Can evidential pluralism reconcile experimentalists and structuralists?

I Introduction

A big disagreement about methodology tears apart economics (Keane, 2010). On one side is the structuralist camp which advocates for empirical work informed by theory. On the other side is the experimentalist camp which advocates for a theory-light approach grounded in good research designs. The credibility revolution refers to the growing popularity of experimentalist methods at the expense of structuralist ones. Both camps agree that causal claims are essential for good economic analysis but fiercely disagree about *how* to make these claims. How can these disagreements be resolved? In this paper, I offer an answer by drawing on the emerging literature on evidential pluralism in philosophy of social science (Shan and Williamson, 2022).

To this end, I make three claims. First, I show that the debate between structuralists and experimentalists boils down to a disagreement about how to identify causal effects. While structuralists argue for using economic theory, experimentalists advocate for better research designs.

Second, in response to the credibility revolution, structuralists have inflicted a

mortal wound on experimentalists. They have correctly pointed out that experimentalists tell us nothing about *why* causal effects obtain and *how* they can be exported to new environment. However, these are genuine problems for *all* empirical work in economics, i.e., they also apply to the structuralist approach. Unfortunately, structuralists do not offer satisfactory solutions to these problems, despite claiming otherwise.

Third, I offer an alternative approach called *eliminative evidential pluralism* which accommodates structuralists' critique but utilises experimentalist methods. Eliminative evidential pluralism starts from a wide hypothesis space which contains all possible mechanisms that can explain a particular phenomena. We progressively apply more and more restrictive methods to eliminate less plausible hypotheses. A method is more restrictive if it imposes more assumptions on economic behaviour. In this approach, experimentalist methods are often the first line of attack when restricting our hypothesis space because they have some of the least restrictive assumptions. For this reason, experimentalist methods are essential for eliminative evidential pluralism.

I begin by discussing the econometric crisis in the 1980s, from which the debate between experimentalists and structuralist arouse (Sections II). Next, I examine how experimentalists and structuralists responded to the econometric crisis (Section III) and structuralists' objections to experimentalists (Section IV). I then present the eliminative evidential pluralism (Section V).

II Econometric crisis

To understand the contemporary debate between structuralists and experimentalists, we need to examine its origins in the crisis in applied econometrics in the 1980s (Hendry, 1980; Leamer, 1983; Heckman, 2000). Following Haavelmo's pioneering work (1943; 1944), cutting-edge empirical work was conducted within the Structural Equation Modelling (SEM) framework (Heckman, 2000). SEM represents economic relations as systems of linear equations with probabilistic error terms. Let us illustrate how SEM works in practice by examining neighbourhood effects which I use as a running example below.

Social scientists are very interested in studying how my neighbourhood as a child affects my income as an adult (Sampson et al., 2002; Chyn and Katz, 2021). We can use SEM to model such neighbourhood effects via a simplified version of the model in Datcher (1982). In the model, neighbourhood affects income both directly (2) and indirectly via education (1):

$$Education = \alpha_1 + \alpha_2 Neighbourhood + u_E \tag{1}$$

$$Income = \beta_1 + \beta_2 Education + \beta_3 Neighbourhood + u_I \tag{2}$$

where u_E and u_I are probabilistically independent error terms. In system (1)-(2), exogenous variables are determined outside of the system whereas endogenous variables are determined within the system. So, $Neighbourhood$ is exogenous and $Education$ and $Income$ are endogenous. Parameters α_1 and β_1 are the intercepts whereas α_2, β_2 and β_3 represent the causal effects.

We can estimate neighbourhood effects, if we represent the endogenous variables in terms of the exogenous variables.[1] So, we do not need data on education and do not have to estimate the whole system in (1)-(2). If we substitute education in (2) with (1), we get $Income = (\beta_1 + \beta_2\alpha_1) + (\beta_2\alpha_2 + \beta_3)Neighbourhood + (\beta_2 u_E + u_I)$. This gives us $Income$ (an endogenous variable) expressed in terms of $Neighbourhood$ (an exogenous variable). We can then recover the total neighbourhood effect, i.e., the sum of the direct and indirect effect ($\gamma_2 = \beta_2\alpha_2 + \beta_3$), by estimating:

$$Income = \gamma_1 + \gamma_2 Neighbourhood + \epsilon_I \tag{3}$$

In other words, SEM allows us to estimate neighbourhood effects by running a simple linear regression model! This is extremely helpful because the least square estimates of γ_1 and γ_2 can be easily derived. They have many desirable properties such as being unbiased[2] under some additional assumptions (Hayashi, 2011, Proposition 1.1.).

