

## Virtual Model Validation for Economics

David K. Levine, [www.dklevine.com](http://www.dklevine.com), September 12, 2010

White Paper prepared for the National Science Foundation, Released under a Creative Commons *Attribution Non-Commercial Share Alike* license

**Abstract:** *How can economic policies lead us to greater wealth, welfare and happiness? There is no bigger question in economics. The answer lies in correct economic theories that capture the causality linking policies to outcomes. Economic theories are a dime a dozen – we have more theories than we have human beings. The key need to answer any economic question lies in our ability to validate theories. Do we live in an Austrian world? In a Keynesian world? A world of rational expectations? This White Paper proposes that major advances in simulating virtual economies is possible and can form the basis for rapid and accurate assessment of current and future economic models. I make general proposals for developing infrastructure, as well as presenting specific ideas about the nature of models of sophisticated expectations that are needed to allow artificial agents to mimic the behavior of real human beings.*

One of the most essential needs for developing better economic theories and policy prescriptions are improved methods of validating theories. Originally economics depended on field data gathered from surveys. The introduction of laboratory experiments added a new dimension: a good theory ought to be able to predict outcomes in the artificial world of the laboratory. Modern economics has extended this in two directions: to field experiments, keeping many of the controls of laboratory experiments, while conducting experiments in more natural environments, and through internet experiments, extending the size and scope of the populations used in experiments. The importance of these innovations is great, and have been discussed in depth by List, among others. Not only is it easier, faster, and more practical to validate theories, but through greater control, issues of causality that are difficult to analyze with field data can be addressed. On the other hand, laboratory, field and internet experiments all have important limitations. Even the largest internet experiment is orders of magnitudes smaller than a small real economy: thousands of subjects rather than millions of real decision makers. Experiments are faster than waiting for new data to arrive, but are still time-consuming – the more so with the National Institute of Health trying to apply inappropriate medical ethics to harmless economics experiments. Subjects are expensive to pay, especially in large scale experiments. Finally, control in experiments is still and necessarily imperfect. In particular, it is not possible to control for either risk aversion or social preferences.

An alternative method of validating theories is through the use of entirely artificial economies. To give an example, imagine a virtual world – something like Second Life, say – populated by virtual robots designed to mimic human behavior. A good theory ought to be able to predict outcomes in such a virtual world. Moreover, such an environment would offer enormous advantages: complete control – for example, over risk aversion and social preferences; independence from well-meant but irrelevant human subjects “protections”; and great speed in creating economies and validating theories. If we were to look at the physical sciences, we would see the large computer models used in testing nuclear weapons as a possible analogy. In the economic setting the great advantage of such artificial economies is the ability to deal with heterogeneity, with small frictions, and with expectations that are backward looking rather than determined in equilibrium. These are difficult or impractical to combine in existing calibrations or Monte Carlo simulations.

The notion of virtual economies is not new: the general concept has become known as agent-based modeling. Yet, despite three decades of effort, agent-based models are largely limited to studying phenomena such as traffic patterns. In economics, the most influential work has been that of Nelson and Winter examining the evolution of growth and change. Yet this work has not had a substantial impact on our understanding of economics. The problematic aspect of agent-based modeling has been the focus on frameworks for agents interacting – the development of languages such as SWARM or Cybele – and the fact that agents are limited to following simple heuristic decision rules. Agent-based models are interesting from the perspective of modeling order arising from the interaction of many simple decision rules – along the lines of Becker’s observation that demand curves slope downwards when people choose randomly along the budget line. These models are also useful in constructing examples to illustrate special points. However, existing agent-based models are too primitive to be used either for evaluating economic policies or for validating economic theories.

Although some argue that people are simple-minded and follow simple rules, the practical problem is that people are far better learners and vastly more sophisticated than existing computer models. Simple rules are not a good representation – for example – of how stock market traders operate. What is needed are agents who use sophisticated algorithms. Real people in the laboratory and the field are able to recognize sophisticated patterns and anticipate future events. One of the simplest examples is the learning that takes place in the laboratory when subjects discover the idea of dominated strategies.

The key to developing useful virtual economies is modeling inferences about causality. A useful place to start thinking about the issues is with Sargent's *The Conquest of American Inflation* and the follow on papers with Cogley. There the Federal Reserve is modeled as a sophisticated Bayesian learner equipped with powerful econometric methods and sophisticated intertemporal preferences – but limited to the data on hand. Dynamic Bayesian optimization including the use of policy experiments enables the Fed to learn the true relationship between unemployment and inflation leading over time to superior monetary policy. The model is validated against the last 50 years of data on monetary policy, inflation and unemployment.

Notice that in the Sargent-Cogley world, the decision problem is narrowly circumscribed: how best to choose the rate of monetary expansion. The incoming data is also circumscribed, and issues such as learning by analogy do not arise. Moreover, they assume one of the underlying models is correct: in an environment where none of the underlying models are correct, Bayesian methods are not so useful.

