

Computing, the Bootstrap, and Economics

by

Russell Davidson

Department of Economics and CIREQ
McGill University
Montréal, Québec, Canada
H3A 2T7

AMSE–GREQAM
Centre de la Vieille Charité
2 rue de la Charité
13236 Marseille cedex 02, France

email: russell.davidson@mcgill.ca

Abstract

A major contention in this paper is that scientific models can be viewed as virtual realities, implemented, or rendered, by mathematical equations or by computer simulations. Their purpose is to help us understand the external reality that they model. In economics, particularly in econometrics, models make use of random elements, so as to provide quantitatively for phenomena that we cannot or do not wish to model explicitly. By varying the realisations of the random elements in a simulation, it is possible to study counterfactual outcomes, which are necessary for any discussion of causality.

The bootstrap is virtual reality within an outer reality. The principle of the bootstrap is that, if its virtual reality mimics as closely as possible the reality that contains it, it can be used to study aspects of that outer reality. The idea of bootstrap iteration is explored, and a discrete model discussed which allows investigators to perform iteration to any desired level.

Keywords: Bootstrap, bootstrap iteration, causality, counterfactuality, identification

JEL codes: C10, C12, C15

The first version of this paper was prepared for the Presidential Address given at the 49th Annual Conference of the Canadian Economics Association, held in May 2015 at Ryerson University, Toronto. The work on which the address was based was supported by the Canada Research Chair program (Chair in Economics, McGill University), and the Fonds de recherche du Québec – Société et culture (FRQSC).

May 2015

1 Introduction

The use of models, explicit or implicit, is universal in scientific disciplines. It is by studying models that science helps us to understand the world, or, if we are more ambitious, the universe. Scientific theories are embodied in models, and the structure of theoretical models is such that they yield explanations of that aspect of the world or universe that they model. In this paper, I develop the idea that models, of all sorts, can be viewed as virtual realities that mimic, as best they can, those properties of the external reality of the world that are the domain of the theory.

Explanation usually takes the form of laying out the causal relations among observed events. I discuss the nature of causality, distinguishing necessary and sufficient causality, and pointing out that, in econometrics, it is only sufficient causality that is a useful concept. It is not possible to give meaning to the idea that something causes something else without being able to consider what might have happened if things were different, that is, by considering counterfactual scenarios. The role of a model in discussions of causality is to define precisely what these counterfactual scenarios are in any particular context, and to delimit the extent to which they may diverge from observed reality.

My major research preoccupation, now and for the past nearly twenty years, is the bootstrap, and its applications to econometrics. The bootstrap can be viewed as virtual reality. As with models more generally, its purpose is to mimic an enclosing reality, which may be either another virtual reality or external reality itself. For the purposes of statistics, the virtual reality of the bootstrap can be thought of as an estimate of the enclosing reality, and used to obtain estimates of interesting properties of it.

Since virtual realities can exist within other virtual realities, one can conceive of bootstrapping the bootstrap, that is, mimicking what the bootstrap does in mimicking its containing reality. Although it is a known fact that thinking about this sort of thing can make one's head spin, it is quite possible to define formally what is meant by bootstrap iteration, as I do in this paper. I believe that this concept is potentially enormously useful in econometrics, and in statistics more generally.

In the [next section](#), I develop the idea of models as virtual reality, with particular discussion of economic models. [Section 3](#) explores how models interact with notions of causality, and discusses how counterfactual situations can be defined and used in econometric analyses. [Section 4](#) provides a short discussion of parallel computing, and points out analogies with the parallel universes that are postulated in some versions of modern quantum theory. Then, [section 5](#), I discuss the bootstrap, beginning with a formal statement of the bootstrap principle. The so-called Golden Rules of bootstrapping, for the purposes of statistical inference, are then enunciated, and this leads to the definition of bootstrap iteration. Some recent work of mine is outlined in [section 6](#). This work exploits the fact, mentioned in [section 2](#), that digital computers treat everything as discrete, by making use of a discrete, and finite, setup in which it is possible to study bootstrap iteration without running into the insuperable computational difficulties encountered in previous work on the topic.

2 Scientific Models as Virtual Reality

The capacities of modern computers have made **virtual reality** something that we can experience in new ways, enabled by new technology. We hear of flight simulators, and the younger generation seems to spend a lot of time in the virtual reality of computer games. But people have been inventing virtual realities for as long as there have been scientists.

In most scientific disciplines, **models** play an essential role. Scientific models are often mathematical, but they need not be so. A mathematical model does, however, make clear the sense in which a model is a sort of virtual reality. Mathematics is nothing if not an abstract discipline; so much so that some have claimed that mathematics, pure mathematics anyway, has no meaning or substantive content. What is true, though, is that we can *give* mathematical constructions interpretations that imply much substantive content. This is just as true in economics as it is in physics.

Why is this? The aim of science is not only to acquire knowledge of the world, although the etymology of the word “science” – from the Latin *scire*, to know – implies knowledge only, but to *understand* the world. Science provides such understanding by *explaining* our experiences. Science advances when it provides better explanations. What constitutes an explanation? Well, a theory. That’s just terminology, and so the question has merely been reformulated as: What constitutes a theory?

A theory is embodied in a model, and the model constitutes a virtual reality. But not all models count as theories, as I will explain later. However, we can conclude at present that virtual realities can give us understanding of the world, through the explanations that they may embody. Of course, some models mimic external reality, as we observe it, better than others, and so they provide better explanations. Scientific controversies are about which explanations are better.

What is there about a theory that provides an explanation? Think, if you will, of Keynes’s celebrated *General Theory*. The theory implies a model of the macroeconomy, the macroeconomy in virtual reality, and within this model, there are relationships among the macroeconomic variables – relations that can be expressed mathematically, and are justified by the arguments that Keynes makes, showing that these relations mimic what we observe of the macroeconomy. When we observe that interest rates fall, the Keynesian model explains the economic mechanisms that led to this fall.

