Socio-dynamic discrete choice: Theory and application

Dugundji, E.R.

3 SOCIO-DYNAMIC BINARY LOGIT: THEORY

The past two decades have seen noteworthy examples of concepts from statistical physics being applied to further the understanding of complex socio-economic systems, particularly with respect to non-market interactions in the sense of conventions, network externalities, neighborhood or group effects, or interactive agents. Notwithstanding Manski's seminal critique "Identification of endogenous social effects: the reflection problem" (1993), early examples include work by Brock (1993), Durlauf (1991, 1993), Blume (1993), Aoki (1995), many of which rely to some extent on mean field approximations in derivation of analytical results. Furthermore principles from the mathematical theory of random perturbations of dynamical systems have added insight to the area of learning and evolutionary game theory, and adaptive economic dynamics more generally. Forerunners in this field include Young (1993, 2010), Kandori, Mailath and Rob (1993), Ellison (1993) from the game theoretic tradition, applying results of Russian mathematicians Freidlin and Wentzell (1979, 1998).[1] Aoki (1995) and Blume and Durlauf (2003) have examined dynamical aspects of social interactions models with a discrete choice theoretic approach.

The traditional discrete choice framework reviewed in section 2 assumes independent individuals. Brock and Durlauf (2001a) and Aoki (1995) relax this assumption. Their approach is to assume that the otherwise independent individuals are influenced by an aggregate of all other choices in the community. There is an inherent dynamic because each individual re-evaluates its choice based on the choices made by other individuals. This implies an implicit time-trajectory of repeated choices that defines the dynamics of the system. It is in this sense that we call this a *socio-dynamic* model: the dynamics are driven by social influence, albeit global social influence at this stage. Brock and Durlauf (2001a) derive results for the equilibrium state of this system. Aoki (1995) goes a step further and also characterizes the dynamics of the transition to the equilibrium as a continuous time discrete state Markov process. The master equation will be given in this

1 Other notable examples of social interactions models include: Hildenbrand (1971), Weidlich (1971, 1994, 2000, 2002), Weidlich and Haag (1983), Weidlich and Braun (1992), Ceccato and Huberman (1989), Eckstein and Wolpin (1989), Kirman (1992), Aoki (1996, 2002), Aoki and Yoshikawa (2007), Helbing (2010); Binmore, Samuelson and Vaughn (1995), Glaeser, Sacerdote and Scheinkman (1996), Glaeser and Scheinkman (2002), Benaim and Weibull (2003); Bayer and Timmins (2005, 2007), Ekelund, Heckman and Nesheim (2004), Topa (2001), Conley and Topa (2002, 2003, 2007), Epple and Sieg (1999), Hoff and Sen (2005); Mobius (2000), Schelling (1971); Akerlof (1997), Berry (1994), Berry, Levinsohn and Pakes (1995, 2004), Heckman, Matzkin and Nesheim (2010).

section and the steady states of this dynamic process will be derived. The model discussed here indeed has a "network", but at this stage it is fully-connected. This is how we are able to calculate the analytical results as a benchmark to verify the programming implementation of the agent-based model at the outset of our consideration of the case study in Chapter 6. Thereafter, we then proceed to further relax the condition of the fully-connected network, and consider explicitly the case of heterogeneous local social influence, instead of global social influence.

For reasons of simplicity, in this chapter we first introduce the socio-dynamic model for binary choice, and limit our initial discussions to the binary logit model. In Chapters 4 and 5, using a different approach, we will subsequently derive analytical results for the trinary multinomial logit model and the trinary nested logit model as a benchmark to verify the programming implementation of the agent-based model for the case study in Chapter 7.

For all thoroughness in incremental scientific development of results as well as clarity of exposition, in Appendices A through C, we also revisit the binary logit model and derive additional analytical results for the case of the binary logit model with constant bias as well as additional analytical results for the trinary multinomial logit model with constant bias using the same methodological approach as in Chapters 4 and 5.

The interested reader is highly encouraged to especially review Appendix A in parallel with reading this chapter, or directly hereafter, before proceding to Chapter 4.

3.1 FIELD EFFECTS MODEL

Let us begin our analysis with the binary case of choice with the universal set containing only two alternatives, $C = i, j$, say, choice of travel by car versus by public transit. Thus, we have the simplification:

$$\begin{aligned} P_{in} &= \frac{e^{\mu V_{in}}}{e^{\mu V_{in}} + e^{\mu V_{jn}}} = \frac{e^{\mu(V_{in}-V_{jn})}}{e^{\mu(V_{in}-V_{jn})}+1} \\ P_{jn} &= \frac{e^{\mu V_{jn}}}{e^{\mu V_{in}} + e^{\mu V_{jn}}} = \frac{1}{e^{\mu(V_{in}-V_{jn})}+1} \end{aligned} \tag{3.1}$$

where we again for convenience we make the arbitrary assumption that the positive scale parameter $\mu = 1$.

