Causality in Economics: A Menu of Approaches

Eduardo Pol*

School of Economics, Faculty of Business, University of Wollongong, Wollongong, NSW 2522, Australia

Abstract: Causality is a notion that occurs often in economics. In using the words 'cause' and 'effect,' economists seek to distinguish causation from association, recognizing that causes are responsible for producing effects, whereas non-causal associations are not. The identification of causes is accorded a high priority because it is viewed as the basis for understanding economic phenomena and developing policy implications. In this survey we look at different approaches to causality in economics and set out the general principles of each approach, so as to assist in the communication and teaching role. Specifically, we confine attention to five approaches to causality in economics (narrative, comparative statics, theoretical, structural, and experimentalist) and elucidate their distinctive characteristics without entering into philosophical discussions. In particular, we pay close attention to the debate between the structuralist and experimentalist schools because this controversy has been extremely useful to clarify a number of fundamental points concerning causality in economics.

Keywords: Causality in economics, causes of effects, counterfactuals, hypotheticals, effects of causes, local average treatment effect.

1. INTRODUCTION

Causality is a notion that occurs often in economics. In using the words 'cause' and 'effect,' economists seek to distinguish causation from association, recognizing that causes are responsible for producing effects, whereas non-causal associations are not. The identification of causes is accorded a high priority because it is viewed as the basis for understanding economic phenomena and developing policy implications.

The idea that causality is central to economics is at least as old as Adam Smith's foundational work. Indeed, the full title of Smith's book, *An Inquiry into the Nature and Causes of the Wealth of Nations,* signals that one of the most important tasks of economics is the search for explanations involving causal connections. The critical question is: how could we infer the existence of causal relations from observations? To be sure, the answer to this question presupposes a definition of the term 'cause.'

The loosest possible definition of 'cause' was proposed by Locke (1960:180). Locke's definition can be paraphrased as follows: A cause is a factor that produces a particular phenomenon called effect. For example, what causes long-term inflation? It is generally agreed that the rapid growth in the quantity of money is the cause of long-term inflation.

The use of the term 'cause' can be deceptive when there are multiple causes that can produce the same effect. It does not necessarily follow from the preceding definition that when the factor in question is absent the phenomenon does not happen. For example, the existence of a monopoly provokes a deadweight loss for society as a whole because output would be lower than under competition, and probably, the monopolist would earn abnormal profits. Quite obviously, if monopoly is nonexistent it does not follow that deadweight losses for society are nonexistent. The inflation tax, for example, also causes deadweight losses for society, as people waste scarce resources trying to avoid it.

Causality (the relation between cause C and effect E) in economics has always been an awkward –and elusive– topic. There are at least four sources of difficulties: first, the use of language with respect to causality can be very confusing (there is *no* single definition of causality); second, assuming that we have a plausible definition of causality, the definition has to be operational (we must be able to *test* the proffered definition); third, the causality definition should capture the notion of controllability (if there is a cause, it can be used to *control* perhaps in combination with other causes); and finally, when events have several plausible candidate causes, as they usually do in the real economic world, the causality tests tend to be difficult to implement. Having said this, it is clear that many economists do not perceive these difficulties as insurmountable barriers, but as challenges to be prevailed over. To unravel cause and effect, economists use common sense, elementary logic, economic models, data, and experiments.

Broadly speaking, there are two kinds of causation. First, the analysis may refer to instances of the same

*Address correspondence to this author at the School of Economics, Faculty of Business, University of Wollongong, Wollongong, NSW 2522, Australia; Tel: + 61 2 4221 4025; Fax: + 61 2 4221 3725; E-mail: epol@uow.edu.au

E-ISSN: 1929-7092/13

phenomenon. The causal analysis of a reproducible phenomenon is called *general* causation. Typical economic examples of this kind of causality include: the effects of a firm's input on its output, the effect of education on earnings, and the effects of employment training programs on subsequent labour market histories. The second kind of causation analysis focuses on a dated and non-replicable phenomenon occurring at a particular location such as the financial crisis 1930-33 in the United States. Causation of the second kind is said to be *singular*. In this case, facts do not permit the type of replicability that is present in much scientific enquiry. Counterfactual reasoning is typically used to explore singular causality. A counterfactual argument requires the analyst to posit: "What would have happened if … had happened (or not had happened)."Although the general/singular dichotomy occasionally becomes blurred, it helps to begin by thinking in such terms.

Causality is not only a deep methodological problem in economics but also a most complex philosophical issue. The nature of causality has consumed the attention of philosophers at least since Aristotle, without resolution. For example, Nancy Cartwright (2004) discusses a variety of definitions of causality from a philosopher's perspective and argues that causation is not a single, monolithic concept. The message send by Cartwright (2004) is that the dominant accounts of causation "do not succeed in treating the exemplars employed in alternative accounts."

In this survey we confine attention to five approaches to causality in economics and elucidate their distinctive characteristics without entering into philosophical discussions. In order to facilitate the identification of a particular methodology, we call the approaches in question 'narrative approach (Smith)', 'comparative statics approach (Marshall)', 'theoretical approach (Hicks)', 'structural approach (Heckman)', and 'experimentalist approach (Angrist-Imbens),' after those individuals most closely identified with the approach. Generally, each approach has its origins further back in time, and is the outcome of research methodologies that have had several contributors apart from the named authors. But the references Smith (2010), Marshall (1966), Hicks (1979), Heckman (2000), and Imbens and Angrist (1994) are representative of the material, and therefore it appears appropriate to employ the chosen names.

There are other procedures to deal with causality in economics that are conceptually distinct from the above mentioned approaches. For example, time-series notions of causality as developed by Granger (1969) and Sims (1972) originated an important literature that will not be surveyed in this paper only because of space limitation. Clive Granger (1969) offered an operational definition of causality which turned out to be good enough for many members of the econometrics profession. Granger causality is more a forecasting technique than a method to detect causes. A 'cause' is identified on the basis of its ability to predict an effect.[1]

The organization of this paper is as follows. There is one section allocated to the description of each approach to causality in economics (Sections 2 to 6). In Section 7 we pay close attention to the debate between the structuralist and experimentalist perspectives because this controversy has been extremely useful to clarify a number of fundamental points concerning causality in economics. There is a brief concluding section, Section 8, which succinctly summarizes the menu of approaches.

2. NARRATIVE APPROACH (SMITH)

One of the strongest traditions in economics has been a literary methodology whereby causality problems are explored almost exclusively with words, but without an explicit definition of 'cause.' This sort of causality analysis goes back at least to Adam Smith (2010). The *narrative approach* consists of a chain of reasoning that it is both theoretical and contextual.

More precisely, the narrative approach exhibits two distinguishing features: first, the approach uses descriptive economics to provide reasons that certain factors (say, F_1, F_2, ... , F_m) provoke a particular effect E but it is not necessarily based on a comprehensive internally consistent formal model showing that when the factors F_1, F_2, ... , F_m are present the effect E must follow; and second, the approach does not have an explicit definition of 'cause.' This approach can be used to address causality issues that fall into either category of causation (general or singular).

When there are several causative factors, it is necessary to distinguish 'separable' from 'non-separable' (candidate) causes. If F is one of the causes

[1]Somewhat roughly, a variable y is said to 'Granger-cause' another variable z if prediction of z is improved by using past values of y. The time-series notion of causality uses time dating of variables to detect empirical causes. Granger causality is data-driven and does not depend on theoretical economics.

of the effect E, then F can be a cause of E by itself. In this case, F is called a *separable* cause. But there may also be non-separable causes of E. It is said that a factor G is a *non-separable* cause of E if either the existence of G presupposes that F must be present or G is brought about by one or more separable causes.

It is not hard to point out the relevance of the narrative approach for a large number of economic problems. To see the possibilities and limitations of this methodology, we briefly developed two specific examples.

2.1. The Causes of the Wealth of Nations[2]

Smith was convinced that it was possible to discover the causes of the wealth of nations by means of scientific investigation. His arguments are more than just theoretical description; they are also contextual in that the reasoning is based on his observations of the existing historical and institutional circumstances.

According to Adam Smith, the cause of the wealth of a nation is the accumulation of capital. The chain of reasoning that establishes this result can be condensed as follows:

The wealth of a nation W depends on two variables: productivity of labour (X_1) and the ratio of productive and unproductive labour (X_2); in symbols, $W = F(X_1, X_2)$;

- X_1 depends on the division of labour (Y_1); in symbols, $X_1 = X_1(Y_1)$;
- Y_1 in turn depends on the extent of the market (Z_1) and capital accumulation CA; in symbols, $Y_1 = G(Z_1, CA)$;
- Z_1 in turn depends on capital accumulation; in symbols, $Z_1 = Z_1(CA)$; and finally,
- X_2 depends on capital accumulation as well; in symbols, $X_2 = X_2(CA)$.

