Claims and confounds in economic experiments☆

Daniel John Zizzo [*],[1]

School of Economics and CBESS, Norwich Research Park, University of East Anglia, Norwich NR4 7TJ, United Kingdom

ARTICLE INFO

ABSTRACT

We present a distinctiveness, relevance and plausibility (DRP) method for systematically evaluating potential experimental confounds. A claim is a statement being inferred on the basis of experimental data analysis. A potential confound is a statement providing a prima facie reason why the claim is not justified (other than internal weakness). In evaluating whether a potential confound is problematic, we can start by asking whether the potential confound is *distinctive* from the claim; we can then ask whether it is *relevant* for the claim; and we can conclude by asking whether it is *plausible* in the light of the evidence.

1. Introduction

Assume that an experimental paper makes a claim A. What happens if there is a potential experimental confound B that may affect behavior in the experiment and hence the validity of the claim? Examples are confusion effects (Ferraro and Vossler, 2010), experimenter demand effects (Zizzo, 2010), framing effects (Cookson, 2000), house money effects (Harrison, 2007), demographic effects (Casari et al., 2007), wealth effects (Armantier, 2006), incentives size effects (Slonim and Roth, 1998), task order effects (Hogarth and Einhorn, 1992), sample selection effects (Harrison et al., 2009), risk aversion effects (Vieider, 2011), behavioral noise effects (Hey, 2005), lack of credibility of experimental instructions due to the use of deception (Hertwig and Ortmann, 2001), or lack of control for social preferences explanations (Gächter et al., 2012) or for expectations about the coplayer (Ashraf et al., 2006).

The aim of this paper is to provide a method to identify when potential confounds are a problem for the claims made in an experimental paper, and, if they are a problem, what are valid ways of addressing it. Experimental economics textbooks and, more generally, research methods textbooks have a general discussion of standard responses to the problems of experimental confounds (e.g., Davis and Holt, 1993; Friedman and Sunder, 1994; Jackson, 2008). While there is plenty of econometric

* Tel.: +44 1603 593668; fax: +44 1603 456259.
E-mail address: d.zizzo@uea.ac.uk

guidance in the literature,[2] what is missing is a general methodological analysis of how to handle experimental confounds that goes beyond what can be found in the introductory textbooks, and that may be useful especially for junior researchers designing or evaluating experiments.

We try to ask the question of *when* experimental economists should be worried about potential confounds. We are interested in providing an evaluation methodology of interest especially to junior researchers designing and running experiments and, when an experiment has been run and a paper completed, of interest to editors, referees and indeed the wider readership in evaluating an experimental paper, or to the authors of such paper in determining what to do next.

Our method centers around the notions of *distinctiveness, relevance and plausibility* (DRP) of a potential confound. We discuss the use of these concepts in the light of a number of practical examples and show how our DRP method can be employed to systematically evaluate potential confounds criticisms. We also review, in the light of the DRP method, other arguments that have been made or can be made to defend experimental designs against potential confound criticisms.

Section 2 provides the general conceptual framework and background, Sections 3, 4 and 5 respectively introduce the notions of distinctiveness, relevance and plausibility of potential confounds. Section 6 considers the implications of the DRP method as a whole, summarizes the method with a table and discusses what the method can and cannot be used for. Section 7 employs the DRP method to evaluate other arguments on potential confounds. Section 8 concludes.

2. The conceptual framework

Define a *claim* as a statement being inferred in a paper P on the basis of experimental data analysis E. We consider one such claim, which we label as A, and we define a claim D which is different from A.

One reason A may not be justified is that it may not be supported by E *as stated in the paper*, which we can label as E|P. If (E|P $\Rightarrow$ A) is not justified, that is if the experimental data analysis as stated in the paper does not imply A, then the experimental data analysis does not support the claim A, and so the claim A is in itself not justified.[3] We can label this situation as one of *internal weakness*.[4]

Define a *potential confound* as a statement providing a prima facie reason why A is not justified other than internal weakness.[5] Let us consider one such potential confound, which we label as B; examples have been provided in the introduction. That is, it is possible that E|P $\Rightarrow$ A, but nevertheless B $\Rightarrow$ D.

We now need to identify a parsimonious and yet comprehensive procedure by which we can test for whether a potential confound should be considered a problem. The first question we can ask is whether A and B are essentially the same. That is to say, the statement (B $\Rightarrow$ D) is criticized on the ground of the relationship between A and B. If B is equivalent to A, then B is not a problem for A: that is, B $\Leftrightarrow$ A. If B is truly different (distinctive) from A, then based on this test we cannot rule out that (B $\Rightarrow$ D). This is a *distinctiveness* test of the potential confound.

Assume that B is distinctive from A. The second question we can ask is whether B negates A in principle. That is to say, the statement (B $\Rightarrow$ D) is criticized on the ground of the relationship between A and D. If the confound has passed the distinctiveness test, we know that D is different from A, but this does not mean that D necessarily contradicts A. The potential confound B may imply D, but may not be relevant for whether A holds. If B is irrelevant, D can be true but at the same time A can also be true. This is a *relevance* test of the potential confound.

Assume that B is both distinctive from A and relevant for A. The third and final question we can ask is whether, based on all available evidence, it is plausible to believe that (B $\Rightarrow$ D) is true in practice. This is no longer a logical test in the sense of distinctiveness and relevance, but rather a consideration of whether the potential confound is plausible from an empirical viewpoint. This is a *plausibility* test of the potential confound.

The next three sections provide examples of the use of the distinctiveness, relevance and plausibility tests. We have simplified our analysis to the existence of a single potential confound B, but our analysis easily extends to considering multiple potential confounds.[6] If a potential confound is distinctive, relevant and plausible, then it is a problem that needs to be addressed and Section 6 provides a discussion.

We believe the procedure to be comprehensive in the sense that evaluating B means to evaluate the relationship (B $\Rightarrow$ D), and this can either mean to question each part of this relationship with A (i.e. B with A, B with D), which is what the distinctiveness and relevance tests do, or to question the validity of the whole of the relationship, which is what the plausibility test does. Obviously different evaluation procedures might be possible and this paper does not preclude research on further

---

[2] Experimental designs can be run to ameliorate the problem; random samples can be used; various sources of confounding can be controlled for using covariates or using suitable econometric tools such as instrumental variables. See Davidson and MacKinnon (1993) for a standard econometric analysis and Greenland et al. (1999) for an example of analysis from a statistician's perspective on confounding.

[3] An example is if a claim is made that a specific frame induces more cooperation in a public good contribution game when in fact, in looking at the regression analysis on contribution the data analysis section of the paper is centered upon, the sign is statistically significant but negative, i.e. in the opposite direction of the claim.

[4] We label it as internal since the lack of justification is based on the data analysis in paper P which is supposed to underpin the claim A made also in paper P.

[5] For examples of definitions of confounds, see Mill (1843/2009), Patten (2007) and Jackson (2008).

[6] In Section 7 we shall consider an argument that will require us to consider two potential confounds at the same time.