baseline case. We assume that the time scale of the duration of interest is short relative to the time scale over which data are collected, so that all duration “spells” are completed. Formally, this assumption means that there is no “right censoring” of the data. This will not be appropriate for out-of-wedlock birth data, since of course not all unmarried females experience the event; nevertheless, the assumption is useful for exposition. Let \( t_i \) denote the time of first birth for individual \( i \). If this timing is associated with probability density \( f(\cdot) \), then the joint density for the \( I \) times is \( \prod_i f(t_i) \). This joint probability will be determined by the individual hazards \( \lambda_i \).

As before, we assume that the hazard for each individual under analysis depends on individual characteristics \( X_i \), neighborhood characteristics \( Y_{n(i)} \) and an expected neighborhood behavioral measure \( m^e_{n(i)} \). In this context, \( m^e_{n(i)} \) may be the expected value of either the within-neighborhood duration or the median group duration. We therefore assume that for each individual

\[
\lambda_i = \lambda(X_i; Y_{n(i)}; m^e_{n(i)})
\]

(94)

so that the associated density for the duration is

\[
f(t | X_i; Y_{n(i)}; m^e_{n(i)}) = \lambda(X_i; Y_{n(i)}; m^e_{n(i)}) \exp(-\lambda(X_i; Y_{n(i)}; m^e_{n(i)})t)
\]

(95)

The expected duration for individual \( i \), conditional on these controls is

\[
E(t | X_i; Y_{n(i)}; m^e_{n(i)}) = \lambda(X_i; Y_{n(i)}; m^e_{n(i)})^{-1}
\]

(96)

and the median of the duration is given by the solution \( t^* \) to \( F(t^*) = \frac{1}{2} \) which implies that

\[
\frac{1}{2} = 1 - F(t^*) = \exp(-\lambda(X_i; Y_{n(i)}; m^e_{n(i)})t^*)
\]

(97)
which implies that $t^*$ solves

$$
\log 2 = \lambda(X_i; Y_{n(i)}, m_{n(i)}) t^*
$$

(98)

Equations (97) and (98) allow us to define self-consistent solutions for this model. Self-consistency with respect to expected duration times requires that

$$
m^e_{n(i)} = m_{n(i)} = \int \lambda(X_i; Y_{n(i)}, m_{n(i)})^{-1} dF_{\hat{X}}
$$

(99)

where as before $F_{\hat{X}}$ is the probability distribution of characteristics within neighborhood $n(i)$. Similarly, self-consistency with respect to the neighborhood median requires that

$$
m^e_{n(i)} = m_{n(i)} = \log 2 \int \lambda(X_i; Y_{n(i)}, m_{n(i)})^{-1} dF_{\hat{X}}.
$$

(100)

These two expressions only differ by a constant of proportionality. Notice that we have assumed that each individual references on her entire neighborhood. It is possible to consider cases where the reference group is smaller, so that for example, one only references on individuals with similar individual characteristics. As we have already seen in the discussion of other models, the “width” of each individual’s reference group plays a key role in identification.

We first consider identification in the parametric case under the assumption that the expected value of the duration time within a group is the relevant endogenous interaction. Following treatments such as Amemiya (1985) section 11.2.3, we assume that the hazard function for individual $i$ is exponential, so that

$$
\lambda_i = \exp(c'X_i + d'Y_{n(i)} + Jm^e_{n(i)}).
$$

(101)

We assume that $X_i$ contains a constant term (Amemiya (1985) eq. 11.2.26). The
associated likelihood function for the data will therefore be

\[ L = \prod_i \exp(c' \mathbf{X}_i + d' \mathbf{Y}_{n(i)} + J M_{n(i)}^e) \exp(- \exp(c' \mathbf{X}_i + d' \mathbf{Y}_{n(i)} + J M_{n(i)}^e) t_i). \tag{102} \]

For this model, choosing parameter estimates for \( c, d, \) and \( J \) to maximize (102) without imposing eq. (99) corresponds to the naive estimator we have described in the binary choice and linear-in-means cases.