Given this insight, SEM became extremely popular in empirical work (Hoover, 2014). Eventually, economists came to regard SEM as an 'empirical failure' (Heckman, 2000). By the 1980s, 'hardly anyone [took] anyone else's data analyses seriously' (Leamer, 1983). Why was that? Economists in different fields have pointed to various problems with SEM. While Lucas (1983) attacked SEM in macroeconomics, Sims (1980), Hendry (1980) and Leamer (1983) objected to SEM from an econometric perspective. The main problem was that estimating SEM models requires strong

[1] Technically, this is the *reduced-form* of the system. In contrast, the original system (1)-(2) is the *structural-form* (Heckman, 2010, p.356n).

[2] By unbiasedness I mean asymptotic consistency.

identifying assumptions which usually do not hold.[3] One necessary condition for identification in (3) is *Orthogonality*:

Assumption 1 (*Orthogonality*)**.** *There is zero (partial) correlation between the errors ϵ_I and the exogenous variable $Neighbourhood$, potentially after controlling for other variables.*

In practice, *Orthogonality* usually fails. For example, family income probably affects both $Neighbourhood$ quality and $Income$. Its omission from our system results in omitted variable bias and so a biased estimate of neighbourhood effects.

Unfortunately, this is not the only problem. SEM allowed researchers to fit numerous versions of their models (e.g., depending on which control variables are included) and cherry-pick the one that best supports their own beliefs (Hendry, 1980; Leamer, 1983). Because of such practices, Hendry (1980) asked if econometrics should be considered as 'alchemy or science'. Luckily, econometricians developed techniques that prevent researchers from cherry-picking their models (Leamer, 1985; Hendry and Doornik, 2014). However, these procedures do not guarantee unbiasedness: there might still be a correlation between the errors and *unobserved* variables causing an *Orthogonality* failure. Purely statistical methods such as SEM cannot deal with such unobservable factors (Hedman and Ham, 2012). So, even if we can stop researchers from cherry-picking models, we need an alternative approach to ensure identification.

[3]Lucas (1983) and Hendry (2002) object to SEM for another reason: assuming time-invariant coefficients.

III Structuralists vs experimentalists

So, applied econometrics experienced a methodological crisis in the 1980s. Interestingly, empirical work presently enjoys more prestige than theoretical work in economics (Backhouse and Cherrier, 2017). How did this happen? To wit, econometricians offered two very different (but ultimately successful) solutions to the identification problems above. However, this also led to a schism between those offering structuralist solutions (Heckman, 2010; Deaton, 2010*b*; Keane, 2010) and those offering experimentalist solutions (Imbens, 2010; Angrist and Pischke, 2010). In this section, we compare these solutions by focusing on neighbourhoods.

The biggest culprit in identifying neighbourhood effects turns out to be *selection bias* (Hedman and Ham, 2012). People who are similar in terms of unobservables tend to select into the same neighbourhoods. For example, unemployed people often move to more deprived areas. Since employment status is correlated with income, we need to take account of this selection mechanism. Otherwise, there might be important unobservables which cause both neighbourhood selection and earnings later in life. This would result in omitted variable bias and so an *Orthogonality* failure. Statistical methods such as SEM cannot eliminate selection bias completely (Hedman and Ham, 2012), so structuralists and experimentalists ventured outside of statistics to achieve this.

III.a Structuralist solution

The structuralist solution is to explicitly model selection into neighbourhoods which allows us to control for it in our estimation. For brevity, we assume from here onwards that *Neighbourhoods* is binary: one lives either in a good or a bad neighbourhood. We begin by building a formal model about how rational agents optimally chooses their neighbourhood. Then, we derive an equation for the probability of choosing a good neighbourhood (over a bad one). We can model this probability with a probit model, from which we can learn what determines neighbourhood choice, e.g., how much location and schools availability matter. Next, we can use the predicted probabilities to control for selection bias in (3) and obtain neighbourhood effects.

Put another way, we apply the two-step procedure of Heckman's selection model (1979) to neighbourhood choice. First, we estimate the coefficients in a probit model for the probability of choosing a good neighbourhood. We are using parents' characteristics here because they (and not the children) chose neighbourhoods. Second, we use the first-stage predicted probabilities (*Selection Correction*) to control for selection:[4]

$$Income = \gamma_1 + \gamma_2 Neighbourhood + \gamma_3 Selection\ Correction + \epsilon_I \tag{4}$$

Under certain additional assumptions (DiTraglia, 2022), the OLS estimate of γ_2 in

[4] *Selection Correction* is officially known as *Heckman Correction*. It is given by the Inverse Mills Ratio $\frac{\phi(\boldsymbol{x}'\hat{\boldsymbol{\delta}_1})}{\Phi(\boldsymbol{x}'\hat{\boldsymbol{\delta}_1})}$. In that equation, $\boldsymbol{x}$ is a vector of observable characteristics, $\phi(.)$ is the probability density function (PDF) and $\Phi(.)$ is the cumulative distribution function (CDF). Both PDF and CDF are calculated relative to the first-stage coefficients $\hat{\boldsymbol{\delta}_1}$ from the probit model given by $Pr(Neighbhourhood = Good|\boldsymbol{x}) = \Phi(\boldsymbol{x}'\boldsymbol{\delta_1})$. DiTraglia (2022) provides more details and discusses identification.