Not every economist is convinced by Keynesian explanations! The opponents of Keynes’s model, or his view of the world, if we are to pay any attention to them, must construct *rival* virtual realities, and argue that the relations that these entail better describe external reality than the Keynesian ones.

The fact that virtual reality is possible is an important fact about the fabric of reality. It is the basis not only of computation, but of human imagination and external experience, science and mathematics, art and fiction.

David Deutsch, *The Fabric of Reality*

The heart of a virtual-reality generator is its computer.

ibid.

I don’t think Keynes ever had anything to do with a digital computer, although it seems likely that, on account of his friendship with Alan Turing, he was acquainted with analogue computers. If I am wrong, since Keynes died in 1946, it would have had to be a very early sort of computer, nothing at all like what we mean by the word nowadays. As an aside, while the Manhattan project was being pursued at Los Alamos during World War II, a “computer” was a person, not a machine, as indeed was a “typewriter” * As another aside, when Claude Shannon, the founder of modern information theory, was asked whether machines could think, his answer was yes, because, he said, “We are machines and we think, don’t we?” (Quoted in Moses (2005).)

Deutsch is making a different point about virtual reality, namely the physical possibility of rendering it, and so I won’t go on with his idea of it. Mathematics can constitute virtual reality as well as computers can. But, as our computers have become more powerful, so our models depend more and more on computer implementations. There are deep philosophical questions concerning whether we, as humans, can really understand something produced by computation rather than logical and mathematical reasoning, especially if one looks forward to what quantum computers may one day be able to do, and can do, in principle, according to the physical theories we have today.

But one thing we can easily say about models implemented on the computer is that everything must be *digital*, and so also *discrete*. That this is no real problem for practical things is evident from the extent we use digital sound recording, digital cameras and so on, and especially digital typography, one of the greatest boons for anyone writing books or papers, like this one. I call it a “paper”, although it need never be printed on paper at all. What exists in the real world is an implementation in virtual reality of a hardcopy paper. Similarly, we all speak of the “slides” for a presentation, although they are just as virtual.

There is in fact no consensus at the present time among theoretical physicists whether space-time is continuous, as I think it’s safe to say is assumed by most current physical models, or rather discrete - quantised, like everything else in quantum mechanics. I take from this that there is no harm in letting our virtual realities be discrete – whether or not they are digital – and, as I hope to show later on, there may be considerable benefits.

Models in Economics

Just as in physics, many economic models assume that space and time are continuous, although in econometrics, for obvious reasons, time, at least, is usually treated as a discrete variable. Unlike many physical models however, econometric models invariably incorporate **random elements**.

There is a considerable philosophical difficulty that arises when we wish to impart any substantive meaning to the mathematics of probability and random variables, *if* we also wish to adhere to a deterministic world view. This is so because, in conventional interpretations of probability, events that have occurred, are occurring, or will (certainly) occur have a probability of one, and events that never occur have a probability of zero. If, as follows

* I am indebted to Samuel Hollander for the information about the typewriter.

from a deterministic view, any event at all either does occur or does not, the mathematics of probability becomes trivial.

But we use probabilistic notions all the time, and not trivial ones either. What in the external world is it that we want to mimic by using randomness? We can all agree that many things in our lives appear to us to be random, but there are many philosophers who, while granting this *appearance* of randomness, still think that, at some fundamental level, the world is deterministic. This leads to a somewhat deeper question. *Why* are there such seemingly random events? To that question, I think the best answer is that we model such events as realisations of random variables because we do not and cannot know everything. Even more to the point, we cannot *explain* everything. Whenever we cannot, or do not wish to, explain events which have an impact on the main objects of interest in our models, we model them as realisations of random variables. That at least is my view of what we do as econometricians, although I suspect that many other econometricians would either disagree or else express things quite differently.

It is not enough to wave our hands and say that we use random elements in our models. We need more than that if we want to consider a model as a virtual reality, probably one to be rendered by the computer. I think the best way to formulate this is to define a model as a set of **data-generating processes**, or DGPs, each of which constitutes a *unique* virtual reality. I like to go further, and specify that a DGP is something that can be simulated on the computer, or that provides a **unique recipe** for simulation. In this way, I am tying the virtual realities of economic models more closely to the computer, just as Deutsch would have it.

What has been missing and now must be introduced is the **distribution** of the random elements. Computers have **random-number generators**, or RNGs, and what they generate are sequences of independent realisations from the uniform distribution on the interval $[0, 1]$. (Or nearly so – I’m not going to talk about that; see Knuth (1997), Volume 2, Chapter 3, for an authoritative discussion.) These random numbers can be transformed into realisations from other distributions we may want to specify; see Devroye (1986). Thus we can indeed incorporate any desired form of randomness that we can specify into the DGPs of a model.

Another feature of economic models is that they involve **parameters**. A model normally does not specify the numerical values of these parameters; indeed a purely parametric model is a set rather than a singleton because the DGPs that it contains may differ in the values of their parameters. Models that are not purely parametric allow the DGPs that they contain to differ also in the **stochastic specification**, that is, the distribution of the random elements.

3 Causal Explanations

Suppose that we have a model of an economic phenomenon that we wish to study. Suppose, too, that it seems to correspond well to what we observe in external reality. Does that mean that we have *explanations*, complete or partial, of what we are studying? Not necessarily. Some models are purely descriptive. A statistical model, for instance, might specify the probabilistic properties of a set of variables, and nothing more. But that may be enough for us to do forecasting, even if our forecasts are not based on any profound understanding. Half a century ago, most physicists thought of quantum mechanics that way, as a mathematical recipe that could be used to predict experimental results. The “interpretations” of quantum mechanics that were then current were very counter-intuitive, and today physicists still argue not only about what interpretation is to be preferred, but about whether *any* interpretation meaningful to the human brain is possible.

However, the positivist approach that has held sway in physics for so long is finally giving way to a thirst for explanations. Perhaps theoretical physics does give better agreement with experimental data than any other discipline, but, some physicists are now asking, does it constitute a true *theory*? A theory must explain, by proposing a mechanism, or in other words a *causal* chain.

