Comments on Identification of Effects of Segregation

Steven N. Durlauf

University of Wisconsin at Madison

HCEO Segregation Conference

2/2/13

Purpose of this talk

1. Indicate some of the identification problems that have been studied in economics of social interactions; problems that apply to evaluation of the consequences of segregation.

2. Argue that segregation, as an equilibrium outcome of endogenous social structure formation, should be modeled with behaviors for which we are concerned that segregation matters.

Social Interactions: Basic Ideas

1. Individual beliefs, preferences, and opportunities are conditioned by group memberships. This conditioning often takes the form of complementarities, so the likelihood or level of an action by one person increases with respect to the behavior (or certain characteristics) of other

-social multipliers created

2. Memberships evolve in response to these interactions. Groups stratify along characteristics which affect outcomes.

-leads to social and economic segregation

3. Inequality and poverty result as members of family dynasties persistently face different interaction environments

-limiting case is a poverty trap

Key Features of this Approach

1. Individual incentives and social structure meld into a more general explanation of individual behavior.

2. Aggregate behaviors such as crime or nonmarital fertility rates emerge through the interactions within a heterogeneous population

3. Segregation effects are manifestations of characteristics and behaviors of group members.

Theory

Consider *I* individuals who are members of a common group *g*. Our objective is to probabilistically describe the individual choices of each *i*, ω_i (a choice that is taken from the elements of some set of possible behaviors Ω_i) and thereby characterize the vector of choices of all members of the group, ω.

It is useful to distinguish between five forms of influences on individual choices.

X_i, $r-$vector of deterministic (to the modeler) individual-specific characteristics,

Y_g, $s-$vector of deterministic (to the modeler) group-specific characteristics

$\mu_i^e(\omega_{-i})$, the beliefs individual i possesses about behaviors of others in the group, expressed as a probability measure over those behaviors.

ε_i, vector of random individual-specific characteristics associated with i, unobservable to the modeler

α_g vector of random group-specific characteristics, unobservable to modeler

Individual choices ω_i are characterized as representing the maximization of some payoff function V ,

$$\omega_i = \mathbf{argmax}_{\lambda \in \Omega_i} V\left(\lambda, X_i, Y_g, \mu_i^e\left(\omega_{-i}\right), \varepsilon_i, \alpha_g\right)$$

The decision problem facing an individual as a function of preferences (embodied in the specification of V), constraints (embodied in the specification of Ω_i) and beliefs (embodied in the specification of $\mu_i^e\left(\omega_{-i}\right)$).

As such, the analysis is based on completely standard microeconomic reasoning to describe individual decisions.

Beliefs

This basic choice model is closed by imposing self-consistency between subjective beliefs $\mu_i^e\left(\omega_{-i}\right)$ and the objective conditional probabilities $\mu\left(\omega_{-i} \mid F_i\right)$.

Self-consistency is equivalent to rational expectations in the usual sense.

From the perspective of modeling individual behaviors, it is typically assumed that agents do not account for the effect of their choices on the decisions of others via expectations formation.

Equilibrium concept: Bayes/Nash

Characterizing Discrete Choice with Social Interactions

Suppose agents face $l \in \{0...L-1\}$ choices. Assume that the utility of choice l is

$$c_l X_i + d_l Y_g + J p^e_{i,l} + \varepsilon_{i,l} + \alpha_g$$

and let

$$h_{il} = c_l X_i + d_l Y_g + \alpha_g$$

Then, the conditional probability of choice l is

$$\mu\left(\omega_i = l \middle| h_{i,j}, p^e_{i,j}\ \forall j\right) =$$

$$\mu\left(\mathrm{argmax}_{j\in\{0\ldots L-1\}}\, h_{i,j} + Jp^e_{i,j} + \varepsilon_{i,j} = l \middle| h_{i,j}, p^e_{i,j}\ \forall j\right)$$

Assume that the random payoff terms $\varepsilon_{i,l}$ are iid and double exponentially distributed

$$\mu\left(\varepsilon_{i,l} \leq \varsigma\right) = \exp\left(-\exp\left(-\beta\varsigma + \gamma\right)\right)$$

where γ is Euler's constant.