(4) gives us an unbiased estimate of the true neighbourhood effects in the presence of selection effects.[5] This examples shows the defining feature of the structuralist approach, i.e., the explicit use of economic theory to avoid the *Orthogonality* failure.

III.b Experimentalist solution

On the experimentalist solution, selection bias is not explicitly modelled. Instead, we use a clever design to eliminate it. Ideally, if we set ethics aside and randomly allocate houses in good and bad neighbourhoods, *Orthogonality* will hold. This approach is akin to a randomised control trial (RCT): $Neighbourhood$ is a treatment which is randomly allocated to control and treatment groups with the aim of learning its average treatment effect.

Individual researchers generally cannot conduct such RCTs. However, in the 1990s the US government implemented a program called Moving to Opportunity (Kling et al., 2007; Chetty et al., 2016). Around 4,600 families in high-poverty areas (bad neighbourhoods) were randomly allocated a voucher for moving to low-poverty areas (good neighbourhoods).[6] This design did not ensure that *neighbourhood choice* itself was uncorrelated with unobservables because not all families assigned a voucher changed their neighbourhood. Probably, families who did in fact move are different from those that did not in important unobservable ways. However, the design did ensure that *voucher assignment* was uncorrelated with unobservables. Luckily, experimentalists have developed procedures that will allow us to get an unbiased esti-

[5] A recent study in demographics (Van Ham et al., 2018) uses a similar procedure. For a study in economics, see Ioannides and Zabel (2008).

[6] I simplify the program for brevity but Chetty et al. (2016, §1) discusses the original design.

mate of the causal effect for a particular group of participating families (Kling et al., 2007, p.91) in such cases. Specifically, we use voucher assignment as an *instrumental variable* that encourages but does not guarantee moving to a good neighbourhood.

More formally, our estimation has two steps (Kling et al., 2007, pp.92-93). First, we regress $Neighbourhood$ on $VoucherAssignment$ which is our instrument. Second, we regress $Income$ on the *predicted* values of $Neighbourhood$ from the first stage:

$$Income = \gamma_1 + \gamma_2 Predicted\ Neighbourhood + \epsilon_I$$

We do not use the *true* value of $Neighbourhood$ because it might be correlated with ϵ_I. In contrast, since $Voucher\ Assignment$ is uncorrelated with ϵ_I due to randomisation, the *predicted* values of $Neighbourhood$ as a linear function of $Voucher\ Assignment$ are also uncorrelated with ϵ_I. Using this method, Chetty et al. (2016) found strong neighbourhood effects for children who moved before turning 13.

While randomisation helped identification here, it is generally not necessary for most experimentalist techniques, including instrumental variables. Widely-used methods such as difference-in-differences and regression discontinuity (Athey and Imbens, 2017) require various identifying assumption but randomisation is not necessary.[7] What matters more than randomisation is finding a natural experiment where the identifying assumption of a given method seem plausible. This is also the biggest strength of the experimentalist approach in my view. In contrast to SEM, experimentalists are open about the assumptions required and these assumptions impose very few restrictions on behavioural mechanisms. In that sense, experimentalist

[7]See Chetty and Hendren (2018) for an example in the context of neighbourhood effects.

results are credible.

IV Evaluating the credibility revolution

We have now set up the debate between structuralists and experimentalists. Both camps agree that we need to ensure *Orthogonality* but disagree about *how* to achieve this. We might ask which camp prevailed. By 2010, the structuralist approach 'has fallen out of fashion' (Keane, 2010, p.18). This also became evident in the type of papers published in top journals (Biddle and Hamermesh, 2017, p.49). The growing popularity of the experimentalist paradigm is known as the credibility revolution (Angrist and Pischke, 2010).[8] However, since 2010, the advent of experimentalist methods into fields such as industrial organisation and macroeconomics has stalled. This trend largely reflects a series of objections raised by structuralists against the experimentalist approach. I outline these objections and argue that whilst these are genuine concerns, structuralists have not offered satisfactory solutions.

The structuralists attack boils down to the idea that experimentalist methods cannot resolve all problems faced by economists. As Leamer (2010, p.33) says, experimentalists cannot just 'wave a clove of garlic and chant "randomization" to solve all our problems'. Consequently, I extracted the following argument from structuralists' work (Deaton, 2010*a*; Heckman, 2010; Low and Meghir, 2017):

P1: We can do good economic analysis only if we make good causal claims.

P2: There exist a set of good causal claims C^* that experimentalist methods

[8] Panhans and Singleton (2017, pp.149-151) discuss several historical factors such as the demands of policy-makers that lead to the credibility revolution.

do not allow us to make *and* that are necessary for good economic analysis.

P3: Structuralist methods allow us to make claims C^*.