Let N_i and N_j be the total numbers of decision-making entities who have chosen respectively alternative i and alternative j at time t. Since we assume the choice set to be mutually exclusive and collectively exhaustive, for the binary case we have $N = N_i + N_j$. Now let $x_i =$

N_i/N and $x_j = N_j/N = (1 - x_i)$ be the global proportions of decision-making entities who have made each choice, and define:

$$x \equiv x_i - x_j = x_i - (1 - x_i) = 2x_i - 1 \tag{3.2}$$

Note that the variable x varies on the range -1 to 1. In the limit where $x = -1$, none of the decision-making entities in the sample have chosen alternative i, that is, all have chosen alternative j. In the limiting case where $x = 1$, all of the decision-making entities in the sample have chosen alternative i, and none have chosen alternative j. In the case where $x = 0$, half of the decision-making entities in the sample have chosen alternative i, and half have chosen alternative j.

Following this approach, we introduce global social dynamics by allowing the term $V_{in} - V_{jn}$ in equation (3.1) to be a linear-in-parameter β function of the proportions x_i and x_j of decision-making entities[2] who have made each choice:

$$V_{in} - V_{jn} \equiv \beta f\left(x_i - x_j\right) = \beta f(x) \tag{3.3}$$

The function $f(x)$ is an arbitrary function of x. The important aspect for the subsequent analytical results is that $V_{in} - V_{jn}$ is linear in the parameter β. In Aoki's original work, he considered an example with $f(x)$ polynomial in x. In our binary case study application in Chapter 6 we consider $f(x) = x$, however the analytical results apply more generally. Substituting equation (3.3) into (3.1) and normalizing the scale parameter $\mu = 1$, we have:

$$P_{in}(x) = \frac{e^{\beta f(x)}}{e^{\beta f(x)} + 1} \tag{3.4}$$

The variable x_i is termed a *field variable*. As motivated by Aoki (1995),

> "Knowledge of a field variable relieves agents (at least partially) of the need for detailed information on interaction patterns. Any macroeconomic variable that serves this decentralizing function is called a field variable."

Such an approach can be particularly useful if other more conceptually restrictive assumptions such as having constant interactions among all possible pairs of microeconomic agents, or interactions only with other agents in neighborhoods in the strict sense of Markov random fields (Kindermann and Snell, 1980), are inappropriate. The field variable provides a way to model an average aggregate social influence.

2 The notation here differs slightly from Aoki (1995).

Note that the parameter β indicates the level of certainty in the model. In the case $\beta f(x) \gg 0$, then the probability that the decision-making entity n chooses alternative i approaches unity and we have effectively deterministic choice:

$$P_{in} = \frac{e^{\beta f(x)}}{e^{\beta f(x)} + 1} \approx 1, \text{ for } \beta f(x) \gg 0 \tag{3.5}$$

That is, it is strongly certain that the utility of alternative i is greater than the utility of alternative j.

In the case $\beta f(x) \approx 0$, then the probability that the decision-making entity n chooses alternative i approaches $1/2$ and we have effectively a "fair coin toss" between the two alternatives:

$$P_{in} = \frac{e^{\beta f(x)}}{e^{\beta f(x)} + 1} \approx \frac{e^0}{e^0 + 1} = \frac{1}{2}, \text{ for } \beta f(x) \approx 0 \tag{3.6}$$

That is, there is uncertainty as to which choice alternative is more profitable.

In the case $\beta f(x) \ll 0$, then the probability that the decision-making entity n chooses alternative i approaches zero. In our binary case, we have again effectively deterministic choice:

$$P_{in} = \frac{e^{\beta f(x)}}{e^{\beta f(x)} + 1} \approx 0, \text{ for } \beta f(x) \ll 0 \tag{3.7}$$

That is, it is strongly certain that the utility of alternative i is less than the utility of alternative j.

The critical difference thus between the two cases $\beta f(x) \gg 0$ and $\beta f(x) \ll 0$ is that in the former, there is a positive influence on the choice alternative i (ie. positive attraction towards the choice alternative i), but in the latter, there is a negative influence on the choice alternative i (ie. negative attraction towards toward the choice alternative i, or otherwise said, a repulsion away from i). Both of the cases $\beta f(x) \gg 0$ and $\beta f(x) \ll 0$ are indeed strongly certain, but the former is a strong and certain attraction towards the choice alternative i and the latter is a strong and certain repulsion away from the choice alternative i.