Note that β indexes the dispersion in the individual-specific unobserved utility terms.

Choices will, under this density assumption, obey the canonical multinomial logit probability structure

$$\mu\left(\omega_i = l \middle| h_{i,j}, p^e_{i,j}\ \forall j\right) = \frac{\exp\left(\beta h_{i,l} + \beta J p^e_{i,l}\right)}{\sum_{j=0}^{L-1} \exp\left(\beta h_{i,j} + \beta J p^e_{i,j}\right)}$$

so the joint probabilities for all choices may be written as

$$\mu\left(\omega_1 = l_1, \ldots, \omega_I = l_I \middle| h_{i,j}, p^e_{i,j}\ \forall i, j\right) =$$

$$\prod_i \frac{\exp\left(\beta h_{i,l_i} + \beta J p^e_{i,l_i}\right)}{\sum_{j=0}^{L-1} \exp\left(\beta h_{i,j} + \beta J p^e_{i,j}\right)}$$

Self-Consistency of Beliefs

Self-consistent beliefs imply that the subjective choice probabilities p_l^e equal the objective expected values of the percentage of agents in the group who choose l, p_l, the structure of the model implies that

$$p_{i,l}^e = p_l = \int \frac{\exp\left(\beta h_{i,l} + \beta J p_l\right)}{\sum_{j=0}^{L-1} \exp\left(\beta h_{i,j} + \beta J p_j\right)} dF_h$$

where F_h is the empirical probability distribution for the vector of deterministic terms $h_{i,l}$.

It is obvious that under the Brouwer fixed point theorem, at least one such fixed point exists, so this model always has at least one equilibrium set of self-consistent aggregate choice probabilities.

Characterizing Equilibria

To understand the properties of this model, it is useful to focus on the special case where $h_{i,l} = 0 \ \forall i, l$. For this special case, the choice probabilities (and hence the expected distribution of choices within a group) are completely determined by the compound parameter βJ.

An important question is whether and how the presence of interdependencies produces multiple equilibria for the choice probabilities in a group.

In order to develop some intuition as to why the number of equilibria is connected to the magnitude of βJ, it is helpful to consider two extreme cases for the compound parameter, namely $\beta J = 0$ and $\beta J = \infty$.

For the case $\beta J = 0$, one can immediately verify that there exists a unique equilibrium for the aggregate choice probabilities such that $p_l = \frac{1}{L}\ \forall l$. This follows from the fact that under the assumption that all individual heterogeneity in choices come from the realizations of $\varepsilon_{i,l}$, a process whose elements are independent and identically distributed across choices and individuals. Since all agents are ex ante identical, the aggregate choice probabilities must be equal.

The case $\beta J = \infty$ is more complicated. The set of aggregate choice probabilities $p_l = \frac{1}{L}$ is also an equilibrium if since conditional on these probabilities, the symmetries in payoffs associated with each choice that led to this equilibrium when $\beta J = 0$ are preserved as there is no difference in the social component of payoffs across choices.

However, this is not the only equilibrium. To see why this is so, observe that for any pair of choices l and l' for which the aggregate choice probabilities are nonzero, it must be the case that

$$\frac{p_l}{p_{l'}} = \frac{\exp(\beta J p_l)}{\exp(\beta J p_{l'})}$$

for any βJ. This follows from the fact that each agent is ex ante identical. Thus, it is immediate that any set of equilibrium probabilities that are bounded away from 0 will become equal as $\beta J \Rightarrow \infty$.

This condition is necessary as well as sufficient, so any configuration such that $p_l = \frac{1}{b}$ for some subset of b choices and $p_l = 0$ for the other $L - b$ choices is an equilibrium. Hence, for the case where $J = \infty$, there exist

$$\sum_{b=1}^{L} \binom{L}{b} = 2^L - 1$$

different equilibrium probability configurations.