C: Therefore, structuralist methods are necessary for good economic analysis.

By economic analysis in this paper, I understand policy evaluations or providing causal explanations for a particular explanandum.[9] The argument starts from the idea that causal claims are essential for economics analysis (P1). Experimentalist methods allow us to make some but not all such claims: there exist causal claims C^* which are required but cannot be established via experimentalist means (P2). For example, experimentalists do not tell us much about the transmission mechanism underlying a causal effect (Deaton, 2010*b*). Luckily, structuralist methods allow us to make claims C^* (P3). Therefore, we need them for good economic analysis (C).

This argument acknowledges the usefulness of research design, e.g., experiments can falsify structural models (Low and Meghir, 2017). The structuralists are not saying that we should completely abandon the experimentalist paradigm. However, we should use experimentalist methods only when appropriate and acknowledge their limitations. Meanwhile, we should continue working within a structuralist paradigm.

The structuralist argument relies on a crucial implicit assumption: structuralist methods are the ***only*** methods available to make C^*. Unless this assumption holds, one can object that there are methods that are neither experimentalist, nor structuralist but allow C^*. This would render the argument invalid. To evaluate the implicit assumption, we need to examine the type of causal claims entering C^*. Structuralists have offered a plethora of such claims (Keane, 2010; Nevo and Whin-

[9]There are non-causal problems in economics such as forecasting the stock markets.

ston, 2010; Low and Meghir, 2017), but the most important ones in my view are related to (i) causal transmission mechanisms (Deaton, 2010*b*) and (ii) extrapolating experimental results to new environments (Heckman, 2010). The reminder of this section shows that there are non-structuralist methods allowing C^* which suggests that structuralist methods are not necessary in economic analysis. My strategy for justifying this claim is to show that supplementing an experimentalist method with non-structuralist and non-experimentalist methods allows for C^*. I justify this strategy in Section V.

IV.a Mechanisms

Deaton (2010*a*,*b*) claims that good economic analysis requires studying causal transmission mechanisms and the experimentalist approach do not tell us much about these mechanisms. Consider the positive correlation between education and earnings. Two well-known mechanisms explaining it are signalling and human capital theory. Perhaps people obtain education in order to *signal* employers that they are good workers and obtain a higher wage (Spence, 1973). On the other hand, education increases people's *human capital*,[10] helping them to get a better paid job (Mincer, 1974). Both mechanisms explain not only the correlation between earnings and education but also many other patterns of economic behaviour. For Deaton, discovering such mechanisms is at the heart of economic analysis because it tells us *why* causal effects obtain. In contrast, experimentalist methods tell us nothing about such *why*-question: they only tell us *if* a causal effect occurs.

[10]By human capital, I mean individual characteristics that are useful in one's workplace, e.g., skills and knowledge (Goldin, 2021).

However, structuralist methods are neither the only way, nor the best way to discover underlying mechanisms. One alternative is causal discovery algorithms (Spirtes et al., 2000) which help us discover how a set of variables is ordered. The output of such algorithms is a directed acyclical graph representing the causal links between the variables of interest. In our case, we can use an experimentalist method to establish if education affects earnings. Then, we can use an algorithm to examine the variables that mediate the effect of education on earnings, e.g., occupation and gender. The resulting graph provides the mechanism through which education affects earnings.

This reasoning does not show that causal discovery algorithms are *better* at examining mechanisms. Deaton's argument would still stand if structuralist methods were the best method to study mechanisms. However, structuralist methods are based on rational choice theory which imposes severe restrictions on the set of acceptable mechanisms. For example, the standard formulation of rational choice allows only for mechanisms that assume transitive and complete preferences (Reiss, 2013). Consequently, behavioural mechanisms which violate these restrictions are excluded. This example illustrates a bigger problem: focusing solely on rational choice models limits the type of factors that we are willing to include in our mechanisms. Prior to the rise of behavioural economics, economists did not model nudges or default options.

While many behavioural mechanisms *can* be modelled via rational choice theory, the problem here is that we cannot discover them in the first place via purely structuralist methods. We need non-structuralist evidence to find behavioural mech-

anisms. Similarly, we need alternative methods to discover other mechanisms that violate the assumptions of rational choice theory. Even if causal discovery algorithms are not unambiguously better at examining mechanisms, there is no reason to grant epistemic priority to structuralist methods given how restrictive they are. So, I agree with Deaton that mechanisms matter but I disagree with him about how to discover them.

IV.b Extrapolation

Heckman (2010) argues that structuralist methods can extrapolate experimentalist results and examine the causal claims which actually interest policy makers.[11] For example, the Moving to Opportunity program discussed above only included 4,600 families but a policymaker might be interested in implementing a similar policy on a much larger scale. Can experimentalists help the policymaker? Heckman does not think so. Evaluating the new policy can only be achieved by embedding the natural experiment within a formal structuralist framework. This framework allows us to vary the number of families that move and to take account of general equilibrium effects, e.g., how residents in good neighbourhoods respond (Heckman, 2010). We can then determine neighbourhood effects for any particular number of moving households and determine if the policy can be scaled up.