In a nutshell, the reasoning involves chains of relationships between economic variables. The wealth of a nation turns out to be a composite function of capital accumulation

$$W = W(CA), \tag{1}$$

where: $W(CA) = F\{X_1[G(Z_1(CA)], CA), X_2(CA)\}$. A change in CA provokes a 'chain reaction' that ends up affecting W.

The foregoing archetypal example of causality analysis in economics displays the two distinguishing features of the narrative approach: first, the argument provides reasons to believe that capital accumulation is the cause of the wealth of nations; and second, there is no explicit definition of 'cause.'

2.2. Hunting the Causes of the Banking Crisis 2007-2008

What caused the USA banking crisis 2007-2008? To answer this question the narrative approach starts with an exposition of the salient aspects of the banking crisis 2007-2008. The purpose is to gain an understanding of the event with a view to identifying the most prominent factors F_1, F_2, ... ,F_m underlying the housing bubble and the credit boom that brought the process of lending to a standstill in 2008.For concreteness, the freeze in the credit markets that occurred in October 2008 is identified here with the effect E in the causal relationship 'F_1, F_2, ... , F_m caused E.'

It goes without saying that discussing the myriad aspects of the banking crisis would take us too far a field. Fortunately, panoramic perspectives of the crisis can be found in Brunneirmeir (2009) and Pol (2012). These overviews provide an event logbook on the banking turmoil in 2007-2008 together with a discussion of the amplification mechanisms that led to the most severe banking crisis since the Great Depression. It is not difficult to see that there are at least nine plausible causative factors that immediately suggest themselves. These factors are gathered together in Table **1**.[3]

The origins of the banking crisis can be traced to two plausible causative factors F_1 (loose monetary policy) and F_2 (global saving glut). In fact, it is generally agreed that persistently low interest rates after the burst of the Internet bubble combined with the high demand for safe securities, especially from Asian

[2]This illustration draws on (Landreth and Colander 2002:86-90).

[3]The fact that factors such as the refinancing ratchet effect, the rating agencies, and the Credit Default Swaps (CDSs) are left out of the picture does not imply that they are irrelevant, merely that their role in the banking crisis was contingent on the presence of one or more of the items enumerated in Table **1**. For example, it is evident that the CDSs delayed the occurrence of the financial collapse and constituted an important magnifying factor of wealth destruction but they cannot be held responsible for the banking crisis. The biggest problem was the fragility of the securities with bad loans in them, not the CDSs.

central banks, engendered a platform for the development of the crisis.

Table 1: Plausible Causative Factors Underlying the Housing Bubble and the Credit Boom

Plausible Causative Factor	Name
F_1	Loose monetary policy
F_2	Global saving glut
F_3	Poor supervision
F_4	Sub-prime market boom
F_5	High securitization activity
F_6	Re-securitization
F_7	Ignorance
F_8	Excessive risk-taking
F_9	Too-big-to-fail

This table reports what appear to be the most important factors conducive to the housing bubble and the credit boom emerging from comprehensive academic research. Each of them suggests itself as a seemingly worthy candidate to be a cause of the USA banking crisis 2007-08.

There is no doubt that F_3 (poor supervision) substantially contributed to worsening the opacity of the financial system. However, it would be hard to argue that this factor in *isolation* provoked the financial catastrophe. Supervision cannot restore transparency if the existing financial regulation is obsolete. For example, the 'shadow banking' system was beyond the control of the U.S. Securities and Exchange Commission.

It is universally recognized that F_4 (sub-prime market boom) was a key factor conducive to the banking crisis. There were three main reasons for the vigorous expansion of the sub-prime market: misaligned incentives of the underwriters, who believed they had little exposure to risk; behaviour of the rating agencies, which did not properly represent risk to investors; and a decline in lending standards which allowed increasingly poor loans to be made. (Dell'Ariccia, Igan and Laeven 2008) establish a link between the credit bubble and the deterioration of the lending standards in the sub-prime market. Their model confirms that with sound lending standards the USA sub-prime mortgage market would have remained relatively small. (Allen and Carletti 2010) argue that loose monetary policy, particularly in the USA, and global imbalances created a bubble in real estate prices in the USA and other developed countries such as Spain and Ireland. These authors also argue that many other factors such as F_5 (high securitization activity) exacerbated the effects of F_1 and F_2 on the sub-prime market.

Re-securitization refers to the process of pooling and tranching a whole set of, for example, mortgages to spread risk differentially implemented by the investment banks. The archetypal example of F_6 (re-securitization) is the creation of *Collateralized Debt Obligations (CDOs)*. The root problem with F_6 was the lack of incentives to monitor the quality of the underlying loans. It should be clear that F_6 cannot exist without securitization because re-securitization presupposes securitization. Therefore, F_6 cannot be separated from F_5.

An efficient financial system presupposes that (a) people have easy access to all relevant information; (b) the availability of information automatically implies a clear understanding of the possibilities and limitations of the products in question; and (c) all risks are recognized *ex-ante*. For lack of a better term, we call this presumption *postulate of full comprehension*. For example, according to this postulate economic agents are perfectly aware of the existence of worst states of the world associated with complex financial products such as the CDOs and they do not ignore the probability of occurrence of the worst states.

If (for whatever reason) some risks associated with financial products are ignored, the factor F_7 (ignorance) is present. (Gennaioli, Shleifer, and Vishny, 2011a) assume that investors and intermediaries did not satisfy the postulate of full comprehension during the unfolding of the banking crisis 2007-08 because they *neglected* tail risks.[4] They argue that, with the neglected risk assumption, new financial products provide false substitutes for truly safe bonds and the financial system is fragile. Specifically, (Gennaioli *et al*. 2011a) state that, "A small piece of news that brings to investors' minds the previously unattended risks catches them by surprise, causes them to drastically revise their valuations of new securities, and to sell them in the market."

As to the second last plausible causative factor F_8 (excessive risk-taking), there is evidence that low interest rates induce imprudent risk-taking. (Maddaloni and Peydro 2011). This suggests that F_8 was brought about by F_1 (loose monetary policy), and therefore, F_8

[4]If investors and intermediaries ignore tail risks, it is said that the assumption of *neglected risk* is met.

cannot be considered as a separable cause of E. On the last causative factor F_9 (too-big-to-fail), it is not inconceivable that the 'too-big-to-fail' phenomenon may have encouraged large and complex financial institutions to take on too much risk. However, few analysts would argue that the expectation that taxpayers would end up footing the bill of bank loss was one of the separable causes of the financial debacle at the end of 2008. It does not appear to be solid evidence that F_9 played a key role in engendering the banking crisis.

The foregoing discussion enables us to reduce the number of plausible candidate causes of the USA banking crisis 2007-08 to six (see Table **2**). It should be stressed that the process of arriving at Table **2** is just reliable inference. In other words, there are reasons to believe that S_1, S_2, S_3, S_4, S_5, and S_6 were key factors conducive to the banking crisis, but we do not have a catch-all formal model *showing* that if S_i (i = 1,..., 6) are present, then there is a banking crisis. This is as far as we can go with the narrative approach.

Table 2: Separable Causative Factors of the Banking Crisis 2007-08

Separable Causative Factor	Name
S_1	Loose monetary policy
S_2	Global saving glut
S_3	Poor supervision
S_4	Sub-prime market
S_5	High securitization activity
S_6	Ignorance

The elimination from Table **1** of F_6 (re-securitization), F_8 (excessive risk-taking), and F_9 (too-big-to-fail) is based on logic connections, empirical evidence, and lack of evidence, respectively. This leads to six separable causative factors, denoted by S_1, S_2, ... , S_6, that were present during the housing bubble and the credit boom.

2.3. Drawbacks

The above examples are but a minuscule sample of specific economic problems which can be explored using the narrative approach. Beyond any doubt, the narrative approach serves as a first approximation because it brings into sharp focus the key factors that should be taken into account to discover cause-and-effect relationships. However, purely verbal analyses tend to be tedious and difficult to formulate consistently especially when there are multiple causative factors inextricably intertwined.

Readers seeking a precise, clear guide to causality analysis within the narrative approach are bound to be disappointed, for they will find none. This approach serves as a good first approximation, but the results should be treated with caution, for two related reasons. First, each separable factor in the causal argument may be a compelling element, but does the argument truly hold? Second, the narrative approach does not have an operational definition of 'cause,' and therefore we cannot test whether the candidate causes are in fact the causes of E.