Recalling that β indexes the density of random utility and J measures the strength of interdependence between decisions, this case, when contrasted with $\beta J = 0$ illustrates why the strength of these interdependences and the degree of heterogeneity in random utility interact to determine the number of equilibria.

These extreme cases may be refined to produce a more precise characterization of the relationship between the number of equilibria and the value of βJ.

Theorem. Multiple equilibria in the multinomial logit model with social interactions

For the multinomial logit model with social interactions, assume that $h_{i,l} = k \; \forall i, l$. Then there will exist at least three self-consistent choice probabilities if $\frac{\beta J}{L} > 1$.

Multinomial Choice under Alternative Error Assumptions

The basic logic of the multinomial model is straightforward to generalize. This can be seen if one considers the preference structure

$$V_{i,l} = h_{i,l} + Jp^{e}_{i,l} + \beta^{-1}\varepsilon_{i,l}$$

This is the same preference structure we worked with earlier, except that β is now explicitly used to index the intensity of choice (in the McFadden sense) rather than as a parameter of the distribution of the random payoff term $\varepsilon_{i,l}$. We assume that these unobserved utility terms are independent and identically distributed with a common distribution function $F_{\varepsilon}(\cdot)$.

The probability that agent *i* makes choice *l* is

$$\mu\begin{pmatrix} \varepsilon_{i,0} - \varepsilon_{i,l} \leq \beta\left(h_{i,l} - h_{i,0}\right) + \beta J\left(p^e_{i,l} - p^e_{i,0}\right),..., \\ \varepsilon_{i,L-1} - \varepsilon_{i,l} \leq \beta\left(h_{i,l} - h_{i,L-1}\right) + \beta J\left(p^e_{i,l} - p^e_{i,L-1}\right) \end{pmatrix}$$

Conditional on a realization of $\varepsilon_{i,l}$, the probability that l is chosen is

$$\prod_{j \neq i} F_{\varepsilon}\left(\beta h_{i,l} - \beta h_{i,j} + \beta J p^{e}_{i,l} - \beta J p^{e}_{i,j} + \varepsilon_{i,l}\right)$$

which immediately implies that the probability of the choice l without conditioning on the realization of $\varepsilon_{i,l}$ is

$$p_{i,l} = \int \prod_{j \neq l} F_{\varepsilon}\left(\beta h_{i,l} - \beta h_{i,j} + \beta J p^{e}_{i,l} - \beta J p^{e}_{i,j} + \varepsilon\right) dF_{\varepsilon}\,.$$

This is a multinomial choice model whose structure is fully analogous to the multinomial logit structure developed under parametric assumptions.

Under self-consistency, the aggregate choice probabilities of this general multinomial choice model are the solutions to

$$p_l = \int\int\prod_{j\neq l} F_{\varepsilon}\left(\beta h_l - \beta h_j + \beta J p_l - \beta J p_j + \varepsilon\right) dF_{\varepsilon} dF_h$$

As in the multinomial logit case, the compound parameter βJ plays a critical role in determining the number of self-consistent equilibrium choice probabilities p_l. This finding is formalized in.

Theorem. Multiple equilibria in the multinomial choice model with social interactions

For the multinomial choice model with social interactions, assume that $h_{i,l} = h \quad \forall i,l$ and $\varepsilon_{i,l}$ are independent across i and l. There exists a threshold T such that if $\beta J < T$, then there is a unique self-consistent equilibrium, whereas if $\beta J > T$ there exist at least three self-consistent equilibria.

The relationship between βJ and the number of equilibria is less precise than was found in the multinomial logit case, as the theorem does not specify anything about the way in which L, the number of available choices, affects the number of equilibria.

This lack of precision is to be expected since we cannot exploit the error distribution structure.

Where is Segregation?

Segregation will be manifested in differences in $F_{h,g}$ across groups. Note that this implicitly involves the distribution of the X_i's, which can include indicators for race, measures of income. Note that in general, one expects X_i to affect Y_g; much of the social interactions literature as assumed that Y_g is the average of the within-group X_i's

A question that has not been studied, at least using the framework I have set up, is the relationship between different degrees of segregation on various elements of X and the number of equilibria.