More generally, Heckman (2010; 2022) differentiates two main types of causal claims that policy-makers need. First, *extrapolation* claims are about verifying if a causal claim made in one context holds in a new context such as another state or a

[11]Following Deaton and Cartwright (2018, p.10), I do not discuss the related concept of external validity.

different level of geographical aggregation (e.g., when scaling-up policies). Second, *speculative* causal claims concern the effects of a completely new policy in location L. While similar policies might have been implemented in L or somewhere else, we have not observed exactly the new policy. For instance, we have never observed a complete tobacco ban in a developed country, even though most countries have some restrictions on smoking.[12] So, we have never observed its impact on health. Heckman's claim is that structuralist methods permit both extrapolation and speculative claims whereas experimentalist methods permit neither.

This claim needs to be qualified. Structuralist models cannot examine *all* speculative causal claims because the set of *speculative* causal claims is simply too large. We need to impose some restrictions on this set. This will help us reach a crucial insight: most speculative claims either reduce to extrapolation causal claims or consider *extreme* situations, in which structuralist method cannot help us.

In *extreme* speculative cases, we consider large-scale policies such as Brexit which might have no precedent. Such policies often shift economic fundamentals in a way unrelated to households' utility-maximisation problems or completely change the structure of these problem (Hendry, 2002). In any case, there will be a lot of uncertainty about how agents respond. Consequently, it is unlikely that we can credibly predict all changes, build a reliable formal model as a basis for structuralist methods and make quantitative predictions. Using qualitative methods such as process tracing (Beach and Pedersen, 2019) might be more useful: there are simply too many ways, in which the situation might change.

[12]Bhutan and Turkmenistan implemented complete tobacco bans, but it is unclear why results from there should apply to developed countries.

On the other hand, most *speculative* claims we might study are closer to *extrapolation* claims than to *extreme speculative* claims. The difference is that we are extrapolating in two dimensions: (i) another location and (ii) a small change in policy. We might have experimental results for a nearly complete tobacco ban in New Jersey but be interested in a complete ban in Connecticut. Such situations look more similar to an extrapolation rather than an extreme speculation. So, structuralist methods cannot help us with extreme speculative claims whereas less extreme speculative claims reduce to extrapolation claims in two dimensions. If we can use non-structuralist methods to extrapolate in two dimensions, we would have disproved Heckman's version of the structuralist argument.

How can we extrapolate experimentalist results? To achieve this, we need to learn about two different things: (i) causal transmission mechanisms and (ii) support factors. Since we discussed (i) above, I focus on (ii). I define a support factor as *an insufficient, but necessary part of an unnecessary but sufficient* condition for the obtainment of a causal effect. Without a support factor S, a causal link would not obtain in the original location (unless we change other support factors). On this definition, there might exist a sufficient set of factors which do not include S. In any case, to export a causal effect, we need to ensure that there is a sufficient number of support factors in the new location.

Unfortunately, support factors are not easy to identify since they might be unobserved. An illuminating example of an unobserved support factors comes from the failure to export the positive effects of class-size reduction from Tennessee to California. The problem was the absence of enough qualified teachers residing in California

(Jepsen and Rivkin, 2009): the number and quality of teachers was an unobserved supporting factor. To uncover support factors, we can use qualitative methods such as process tracing based on historical case studies. Similarly to meta-analysis, we can also examine other studies that have investigated similar research question.

After we establish the transmission mechanism and the support factors, we can make extrapolation claims. We first check the relevant support factors and if the causal transmission mechanism can occur. If these conditions are satisfied, we can make the extrapolation claim. Going back to Heckman's concerns, this approach also helps a policy-maker determine if the causal mechanism associated with Moving to Opportunity program is likely to obtain when scaled up. Since this procedure does not use structuralist methods, it shows that these methods are not necessary for good economic analysis which disproves the implicit assumption required by the structuralist argument.

Before moving forward, two things should be noted. First, my argument differentiates extreme and less extreme speculative claims. While this distinction is vague, all speculative claims examined by Heckman seem to me as being less extreme. Second, even if we reject the structuralist approach, we can still use economic theory to investigate mechanisms. The use of a formal model is not necessarily a structuralist method: our approach does not have to estimate that model using data (e.g. see Ansell, 2008) which is a defining feature of the structuralist approach.

V Eliminative evidential pluralism

Structuralists raised some serious objections against experimentalists. The latter do not tell us much about mechanisms or extrapolation, despite their significance for economic analysis. While experimentalists have suggested methods to tackle these issues (Wager and Athey, 2018), these methods have not gained a lot of popularity among economists.[13] There is more bad news. Structuralists also claim to provide methods that allow extrapolation and studying mechanisms. However, their methods are neither the only methods, nor the best methods for these purposes. These considerations raise the question if we can ever conduct good economic analysis that both provides credible estimates of causal effect and allows for mechanisms and extrapolations.