3.2 SCALAR AUTONOMOUS EQUATION

Aoki models decision-making entities as *jump Markov* processes, and the dynamics of interactions (among these entities) as *birth-and-death* stochastic processes where they switch their choices randomly and asynchronously with transition rates that are functions of the aggregate situations summarized by the proportion of decision-making entities who have taken the same choices. As remarked by Blume and Durlauf (2003) in an analogous continuous time Markov process changing state in discrete jumps with a uniform global interactions model (that is, a model with interactions of constant strength between all pairs of decision-making entities),

> "Implicit... is the fact that players are *myopic* in (stochastically) best-responding to the current play of the population rather than some forecast of future paths of play."

In this dissertation, we will accept this myopic assumption for our exogenous network case with transportation mode choice, however in future work, particularly for the endogenous network case with residential choice, this assumption may be worth re-visiting.

Let us formalize these assumptions by considering the aggregate behavior of the population of N decision-making entities instead of the behavior of an individual decision-making entity. Let $P(N_i, t)$ denote the probability that N_i number of decision-making entities have chosen alternative i at time t. The total number of possible states of the population of N decision-making entities is $N+1$, since the number of decision-making entities choosing alternative i can range from 0 to N, and the number of decision-making entities choosing alternative j is fully-determined given the number choosing alternative i, for our binary choice case. Let $W_{N_i, N_i'}$ denote the transition rate between the states of the population with N_i and N_i' number of decision-making entities choosing alternative i, and let W_{N_i', N_i} be the rate of the inverse transition. Aoki uses the *backward Chapman-Kolmogorov* equation, or so-called "master equation," to govern the time evolution of the probability density.[3] The master equation is fully specified once the transition rates are given between the states.

$$\begin{aligned}\frac{\partial P\left(N_i, t\right)}{\partial t} &= \sum_{\forall N_i' \neq N_i} P\left(N_i', t\right) W_{N_i', N_i} - \sum_{\forall N_i' \neq N_i} P\left(N_i, t\right) W_{N_i, N_i'} \\ &= \sum_{\forall N_i' \neq N_i} \left\{P\left(N_i', t\right) W_{N_i', N_i} - P\left(N_i, t\right) W_{N_i, N_i'}\right\}\end{aligned} \tag{3.8}$$

As remarked in Reif (1965),

> "Note that all terms... are real and that the time t enters linearly in the first derivative. Hence the master equation does not remain invariant as the sign of the time t is reversed from t to $-t$. This equation describes, therefore, the irreversible behavior of a system."

Nonetheless, as motivated by Reif, there is assumed to be a symmetry property relating a transition to its inverse:

$$W_{N_i', N_i} = W_{N_i, N_i'} \tag{3.9}$$

3 For the origin of the word master, see for example Aoki, 1995, p.153, van Kampen, 2007, p. 97; for more about the master equation in equilibrium situations, see Reif, 1965, p. 550.

In general, we have that the probability density $P(N_i, t)$ tends to increase with time because the population transitions from other states to the given state with N_i number of decision-making entities choosing alternative i, and the probability density tends to decrease with time because the population in the given state transitions to other states. For an assumption of asynchronous choices of the decision-making entities, however, we have a convenient simplification, since the only states to which the population in the given state with N_i number of decision-making entities choosing alternative i can possibly transition to, are the states with $N_i' = N_i + 1$ and/or $N_i' = N_i - 1$ number of decision-making entities choosing alternative i. In short, in the birth-and-death processes of this thesis, the transition rates are non-zero only for N_i' which is either $N_i + 1$ (a so-called "birth") or $N_i - 1$ (a so-called "death"). We can thus simplify the master equation for this continuous time discrete state Markov process:

$$\begin{aligned}\frac{\partial P(N_i, t)}{\partial t} &= \{P(N_i+1, t)\, W_{N_i+1, N_i} - P(N_i, t)\, W_{N_i, N_i+1}\} \\ &\quad + \{P(N_i-1, t)\, W_{N_i-1, N_i} - P(N_i, t)\, W_{N_i, N_i-1}\} \\ &= \{P(N_i+1, t) - P(N_i, t)\} W_{N_i, N_i+1} \\ &\quad + \{P(N_i-1, t) - P(N_i, t)\} W_{N_i, N_i-1}\end{aligned} \tag{3.10}$$

The assumption that only one decision-making entity revises its choice per unit time may be reasonable for analytical purposes if we consider an arbitrarily small time unit. In practical situations, particularly with very large populations however, we can also imagine that there can be a non-negligible time-lag in the spread of information in the population, whereby multiple decision-making entities may revise their choices per unit time interval, before the knowledge about changes in the system is disseminated. Although not explored here, in the multi-agent simulation implementation of the model we do allow for the possibility to relax this assumption of asynchronous choices, and to explicitly allow for revisions by multiple decision-making entities per unit time via an external parameter that can be set by the researcher.