What is known is the following. If one segregates individuals so that the "fundamentals" are weak (X's near 0), socially inefficient equilibria can arise.

Further, if one integrates positive and negative X's so as to reduce associated Y_g values, then one can move communities from unique equilibria into the multiple equilibrium range.

Identification

Suppose that

1. $\alpha_g = 0 \ \forall g$

2. Groups are exogenously determined.

Under these assumptions

Theorem. Identification for the multinomial choice model with social interactions

Under the normalization $k_0 = 0$, $c_0 = 0$, $d_0 = 0$, $J_0 = 0$ and β=1, if

1. the joint support of $X_i, Y_{g(i)}$ is not contained in a proper linear subspace of R^{r+s}
2. the support of $Y_{g(i)}$ is not contained in a proper linear subspace of R^s,
3. no linear combination of elements of X_i and $Y_{g(i)}$ is constant,
4. for each choice l, there exists at least one group g_l such that conditional on Y_{g_l}, X_i is not contained in a proper linear subspace of R^r,
5. none of the elements of $Y_{g(i)}$ possesses bounded support,
6. $p_{g(i),l}$ is not constant across groups,
7. $\varepsilon_{i,l}$, the random utility terms for each individual, are independent of his associated X_i and $Y_{g(i)}$ and independent and identically distributed across choices and individuals.

the model parameters $\left(k_1, c_1, d_1, J_1, \ldots, k_{L-1}, c_{L-1}, d_{L-1}, J_{L-1}\right)$ is identified relative to any distinct alternative.

What the theorem in essence requires is three things.

First, there must be sufficient intragroup variation within at least one group to ensure that c_l is identified $\forall l$.

Second, there must be enough intergroup variation in $Y_{g(i)}$ to ensure that c_l, d_l and J_l are identified $\forall l$ because of the nonlinear relationship between contextual effects and endogenous effects.

Third, there cannot be collinearity between the regressors contained in X_i and $Y_{g(i)}$, so that individual and contextual effects may be distinguished.

Simultaneity/Reflection Problem

This theorem implies that the so-called reflection problem does not arise. The reason is that the system is nonlinear.

On the other hand, it does assume something about network structure within groups: unweighted averaging. Work with Blume, Brock, and Jayaraman addresses identification for linear models when network structure is only partially observed. No work yet on discrete choice case.

What happens when assumptions 1 and 2 are relaxed?

Unobserved Group Effects

The import of unobserved group effects depends on the nature of the data available. With panel data one can implement variants of differencing to address in discrete choice contexts.

However, except for binary choice case, this has not been worked out formally for social interactions models.

Partial Identification.

For binary choice models, one can develop evidence of social interactions for cross-section data even in the presence of group-level fixed effects even if one is restricted to cross section data. (Multinomial choice has yet to be studied.)

The reason why cross-section data on binary choices may produce evidence in support or against social interactions is that the binary choice model can produce multiple equilibria only if endogenous social interaction effects are present. If the available data require the existence of multiple equilibria, this in turn implies the existence of endogenous social interactions.

To develop this argument, we assume that there is random assignment of individuals across groups

$$F_{X|g} = F_X .$$

Even for this case, the translation of multiple equilibria into data restrictions is somewhat complicated.

An intuition as to why multiple equilibria are associated with endogenous social interactions is that the multiple equilibria can produce what Brock and Durlauf refer to as pattern reversals.

Assume that $d > 0$ so that increasing any element in Y_g increases, other things equal, the probability that an individual in g chooses 1. One can always measure the elements of Y_g this way, so long as one knows the direction of the effects of its elements).

A pattern reversal occurs for groups g and g' if

$$Y_g < Y_{g'} \text{ and } m_g > m_{g'}.$$

Recall that m_g can be computed, since it is the conditional expectation of the same average of within-group choices $\bar{\omega}_g$, so pattern reversals represent restrictions on data.

For the identification of social interactions, pattern reversals are important because they may derive from the presence of endogenous social interactions producing multiple equilibria.