In my view, the answer is yes. We can use the *eliminative evidential pluralism* developed in this section. It proceeds by applying successively more and more demanding methods in order to eliminate less promising answers to a specific causal question. I firstly situate it within the emerging literature on evidential pluralism in social science (reviewed in Shan and Williamson, 2022) and then describe it in more depth.

V.a Evidential pluralism in social science

In its original formulation, *evidential pluralism* postulates that establishing causality requires both (i) evidence of difference-making and (ii) evidence of mechanisms (Shan and Williamson, 2021, p.4). We can obtain (i) from associational studies

[13]Most citations of Wager and Athey (2018) do not come in papers by economists.

(including RCTs) whereas (ii) is provided by mechanistic studies (including clinical expertise). While evidential pluralism was developed in the context of biomedical sciences, Shan and Williamson (2021) have recently used it to discuss causality in social sciences which unleashed a wave of papers on the topic (Beach, 2021; Maziarz, 2021; Runhardt, 2021; Shan and Williamson, 2022). To show that neighbourhood effects matter, for example, we need (i) evidence that better neighbourhoods increase incomes and (ii) evidence about how neighbourhoods affect income. Why is evidential pluralism an attractive approach specifically in economics? Maziarz (2021) suggests that it can resolve many empirical disagreements. Often, valid econometric approaches reach contrasting conclusions about difference-making. However, by examining the corresponding mechanisms, evidential pluralism can tell us which side is making better causal claims.

In my view, applying directly evidential pluralism to economics does not work because mechanisms there are very different from mechanisms in biomedical sciences. Mechanisms in economics are much less homogeneous and depend much more strongly on support factors. If we find a mechanism for how smoking causes cancer in Tennessee, it probably also holds for people in California because most people share similar biological mechanisms. However, we saw that this is not the case with class-size reductions. When we try to export the mechanism from Tennessee, we have to consider (i) if the support factors hold in California and (ii) if class-size reduction affects students through the same mechanism. This exercise seems much more demanding than in the biomedical case.

Admittedly, this is just a single example, but a similar point is made by Beach

(2021). If we do not supplement evidential pluralism with a better picture of economic mechanisms, we might be unable to compare two similar studies in economics because the mechanisms might be incomparable. In other words, we cannot use Maziarz (2021)'s argument for using evidential pluralism in economics. This is where my eliminative evidential pluralism enters. It not only takes seriously the specifics of economic mechanisms but can also resolve the debate between structuralists and experimentalists.

V.b An evidential pluralism for economics

On eliminative evidential pluralism, we begin with a set of hypothesis that can explain the observed phenomena. Subsequently, we start restricting this set by using different methods. Ideally, we end up with a relatively small final set of potential mechanisms. From the final set, we can examine in greater depth the causal transmission mechanisms that seem most plausible and determine if they can be exported to new environments.

To gain more insight into eliminative evidential pluralism, we can see how it can be used to examine why education and earnings are related. The complete set of mechanisms explaining this relationship is H_0. Some mechanisms in H_0 have already been studied but others might be purely speculative and we might be unaware of them. The only restrictions on the hypotheses entering H_0 is that they they postulate causal links between earnings and education and that they have some social science justification. Importantly, each mechanism *can* tell us about both (i) support factors and (ii) causal transmission mechanisms. Of course, we do not have to specify all the

details of these mechanisms. At this stage, this would be too demanding but what matters is that we recognise the importance of support factors and mechanisms for each hypothesis in H_0.

We begin implementing our procedure by using experimentalist methods such as RCTs or instrumental variables. These methods are our starting point since they rely on fewer assumptions about transmission mechanisms than other methods. While they tell us little about mechanisms (Deaton, 2010*b*), this is also a benefit: they work independently of any specific mechanism. If we replicate the experimentalist study by Card and Krueger (1992) about the returns to education, we will find that education causally affects earnings. So, we can eliminate from H_0 all mechanisms indicating no causal effect of education on earnings to form the new H_1.

Next, we utilise a causal discovery algorithm to examine the mediating mechanism between education and earnings. Such algorithms rely on stronger assumptions than experimentalist methods but tell us more about the mediating mechanism and the support factors (Weinberger, 2022, §3), e.g., if gender and occupation are the only mediators. As shown above, because the algorithms do not say much about unobservable support factors, we might complement them with qualitative methods such as process tracing.

At this point, H_1 reduces to H_2 after we eliminate all mechanisms incompatible with the experimentalist results, the mechanisms found by causal discovery algorithms and the required support factors. We can then start examining specific transmission mechanisms within H_2. To achieve this, we can review the literature to identify mechanisms that are in line with the restrictions on H_2. We can also

examine case studies to delve into specific causal links. We can now consider if these mechanisms are exportable to a new location. To that aim, we examine whether in the new location the support factors hold up and the causal transmission mechanism obtains. This shows that eliminative evidential pluralism can be utilised to tackle the problems identified by Deaton and Heckman but without using structuralist methods.