3. COMPARATIVE STATICS APPROACH (MARSHALL)

The most explicit recognition of the importance of causality in economics was made by Alfred Marshall in his *Principles of Economics*. Marshall's notion of cause-and-effect relation revolves around the idea of *ceteris paribus* change:

> It is sometimes said that the laws of economics are "hypothetical." Of course, like every other science, it undertakes to study the effects which will be produced by certain causes, not absolutely, but subject to the condition that *other things are equal*, and the causes are able to work out their effects undisturbed. Almost every scientific doctrine, when carefully and formally stated, will be found to contain some proviso to the effect that other things are equal: the action of the causes in question is supposed to be isolated; certain effects are attributed to them, but only *on the hypothesis* that no cause is permitted to enter except those distinctly allowed for.

Marshall (1966:30) [*Italics in original*]

Marshall's conception of causality in economics gave rise to the development of comparative statics.[5]

The intuition behind the comparative statics procedure is easily expressed in words. The methodology consists of comparing one equilibrium situation with another where the equilibrium situations are associated with different sets of values of exogenous variables. Comparative statics is typically used to generate testable hypotheses of cause and

[5]According to Joseph A. Schumpeter, the term 'comparative statics' was coined by Franz Oppenheimer (1864-1943) but the insight was first introduced by David Ricardo. Schumpeter (1954:494 and 965).

effect. For example, if a government uses price controls (cause) to keep the price of food low during a drought, food shortages (effect) would occur.

3.1. Mathematical Treatment

The translation of the comparative statics methodology into the mathematical language was accomplished by Paul A. Samuelson in his *Foundations*. Samuelson (1965, especially Chapters II and III). In this context, causation is formally expressed in terms either of partial derivatives or finite changes. Using differential calculus, it is not difficult to sketch the strategy for deriving cause-and-effect propositions.

Typically, an economic model or theory involves three constituent parts: a collection of endogenous variables $x = (x_1, x_2,\ldots,x_n)$; a collection of exogenous variables $a = (a_1, a_2, \ldots, a_m)$; and finally, a system of equations

$$f_i(x, a) = 0 \qquad (i = 1, 2,\ldots,n) \tag{2}$$

The equilibrium system (2) consists of n independent functions reflecting a set of economic assumptions. Under certain regularity conditions imposed to the functional relationship

$$f(x, a) = (f_1(x, a), f_2(x, a),\ldots,f_n(x, a)), \tag{3}$$

the implicit function theorem guarantees that the endogenous variables are determined by the exogenous variables in the following way: in correspondence with a pre-assigned vector a^0, there is a unique set of values for the endogenous variables x^0 such that $f(x^0, a^0) = 0$. Or, to put it differently, the implicit function theorem implies the (local) existence of n explicit functions

$$x_i = g_i(a_1, a_2, \ldots, a_m) \qquad (i = 1, 2,\ldots,n) \tag{4}$$

The changes in the exogenous variables are the *causes* of changes in the endogenous variables and $\partial x_i/\partial a_j$ gives the (marginal) causal effect of the *ceteris paribus* change in a_j on the variable x_i.

3.2. Miscellaneous Comments

It should be clear that the identification of a cause depends on the theoretical model. For example, in the familiar model of constrained utility maximization *ceteris paribus* changes in the product prices typically cause changes in the quantity demanded of a particular product. But in a Walrasian general equilibrium model the product prices are endogenous variables. It should also be clear that difficulties of interpretation arise when several parameters change simultaneously:

> The only sense in which the use of the term causation is admissible is in respect to changes in external data or parameters. As a figure of speech, it may be said the changes in these *cause* changes in the variables of our system. An increase in demand, i.e., a shift in the demand function due to a change in the data, tastes, may be said to cause an increased output to be sold. Even here, when several parameters change simultaneously, it is impossible to speak of causation attributable to each except in respect to limiting rates of change (partial derivatives).

Samuelson (1965:9-10) [*Italics in original*]

Five additional comments are in order. First, in this conceptual scheme the notion of equilibrium establishes a causal nexus between exogenous and endogenous variables. Second, it should be noticed that the partials $\partial x_i/\partial a_j$ are obtained through implicit differentiation in order to capture the properties of the equilibrium system (2). Third, when it is possible to determine unambiguously the sign of the effect of a *ceteris paribus* change in a_j on x_i, it is customary to use modes of expressions that invoke causation. For example, if we know that $\partial x_i/\partial a_j > 0$, then economists say that (within the model) changes in a_j *ceteris paribus* cause, or lead to, changes in x_i in the same direction. Fourth, nothing is explicitly said in regard to the empirical testing of the definition of causality. Finally, consistency (within the model) requires that the new equilibrium situation corresponding to a *ceteris paribus* change of a_j be asymptotically stable.[6]

3.3. Summary

To sum up, the familiar methodology of comparative statics can be thought of as an approach to causality analysis in economics where the exogenous variables are the causes and the changes in the endogenous variables are the effects provoked by a disturbance of one or more exogenous variables. The cause-and-effect relationships are inextricably linked to the existence and stability of the equilibrium. This approach

[6]Loosely speaking, a dynamic adjustment process is asymptotically stable if it restores equilibrium should the equilibrium be disturbed.

is flexible enough to deal with both general and singular causation.

4. THEORETICAL APPROACH (HICKS)

One approach to causality in economics dealing with singular causation is due to Hicks (1979). For reasons that will become apparent in a moment, Hick's causality methodology can be termed 'theoretical' approach to causality. In rough outline, the Hicksian approach to causality involves three steps: first, formulation of a counterfactual definition of 'cause'; second, given a collection of plausible causative factors, distinction between separable and non-separable causative factors; and finally, causality tests to identify separable causes.

4.1. Single Cause (Strong Causation)

According to Hicks (1979:12), to assert that 'C caused E' presupposes that both C and E existed and involves positing that

'If C had not have happened, E would *not* have happened.' (5)

This definition assumes that C is the *sole* cause of E (*strong* causation), that is, there is no other factor which is admitted to be a cause of E. Hicks (1979:13).

To test this basic definition, one must construct the hypothetical situation

'C did not exist,' or briefly, 'not-C.' (6)

The reason is easily seen. We know that C did happen, but we do not know what would have happened if C had not happened. In turn, the not-C situation requires a model or theory of the way C and E are connected. The model should provide reasons for thinking that events in reality may have been connected in the hypothesized way.

If we represent the expression 'E would *not* have happened' by the symbol $\ominus$, that is,

$\ominus \equiv$ E would *not* have happened, (7)

the foregoing definition of cause can be slightly reformulated as follows:

'C caused E' if 'not-C produces $\ominus$.' (8)

In the case of a single cause, the causality test is fairly obvious:

- If not-C produces $\ominus$, then C caused E; and
- If not-C produces $\oplus$, then C did not cause E,

where the symbol $\oplus$ is defined as

$\oplus \equiv$ 'E would have happened' (9)

Suppose that the event E represents the USA banking crisis 2007-08 and ask: What caused E? One can always answer this question specifying a cause that is so general as to be useless. For instance, we can claim that 'capitalism' was the cause of the financial crisis 2007-08. The sheer number and variety of elements characterizing 'capitalism' makes the claim both trite and trivial. The 'cause' just mentioned reminds us what Hayek (1967) suggested many years ago. The complexity of the economy makes it possible for the economist to seek only the most prominent causative factors at work prior to the occurrence of E.

4.2. Multiple Causes (Weak Causation)

It should be clear that the definition of strong causation is very stringent. Typically, there are several separable causes (say, the factors C_1, C_2, C_3, etc.) operating jointly to produce the effect E. When there are multiple causes, the statement 'C caused E' means that C was one of the causes of E, and causation is said to be *weak*. The importance of weak causation is evident because in the economy there are typically quite a number of factors operating together to produce an effect.

We now consider the simplest case of weak causation (two separable factors C_1 and C_2) and confine attention to two situations:

(1) The effect E does not occur when either C_1 is absent or C_2 is absent or both C_1 and C_2 are absent; or

(2) The effect E occurs if either factor is present, but it does not occur when both C_1 and C_2 are absent.

The first situation can be symbolically represented as follows:

Situation 1

not-C_1 and C_2 is present, *ceteris paribus*, produces $\ominus$;

not-C_2 and C_1 is present, *ceteris paribus*, produces $\ominus$;

not-C_{12} *ceteris paribus* produces $\ominus$, where the symbolism not-C_{12} denotes a theoretical construction in which both C_1 and C_2 are absent. If situation 1 holds, it is said that C_1 and C_2 are *additive causes*. Hicks(1979:15). The present situation can be visualized with the help of Table **3**.