What is a cause?

This subsection draws heavily on the insights in Chapter 3 of Dennett (2003). Consider two events, A and B . An intuitive definition of the proposition that A causes B is:

- (i) A and B are real, or true;
- (ii) If A is not real or true, then neither is B ; and
- (iii) A precedes B in time.

This definition raises a number of issues. What do we mean by an “event”? There are several admissible answers: an action, a fact of nature, among others. A fact is true or not, and action is performed (it is real) or not. Our tentative definition is general enough to allow for various different possibilities.

In order to steer clear of some trivial cases, we want to suppose that the events A and B are logically *independent*. Thus we don’t want to say that the conclusion of a mathematical theorem is *caused* by the premisses of the theorem.

It is important to distinguish between causal **necessity** and causal **sufficiency**. Necessity means that:

not A (written as $\neg A$) implies $\neg B$.

In words, without A , there can be no B . Logically, the condition is equivalent to the condition that B implies A ; that is, A is a *necessary* condition for B . This is our condition (ii).

Sufficiency means that:

A implies B , or $\neg B$ implies $\neg A$.

In words, every time that A holds, unavoidably B holds as well; that is, A is a *sufficient* condition for B . Sufficiency is logically quite distinct from necessity. Necessity leaves open

the possibility that A holds without B . Sufficiency leaves open the possibility that B holds without A .

It is easy enough to see how we might study these two types of causality when the events A and B are repeated, as with coin tosses or the roulette wheel, where we don't *a priori* expect to find any causality at all, or when an experiment is undertaken in which both A and $\neg A$ can occur, and possibly also B and $\neg B$.

But if A and B are unique, not repeated, events, what sense can we make of the assertion that A caused B ? I suppose here that condition (i) is satisfied, so that A and B both occurred. In order to make any sense of the statement about causality, we have to admit to our discussion *imaginary worlds* or even *universes*. We call such worlds or universes **counterfactual**. Without considering them, it is impossible to know what *might* have occurred if A did not, or if B did not occur.

But this remark gives rise to as many problems as answers. What is the set of universes that these counterfactual universes inhabit? How can we delimit this set? Let's denote the set by \mathcal{X} . Then we have a number of reasonable choices:

- (a) \mathcal{X} is the set of *logically* possible universes, that is, all universes that are not logically self-contradictory;
- (b) \mathcal{X} is the set of universes compatible with the laws of physics, as we know them;
- (c) \mathcal{X} is the set of logically and physically admissible universes that are sufficiently *similar* or *close* to the real world.

The last choice is no doubt the best, but, in order to implement it, what topology can we use to define a *neighbourhood* of the real world?

Causality in econometrics

In all scientific disciplines, progress comes from the result of an experiment, or an observation, that leads us to *reject* a hypothesis. It is therefore important to be able to demonstrate *non-causality*, that is, the absence of any relation of cause and effect between two events, or types of events.

In econometrics, most of the time we deal with continuous variables, which means that the event B (the effect) must be replaced by a quantitative measure of one or more variables. Similarly for the cause, A . This makes it simpler to define what we mean by **non-causality**. We say that a variable X does not cause another variable Y if the earlier values of X have no influence on the later values of Y . This sort of causality is called **Granger causality**, having been introduced by Granger (1969). A similar related approach is due to Sims (1972).

The random elements, shocks, disturbances, in an econometric model allow us to introduce the required neighbourhood of circumstances (worlds, universes) that surround the observed trajectories of X and Y . We no longer need to invent imaginary trajectories that *might* have existed in the real world. It is enough to vary the realisations of the random elements in order to create, within a virtual reality, all the relevant circumstances needed to reject causal sufficiency.

Why not causal necessity? We said that A is a necessary cause of B if B implies A . In propositional logic,

$$B \Rightarrow A \quad \Leftrightarrow \quad A \vee \neg B,$$

of which the negation is

$$\neg(A \vee \neg B) \quad \Leftrightarrow \quad \neg A \wedge B.$$

Causal necessity is rejected if B occurs in the absence of A . If we translate this into econometric terminology, it would mean that the variable Y (associated with the event B) varies without any variation of X (associated with A). But normally Y has several determinants, which implies that Y can perfectly well vary with no change in the value of X . This is enough to reject causal necessity.

This seems too easy. In economic theory, we often say *ceteris paribus*, other things being equal, *toutes choses gales par ailleurs*. In virtual reality, we can arrange things so that all the other variables, and also the realisations of the random elements, do not vary. This is a way to restrict the set of circumstances we consider for the purpose of establishing the existence or the non-existence of a causal link. However, if nothing but the two variables X and Y can move, then there is a deterministic functional relation between the two variables. In that case, we would never be able to reject causal necessity. Conclusion: causal necessity is not a useful concept in econometrics.

Causal sufficiency is the proposition that A implies B . Propositional logic tells us that the negation of this proposition is $\neg B \wedge A$. Translating this, we see that this means that X varies without producing the effect of a variation of Y . Once again, then, we can reject causal sufficiency if, *ceteris paribus*, Y takes on the same value whatever the value of X . This would mean that the deterministic relation between the two variables introduced by the *ceteris paribus* assumption admits one and only one value for Y .

This has finally led us to a testable proposition. The null hypothesis specifies a no doubt complicated relation among the full set of variables considered relevant for the model, along with a set of random elements. This specification has the property that, for any configuration of the variables other than X and Y , and for any realisation of the random elements, the value of Y is uniquely determined, whatever the value of X . The alternative hypothesis allows the value of X to have an influence on that of Y .