VI Conclusion

This paper investigated the debate between structuralists and experimentalists about the credibility revolution. It proposed eliminative evidential pluralism as one approach which can reconcile the two sides. Broadly, evidential pluralism proposes that we can learn about causal effects by progressively imposing more restrictions on the starting set of hypotheses. This approach should be attractive to experimentalists because it acknowledges the importance of credible identification of causal effects. It should also be acceptable to structuralists because it takes seriously their concerns about mechanisms and extrapolation.

While eliminative evidential pluralism is fundamentally an empirical approach, it also takes theory seriously. Theory provides some of the mechanisms in the starting set H_0 and is also essential for evaluating the different assumption we impose in restricting H_0. Thus, eliminative evidential pluralism takes a middle ground between the theory-light experimentalist methods and the theory-heavy structuralist methods.

References

Angrist, J. D. and Pischke, J.-S. (2010), 'The credibility revolution in empirical economics: How better research design is taking the con out of econometrics', *Journal of economic perspectives* **24**(2), 3–30.

Ansell, B. W. (2008), 'Traders, teachers, and tyrants: democracy, globalization, and public investment in education', *International Organization* **62**(2), 289–322.

Athey, S. and Imbens, G. W. (2017), 'The state of applied econometrics: Causality and policy evaluation', *Journal of Economic perspectives* **31**(2), 3–32.

Backhouse, R. E. and Cherrier, B. (2017), 'The age of the applied economist: the transformation of economics since the 1970s', *History of Political Economy* **49**(Supplement), 1–33.

Beach, D. (2021), 'Evidential pluralism and evidence of mechanisms in the social sciences', *Synthese* **199**(3), 8899–8919.

Beach, D. and Pedersen, R. B. (2019), *Process-tracing methods: Foundations and guidelines*, University of Michigan Press.

Biddle, J. E. and Hamermesh, D. S. (2017), 'Theory and measurement: Emergence, consolidation, and erosion of a consensus', *History of Political Economy* **49**(Supplement), 34–57.

Card, D. and Krueger, A. B. (1992), 'Does school quality matter? returns to edu-

cation and the characteristics of public schools in the united states', *Journal of political Economy* **100**(1), 1–40.

Chetty, R. and Hendren, N. (2018), 'The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects', *The Quarterly Journal of Economics* **133**(3), 1107–1162.

Chetty, R., Hendren, N. and Katz, L. F. (2016), 'The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment', *American Economic Review* **106**(4), 855–902.

Chyn, E. and Katz, L. F. (2021), 'Neighborhoods matter: Assessing the evidence for place effects', *Journal of Economic Perspectives* **35**(4), 197–222.

Datcher, L. (1982), 'Effects of community and family background on achievement', *The Review of Economics and Statistics* **64**(1), 32–41.

Deaton, A. (2010*a*), 'Instruments, randomization, and learning about development', *Journal of economic literature* **48**(2), 424–55.

Deaton, A. (2010*b*), 'Understanding the mechanisms of economic development', *Journal of Economic Perspectives* **24**(3), 3–16.

Deaton, A. and Cartwright, N. (2018), 'Understanding and misunderstanding randomized controlled trials', *Social science & medicine* **210**, 2–21.

DiTraglia, F. J. (2022), 'Introductory graduate econometrics: Lecture notes on limited dependent variable', *Oxford MPhil Economics Lecture Note* .
URL: *https://www.economictricks.com/limdep-notes.pdf*

Goldin, C. (2021), Human capital, *in* C. Diebolt and M. Haupert, eds, 'Handbook of Cliometrics', Springer Berlin Heidelberg, pp. 1–27.

Haavelmo, T. (1943), 'The statistical implications of a system of simultaneous equations', *Econometrica* **11**(1), 1–12.

Haavelmo, T. (1944), 'The probability approach in econometrics', *Econometrica* **12**(Supplement), iii–115.

Hayashi, F. (2011), *Econometrics*, Princeton University Press.

Heckman, J. J. (1979), 'Sample selection bias as a specification error', *Econometrica* **7**(1), 153–161.

Heckman, J. J. (2000), 'Causal parameters and policy analysis in economics: A twentieth century retrospective', *The Quarterly Journal of Economics* **115**(1), 45–97.

Heckman, J. J. (2010), 'Building bridges between structural and program evaluation approaches to evaluating policy', *Journal of Economic literature* **48**(2), 356–98.

Heckman, J. J. (2022), 'Interview with james heckman', *Observational Studies* **8**(2), 7–22.

Hedman, L. and Ham, M. v. (2012), Understanding neighbourhood effects: Selection bias and residential mobility, *in* 'Neighbourhood effects research: New perspectives', Springer, pp. 79–99.