Table 3: Additive Causes

Not-C_1 (C_2 present)	Not-C_2 (C_1 present)	Not-C_{12}
⊖	⊖	⊖

The not-C_1 construction is a model that allows the evaluation of the logical connection between the statement "C_1 absent and C_2 present, *ceteris paribus*" and the mutually exclusive outcomes "⊖ (non-occurrence of E)" and "⊕ (occurrence of E)." According to the first column of this table, if the statement "C_1 absent and C_2 present, *ceteris paribus*" is inserted into the model the effect E does not occur. A similar interpretation can be given to the not-C_2 and not-C_{12} constructions. For example, the last column of this table asserts that within the model "C_1 absent and C_2 absent, *ceteris paribus*" implies the non-occurrence of E. All in all, this table asserts that the effect E will not happen unless both causes are present.

Situation 2 captures the possibility that additivity breaks down. In symbols,

Situation 2

not-C_1 and C_2 is present, *ceteris paribus*, produces ⊕;

not-C_2 and C_1 is present, *ceteris paribus*, produces ⊕; and

not-C_{12}*ceteris paribus* produces ⊖

This is the case of *overlapping causes*. Hicks (1979:15). Table **4** shows symbolically the essence of overlapping causes.

Table 4: Overlapping Causes

Not-C_1	Not-C_2	Not-C_{12}
⊕	⊕	⊖

The interpretation of the symbolisms not-C_1, not-C_2 and not-C_{12} is the same as in Table **3**. The first and second columns of this table assert that the absence of only one factor (C_1 or C_2) is not enough to prevent the occurrence of the effect E. The message conveyed by this table is that the effect E will occur unless both C_1 and C_2 are absent.

4.3. Identifying the Causes of the Banking Crisis 2007-08[7]

What is the use of all these mental gymnastics? We can revert to the USA banking crisis 2007-08 to provide insight into the answer to this general question.

The narrative approach ended with a list of six separable causative factors (see Table **2**). According to the theoretical approach, the proof that S_1 (loose monetary policy), S_2 (global saving glut), S_3 (poor supervision), S_4 (sub-prime market), S_5 (high securitization activity), and S_6 (ignorance) are separable causes of E (freeze of the credit markets in October 2008) would require at least one theoretical construction capturing these key factors. In reviewing the literature on the banking crisis 2007-08, we have not uncovered any work that incorporates all these six factors into a single model. It is likely that a model explicitly involving such large number of factors would be intractable or ambiguous.

Faced with this stumbling block, we can posit a hierarchical classification of the causative factors. The *first causes* of E are those separable causative factors that establish a platform for the unfolding of the banking crisis. (Archaya and Richardson 2009) argue that loose monetary policy, global imbalances, poor supervision, and the sub-prime market were the first causes of E. The collection of these causative factors can be thought of as a 'crisis environment' necessary – but not sufficient– for E to occur. To guarantee the occurrence of E, there has to be supplementary and powerfully operative factors. We call these high-powered factors which have to be embedded into the crisis environment to provoke the freeze of the credit markets, the *preponderant causes* of E.

Few economists would deny that C_1 (securitization) and C_2 (ignorance)[8] were powerfully operative factors at work prior to the occurrence of E. However, to argue that there are reasons to believe that C_1 and C_2 constituted important separable factors in the context of the banking crisis is not the same as showing that they were, in fact, the preponderant causes of the crisis. We can use the Hicksian methodology for causality analysis together with a model of shadow banking developed by (Gennaioli, Shleifer and Vishny 2013) to show that C_1 and C_2 were the (additive) preponderant causes of the USA banking crisis 2007-08. To this end, we must show that the following conditions are satisfied:

Condition ①: C_1 is absent while C_2 is present *ceteris paribus* imply

E does not happen;

Condition ②: C_2 is absent while C_1 is present *ceteris paribus* imply

[7]This illustration draws on Pol (2012).

[8]The change in notation from S_5 and S_6 to C_1 and C_2, respectively, is just to facilitate contact with the notation in the present section.

Table 5: Schematic View of the Causality Proof

Causes \ Cases	Case ①	Case ②	Case ③
C_1 (Securitization)	Absent	Present	Absent
C_2 (Ignorance)	Present	Absent	Absent
E (Credit Market Freeze)	⊖	⊖	⊖

The proof that securitization and ignorance were the preponderant causes of the USA banking crisis 2007-08 is based in the Gennaioli-Shleifer-Vishny model of shadow banking. The last row in Table **5** shows that inserting into this model the assumptions corresponding to each case implies that the credit market freeze does not occur. For example, Case ① (second column in Table **5**) says that if securitization is absent and ignorance is present, the crisis does not happen. All in all, the message conveyed by Table **5** is that the crisis does not happen unless both C_1 and C_2 are present.

E does not happen; and

Condition ③: Both C_1 and C_2 are absent *ceteris paribus* imply

E does not happen.

The proof is briefly sketched in Table **5**.

4.4. Limitations

There can be little doubt that the Hicksian approach is a coherent methodology to organize thinking about singular causation. Its principal message is that theoretical economics is a key tool for hunting the causes of singular events that cannot be replicated. But praise does not imply perfection. Like most methodologies this approach has weaknesses.

Researchers employing the theoretical approach to causality may encounter two remarkable limitations. First, the methodology may be inconclusive. Suppose that two well-specified models M_a and M_b are suitable to accommodate the not-C construction. Suppose, in addition, that model M_a produces ⊖ but model M_b produces ⊕. In this hypothetical situation, Hicks' account of causation turns out to be inconclusive. We are left, in principle, with an indeterminate outcome. Second, when the number of potential causes increases the logical complexities of the Hicksian approach are formidable. For example, the test for three potential causes, say C_1, C_2, and C_3, involves the analysis of seven hypothetical situations:

Not-C_1: C_1 was absent while C_2 and C_3 were present;

Not-C_2: C_2 was absent while C_1 and C_3 were present;

Not-C_3: C_3 was absent while C_1 and C_2 were present;

Not-C_{12}: C_1 and C_2 were absent while C_3 was present;

Not-C_{13}: C_1 and C_3 were absent while $C_2$was present;

Not-C_{23}: C_2 and C_3 were absent while $C_1$was present;

Not-C_{123}: C_1, C_2 and C_3were absent.

5. STRUCTURAL APPROACH (HECKMAN)

The traditional approach to causality in econometrics goes back to Haavelmo's (1943) and is known as *structural approach* because it is based on structural equations models. These models rely on the specification of systems of equations representing behavioural relationships between endogenous variables and exogenous variables, and the method of controlled variation.[9]

Conceptually speaking, the structural approach to causality is a refinement of the comparative statics approach. According to James J. Heckman (2000), the structural approach to causal modelling is the major contribution of twentieth century econometrics. To quote Heckman extensively,

> Econometric theory was developed to analyse and interpret economic data. Most econometric theory adapts methods originally developed in statistics. The major exception to this rule is the econometric analysis of the identification problem and the companion analyses of structural equations, causality, and economic policy evaluation. Although an economist did not invent the phrase "correlation does not imply causation,"

[9]'Controlled variation' means 'variations in treatment holding other factors constant.' This method goes back to Marshall (1966) who repeatedly used the method of controlled variation in his *ceteris paribus* clauses.

> economists clarified the meaning of causation within well-specified models, the requirements for a causal interpretation of an empirical relationship, and the reasons why a causal framework is necessary for evaluating economic policies.

Heckman (2000:45-46)

5.1. Definition of Cause and Hypotheticals

One essential distinguishing feature of the approach under consideration is that a definition of 'cause' based on 'hypotheticals' emerges naturally from a well-posed economic model M as an automatic by-product. An outcome within M is termed *hypothetical*. Hypotheticals may refer to the effects of policies never previously experienced or products never previously consumed or may be contrary to certain facts. Consequently, a counterfactual is a hypothetical but the converse is not necessarily true.

Any definition of causality involves two ingredients: first, a set of hypotheticals generated by a system of functions depending on 'determinants'; and second, a 'manipulation' where one or more determinants are changed. The *ceteris paribus* change in outcome associated with manipulation of the varied determinant is called a 'causal effect' of the manipulated determinant. In a nutshell, a 'cause' is a 'manipulated' determinant within a model M.

5.2. Causal Effects Between Endogenous Variables[10]

As an illustration of the structural approach, we consider the problem of defining and estimating the causal effect of one endogenous variable on another. Identification of causal effects can be defined by exclusion of variables. In rough outline, the procedure can be sketched as follows. Consider a model M represented by a set of non-linear simultaneous equations and call this set structural system (S). To simplify things, we confine attention to the following version of (S):

$$x_1 = G_1(x_2, a_1, a_2, U_1) \tag{10}$$

$$x_2 = G_2(x_1, a_1, a_2, U_2) \tag{11}$$

where x_1 and x_2 are endogenous variables, and a_1, a_2, U_1, and U_2 are exogenous variables specified independently of (S).