In the simplest birth-and-death processes, the transition rates W_{N_i, N_i+1} and W_{N_i, N_i-1} are given by:

$$\begin{aligned}W_{N_i, N_i+1} &= \kappa (N - N_i) = N\kappa \left(1 - \frac{N_i}{N}\right) \\ W_{N_i, N_i-1} &= \lambda N = N\lambda \frac{N_i}{N}\end{aligned} \tag{3.11}$$

More generally we can express the transition rates W_{N_i, N_i+1} and W_{N_i, N_i-1} as expansions in powers of N^{-1}:

$$\begin{aligned} W_{N_i,N_i+1} &= g(N)\left[\gamma_0(N_i/N) + (1/N)\gamma_1(N_i/N) + O\left(N^{-2}\right)\right] \\ W_{N_i,N_i-1} &= g(N)\left[\rho_0(N_i/N) + (1/N)\rho_1(N_i/N) + O\left(N^{-2}\right)\right] \end{aligned} \tag{3.12}$$

Dropping all terms of order $N-1$ and higher, and making the simplifying assumptions that:

$$g(N) = N \tag{3.13}$$

and that the "birth" transition rate W_{N_i, N_i+1} is linear in the individual choice probability P_{in} that alternative i is superior to alternative j, and the "death" transition rate W_{N_i, N_i-1} is linear in the individual choice probability P_{jn} that alternative j is superior to alternative i, we have instead of equation (3.11):

$$\begin{aligned} W_{N_i,N_i+1} &= N\gamma_0(N_i/N) = N\kappa\left(1-\frac{N_i}{N}\right)P_{in}(x) = N\kappa\frac{1-x}{2}P_{in}(x) \\ W_{N_i,N_i-1} &= N\rho_0(N_i/N) = N\lambda\left(\frac{N_i}{N}\right)P_{jn}(x) = N\lambda\frac{1+x}{2}P_{jn}(x) \end{aligned} \tag{3.14}$$

Aoki (1995) shows that the mean φ of the field variable x is governed by the deterministic differential equation:[4]

$$\frac{d\varphi}{dt} = \kappa\frac{1-\varphi}{2}P_{in}(\varphi) - \lambda\frac{1+\varphi}{2}P_{jn}(\varphi) \tag{3.15}$$

Substituting (3.4) into (3.15) and normalizing $\kappa = 1$ and $\lambda = 1$, we have:

$$\begin{aligned} \frac{d\varphi}{dt} &= \frac{1-\varphi}{2}\left(\frac{e^{\beta f(\varphi)}}{e^{\beta f(\varphi)}+1}\right) - \frac{1+\varphi}{2}\left(\frac{1}{e^{\beta f(\varphi)}+1}\right) \\ &= \frac{e^{\beta f(\varphi)} - \varphi e^{\beta f(\varphi)} - 1 - \varphi}{2\left(e^{\beta f(\varphi)}+1\right)} \\ &= \frac{1}{2}\frac{e^{\beta f(\varphi)}-1}{e^{\beta f(\varphi)}+1} - \frac{\varphi}{2}\frac{e^{\beta f(\varphi)}+1}{e^{\beta f(\varphi)}+1} \\ &= \frac{1}{2}\frac{e^{\frac{1}{2}\beta f(\varphi)} - e^{-\frac{1}{2}\beta f(\varphi)}}{e^{\frac{1}{2}\beta f(\varphi)} + e^{-\frac{1}{2}\beta f(\varphi)}} - \frac{\varphi}{2} \\ &= \frac{1}{2}\left(\tanh\frac{1}{2}\beta f(\varphi)\right) - \frac{\varphi}{2} \end{aligned} \tag{3.16}$$

4 For the origin and derivation of the Fokker-Planck equation, see for example van Kampen, 2007, Chapter 8, or Reif, 1965, Section 15.11-15.1.

Stationary points are zeros of $d\varphi/dt$. Thus the key equation to determine local equilibria is: [5]

$$\frac{d\varphi}{dt} = 0: \ \varphi = \tanh \frac{1}{2}\beta f(\varphi) \tag{3.17}$$

This scalar autonomous equation can be solved conveniently graphically, for example, by plotting the left-hand-side and the right-hand-side on a graph and finding their intersection. Depending on the specification of $f(\varphi)$ and the value of β, this equation may have more than one solution.