Hendry, D. F. (1980), 'Econometrics-alchemy or science?', *Economica* **47**(188), 387–406.

Hendry, D. F. (2002), 'Forecast failure, expectations formation and the lucas critique', *Annales d'Economie et de Statistique* pp. 21–40.

Hendry, D. F. and Doornik, J. A. (2014), *Empirical model discovery and theory evaluation: automatic selection methods in econometrics*, Mit Press.

Hoover, K. D. (2014), 'On the reception of haavelmo's econometric thought', *Journal of the History of Economic Thought* **36**(1), 45–65.

Imbens, G. W. (2010), 'Better late than nothing: Some comments on deaton (2009) and heckman and urzua (2009)', *Journal of Economic literature* **48**(2), 399–423.

Ioannides, Y. M. and Zabel, J. E. (2008), 'Interactions, neighborhood selection and housing demand', *Journal of urban economics* **63**(1), 229–252.

Jepsen, C. and Rivkin, S. (2009), 'Class size reduction and student achievement the potential tradeoff between teacher quality and class size', *Journal of human resources* **44**(1), 223–250.

Keane, M. P. (2010), 'Structural vs. atheoretic approaches to econometrics', *Journal of Econometrics* **156**(1), 3–20.

Kling, J. R., Liebman, J. B. and Katz, L. F. (2007), 'Experimental analysis of neighborhood effects', *Econometrica* **75**(1), 83–119.

Leamer, E. E. (1983), 'Let's take the con out of econometrics', *The American Economic Review* **73**(1), 31–43.

Leamer, E. E. (1985), 'Sensitivity analyses would help', *The American Economic Review* **75**(3), 308–313.

Leamer, E. E. (2010), 'Tantalus on the road to asymptopia', *Journal of Economic Perspectives* **24**(2), 31–46.

Low, H. and Meghir, C. (2017), 'The use of structural models in econometrics', *Journal of Economic Perspectives* **31**(2), 33–58.

Lucas, R. (1983), Econometric policy evaluation: A critique, *in* K. Brunner and A. Meltzer, eds, 'Theory, Policy, Institutions: Papers from the Carnegie-Rochester Conferences on Public Policy', Vol. 1, pp. 19–42.

Maziarz, M. (2021), 'Resolving empirical controversies with mechanistic evidence', *Synthese* **199**(3), 9957–9978.

Mincer, J. (1974), *Schooling, Experience, and Earnings*, National Bureau of Economic Research.
URL: *https://www.nber.org/books-and-chapters/schooling-experience-and-earnings*

Nevo, A. and Whinston, M. D. (2010), 'Taking the dogma out of econometrics: Structural modeling and credible inference', *Journal of Economic Perspectives* **24**(2), 69–82.

Panhans, M. T. and Singleton, J. D. (2017), 'The empirical economist's toolkit: from models to methods', *History of Political Economy* **49**(Supplement), 127–157.

Reiss, J. (2013), *Philosophy of economics: A contemporary introduction*, Routledge, chapter 3.

Runhardt, R. W. (2021), 'Evidential pluralism and epistemic reliability in political science: Deciphering contradictions between process tracing methodologies', *Philosophy of the Social Sciences* **51**(4), 425–442.

Sampson, R. J., Morenoff, J. D. and Gannon-Rowley, T. (2002), 'Assessing "neighborhood effects": Social processes and new directions in research', *Annual review of sociology* **28**, 443–478.

Shan, Y. and Williamson, J. (2021), 'Applying evidential pluralism to the social sciences', *European Journal for Philosophy of Science* **11**(4), 1–27.

Shan, Y. and Williamson, J. (2022), 'Evidential monism, evidential pluralism, or evidential contextualism? an introduction to evidential diversity in the social sciences', *Synthese* **200**(4), 1–6.

Sims, C. A. (1980), 'Macroeconomics and reality', *Econometrica* **48**(1), 1–48.

Spence, M. (1973), 'Job market signaling', *The Quarterly Journal of Economics* **87**(3), 355–374.

Spirtes, P., Glymour, C. N., Scheines, R. and Heckerman, D. (2000), *Causation, prediction, and search*, MIT press.

Van Ham, M., Boschman, S. and Vogel, M. (2018), 'Incorporating neighborhood choice in a model of neighborhood effects on income', *Demography* **55**(3), 1069–1090.

Wager, S. and Athey, S. (2018), 'Estimation and inference of heterogeneous treatment effects using random forests', *Journal of the American Statistical Association* **113**(523), 1228–1242.

Weinberger, N. (2022), 'Comparing Rubin and Pearl's causal modelling frameworks: a commentary on Markus)', *Forthcoming in Economics & Philosophy* pp. 1–9.
URL: *https://drive.google.com/file/d/1DAZFZD9JlUf-gGCfCSspgRvRMrzoIao/view*