries, drafted soldiers, randomly assigned judges.

But choice is the essence of economics. In the real world, people select into treatment based on their own sense of what they stand to gain from it, and that selection process – not just its aftermath – is what an economist is actually meant to try and explain.

Take the topical issue of remote work: we can estimate an ‘effect’ of working from home through a randomised trial, but in the real world nobody is randomly assigned to remote work. Employers and employees sort into the arrangement based on what each expects to gain. Understanding that sorting is the economics.

This is why when James Heckman sought to formalise the marginal treatment effect across different individuals, he went back to the Roy model – a canonical model of discrete choice, where people select occupations, based on their perceived benefits from different choices.

When we generate a treatment effect from a quasi-random assignment, we need to ask the question: whose decision is the estimated treatment effect actually meant to inform?

If we could establish, cleanly, the effect of a charter school on student results, who is meant to use that number?

The individual parent? They would need heroic powers of introspection to know whether they are equivalent to the ‘complier’ group the estimate describes.

Is it the government, pondering whether to introduce charter schools? Perhaps, but of course, they will not be allocating enrolment by lottery. They need to understand who opts in and who opts out, and why, which is a question about choice mechanisms, not really about average treatment effects under quasi-randomisation.

Where pure causal inference is most useful, in other words, tends to be exactly the place with the least room for choice by humans. Often it's about the effectiveness of a program which involves limited choice on the part of participants.

This can be genuinely useful, but it subtly pulls us back into the world of social engineering. A world where the role of the economist is to arm the social planner with policies to administer, like medicine, for the benefit of the population.

As I have tried to argue in this speech, economists should generally steer away from that mindset.

Returning to my trivial example of dog walking, perhaps someone will take my statement about the inability to randomise dog ownership as a motivation – and find an

instrument or some other quasi-experimental approach to test a genuine causal treatment effect.

But I would still contend that the economics of this is not particularly interesting.

In fact, properly conceived, there is no economics in it at all.

That's aside from the fact that it's also a relatively trivial question.

One of the risks of the credibility revolution is that it drives the economics profession towards answering narrow questions very precisely.

Questions that are answerable with the data we have, or because a source of exogenous variation suggests itself. Not because anyone was asking the question with any particular urgency.

Whatever the intellectual merit of that pursuit, I would contend that it is not a path to policy relevance.

Policy makers don't get to choose their questions. The questions come at them – big, complicated, often genuinely ugly – and the job is to respond to it as well as possible: understand the core drivers, work through the trade-offs, rule out the worst options, buy time where you can.

Doing it well

Is there a middle ground?

A way to use rigorous empirical work to help deal with the biggest issues: explaining drivers, highlighting trade-offs and thereby improving the debate?

To use that empirical toolkit not to answer narrow questions precisely but respond to important ones as well as possible.

The starting point is to look at a policy question that actually matters and then reach for whichever tool delivers the most credible partial answer to that question, rather than the cleanest answer to a nearby one.

e61's own practice offers a few examples of what this looks like.

One is distributional microsimulation: using administrative microdata to recompute, for the whole population, what a policy change would mean under the existing rules of behaviour – who gains, who loses, by how much, and where.

We have done that to demonstrate the effects of forgiving 20 per cent of HELP loans, Highlighting the proportion of the benefit that goes to the highest income earners.

We have done it to highlight the downsides of the 50 per cent CGT discount, which has very different effects across the full distribution of real returns actually experienced by asset owners.

It doesn't attempt to estimate behavioural responses, but it can nonetheless make important points about the distributional impacts of policies.

A second is event study work that deliberately targets large, whole-of-population events, rather than a natural experiment confined to one place or one narrow group. Studying everyone who loses a job, or everyone who joins the NDIS for the first time, trades away some of the causal cleanliness of a natural experiment but achieves much more policy salience in return.

A third is decomposition: taking an observed aggregate phenomenon – as we did with the gender pay gap – and partitioning it into its component drivers.

How much reflects differences in pay within the same occupation, and how much reflects men and women sorting into different occupations in the first place? We found that a surprisingly high share of the gender pay gap (around 80 per cent) was associated with differences within occupations, rather than men and women choosing different occupations.

Again, it doesn't identify a single causal effect. But it does tell you where to look. All of these involve big issues.

And all make a similar trade-off: each gives up some causal precision in exchange for salience and meaning.

Close

So, we return to where we started. In a fragmenting, disrupted world, the social scientist needs to be part of the response – but that means wrestling with the big, messy problems, not retreating into the workshop for tidy answers to narrow ones.

The credibility revolution has been good for economics, in real and durable ways. It has made us more grounded in the real world, more disciplined about causation, harder to fool with a loose correlation dressed up as a causal story. None of that should be given up.

But rigour applied to the wrong question is not the same thing as being useful. An estimate that is precisely right about a narrow slice of the world is not the same as an answer that actually helps a community, or the people who govern it, work through a genuinely hard choice.

So, the task is not to abandon the tools of the credibility revolution, but to point them at the questions that matter, and to be honest about what they can, and cannot, tell us once we do.

We are more like cartographers than captains.

Our job is to draw the map honestly: to mark the cliffs, the currents, the trade-offs nobody wants to hear about – and then to get out of the way, while the community, and the people who govern it, decide where to sail.