In econometrics, causality is most frequently studied in the context of VAR models, where "VAR" stands for "vector autoregression". In a model of this type, the current values of a set of endogenous variables are determined by the lagged values of the same set of variables and by the realisations of a set of random elements. In the current state of the art, one almost always postulates a linear relation among the variables. Here is an illustrative example that makes use of the household consumption function.

$$\begin{aligned} c_t &= \alpha_1 + \beta_{11}c_{t-1} + \beta_{12}y_{t-1} + u_{t1}, \\ y_t &= \alpha_2 + \beta_{21}c_{t-1} + \beta_{22}y_{t-1} + u_{t2}. \end{aligned}$$

The two variables are c , household consumption, or, more likely, the logarithm of household consumption, and y , disposable income of households, or its log. This is a macroeconomic

relation. The variables c and y are aggregate variables, and they represent flows. The time index t refers to a period of some given duration, typically a year, a quarter, or a month. The random elements u_{t1} and u_{t2} are realisations of a bivariate distribution with zero expectation. We may wish to suppose that the pair (u_{t1}, u_{t2}) is independent of all other pairs (u_{s1}, u_{s2}) , with $s \neq t$. The quantities denoted by $\alpha_i, \beta_{ij}, i, j = 1, 2$, are the model parameters, which are treated as deterministic constants.

If the parameter values are known, along with the bivariate distribution of the random elements, we can undertake a stochastic simulation – virtual reality – if we have the initial condition (c_1, y_1) . This amounts to specifying a DGP, a unique recipe for simulation, and that is enough for us to be able to study all the statistical properties of the variables that the DGP can generate.

According to elementary macroeconomic theory, disposable income causes consumption. Normally, except for some sophisticated models, we don't imagine that consumption causes income. This hypothesis corresponds formally to the hypothesis that $\beta_{21} = 0$, and we have several ways in which we could test this hypothesis. We would refer to this hypothesis as that of Granger non-causality.

Although the concept of Granger causality and its implementation by means of VAR models allows us to formulate hypotheses of non-causality, and possibly to reject them, it does not, or not always, satisfy our desire for understanding and explaining economic mechanisms by means of causal chains. For that, it is necessary to base the models used for estimation and inference on **economic theory**. Granger's approach has come in for very little criticism on this ground however, because Granger always maintained that the goal of his methodology is to help economic forecasting. To the extent that this goal is attained, the methodology must be justifiable on some level.

But econometricians have always had a preference for **structural models**, in which the relation between the formal model and the underlying economic theory is clear. This preference led to the seminal work of the Cowles Commission. Historically, this work led to models that were often thoroughly incompatible with the data, and the realisation of this led to a great many advances in econometric theory. In particular, since statistical models with little or no explanatory power often gave much better fits than models supposedly based on economic theory, econometricians became more concerned with testing the statistical reliability of their models, and less concerned with the relation of these models to economic theory.

But it is probably fair to say that structural models returned as the main focus of interest of many econometricians with the advent of the twenty-first century. Problems associated with the identification of such models and of their parameters assumed considerable importance, and stimulated much work intended to elucidate the nature of these problems, and ways of solving them. However, whether a model is structural or not makes little difference to how we can perform inference about causal sufficiency. The essential element is to be able to set up counterfactual situations by means of the model.

Counterfactual econometrics

In biostatistics and medicine, emphasis is often put on **randomised trials**, in which two groups of subjects are treated differently. One usually speaks of a control group, the members of which are not **treated**, and a treatment group, for which a particular treatment is prescribed. After some definite period, the members of both groups are examined for some particular property, which is thought of as the effect of being treated or not. Clearly, the idea is to be able to see whether the treatment causes the effect, and, perhaps, to reject the hypothesis that it does so. Here, if one can select the members of the two groups quite randomly, in a way totally unrelated to the treatment or the effect, then the distribution of effects within each group serves as the counterfactual distribution for the other.

Even in medicine, a truly randomised trial can be difficult to achieve, for both practical and ethical reasons. In econometrics, it is even more difficult, although not completely impossible. However, “natural experiments” can arise for which an econometrician may be able to identify two groups that are “treated” differently, perhaps by being subject to some government program, and to measure some effect, such as wages, that might be affected by the treatment. This can be fruitful, but, naturally enough, it requires the use of sophisticated statistical and econometric techniques.

In a polemical essay, Heckman (2001) maintains that econometrics has suffered as a result of too great an application of the methodology of mathematical statistics. He says that

Statistics is strong in producing sampling theorems and in devising ways to describe data. But the field is not rooted in science, or in formal causal models of phenomena, and models of behavior of the sort that are central to economics are not a part of that field and are alien to most statisticians.

This is a strong statement of what I have called the preference of econometricians for structural models.

Whether or not they go along completely with Heckman on this point (and I believe that I do), econometricians, even sometimes in company with statisticians, have developed techniques for getting indirectly at information about counterfactual worlds. Of these, the method called **difference in differences** is probably the best known and the most used; an early example in the econometrics literature is Ashenfelter and Card (1985). Since counterfactual worlds are never realised, *some* assumptions must *always* be made in order to invent a virtual reality in which they can be rendered. Often, an assumption is made implying constancy in time of some relations; other times the assumption might be, as with randomised trials, that two or more groups are homogeneous. To say that we always need some assumption(s) is to say that there must always be a model, rich enough in its explanatory power to render credible counterfactual, and so virtual, realities.

One development of this sort is found in Athey and Imbens (2006). They extend the idea behind the difference-in-differences method to a method called change-in-changes. The name does not make clear what I regard as the chief virtue of their method, namely that, instead of limiting attention to *average* treatment effects, it considers the entire distribution of these effects. Average effects may be enough for biostatisticians; not for econometricians.

4 Parallel Computing; Parallel Universes

We will all have heard that the future of computing lies in parallel computing, now that Moore’s Law is reaching the end of its useful life, as computer chips come up against the quantum nature of the physical world. With parallel computing, a computer program makes use of several CPUs, or cores, simultaneously. This is what we call Concurrent Programming, which lets computers perform concurrent processing.

Is there a counterpart in the external world to parallel or concurrent processing in the virtual reality of computers? That depends on which physicist you talk to! Some, like David Deutsch (1997), believe that the quantum theory implies the existence of what he calls “parallel universes”, which together make up the “multiverse”. Others find it unnecessary to suppose anything of the sort, but then they are obliged to abandon a deterministic world view, and assume that there really is some irreducible randomness in nature. If I understand him correctly, Deutsch maintains a deterministic world view, but only at the level of the whole multiverse, not in any universe that we might live in and observe.