3.3 STABILITY ANALYSIS

A stationary point of the mean φ for the field variable x is locally stable in perturbations of the mean if the derivative $d^2\varphi/d\varphi dt$ is negative:

$$\begin{aligned}
\frac{d}{d\varphi}\left(\frac{d\varphi}{dt}\right)\Big|_{\varphi=\tanh\frac{1}{2}\beta f(\varphi)} &= \frac{d}{d\varphi}\left(\frac{1}{2}\left(\tanh\frac{1}{2}\beta f(\varphi)\right) - \frac{\varphi}{2}\right)\Big|_{\varphi=\tanh\frac{1}{2}\beta f(\varphi)} \\
&= \frac{1}{2}\left(1-\tanh^2\frac{1}{2}\beta f(\varphi)\right)\left(\frac{1}{2}\beta f'(\varphi)\right) - \frac{1}{2}\left(\frac{d\varphi}{d\varphi}\right)\Big|_{\varphi=\tanh\frac{1}{2}\beta f(\varphi)} \\
&= \frac{1}{4}\left(1-\varphi^2\right)\beta f'(\varphi) - \frac{1}{2}
\end{aligned} \tag{3.18}$$

Thus we have the condition for local stability:

$$\frac{d}{d\varphi}\left(\frac{d\varphi}{dt}\right)\Big|_{\varphi=\tanh\frac{1}{2}\beta f(\varphi)} < 0: \ \left(1-\varphi^2\right)\beta f'(\varphi) < 2 \tag{3.19}$$

If the derivative $f'(\varphi)$ is non-positive with β any non-negative value, or if the derivative $f'(\varphi)$ is non-negative with β any non-positive value, then this local stability condition is always satisfied, since we have defined $x \equiv x_i - x_j$ on the interval [-1,1], and thus $\varphi^2 = [E(x)]^2$ will always have a value between 0 and 1, so that the term $(1-\varphi^2)$ is always non-negative. If however, we have a case where $\beta f'(\varphi) \gg 0$, the inequality in equation (3.19) may be violated, with the equilibrium becoming unstable.

3.4 CONCLUSIONS AND REFLECTIONS

In this chapter we have reviewed the important theoretical concept of a "field" variable which we will apply throughout the remainder of this dissertation. We have then reviewed Aoki's derivation (1995)

5 Another quicker derivation considers aggregate dynamics in the limit of the number of agents going to infinity and uses an infinitesimal generator for jump Markov processes, see for example Ethier and Kurtz, 1986, Chapter 11.

of the scalar autonomous equation describing the dynamics of socio-dynamic binary logit model. We have noted that this equation can be solved conveniently graphically. We have also reviewed how the stability of the steady-state solutions at equilibrium can be determined. We will use these theoretical results in this chapter as a benchmark for our further exploration of the socio-dynamic binary logit model later in Chapter 6.

Perhaps not surprisingly in light of the similarity noted in Chapter 1 between the multinomial logit model and the canonical distribution, the binary discrete choice problem with social interactions presented thus far also has a direct analogy in statistical physics. Huang (1997) introduces the topic of ferromagnetism:

> "One of the most interesting phenomena in the physics of the solid state is ferromagnetism. In some metals, e.g. Fe and Ni, a finite fraction of the spins of the atoms becomes spontaneously polarized in the same direction, giving rise to a macroscopic magnetic field. This happens, however, only when the temperature is lower than a characteristic temperature known as the Curie temperature. Above the Curie temperature, the spins are oriented at random, producing no net magnetic field."

In terms of the discrete choice problem, we have an analogy between spins of atoms and the choices of decision-making entities. With social feedback, these choices can become spontaneously polarized toward one alternative or another giving rise to observed aggregate preferences in blocks of a sampled population. As we saw in Aoki's derivation, this only happens however, when the parameter β, the coefficient characterizing the interaction effect, is higher than a certain critical value. Below this critical value, the choices are oriented at random when there are no other explanatory variables in the model, producing no net aggregate preferences.

If we are only interested in equilibrium aspects of the problem, there is a more straightforward approach yielding the same result as in (3.17). This alternative approach is well known in statistical physics: the self-consistency condition applied to a heuristic mean field theory of ferromagnetism. We will apply the self-consistency approach in the next chapter.

For deeper appreciation of the research presented in the next chapter, the interested reader is highly encouraged at this point to first review Appendix A where the results in this chapter are re-derived using the self-consistency approach.