Why? Intuitively, multiple equilibria can produce a pattern reversal because group g can coordinate on a high m_g equilibrium whereas group g' does not so that the effect of the higher value of Y on the average outcome in the group is negated.

The difficulty with using this heuristic argument is that without any restrictions on α_g, pattern reversals can occur without multiple equilibria being present.

Brock and Durlauf (2007) attempt to identify weak restrictions associated with α_g such that pattern reversals imply the existence of multiple equilibria and hence endogenous social interactions.

This type of argument does not identify the value of the endogenous social interactions parameter J, rather it shows that the value is nonzero and large enough to produce multiple equilibria.

As such, it is a form of partial identification.

What sorts of assumptions allow for partial identification of J via pattern reversals?

One assumption is a stochastic monotonicity restriction on the group level unobservables. Suppose that if $Y_g > Y_{g'}$, then the conditional distribution of unobservables in g', $F_{\alpha_{g'}|Y_{g'}}$, is first order stochastically dominated by $F_{\alpha_g|Y_g}$.

In this case a pattern reversal will imply that endogenous social interactions exist.

Another route towards partial identification of social interactions is via unimodality versus multimodality comparisons.

Suppose that Y_g is constant across groups, X_i is constant across all individuals within and across groups and that $\alpha_g = 0$. In this case, it is easy to see that m_g will take on a single value when there are no endogenous social interactions and will take one of a finite set of values when there are multiple equilibria due to social interactions.

In this case, m_g will be multimodal, with each equilibrium representing a possible value. This leads to the intuition that multiple equilibria may occur when one relaxes the assumption that Y_g and X_i are constant.

The translation of this intuition into data restrictions turns out to be fairly hard.

One reason for this is straightforward: even if α_g exhibits multimodality, then there is no link between multiple equilibria and unimodality of the other variables.

The reason for this that the relationship between m_g and Y_g is nonlinear and this nonlinearity can induce multimodality.

Brock and Durlauf (2007) overcome this problem by considering $dF_{Y_g|m_g}$ rather than $dF_{m_g|Y_g}$.

Specifically, this paper shows that unimodality of $dF_{\alpha_g|Y_g}$ implies that there must exist a vector π such that

$$dF_{\pi Y_g|m_g} \text{ is unimodal}$$

if there are no endogenous social interactions.

Frontier: Social Interactions and Self-Selection

One does not think of families as being randomly allocated across neighborhoods; rather, families choose neighborhoods subject to constraints such as rent levels and personal income.

For environments in which self-selection is present, the consistency of various statistical methods for estimating social interactions may be affected.

Specifically, the presence of self-selection can mean that the expected value of the random term ε_i, conditional on the individual's characteristics and group memberships, may no longer be zero.

If one ignores self-selection in estimation, then it is obvious how one can produce spurious evidence of social interactions.

For example, if poorer neighborhoods tend to contain relatively less ambitious parents than affluent neighborhoods, and if lack of ambition leads to lower educational performance by children, then the failure to account for this self-selection could lead to the false conclusion that poor neighborhoods causally affect education.

Needed: more theory and more econometrics.

Comment: control function approach needs to be developed for discrete choice with social interactions environments in ways analogous to what has been done for linear models.

A Nested Choice Approach to Integration of Behaviors and Group Memberships

Group memberships may be developed using the nested logit framework The basic idea of this framework is the following. An individual is assumed to make a joint decision of a group $g \in \{0,...G-1\}$ and a behavior $l \in \{0,...L-1\}$. We will denote the group choice of i as δ_i.

The structure of this joint decision is nested in the sense that the choices are assumed to have a structure that allows one to decompose the decisions as occurring in two stages: first, the group is chosen and then the behavior.