Recall that (mathematical) probability originated in human thought as an idealisation of the notion of the **frequency** with which a repeated action or experiment yields the different results of which it is capable. Another aside: Jakob (or Jacques, or James) Bernoulli lived from 1655 to 1705; Thomas Bayes from 1701 to 1761, and anyway he wasn’t a Bayesian!

Deutsch’s view is that, in quantum mechanics at least, the probability of the different results that an observation may yield is indeed a frequency, or, more precisely, the **proportion** of universes in the multiverse in which that result is found. And this corresponds exactly with what happens in a computer simulation that makes use of random numbers. Whether the program that implements the simulation is run sequentially or concurrently, the same code is run many times with different random inputs, and, at the end of the simulation, we estimate the theoretical probabilities that are defined in our virtual reality by the frequencies, or proportions, of the repetitions that gave the various outcomes.

If we adopt my point of view about why there are random elements in economic models, then we see why it is of interest to perform simulations with random numbers. Yes, the goal of our models is to understand through explanation, and calling things random explains nothing, but, even so, models with random elements can help us understand economic phenomena by giving *partial* explanations of economic mechanisms. Another conclusion from this reasoning is that some virtual realities may be quite imperfect renderings of the real world. Maybe flight simulators are pretty good these days, but they weren’t always, and video games don’t even try to mimic the real world.

5 The Bootstrap

A virtual reality may be contained in, or contain, other virtual realities. The bootstrap is an example of this. Within the context of a model (the outside virtual reality), for which we do not know the specific DGP that may have generated the data we wish to analyse, we create another virtual reality, often called the “bootstrap world”, in order to

test hypotheses or construct confidence sets. The bootstrap has other uses, but I won’t discuss them here; see for instance Davison and Hinkley (1997).

The bootstrap makes no use of asymptotic considerations, but, as with much econometric theory, current bootstrap theory relies heavily on asymptotics. This is an undesirable state of affairs, if for no other reason than that the choice of an asymptotic construction is inevitably somewhat arbitrary.

Asymptotic arguments rely on sequences of random variables that converge, in probability or in distribution, to some desirable limit. Since bootstrap inference is exact only in rare instances, any justification of it must also rely on some sequence of random variables with a desirable limit.

The approach I outline here involves the convergence or otherwise of a sequence of bootstrap P values obtained by **iterating** the bootstrap. The first iterate, called the double bootstrap, is in many cases feasible; when it is, it is presumed that it can provide more reliable statistical inference than either conventional asymptotics or the single uniterated bootstrap.

It is certainly tempting to suppose that the only barrier to still more reliable inference *via* bootstrap iteration is computational infeasibility. Here I discuss some preliminary work that shows that bootstrap iteration can indeed improve reliability of inference. In order to do so, a procedure of discretisation is used, by means of which the model under test is described by a finite three-dimensional array of probabilities.

Definitions and notations

Recall that what I mean by a model is a collection of DGPs. Let \mathbb{M} denote a model. Then \mathbb{M} may also represent a **hypothesis**, namely that the true DGP, μ say, belongs to \mathbb{M} . I denote by \mathbb{M}_0 the set of DGPs that represent a null hypothesis we wish to test, using a test statistic τ . It is conventional to suppose that τ is defined as a random variable on some suitable probability space, on which a different probability measure is defined for each different DGP. Rather than using this approach, I define a probability space (Ω, \mathcal{F}, P) , with just one probability measure, P . Then the test statistic τ is treated as a stochastic process the index set of which is the set \mathbb{M} . We have

$$\tau : \mathbb{M} \times \Omega \rightarrow \mathbb{R}. \quad (1)$$

Since we are in virtual reality, the probability space can be taken to be that of a random number generator. A realisation of the test statistic is therefore written as $\tau(\mu, \omega)$, for some $\mu \in \mathbb{M}$ and $\omega \in \Omega$.

This approach, rather than the conventional one, corresponds precisely to what we do in a simulation experiment. The random elements of the simulation all come, directly or indirectly, from the random number generator, and the DGP takes the form of a part of the computer program, in which data are generated by *deterministic* transformations of the random numbers and the parameters that correspond to that DGP. It is sometimes possible, and sometimes essential, to use the same random numbers in combination with

different specifications of the DGP, and this is captured quite precisely by the notation defined in (1).

For notational convenience, we suppose that the range of τ is the $[0, 1]$ interval rather than the whole real line, and that the statistic takes the form of an approximate P value, which thus leads to rejection when the statistic is too small. Let $R_0 : [0, 1] \rightarrow [0, 1]$ be the cumulative distribution function (CDF) of τ under any DGP $\mu \in \mathbb{M}$:

$$R_0(\alpha, \mu) = P\{\omega \in \Omega \mid \tau(\mu, \omega) \leq \alpha\}.$$

For $\mu \in \mathbb{M}_0$, the random variable $R_0(\tau(\mu, \omega), \mu)$ follows the uniform distribution $U(0,1)$ if its distribution is continuous on $[0, 1]$. This property is what allows the Monte Carlo tests of Dwass (1957) to give exact inference when τ is pivotal with respect to \mathbb{M}_0 , by which is meant that the random variable $\tau(\mu, \omega)$ has the *same* distribution for all $\mu \in \mathbb{M}_0$. See also Dufour and Khalaf (2001) for more information on Monte Carlo tests.

Bootstrap Principle

The principle of the bootstrap is that, when we want to use some function or functional of an unknown DGP μ , we use an estimate $\hat{\mu}$ in place of μ . The DGP μ could be either external reality or an outer virtual reality. The estimate $\hat{\mu}$, which is called the **bootstrap DGP**, is a virtual reality embedded in whatever reality contains μ .