The critical issue is to define the causal effect of one endogenous variable on another in the context of an interdependent system such as (S). The causal effect of x_2 on x_1 is defined as

$$\partial G_1/\partial x_2 \tag{12}$$

It is clear that one can define the causal effect of x_1 on x_2 by analogy, that is,

$$\partial G_2/\partial x_1 \tag{13}$$

What may not be as obvious is the fact that we can empirically estimate the causal effects between endogenous variables using the partials of the reduced system with respect to a_j ($j = 1, 2$).

To make contact with the empirical evidence, we solve the endogenous variables in terms of the exogenous variables to obtain the reduced system (R):

$$x_1 = H_1(a_1, a_2, U_1, U_2) \tag{14}$$

$$x_2 = H_2(a_1, a_2, U_1, U_2), \tag{15}$$

where a_1 and a_2 are assumed statistically independent of U_1 and U_2.These functions can be determined from the empirical data. Taking the partial derivative of x_1 with respect to a_1, yields

$$\partial x_1/\partial a_1 = \partial G_1/\partial x_2 \times \partial x_2/\partial a_1 + \partial G_1/\partial a_1 \tag{16}$$

Assuming exclusion with respect to a_1, that is,

$$\partial G_1/\partial a_1 \equiv 0, \tag{17}$$

yields

$$\partial G_1/\partial x_2 = \partial H_1/\partial a_1 \div \partial H_2/\partial a_1 \tag{18}$$

The ratio of partials on the right hand side of (18) can be quantified using the reduced system (R).

The intuition behind this methodology to define causal effects by exclusion of variables can be easily seen by considering the linear case:

$$x_1 = \theta_{12}x_2 + b_{11}a_1 + b_{12}a_2 + U_1 \tag{19}$$

$$x_2 = \theta_{21}x_1 + b_{21}a_1 + b_{22}a_2 + U_2 \tag{20}$$

Using simple algebra (for example, Cramer's rule) allows us to solve out for the endogenous variables (x_1, x_2) as a function of the exogenous variables:

[10]This illustration draws on Heckman (2008).

$$x_1 = \rho_{11}a_1 + \rho_{12}a_2 + V_1 \quad (21)$$

$$x_2 = \rho_{21}a_1 + \rho_{22}a_2 + V_2 \quad (22)$$

In particular,

$$V_1 = (U_1 + \theta_{12}U_2)/\Delta,\ V_2 = (U_2 + \theta_{21}U_1)/\Delta,\ \Delta = 1- \theta_{12}\theta_{21}, \quad (23)$$

and (a_1, a_2) is assumed to be statistically independent of (V_1, V_2).

We can estimate the ρ_{ij} (i = 1,2 and j = 1,2) by ordinary least squares. Exclusion with respect to a_1 means $b_{11} = 0$, that is, a_1 affects x_1 only through its effect on x_2. Using (18), it follows at once that

$$\partial G_1/\partial x_2 = \theta_{12} = \partial H_1/\partial a_1 \div \partial H_2/\partial a_1 = \rho_{11} \div \rho_{21} \quad (24)$$

5.3. Summary and Limitations

To summarize, the definition of causal effects between endogenous variables requires (a) the formulation of a well-posed model M; (b) the exogenous variables can be manipulated; (c) exclusion of one or more exogenous variables; and (d) for the given model M, the functions (10) and (11) do not shift when the exogenous variables are manipulated. It should be clear that all attempts to define causality using well-posed models require some imposed structure. *A priori* assumptions reflecting postulated causal mechanisms make the estimation of the parameters of the model possible.

All in all, the important messages conveyed by the structural approach to causality are the following seven: first, causality is a property of a model; second, hypotheticals are an ingredient of causal models; third, causal models also specify a mechanism by which the causal variables are externally manipulated; fourth, ambiguity in model specification implies ambiguity in the notion of causality; fifth, the models used are provisional and depend on *a priori* assumptions (this is what Heckman calls 'provisional nature of causal knowledge'); sixth, many models are consistent with the same data; finally, different economists may construct different (well-posed) models, and therefore, identify different factors as causes, while ignoring others.

In reviewing research on the structural approach to causality, we have found that the main criticisms to this methodology are the following two: the structural econometric modelling relies on many strong *a priori* statistical and economic assumptions, and entails formidable computational complexities. Keane (2010a).

6. EXPERIMENTALIST APPROACH (ANGRIST-IMBENS)

Frustration with the fact that the structural approach to causality hinges on "too many assumptions" and implies overly complex computational procedures gradually led econometricians to eschew the structural approach in favour of experimentation. Since the early 1990s, the experimentalist approach to causality has been in vogue. The experimentalist perspective seeks to infer causes from data without modelling the causes of effects. Somewhat roughly, the idea is that we can learn interesting things about causality without using too much theoretical economics – the guiding principle of the experimentalist camp in econometrics is "just let the data speak."

6.1. Instrumental Variables and Rubin Causal Model

The experimentalist (or 'a theoretical') approach to causality in economics emerged from the conflation of the instrumental variable (IV) notion and the *Rubin Causal Model (RCM)*.[11] Instrumental variables are variables excluded from some equations and included in others, so that they are correlated with some outcomes only through the effect on other variables. The RCM is the key tool in the area of statistical causality. This model uses the terms 'cause' and 'treatment' interchangeably.[12]

The statistical approach draws on experiments to deal with causality and suggests that random assignment is the most convincing way to deal with causality. In essence, the statistical approach maintains that causality can only be determined by randomization and thereby reify randomization as the gold standard of causal inference.

In the RCM the causal effect of a treatment (e.g. military service) is the comparison between the value of the outcome if the unit is treated and the value of the outcome if the unit is not treated. The average causal effect is defined as the average difference between treated and untreated outcomes across all units in a population or some subpopulation.

[11]The term Rubin Causal Model (RCM) was coined by Holland (1986) as referring to a model for causal inference where causal effects are defined by comparing potential outcomes.
[12]For a discussion of the statistical approach to causal inference and what can be a cause, see Holland (1986).

6.2. Local Average Treatment Effect

Several key papers marked the rising popularity of the experimentalist approach. In the list of seminal contributions one should include Angrist (1990), (Imbens and Angrist 1994), and (Angrist, Imbens and Rubin 1996), to mention only the most prominent.

The parameter *LATE (local average treatment effect),* introduced in the econometrics literature by (Imbens and Angrist 1994), plays a central role in this approach, and is defined as follows: "LATE is the average treatment effect for individuals whose treatment status is influenced by changing an exogenous regressor that satisfies an exclusion restriction." (Imbens and Angrist, 1994:467). Angrist, Imbens, and Rubin showed that the IV estimator can be embedded within the RCM and this estimator is the average causal effect for a subgroup of units, the compliers. Specifically, it is shown in (Angrist *et al*. 1996) that the average causal effect for compliers is the LATE.

6.3. The Thrust of the Estimation Procedure[13]

Without striving for rigour, the experimentalist approach to causality can be described as follows. In an equation

$$x_1 = F(x_2, U) \tag{25}$$

with endogenous variables x_1 and x_2, and error U (unobservables that also affect x_1), the analyst is interested in estimating the causal effect of x_2 on x_1. For example, if x_1 represents future earnings, and x_2 denotes schooling, then $F(x_2, U)$ can be thought of as an earnings function where U captures factors such as inborn talents; the problem is how to measure the causal effect of an additional year of education on future earnings. The solution to this problem is far from obvious because people with different education levels tend to exhibit different levels of U.

All in all, the experimentalist approach stresses the importance of looking for exogenous sources of variation in potential causes. It defines causality using experimental manipulations. Specifically, a causal effect is defined by a randomization. The procedure to estimate the causal effect of x_2 on x_1 can be presented in three steps.

[13]This step-by-step explanation draws on Keane (2010a).

Step 1: Find a natural experiment

A natural experiment is an event or situation that generates exogenous variation in certain variables that would otherwise be endogenous.

Step 2: Identify an instrument

An instrument generates random assignment. Specifically, the analyst identifies an event that affects a random subset of the population inducing some members in this subset to choose a lower (higher) level of x_2 than they would have otherwise. Typically, forces of nature and human institutions provide random assignment.

Step 3: Use an IV procedure

Finally, an instrumental variable procedure is used to estimate the causal effect of x_2 on x_1.