The key feature of this type of model is the assumption that choices at each stage obey a multinomial logit probability structure. For the behavioral choice, this means that

$$\mu\left(\omega_i = l \,\middle|\, h_{i,l,g}, p^e_{i,l,g}, \delta_i = g\right) = \frac{\exp\beta\left(h_{i,l,g} + Jp^e_{i,l,g}\right)}{\sum_{j=0}^{L-1}\exp\beta\left(h_{i,l,g} + Jp^e_{i,l,g}\right)}$$

which is the same behavioral specification as before. Group membership choices are somewhat more complicated. In the nested logit model, group choices are assumed to obey

$$\mu\left(i \in g \,\middle|\, h_{i,l,g}, p^e_{i,l,g} \forall l, g\right) = \frac{\exp(\beta_g Z_{i,g})}{\sum_g \exp(\beta_g Z_{i,g})}$$

where

$$Z_{i,g} = E(\max_l h_{i,l,g} + Jp^e_{i,l,g} + \varepsilon_{i,l,g})$$

A standard result (e.g. Anderson, de Palma and Thisse (1992, pg. 46)) is that

$$E\left(\max\left(h_{i,l,g} + Jp^e_{i,l} + \varepsilon_{i,l,g} \left| h_{i,l,g}, p^e_{i,l,g} \forall l,g \right.\right)\right) = \beta^{-1}\log\left(\sum_l \exp\beta\left(h_{i,l,g} + Jp^e_{i,l,g}\right)\right)$$

Combining equations, the joint group membership and behavior probabilities for an individual are described by

$$\mu\left(\omega_i = l, \delta_i = g \,\middle|\, h_{i,l,g}, p^e_{i,l,g} \forall l, g\right) =$$

$$\frac{\exp\left(\beta_g \beta^{-1} \log\left(\sum_l \exp\beta\left(h_{i,l,g} + Jp^e_{i,l,g}\right)\right)\right)}{\sum_g \exp\left(\beta_g \beta^{-1} \log\left(\sum_l \exp\beta\left(h_{i,l,g} + Jp^e_{i,l,g}\right)\right)\right)} \frac{\exp\beta\left(h_{i,l,g} + Jp^e_{i,l,g}\right)}{\sum_{j=0}^{L-1} \exp\beta\left(h_{i,l,g} + Jp^e_{i,l,g}\right)}$$

Some variants of the choice side of this model have appeared in the literature, e.g. Bayer and Timmins (2007), but these have been special cases, which ignore two levels of choice, endogenous effects, etc.

This probabilistic description may be faulted in that it is not directly derived from a utility maximization problem. In fact, a number of papers have identified conditions under which the probability structure is consistent with utility maximization.

A simple condition that renders the model compatible with a well posed utility maximization problem is $\beta_g \leq \beta$, which in essence requires that the dispersion of random payoff terms across groups is lower than the dispersion in random payoff terms across behavioral choices within a group.

Comment: function form assumption relaxation is important for empirical work.

Next Steps in this Specification

The model assumes that agents know the values of $h_{i,l,g}$ are known to agents. However, these will depend on who is a member of each group, when the values of Y_g depend on with within-group distribution of X_i. Hence the fixed point problem is qualitatively different from the initial environment I described.

This raises a first question as to the specification of an equilibrium configuration of individuals across groups. In general, one needs to attach prices to the group memberships for existence to be possible. The intuitive problem is that two agents may "disagree" as to the desirability of one another as fellow group members.

Further, the mapping from X_i to Y_g may exhibit problems. This is natural if the groups are school districts and voting will determine public education investment. When a private school option exists, voters may not have single peaked preferences over tax rates, which means a voting equilibrium does not exist and hence Y_g .

Comment: this is an example of a weakness of much of the social interactions literature-mechanisms for social influences are *not* explicitly described.

Comment: theoretical urban economics is very hard because of the commonality of nonexistence results.

Unexplored Possibilities

1. With the exception of an unpublished thesis by Lars Nesheim, there has been no systematic investigation of the informational content of prices in identification of social effects. Work by Heckman with Ekeland, Matzkin, and Nesheim on identification in hedonic prices models suggests that this is a natural avenue to explore.

2. Configurations themselves contain information. Consider the Becker discrimination model. It is possible that the wage premium between blacks and whites is zero, yet information on discriminatory preferences of some employers could be obtained from excess segregation of employees across firms.