Suppose that we have a statistic computed from a data set that may or may not have been generated by a DGP $\mu \in \mathbb{M}_0$. Denote this statistic by t . We define the DGP-valued process

$$\beta : \mathbb{M} \times \Omega \rightarrow \mathbb{M}_0.$$

The bootstrap DGP that serves as the estimate of μ is $b = \beta(\mu, \omega)$, where ω is the *same* realisation as in $t = \tau(\mu, \omega)$, since both t and b are computed using the same data set. Then, following the bootstrap principle, we define the bootstrap P value, expected to follow the $U(0,1)$ distribution *approximately*, to be $R_0(t, b) = R_0(\tau(\mu, \omega), \beta(\mu, \omega))$.

The bootstrap is a very general statistical technique. The properties of the true unknown DGP that one wants to study are estimated as the corresponding properties of the bootstrap DGP. In practice, although not in principle, these properties have to be studied by means of a simulation experiment.

The Golden Rules of Bootstrapping

My “Golden Rules” for bootstrapping, extensions and reformulations of “guidelines” for bootstrap hypothesis testing found in Hall and Wilson (1991), are these:

Golden Rule 1:

The bootstrap DGP b must belong to the model \mathbb{M}_0 that represents the null hypothesis.

This is because what we want for a test is an estimate of the distribution of the test statistic under the null hypothesis. The power of a test is related to how different the distribution

of the statistic is under the null and alternative hypotheses. One expects serious loss of power if the realised statistic is compared to distribution under the alternative.

Golden Rule 2:

Unless the test statistic is pivotal for the null model \mathbb{M}_0 , the bootstrap DGP should be as good an estimate of the true DGP as possible, under the assumption that the true DGP belongs to \mathbb{M}_0 .

Although some econometricians still have doubts about the last part of the statement of this rule, fearing that it may lead to loss of power, it makes sense to exploit the fact that imposing the restrictions of a true null leads to greater efficiency of estimation. Further, under local alternatives, it is shown in Davidson and MacKinnon (2006) that there is no power loss by estimating the bootstrap DGP imposing the restrictions of the null.

Inference

If τ is not pivotal, exact inference is no longer possible, because the true DGP μ is unknown. The bootstrap principle tells us to replace it by an estimate, namely the bootstrap DGP b , and obtain the **bootstrap P value** $R_0(t, b)$. In order to estimate it by simulation, we make the definition

$$\hat{R}_0(\alpha, \mu) = \frac{1}{B} \sum_{j=1}^B \mathbf{I}(\tau(\mu, \omega_j^*) < \alpha),$$

where the ω_j^* are independent. Each ω_j^* can be thought of as a set of those random numbers needed to generate a realisation of the statistic. Then, as $B \rightarrow \infty$, $\hat{R}_0(\alpha, \mu)$ tends almost surely to $R_0(\alpha, \mu)$. Accordingly, we estimate the bootstrap P value by $\hat{R}_0(t, b)$, for some suitable choice of B , the number of bootstrap repetitions.

Bootstrap Iteration

The bootstrap P value $R_0(t, b)$ is a realisation of a random variable $p_1(\mu, \omega)$, where the new function $p_1 : \mathbb{M} \times \Omega \rightarrow [0, 1]$ is defined as follows:

$$p_1(\mu, \omega) \equiv R_0(\tau(\mu, \omega), \beta(\mu, \omega)).$$

We denote the CDF of $p_1(\mu, \omega)$ by $R_1(\cdot, \mu)$. The random variable $R_1(p_1(\mu, \omega), \mu)$ is, by construction, distributed as $U(0,1)$. But this fact is not enough to allow exact inference, because the actual μ that generates the data is unknown outside the context of a simulation experiment.

However, the bootstrap principle can again be applied, and the unknown μ replaced by the estimate b . This leads to the double bootstrap, of which the P value, for realisations t and b , can be written as

$$R_1(R_0(t, b), b),$$

where R_1 can be estimated just like R_0 , but by means of a much costlier simulation experiment.

Under the assumption that $\mu \in \mathbb{M}_0$, the double bootstrap P value is the probability mass in the distribution of the single bootstrap statistic to the left of $R_0(t, b)$. Expressed as a random variable, it is

$$p_2(\mu, \omega) \equiv R_1(R_0(\tau(\mu, \omega), \beta(\mu, \omega)), \beta(\mu, \omega))$$

If we write the right-hand side above as $R_1(p_1(\mu, \omega), \beta(\mu, \omega))$, we see that the double bootstrap effectively bootstraps the single bootstrap P value.

From that observation, it is clear that we can define iterated bootstraps as follows. For $r = 0, 1, 2, \dots$, we define

$$\begin{aligned} R_r(\alpha, \mu) &= P\{\omega \in \Omega \mid p_r(\mu, \omega) \leq \alpha\}, \\ p_{r+1}(\mu, \omega) &= R_r(p_r(\mu, \omega), \beta(\mu, \omega)), \end{aligned}$$

where we initialise the recurrence by the definition $p_0(\mu, \omega) = \tau(\mu, \omega)$. Thus $p_{r+1}(\mu, \omega)$ is the bootstrap P value obtained by bootstrapping the r^{th} order P value $p_r(\mu, \omega)$. It estimates the probability mass in the distribution of the r^{th} order P value to the left of its realisation.

In order for bootstrap iteration to be useful, it is necessary for the sequence $\{p_r(\mu, \omega)\}$ of iterated bootstrap P values to converge as $r \rightarrow \infty$, and the limit of the sequence to be distributed as $U(0,1)$ for all $\mu \in \mathbb{M}_0$. Note that it is possible to have convergence to a distribution quite different from $U(0,1)$. Indeed, this is necessary for test power when $\mu \notin \mathbb{M}_0$.

With a resampling bootstrap, since with high probability each resample does not contain some of the observations of the original sample, repeated iteration leads to iterated bootstrap DGPs that have only one observation out of the original sample, repeated as many times as the original sample has observations. That is one reason for which our notation has so far implicitly assumed that all distributions are absolutely continuous.