6.4. Illustration: Military Service and Earnings

To illustrate the essence of the experimentalist approach to causality, we use the highly influential paper by Angrist (1990). This paper takes advantage of the Vietnam era lottery numbers to set up a natural experiment that randomly influenced those who served in the military and estimates the average causal effect of military service on subsequent earnings. Note that the draft lottery number affects x_2 (schooling) but is uncorrelated with the unobservables U (e.g. innate ability) that also affect x_1 (subsequent earnings).

The starting point was a sample of the men born between 1950 and 1953. Each man was assigned a number from 1 to 365 based on random drawings of birth dates. Only those below a certain ceiling were draft eligible. For each cohort, this selection procedure generated a random subset of the corresponding populations. The subset of draft eligible men who were actually drafted into the military was determined after various tests and physical exams. The instrumental variable is the randomly assigned lottery number.

'Compliers' are men who were induced by the draft lottery to serve in the military.[14] The potential outcomes are earnings if men served in the military or if they do not. There are three necessary assumptions to identify the average causal effect of serving in the military on

[14]Individuals who would volunteer irrespective of the lottery number are called 'always-takers' and those who would not serve irrespective of their lottery numbers are identified as 'never-takers.'

the subpopulation of compliers. First, draft eligibility is exogenous. Second, there is no direct effect of the instrument on the outcome (exclusion restriction). And finally, it is assumed that any man who would serve if not draft eligible, would also serve if draft eligible (monotonicity assumption).[15]

For each cohort, Angrist runs a regression of earnings in some subsequent year (1981 through 1984) on a constant and a dummy variable for veteran status and concludes that military service reduced annual earnings for whites by about $1,500 and $3,000 in 1978 dollars. Thus, for compliers average earnings were 15% lower in the early 1980s than they would have been otherwise. In brief, Angrist (1990) concluded that the effect of military service was to reduce earnings of whites, but not non-white veterans.

6.5. Two-Stage Procedure

Returning to the experimentalist paradigm, one obvious point should be mentioned. Like most researchers, experimentalists have an interest on the effects of causes as well as the economic mechanisms producing the effects. As a result, it is clear that the experimentalist paradigm involves two stages of research. In the first stage, experimentalists “let the data speak” without imposing *a priori* economic mechanisms on the data. The second stage consists of finding an economic explanation for the causes of effects.

The Angrist (1990) draft lottery paper provides a good illustration of the two-stage procedure. Angrist (1990) first shows that there is an adverse effect of military experience on earnings ignoring the mechanism whereby military service affects earnings, and then he proposes an explanation for the loss of earnings of white veterans, namely: they earn less because their military experience is a poor substitute for lost civilian labour market experience.

6.6. Credibility Revolution

Thirty years ago Edward E. Leamer published his famous “con out of econometrics” article on the state of empirical work in economics. Leamer’s pessimistic assessment of the state of affairs in econometrics was categoric:

> This is a sad and decidedly unscientific state of affairs we find ourselves in. Hardly anyone takes data analysis seriously. Like elaborated plumed birds who have long since lost the ability to procreate but not the desire, we preen and strut and display our *t*-values.

Leamer (1983:37)

Experimentalists claim that their approach to causality has produced a credibility revolution in econometrics. The approach has found widespread appeal in the economics profession (as mentioned before, this approach has been in vogue since roughly 1990). (Angrist and Pischke 2010) present a catalogue of examples of practical applications of the experimentalist approach to support the view that “empirical microeconomics has experienced a credibility revolution, with a consequent increase in policy relevance and scientific impact.”

Beyond any doubt, the experimentalist approach has made a valuable contribution to the study of causality in economics. However, it is the limitations of the approach that have excited the most heated debate. Experimentalists believe that their approach is revolution –not evolution– and thereby, criticism should be expected as a natural reaction. Or, to put it in the words of (Angrist and Pischke 2010) “The rise of the experimentalist paradigm has provoked a reaction, as revolutions do.”

7. STRUCTURAL *VERSUS* EXPERIMENTALIST APPROACH

The experimentalist school proffers that randomized control trials have a special ability to produce more credible knowledge than other methods. In particular, experimentalists assert that the value of econometric methods can be assessed by how closely they approximate randomized controlled trials, that is, randomization is a gold standard. Furthermore, the experimentalist movement asserts that the structural approach produces results that rely on too many assumptions to be credible.

Not surprisingly, the experimentalist movement met sharp criticism. Opponents to the movement maintain that the relevance of the experimentalist conceptual framework has been oversold by making overbroad claims for its favoured methodologies. As will become apparent in a moment, a powerful criticism of this body

[15]These three assumptions are not sufficient to identify the average effect on the full population. Keane (2010a:5).

of thought is that the experimentalist perspective is potentially misleading. A passionate discussion about the pros and cons of the experimentalist approach can be found in the essays in the symposium “Con Out of Economics,” *Journal of Economic Perspectives*, vol. 24, Spring 2010.

To understand the difference between the structural approach and the experimentalist approach to causality, it is useful to focus attention on the difference between the ‘causes of effects’ and the ‘effects of causes.’ While (traditional) causality analysis in economics focuses on the ‘causes of effects,’ statistical causality focuses on the ‘effects of causes.’ This is not a play on words:

> The emphasis here will be on measuring the effects of causes because this seems to be a place where statistics, which is concerned with measurement, has contributions to make. It is my opinion that an emphasis on the effects of causes rather than causes of effects is, in itself, an important consequence of bringing statistical reasoning to bear on the analysis of causation and directly opposes more traditional analyses of causation.

(Holland 1986:945)

Or, to put it differently, the statistical approach models the effects of causes without modelling the causes of effects.

The debate between the structuralist and experimentalist camps revolves around five dimensions related to each other: (1) definition of causality in econometrics; (2) the necessity of economic models to interpret data; (3) the parameter LATE; (4) internal/external validity; and (5) the correct way to carry out empirical policy analysis. We consider briefly these dimensions and complete the present section by reverting to the issue of taking the “con” out of econometrics.

7.1. Definition of Causality in Econometrics

Heckman (2000, 2008) advocates an approach to causality in econometrics based on structural modelling. He argues that causality in economics is all about constructing models of the causes of effects, not studying the effect of causes without a clear model of how the phenomenon being described is generated. Economic empiricists want something for nothing: they want to model the effects of causes without modelling the causes of effects.

Advocates of the experimentalist approach posit that causality analysis should begin by the measurement of effects of causes rather than defining what the cause of a given effect is. The conflation of the tasks of defining causality and identifying causal parameters from data is the distinguishing feature of the experimentalist school. It should be emphasized that –in the context of the experimentalist approach– the definition of causality is formulated without a clearly articulated model of hypotheticals, and thereby, additional research is necessary to establish what mechanisms may have been at work producing the quantitative effects. For example, in Angrist (1990) it is shown that there is an adverse effect of military service on earnings, but it is not clear *what* causes the 15% reduction in future earnings. More concretely,

> It is not clear from Angrist’s estimates what causes the adverse effect of military experience on earnings. Is the return to military experience lower than that of civilian experience, or does the draft interrupt schooling, or were there negative psychic or physical effects for the subset of draftees who served in Vietnam (e.g. mental illness or disability), or some combination of all three? If the work is to guide future policy, it is important to understand what mechanism was at work.

Keane (2010a:5)

7.2. The Necessity of Economic Models to Interpret Data

According to Keane (2010a), the real distinction between the structural approach and the experimentalist approach lies in the assumptions about economic behaviour. While the structural approach postulates that these assumptions must be laid out explicitly, the guiding principle of the experimentalist approach is: leave the key assumptions of economic behaviour implicit and ‘let the data speak.’ This implies that empirical work can exist independently from –or occur prior to– economic theorizing.

Many social scientists believe that the principle “measurement without theory” is methodologically untenable, that there must be a theoretical framework as a precondition to begin the systematic assembly of empirical regularities. Keane (2010a) argues that it is out of the question to learn interesting aspects about economic relationships without behavioural assumptions even when one can use natural experiments:

> Data cannot determine interesting economic relationships without *apriori* identifying assumptions, regardless of what source of idealized experiments, "natural experiments" or "quasi-experiments" are present in that data. Economic models are always needed to provide a window through which we interpret data, and our interpretation will always be subjective, in the sense that it is contingent on our model.
>
> (Keane 2010a:4)

Deaton (2009) and (Heckman and Urzua 2009) argue against an excessive use of experimental and quasi-experimental work in economics in the last decade. In general, they argue that IV methods do not answer interesting questions and they suggest moving toward causal modelling with explicit behavioural assumptions and away from randomized and natural experiments. Imbens (2010) argues that the empirical work is much more credible as a result of the 'natural experiments revolution' and that "Deaton is both formally wrong and wrong in spirit" when he argues that randomized experiments "do not occupy any special place in some hierarchy of evidence."

