Model of Theory Choice (Brock and Durlauf 1999)

Suppose there is a population of scientists who form a collective judgement about the "best" theory for some natural phenomenon. For simplicity, we will consider a case where there are two theories, a standard established theory and a novel theory. We will study the case where the weight of evidence for the novel theory is growing slowly due to new data. The weight of evidence for the two theories could be measured by the posterior probability, though we will not need a Bayesian analysis.

As is usual in economic models, assume that each scientist uses a utility measure to decide whether to support one theory or the other. This is analogous to using a benefit-cost analysis to choose which car to purchase from an array of different makes and models. Using benefit-cost analysis, a person would write down the benefits and costs of each car in some standardized way, and choose the car that maximized the difference between benefits and costs, or utility. In our analysis of two theories, $V_{i,t}(\omega_{i,t})$ stands for the utility of theory ω for scientist i at time t. At time t, the scientist supports the theory with greatest $V_{i,t}(\omega_{i,t})$. The scientist's choice can change over time as new information becomes available.

Suppose that each scientist's utility is composed of three components:

[1] $$V_{i,t}(\omega_{i,t}) = u_{i,t}(\omega_{i,t}) - E[\sum_{j \neq i} J_{i,j,t}(\omega_{i,t} - \omega_{j,t})^2] + \varepsilon(\omega_{i,t})$$

The first term is the deterministic data-based utility. It will be the same for all scientists, in the case when they all have equal access to the same data. The second term is a conformity effect. It measures the expected loss of utility incurred by disagreeing with another scientist, summed over all the other scientists. The squared term measures the difference in theory choice between scientist i and scientist j, J measures the penalty to scientist i of disagreeing with scientist j, and E is the expectation operator to account for scientist i's uncertainty about the beliefs of scientist j. The final term measures an individual random component of theory choice. This represents random variability among individuals in their intuition about the theories, independent of the deterministic data-based component (the first component).

In economic theory, the extreme-value distribution is commonly used for random effects like the third component of equation 1. In the present case, there are only two choices so the extreme-value probability density function is

[2] $$\Pr(\varepsilon(S) - \varepsilon(N) < z) = \frac{1}{1 + \exp(-\beta z)}$$

where S indexes the standard theory, N indexes the novel theory, and the parameter $\beta > 1$. The parameter β measures the diversity of individual theory evaluations in the community of scientists. Small values of β correspond with high diversity, low values of β correspond with low diversity.

Using equations 1 and 2, one can simulate the theory choices of a large number of scientists as the weight of evidence for the novel theory changes over time. Because this simple model has only two theories, we can measure the relative weight of evidence for the novel theory as

$$[3] \qquad h_{i,t} = \frac{1}{2}[u_{i,t}(N) - u_{i,t}(S)]$$

where as before N indexes the novel theory and S indexes the standard theory. As the data-based evidence for the novel theory increases, h increases.

Suppose that h grows gradually (i.e. data-consistent support for the novel theory gradually increases relative to data-consistent support for the standard theory). Suppose that at each level of h there is sufficient time for the scientists' preferences to reach a steady state. Brock and Durlauf (1999) show that a relatively simple equation describes the steady-state distribution of scientists' preferences for a given value of h. The steady state mean theory preference m* for a given level of h is given by solutions of the equation

$$[4] \qquad m^* = \tanh(\beta h + \beta J m^*)$$

In equation 4, tanh is the hyperbolic tangent function and J is the average penalty for nonconformity.

In the computer exercise, we will study the behavior of equation [4].

<u>Reference</u>

Brock, W.A. and S.N. Durlauf. 1999. A formal model of theory choice in science. Economic Theory 14: 113-130.