A useful framework for thinking about this problem is the computer science problem that underlies boosting: the choice among experts. A carefully chosen randomization strategy giving greater weight to experts with better track records can do as well asymptotically as the best expert – this is true even when all the experts are wrong. The framework can be extended to dynamic decision making by putting time into blocks – a technique often used in analyzing repeated games. If the block is long enough the payoff is approximately the same as the infinite present value. While this may be a useful benchmark for learning about causality, it is a weak criterion. First, blocking periods means that the length of time taken to learn is enormous. While the evaluation of dynamic plans requires that those plans be maintained for some period of time, there is little point in sticking with an expert when it is clear that he is doing a poor job. Second, causality between periods is ignored. To take a simple example, imagine a repeated Prisoners' Dilemma game where your opponent plays tit-for-tat starting by not cooperating. An expert who says your opponent will always cheat will lead you to cheat – and his forecasts will be correct. Of course an expert who says you should always cooperate and your opponent will cooperate after the first period is equally correct, and you will do much better following his advice.

If we accept the basic framework of replacing a prior over models with a probability of choice over experts it is possible to outline the issues that need to be resolved. Experts make recommendations that can be evaluated directly – the weak criterion for asymptotic success has already been described. They also provide suggestions of evidence that demonstrate their ability as experts. That evidence needs to be assessed on several dimensions:

1. Calibration – how accurate are the predictions?
2. Precision – are the predictions vague or are they sharp? Does the expert always say “it might rain or shine with equal probability” or does he say half the time “it will rain for sure” and half the time “it will shine for sure.” The latter prediction is more precise.

Calibration and precision are traditional criteria for model evaluation. However, there are additional considerations.

3. Relevance: A molecular biologist may be able to make very accurate forecasts about the formation of molecules – but why should that lead me to take his investment advice? Notice that there is surely heterogeneity among people in evaluating the relevance of forecasts: some may believe that a good molecular biologist can better forecast stock prices than a bad one. Utility is directly related to relevance: two experts may both recommend I not jump off a bridge. One may say “if you jump you will die” while the

other may tell me the speed at which I will hit the water and how far my body parts will be flung. But this additional information is of no use in decision making. Detailed information about inferior plans is not helpful.

4. Scope: experts differ in the number of things they can forecast. It is natural to put more weight on the advice of an expert who can predict a great many things well over one who can predict only a few things well. Notice that this goes in the opposite direction from relevance.
5. Ease of implementation. Some advice may be difficult to follow in practice. Here the analysis of impulsive behavior as in the theoretical models of Fudenberg and Levine or the empirical work of Cunha and Heckman may play a useful role.

Notice that the expert approach gets at several tricky issues. One is the issue of generalization. An expert who makes forecasts in many domains implicitly provides a formula for generalizing results from one domain to another. For example we may want to get at the idea that when someone learns the idea of dominated strategies they do not merely learn not to play a dominated strategy in a particular game, but they learn not to play dominated strategies in any game. This can be done in the expert framework by providing an expert who advises against playing dominated strategies in all games. Second, the framework deals well with the transmission of ideas – experts can be communicated from one person to another – unlike the sending of messages or provision of data there is no issue of the reliability of the information; the recipients can test the ideas implicit in the expert for themselves. However, it deals less well with the need to experiment with “off the equilibrium path” behavior to determine the causal consequences because it does not tell us what is the option value of experimentation.

A large part of the advancement of the science must be the development of these and other learning models, understanding which ones have the best theoretical properties; which ones work best in practice; and which ones are most descriptive of actual behavior. The validation against behavior may benefit from neuro-economic experimental methods such as that of Glimcher or Rustichini. At the extreme, efforts such as the blue-brain project can provide additional paths of validation.

The infrastructure requirements for this project are large. The development and validation of sophisticated agent models is only a part. To combine many agent-models into a single economy requires reliable high speed networking and substantial computer power at each end, as well as thoughtful and well-developed models of production, trade and consumption. Existing agent-based modeling frameworks may provide a starting point, but are not equipped to handle the load that simulating an artificial economy requires. Multi-player computer games have solved many of these problems and may provide an alternate point of departure. Second Life is one such game, but it is a gigantic and open-ended, so not ideal for test-bedding. Smaller more controlled gaming environments such as Capitalism II might be more suitable.

Game environments raise an important issue. We cannot reasonably simulate economies since the beginning of time, nor the process by which people acquire information growing up and in school. So artificial agents will need to be endowed with some knowledge of the environment they are in. Games often have artificial intelligence agents – sometimes quite clever – but very specialized. Economics needs generalist agents, but these agents must be equipped with reasonable initial knowledge and the ability to respond to complicated rule changes. Our robots will neither be able to read the health-care bill; nor to understand the pronouncements by experts about what it means. Some alternatives will need to be developed.

At the human level the infrastructure requires the collaboration of economic theorists and practitioners with computer scientists, psychologists, neuroscientists, and quite possibly computer game developers.

### **References**

Fudenberg, D. and D. K. Levine [2009]: "Learning-Theoretic Foundations for Equilibrium Analysis," *Annual Review of Economics*, 1

Levitt, S. and J. List, "Field Experiments in Economics: The Past, The Present, and The Future," *European Economic Review*, forthcoming, 2009.

Nelson, R. and S. Winter [1982] *An Evolutionary Theory of Economic Change*. Cambridge: Harvard University Press