6 Discrete Bootstrap Iteration

This section attempts to summarise some very recent work, as yet available only as a working paper, Davidson (2015), on bootstrap iteration. My aim is not so much to be able to handle conventional resampling bootstraps, but to find a setup in which I can overcome the immense computational cost of iterating the bootstrap past the first few iterations. It was the idea that everything is necessarily discrete when a virtual reality is rendered by a computer that made me consider going the whole way, and starting from a discrete and finite representation of the bootstrap.

I assume that the statistic, in approximate P value form, can take on only the values i/n , $i = 0, 1, \dots, n$. Further, I assume that there are only m possible DGPs in the null model. Thus I can let the outcome space Ω consist of just $m(n+1)$ points, labelled by two coordinates (i, j) , $i = 0, 1, \dots, n$, $j = 1, \dots, m$. Golden Rule 1 requires the bootstrap DGP to satisfy the null hypothesis, and so any DGP μ we consider is represented by k , say, with $k = 1, \dots, m$.

We can then write

$$\tau(k, (i, j)) = i/n, \quad \beta(k, (i, j)) = j,$$

where the DGP μ is represented by k , and the outcome ω by (i, j) . In this way, the model is completely characterised by the probabilities p_{kij} , $k, j = 1, \dots, m$, $i = 0, 1, \dots, n$, where

$$p_{kij} = P[\tau(k, (i, j)) = i/n \text{ and } \beta(k, (i, j)) = j].$$

We have, for all $k = 1, \dots, m$, that

$$\sum_{i=0}^n \sum_{j=1}^m p_{kij} = 1,$$

and we make the definitions

$$a_{kij} = \sum_{l=0}^{i-1} p_{klj} \quad \text{and} \quad A_{ki} = \sum_{j=1}^m a_{kij} \quad i = 0, \dots, n+1.$$

Then a_{kij} is the probability under DGP k that τ is less than i/n and that $b = j$, while A_{ki} is the marginal probability under k that $\tau < i/n$. Thus we may write

$$R_0(\alpha, k) = A_{k, [\alpha n] + 1}.$$

Note that $a_{k0j} = A_{k0} = 0$ for all $k, j = 1, \dots, m$. Further, $A_{k(n+1)} = 1$ and $a_{k(n+1)j}$ is the marginal probability under k that $\beta = j$, for all $k = 1, \dots, m$.

With this setup, it can be shown that the pattern of bootstrap iteration is follows. Things are initialised by:

$$p_0(k, (i, j)) = A_{ji}^0 = i/n; \quad q_j^0(\alpha) = [\alpha n] \text{ independent of } j.$$

It turns out that A_{ji} is the bootstrap P value for the realisation (i, j) . In the context of bootstrap iteration, it is now denoted as A_{ji}^1 . The subsequent steps of the iteration are as follows. At step r , we have

$$\begin{aligned} p_r(k, (i, j)) &= A_{ji}^r, \\ q_k^r(\alpha) &= \max_{i=1, \dots, n} \{i \mid A_{ki}^r \leq \alpha\}, \\ q_k^r(A_{ki}^r) &= i, \quad \text{and} \\ R_r(\alpha, k) &= \sum_{j=1}^m a_{kq_j^r(\alpha)j}. \end{aligned} \tag{2}$$

The recurrence is then implemented by a trip across the plane defined by k :

$$A_{ki}^{r+1} = \sum_{j=1}^m a_{kq_j^r(A_{ki}^r)j} = R_r(A_{ki}^r, k).$$

The **bootstrap discrepancy** is defined, for a given DGP μ , and a given significance level α , as the difference between the rejection probability of the bootstrap test for DGP μ and level α and α itself. It is therefore equal to $R_1(\alpha, \mu) - \alpha$, in the notation used for the continuous case. For iterated bootstraps, the definition is the same: at level r , the order- r discrepancy is $R_r(\alpha, \mu) - \alpha$.

Suppose that the recurrence (2) converges, in the sense that, for given k and i ,

$$q_j^{r+1}(A_{ki}^{r+1}) = q_j^r(A_{ki}^r) \quad \text{for all } j = 1, \dots, m.$$

Then $A_{ki}^{r+s} = A_{ki}^r$ and $q_j^{r+s}(A_{ki}^{r+s}) = q_j^r(A_{ki}^r)$ for all j and for all positive integers s . Since $R_r(A_{ki}^r, k) = A_{ki}^{r+1} = A_{ki}^r$, it follows that the bootstrap discrepancy is zero for significance level A_{ki}^r if k is the DGP and $\tau = i/n$. If there is convergence for all $i = 0, 1, \dots, n$, the bootstrap discrepancy is zero unconditionally for DGP k for levels A_{ki}^r .

Discussion

The most important advantage of this discrete approach is that it eliminates conventional bootstrapping based on a simulation experiment. In exchange, it is necessary to conduct the probably costly simulation experiment needed in order to estimate the p_{kij} . However, once that experiment is carried out, it serves as a fixed overhead for arbitrary levels of bootstrap iteration. And, as we would expect, the experiment lends itself ideally to concurrent processing. It thus becomes feasible to examine the convergence or otherwise of the sequence of iterated bootstrap P values.

In the discrete case, convergence of the sequence of iterated bootstraps is probably guaranteed. It remains to be seen whether this can be proved, and, if so, under what regularity conditions. However, this need not imply that exact inference is possible at any chosen level, because there is only a finite set of levels for which the bootstrap discrepancy is necessarily zero after convergence. It can turn out that these levels are all equal or close to zero or one.

If we abstract from the simulation noise in the estimation of the p_{kij} , the discrete model is quite nonrandom. We are, in effect, working simultaneously with every point in the outcome space. Convergence, therefore, is to be understood in the ordinary sense of convergence of a sequence of real numbers. In the continuous case, of course, we have to speak of stochastic convergence, which may perhaps be almost sure, or in probability.

If this discrete approach were to be used with real data, it would be necessary to use these data to compute realisations of the quantity being bootstrapped and of the bootstrap DGP, and then to discretise them according to the plan of discretisation in use. If the realised quantity is indexed by i and the realised bootstrap DGP by j , then, for the r -tuple bootstrap, the bootstrap P value is A_{ji}^r .