7.3. The Parameter LATE

The parameter LATE has been the target of harsh criticism. For example, Deaton (2009), and (Heckman and Urzua 2009) find it hard to make sense of the LATE. The difficulty in interpreting the parameter LATE probably lies in the following warning:

> Finally, it is important to note that (under our assumptions) we cannot identify the specific members of the group of compliers... for whom we can identify the average treatment effect. Thus, the local average treatment effect (i.e., the average causal effect for compliers) is not the average treatment effect for either the entire population or for a subpopulation identifiable from observed values.
>
> Angrist *et al*. (1996:449)

According to the critics of LATE, there are two central problems with this parameter. First, LATE is the average causal effect for a subpopulation that *cannot* be identified because it is impossible to label all individuals units in the population as compliers and non-compliers. In other words, LATE is a parameter defined for an unobservable sub-population. And second, the sign of the IV estimator can be different from that of the true causal effect. (Heckman and Urzua 2009: 2 and 19). Moreover, Leamer (2010:35) makes a derisive analogy between LATE and the achievement of justice: "[LATE] is a little like the lawyer who explained that when he was a young man he lost many cases he should have won but as he grew older he won many that he should have lost, so that on the average justice was done."

The defenders of LATE maintain that a key insight from this parameter is precisely that "although one could not identify the average effect for the overall population, one could still identify the average effect for compliers, or the LATE." Imbens (2010:416).

7.4. Internal/External Validity

A particular aspect of the debate centres on the weight econometricians put on 'internal validity' versus 'external validity.' *Internal validity* refers to the validity of inferences about whether observed covariation captures a causal relationship in a given domain (which includes both a specific population and historical experience). Assuming that internal validity has been established there remains the issue of *external validity*, i.e. whether the causal relationship holds over other domains.

It is generally agree that both internal and external validity are important. However, the structural school emphasizes that the credibility of the estimator of the causal effect of interest emerging from the experimentalist approach is at best confined to a given subpopulation and has no validity for other populations. Put another way, the criticism is that randomized experiments may provide insight in terms of internal validity but they are not robust in terms of external validity. Randomized experiments do poorly in terms of external validity relative to structural models. Leamer (2010) uses the USA banking crisis 2007-08 to highlight the pitfall of extrapolating from natural experiments. The rating agencies transfer findings from one historical experience to a setting to which they no longer applied.

(Angrist and Pischke 2010:23) accept that "extrapolation to new settings is always speculative."

Imbens (2010) argues that the experimentalist approach has significantly improved the credibility of empirical work by emphasizing internal validity, but he does not deny the importance of the ability of the estimators to generalize to other populations and settings (external validity): “In order to be useful in informing policy, a study needs to have internal validity (to have a credible causal interpretation for the population it refers to) as well as external validity (to be relevant for the populations the treatment may be extended to).” Imbens (2010:417).

7.5. The Correct way to Carry Out Policy Analysis

Another important dimension of the debate between the structural and experimentalist camps is about the correct way to carry out empirical policy analysis. The emphasis of the experimentalist movement is on recovering causal parameters, not on evaluating the effects of new policies never previously implemented.

James Heckman (2008) argued that it is possible to bridge the gap between the structural approach and the experimentalist approach by combining the best features of both approaches. The bridge is provided by the seminal paper by Marschak (1953). Heckman (2008) formulated a principle that he labels *Marschak’s maxim* in honour of Marschak’s insight. This maxim connects the statistical treatment effect literature with the literature on econometric causality by showing that the statistical approach focuses on one narrow question while the structural approach attempts to answer many questions.

The treatment effect literature has attractive features because it makes fewer statistical assumptions (in terms of functional form, exclusion, exogeneity, etc.) than the structural approach. Marschak’s maxim points out that for many specific questions of policy analysis it is not necessary to identify fully-specified structural models. All that is required is that policies may be forecast using reduced forms, ignoring the full structure. Heckman (2008) views the statistical treatment effect literature as an implicit implementation of the Marschak’s maxim where the goal of policy analysis is restricted to evaluating policies in place and not concerned with forecasting the effects of new policies. “Marschak established that structural models are only a necessary ingredient for evaluating a new policy.” Heckman (2000:70)

Heckman (2010) is a conciliatory paper. In this paper Heckman does not attack or endorse any specific statistical methodology. He compares the structural approach to empirical policy analysis with the program evaluation approach and makes the implicit economics of LATE explicit.

7.6. Reprise

Reverting to the issue of taking the “con” out of econometrics, experimentalists firmly believe that their approach for causal analysis has been important and beneficial in increasing the credibility of empirical work in economics. According to Keane (2010b), Leamer (2010), and Sims (2010), the experimentalist approach overstates the benefits of their methodology, is plagued by exaggerated claims, and has not taken the con out of econometrics. It is still too early to determine whether the innovations in thought of the last two decades will lead to some sort of scientific consensus in econometrics.

8. WRAPPING UP

If there is a vast and complex theme in economics comparable to an intellectual labyrinth, that topic is causality. This paper provides a tourist guide useful to visit the labyrinth and get out with a clear grand view and a positive attitude towards the various causality tools.

There is no universally accepted approach to causality in economics. There are instead several approaches, with no single unifying methodology in sight or even possible. In this paper we have outlined five methods for hunting causes. All of them have merit.

The narrative approach (Smith) is a rough methodology for causality analysis based on both theoretical description and observation of historical and institutional circumstances. This approach is good as a first approximation to identify plausible candidate causes, but it does not have an explicit definition of the term ‘cause.’ It is probably for this reason that the authors who use the narrative approach tend to be elusive at the time of identifying particular factors as concrete causes. For example, Brunnermeir (2009) concludes his illuminating paper about the liquidity and credit crunch 2007-08 asserting “This paper outlined several amplification mechanisms that help explain the causes of the financial turmoil.” However, a precise identification of the causes of the banking crisis 2007-08 cannot be found in Brunnermeir (2009).

A large proportion of economic analysis relies on comparing static equilibria. The equilibrium system formalizes the economic assumptions and establishes the implicit relations involving exogenous and endogenous variables. The general question is: how the equilibrium value of an endogenous variable will change when there is a change in any of the exogenous variables? In the comparative statics approach (Marshall) to causality, the exogenous variables are the 'causes' and the 'effect' is the change provoked by an exogenous variable on an endogenous variable.

The theoretical approach (Hicks) focuses on singular causality and shows that this kind of causality depends on theoretical economics in a fundamental way. In the Hicksian methodology for causality analysis, counterfactual reasoning is of absolutely fundamental importance. This methodology requires the identification of separable causative factors and the use of economic models connecting these causative factors to test for causality. When there are more than two separable causative factors the causality tests are difficult to implement because the construction of models connecting all the plausible candidate causes may be difficult or impossible.

The goal of the structural approach (Heckman) is to understand the causal mechanisms producing effects using both theory and empirical evidence. Heckman (2000) advocates a methodology to causal inference that draws upon structural modelling of outcomes. The concept of causality employed by the structural school is distinct from the notions of causality based on prediction developed by Granger (1969) and Sims (1972). Causality is a property of a model of hypothetical states. This methodology can be used to tackle problems of both general and singular causation. The advocates of the structural approach maintain that it is methodologically improper to estimate causal effects without a model of economic mechanisms. Complex computational methods are required to implement this approach.

The experimentalist approach (Angrist-Imbens) has succeeded in stimulating thought in both the realm of academic research and policy evaluation. This approach focuses on 'effects' defined by experiments or subrogates for experiments without modelling the causes of effects. More concretely, this perspective emphasizes that we must first detect effects by "letting the data speak," and then, find the underlying economic mechanisms. Economists advocating the experimentalist approach dismiss the structural approach as overly complex and not 'credible.'

Touted as a revolution in empirical economics the experimentalist approach has become the centre of a

Table 6: A Menu of Approaches to Causality in Economics

Approach to causality in economics	Definition of cause	Economic model establishing connexions between causative factors and effects
Narrative approach (Smith): Causality with theoretical description combined with historical and institutional circumstances	Tacit	An appreciative model evolves together with causality analysis, but the research endeavour ends with untested conjectures.
Comparative statics approach (Marshall): Causality with theoretical economics	Explicit definition of cause and effect	Causality is defined within an economic model. Nothing is explicitly said about how to test the definition of causality.
Theoretical approach (Hicks): Causality with theoretical economics	Explicit counterfactual definition of cause and effect	The definition of causality is independent of economic models, but the test of the definition requires economic models in a fundamental way.
Structural approach (Heckman): Causality with theoretical economics	Explicit definition of cause based on hypotheticals such as, for example, policies never previously experienced, products never previously consumed, et cetera.	Hypotheticals are an ingredient in a *priori* economic models. These models specify the mechanisms by which causal variables are externally manipulated. Empirical tests are required to test the existence of a cause-and-effect relationship.
Experimentalist approach (Angrist-Imbens): Causality without theoretical economics	Explicit definition of causal effect based on a statistical model. Definition of what the cause of an effect is neither necessary nor desirable.	*A priori* economic models are not required to establish causality. The existence of causality depends on what the data say.