It would be immensely useful to find ways of discretising the set of bootstrap DGPs used in situations that are not purely parametric. While it is easy enough to replace the use of a discrete empirical distribution for resampling by a continuous version, thus avoiding the problem inherent in iterating a conventional resampling bootstrap, it is not obvious how to

make discrete the set of bootstrap DGPs that would be obtained in this way. I conjecture that, when bootstrapping an approximately pivotal statistic, it may be possible to cover the set of bootstrap DGPs rather coarsely and still achieve satisfactory results. How best to do so remains to be seen.

The double bootstrap was introduced by Beran (1987) and (1988). He refers to “pre-pivoting”, meaning making some quantity more close to being pivotal for a model by bootstrapping it, and then bootstrapping the result. This interpretation clearly applies to higher orders of bootstrap iteration. In some sense, the iterative procedure serves to project the original statistic into a space of pivotal statistics. It would be desirable to formalise this intuition. It is also necessary to see to what extent this “projection” may adversely affect the power of a test. Of course, power is not uniquely defined when a non-pivotal statistic is used; see Horowitz and Savin (2000), and Davidson and MacKinnon (2006). But if an iterated bootstrap P value follows the uniform $U(0, 1)$ distribution, it is by definition a pivot.

In discussing bootstrap “validity”, it is conventional to make use of an appropriate asymptotic construction in order to show that the limiting distribution of the quantity considered is the same as the limiting distribution of its bootstrap counterpart. This is of course a very weak requirement. A somewhat better justification for the bootstrap comes from any refinements that can be demonstrated by an asymptotic argument, as in Hall (1992), where he uses Edgeworth expansion – still an asymptotic technique, of course.

It seems to me that convergence of the sequence of iterated bootstrap P values to the uniform distribution is a much richer and more satisfactory means of justifying or validating the bootstrap. No asymptotic argument is involved, so that the potential arbitrariness of the choice of an asymptotic construction is avoided. To the extent that the approach outlined here can be made operational for problems of interest, the approach carries its validity along with it.

Further, the new proposed criterion for validity is by no means equivalent to asymptotic validity. An example of this is when a regression model, the disturbances of which are not necessarily Gaussian, is bootstrapped using a bootstrap DGP that imposes Gaussianity. Under very weak conditions on the asymptotic construction, this bootstrap is asymptotically valid. But it certainly is not, by the criterion of convergence of iterated P values to $U(0, 1)$, for any DGPs in the model the disturbances of which are in fact not Gaussian.

7 Concluding Remarks

This paper has journeyed from some vaguely philosophical notions concerning scientific modelling, virtual reality, computation with digital computers, discreteness, simulation, and the nature of randomness, to a discussion of the bootstrap and bootstrap iteration. It is argued that viewing much scientific endeavour as involving virtual reality or realities can be very fruitful. The bootstrap can readily be interpreted as a form of virtual reality, and this interpretation leads naturally to the idea of bootstrap iteration. It is my hope that the discrete model I propose for the study of bootstrap iteration will lead to significant advances in both the theory and the practice of the bootstrap.

References

- Ashenfelter, O. and D. Card, (1985). "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs", *Review of Economics and Statistics*, **67**, 648–60.
- Athey, S. and Imbens, G. W. (2006). "Identification and Inference in Nonlinear difference-in-differences models", *Econometrica*, **74**, 431–97.
- Beran, R. (1987). "Prepivoting to Reduce Level Error of Confidence Sets", *Biometrika* **74**, 457–68.
- Beran, R. (1988). "Prepivoting test statistics: A bootstrap view of asymptotic refinements", *Journal of the American Statistical Association*, **83**, 687–97.
- Davidson, R. (2015). "A Discrete Model for Bootstrap Iteration", working paper.
- Davidson, R. and J. G. MacKinnon (2006). "The Power of Asymptotic and Bootstrap Tests", *Journal of Econometrics*, **133**, 421–441.
- Davison, A. C., Hinkley, D. V., (1997). *Bootstrap Methods and Their Application*, Cambridge University Press, Cambridge.
- Dennett, D. C. (2003). *Freedom Evolves*, the Penguin Group.
- Deutsch, D. (1997). *The Fabric of Reality*, Penguin Books.
- Devroye, Luc (1986). *Non-Uniform Random Variate Generation*, Springer-Verlag, New York. But see: <http://luc.devroye.org/rnbookindex.html>
- Dufour, J.-M., and L. Khalaf (2001). "Monte Carlo test methods in econometrics", Ch. 23 in *A Companion to Econometric Theory*, ed. B. Baltagi, Oxford, Blackwell Publishers, pp. 494–519.
- Dwass, M. (1957). "Modified randomization tests for nonparametric hypotheses", *Annals of Mathematical Statistics*, **28**, 181–187.
- Granger, C. W. J. (1969). "Investigating Causal Relations by Econometric Models and Cross-spectral Methods", *Econometrica*, **37**, 424-38.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York.
- Hall, Peter, and Susan R. Wilson (1991). "Two guidelines for bootstrap hypothesis testing", *Biometrics* **47**, 757–62.
- Heckman, J. J. (2001). "Econometrics and Empirical Economics", *Journal of Econometrics*, **100**, 3-5.
- Horowitz, J. L., and N. E. Savin (2000). "Empirically relevant critical values for hypothesis tests," *Journal of Econometrics*, **95**, 375–389.
- Keynes, J. M. (1936). *The General Theory of Employment, Interest, and Money*, Palgrave Macmillan, London.
- Knuth, D. E. (1997). *The Art of Computer Programming*, Volume 2, *Seminumerical Algorithms*, 3rd edition, Addison-Wesley. (Recently available as an ebook.)
- Moses, A. M. (2005). *Last Frontiers of the Mind – Challenges of the Digital Age*, Prentice-Hall of India Private Ltd., New Delhi.
- Sims, C. A. (1972). "Money, income and Causality", *American Economic Review*, **62**, 540-52.