This summary table highlights the key differentiating characteristics of the different approaches to causality in economics.

growing controversy. In the special symposia "Taking the Con Out of Economics," Keane (2010b), Leamer (2010), and Sims (2010) contend that the experimentalist movement has failed on deliver the promise of its proponents.

Methodological wrangles often generate more heat than light. The controversy between the structuralist school and the experimentalist school is a counterexample. This debate has thrown light on a significant number of causality issues in economics. All in all, the debate between these camps revolves around the simple, yet fundamental question: is it scientifically acceptable to model the effects of causes without modelling the causes of effects? Whatever one's attitude toward the experimentalist approach, it must be surely agreed that it has evoked a flood of valuable research and writing, in both opposition and application.

The papers dealing with the pros and cons of the experimentalist approach reflect the notable advances in our understanding of the complexities of causality analysis in economics. The past two decades has been marked by a number of important developments in this central field of enquiry. Perhaps, in another decade, we will be able to report more progress on causality in economics.

The message conveyed by our survey is that methodological pluralism in the context of causality is unavoidable because a unified approach applicable to all causal problems that may arise in economics does not exist. This message emerges neatly from Table **6** which highlights the key distinguishing features of the different approaches to causality in economics.

A glance at Table **6** shows that causality in economics is not a monolithic concept. Economists are fortunate to have a rich menu of approaches to causality from which to choose. One needs both technical skill and good judgement in order to use a particular methodology and evaluate the corresponding findings.

REFERENCES

Allen, Franklin and Elena Carletti. 2010. "An Overview of the Crisis: Causes, Consequences, and Solutions. International Review of Finance 10: 1-26. http://dx.doi.org/10.1111/j.1468-2443.2009.01103.x

Angrist, Joshua D. 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records." American Economic Review 80: 313-36.

Angrist, Joshua D. Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." Journal of the American Statistical Association 91: 444-55. http://dx.doi.org/10.1080/01621459.1996.10476902

Angrist, Joshua D. and Jorn-steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better research Design is Taking the Con Out of Econometrics." *Journal of Economic* Perspectives 24: 3-30. http://dx.doi.org/10.1257/jep.24.2.3

Archaya, Viral V. and Matthew Richardson. 2009. Restoring Financial Stability. How to Repair a Failed System. New Jersey: John Wiley & Sons.

Brunnermeier, Markus K. 2009. "Deciphering the Liquidity and Credit Crunch 2007-2008." Journal of Economic Perspectives 23: 77-10. http://dx.doi.org/10.1257/jep.23.1.77

Cartwright, Nancy. 2004. Causation: One Word, Many Things. Philosophy of Science 71: 805-820. http://dx.doi.org/10.1086/426771

Deaton, Angus S. 2009. "Instruments of Development: Randomization in Tropics, and the Search for Elusive Keys to Economic Development." *NBER Working Paper Series*. Available at http://www.nber.org/papers/w14690

Dell'Ariccia, Ciovanni, Denis Igan, LucLaeven. 2008. "Credit Booms and Lending Standards: Evidence from the Subprime Lending Market. International Monetary Fund Working Paper, WP/08/106:1-37.

Gennaioli, Nicola, Andrei Shleifer, and Robert W. Vishny. 2011a. "A model of Neglected Risks, Financial Innovation, and Financial Fragility." Journal of Financial Economics104: 452-468. http://dx.doi.org/10.1016/j.jfineco.2011.05.005

Gennaioli, Nicola, Andrei Shleifer, and Robert W. Vishny. 2013. "A Model of Shadow Banking." Journal of Financial Economics (forthcoming).

Granger, Clive W. 1969. "Investigating Causal Relations by Econometric Models and Cross-spectral Methods." Econometrica 37: 424-438. http://dx.doi.org/10.2307/1912791

Haavelmo, Trygve. 1943. "The Statistical Implications of a System of Simultaneous Equations." Econometrica 11: 1-12. http://dx.doi.org/10.2307/1905714

Hayek, Friedrich. 1967. Studies in Philosophy, Politics and Economics. London: Routledge**.**

Heckman, James J. 2000. "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective." Quarterly Journal of Economics 115: 45-97. http://dx.doi.org/10.1162/003355300554674

Heckman, James J. 2008. "Econometric Causality." International Statistical Review 76: 1-27. http://dx.doi.org/10.1111/j.1751-5823.2007.00024.x

Heckman, James J. 2010. "Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy." Journal of Economic Literature 48: 356-98. http://dx.doi.org/10.1257/jel.48.2.356

Heckman, James J. and Sergio Urzua. 2010. "Comparing IV with Structural Models: What Simple IV Can and Cannot Identify." Journal of Econometrics 156:27-37. http://dx.doi.org/10.1016/j.jeconom.2009.09.006

Hicks, John R. 1979. Causality in Economics. Oxford: Basil Blackwell.

Holland, Paul W. 1986. "Statistics and Causal Inference." Journal of the American Statistical Association 81: 945-60. http://dx.doi.org/10.1080/01621459.1986.10478354

Imbens, Guido W. 2010. "Better LATE than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)." Journal of Economic Literature 48: 399-423. http://dx.doi.org/10.1257/jel.48.2.399

Imbens, Guido W. and Joshua D. Angrist, J. 1994. "Identification and Estimation of the Local Average Treatment Effects." Econometrica 62: 467-476. http://dx.doi.org/10.2307/2951620

Keane, Michael P. 2010a. "Structural vs. Atheoretic Approaches to Econometrics." Journal of Econometrics 156: 3-20. http://dx.doi.org/10.1016/j.jeconom.2009.09.003

Keane, Michael P. 2010b. "A Structural Perspective on the Experimentalist School." Journal of Economic Perspectives 24: 47-58. http://dx.doi.org/10.1257/jep.24.2.47

Landreth, Harry and David C. Colander. 2002. History of Economic Thought. Boston: Houghton Mifflin.

Leamer, E. (1983) Let's Take the Con Out of Econometrics. *American Economic Review* 73: 31-43.

Leamer, E. (2010) Tantalus on the Road to Asymptopia. *Journal of Economic Perspectives* 24: 31-46. http://dx.doi.org/10.1257/jep.24.2.31

Locke, John. 1960. An Essay Concerning Human Understanding. Abridged and Edited by A.S. Pringle-Patisson. Oxford: Clarendon Press. First published in 1690.

Maddaloni, Angela and Jose-Luis Peydro. 2011. "Bank Risk-taking, Securitization, Supervision, and Low Interest Rates: Evidence from the Euro Area and US Lending Standards." Review of Financial Studies 24: 2120-165. http://dx.doi.org/10.1093/rfs/hhr015

Marschak, Jacob. 1953. "Economic Measurements for Policy and Prediction." In Studies in Econometric Methods, W. Hood and T. Koopmans, eds., 1-26. New York: Wiley.

Marshall, Alfred. 1966. Principles of Economics. London: Macmillan. First published in 1890.

Pol, Eduardo. 2012. The Preponderant Causes of the USA Banking Crisis 2007-08. Journal of Socio-Economics 41: 519-528. http://dx.doi.org/10.1016/j.socec.2012.04.019

Samuelson, Paul A. 1965. Foundations of Economic Analysis. New York: Atheneum. Originally published by Harvard University Press in 1947.

Schumpeter, Joseph A. 1954. History of Economic Analysis. London: George Allen & Unwin.

Sims, Christopher A. 1972. "Money, Income and Causality." American Economic Review 62: 540-52.

Sims, Christopher. 2010. "But Economics is Not an Experimental Science." Journal of Economic Perspectives 24: 59-68. http://dx.doi.org/10.1257/jep.24.2.59

Smith, Adam. 2010. An Inquiry into the Nature and Causes of the Wealth of Nations. Chicago: University of Chicago Press. First published in 1776.

Received on 24-06-2013 Accepted on 09-09-2013 Published on 20-09-2013

DOI: http://dx.doi.org/10.6000/1929-7092.2013.02.25