Following the analysis in Amemiya (1985), identification in population requires that the expected value of the Hessian matrix of \( \log L \) is nonsingular at the self-consistent solution (99). Letting \( b = (c', d', J)' \) and \( M_i^e = (X_i', Y_{n(i)}', M_{n(i)}^e)' \), the expected value of the Hessian equals

\[ \mathbb{E}\left( \frac{\partial^2 \log L}{\partial b \partial b'} \mid M_i \right) = - \mathbb{E}\left( \sum_i t_i \exp(b'M_i) M_i M_i' \mid M_i \right). \tag{103} \]

Further, since \( \mathbb{E}(t_i \mid M_i) = \lambda_i^{-1} \), it is further the case that

\[ - \mathbb{E}\left( \sum_i t_i \exp(b'M_i) M_i M_i' \mid M_i \right) = - \sum_i M_i M_i'. \tag{104} \]

which means, dividing both sides by \( I \), that identification asymptotically depends on the linear independence of the controls which constitute \( M_i \). This is the same condition which appeared in both the binary choice and linear-in-means. However, if \( m_{n(i)} \) is the within-group mean, then by eq. (99) \( m_{n(i)} \) is a nonlinear function of \( X_{n(i)} \) and \( F_X \).

A nonlinear relationship of this type is the key condition for global identification in the binary choice model. Extensions of the analysis for binary choice may be made to the exponential hazard model as well as other parametric cases such as the Weibull, log normal and log-logistic distributions and can further be done for cases such as right censoring or for Cox’s partial likelihood approach. A formal characterization of the conditions for identification in these various cases is left for future work. In particular, we believe that analogous conditions for local
identification to those found in Theorem 7 can be developed for the current case.

iii. Nonparametric approaches

a. treatment effects

Our discussion of identification has assumed that a researcher possesses prior information concerning the form of the model under study, so that estimation occurs with respect to a finite set of parameters. In our context, this information has taken the form of both the functional form for individual behavior and, where selection into neighborhoods is an issue, the rules for neighborhood self-selection. Dissatisfaction with the assumption of such strong prior information has led to a vast literature on semi- and non-parametric approaches to estimation. In the context of interactions-based models, one can think of a nonparametric approach to estimation in the context of identifying a role for neighborhood characteristics on individual behavior while making relatively weak assumptions on the functional forms describing individual behavior. In turn, one can think of this question as analogous to the nonparametric identification of treatment effects, where group influences are the "treatment" whose effect we wish to uncover.

In this section, we develop approaches to both point and interval identification of interaction effects under substantially weaker modelling assumptions than we have employed thus far. First, following Heckman (1997), we show how the assumption that neighborhood interaction effects act as a "shifted outcome effect" combined with an exclusion restriction on the determinants of neighborhood membership, can lead to identification of an interaction effect. Second, following Manski (1995) and Manski and Pepper (1998), we show how one may relax this exclusion restriction and nevertheless
obtain an upper bound on the interaction effect.

To make our analysis concrete, suppose that, following the work of Steinberg (1996) we are interested in determining whether a peer group of "brains" (denoted as group 1) versus a peer group of "nonbrains" (denoted as group 0) affects individual student performance. Observations are available from $G$ different schools, each of which contains students who are members of each such group. The variable $\xi_{i,g}$ tracks the group of individual $i$ in school $g$. The goal of the exercise is to determine the effect of membership in group 1 versus group 0 on a continuous outcome variable, $\omega_{i,g,\xi}$. Notice that we index according to both school and group. Membership in the brains groups is therefore our "treatment" and so we wish to measure the treatment effect. We let $X_{i,g}$ denote those observable individual variables which directly determine $\omega_{i,g,\xi}$ and $R_{i,g}$ denote those observable variables which determine whether $i$ is a member of group 1. We refer to the average behaviors in the two groups as $m_{g,0}$ and $m_{g,1}$ with $m_g = (m_{g,0}, m_{g,1})$. We have included $g$ in the subscripts so that each observation refers to both an individual and the school which he attends.

For a given individual $i$, we assume that $\omega_{i,g,\xi}$ obeys

$$\omega_{i,g,\xi} = \phi(\xi_{i,g}, X_{i,g}, R_{i,g}, m_g) + \epsilon_{i,g}(\xi_{i,g})$$

(105)

for some function $\phi(\cdot, \cdot, \cdot, \cdot)$ where

$$E(\epsilon_{i,g}(\xi_{i,g}) \mid \xi_{i,g}, X_{i,g}, R_{i,g}, m_g) = 0.$$  

(106)

The identification question therefore refers to what can be learned about $\Delta_{i,g} = \omega_{i,g,1} - \omega_{i,g,0}$. Following Heckman (1997), one is typically interested in

$$E(\Delta_{i,g} \mid X_{i,g}, R_{i,g}, m_g) = \phi(1, X_{i,g}, R_{i,g}, m_g) - \phi(0, X_{i,g}, R_{i,g}, m_g)$$

(107)

where the equality follows immediately from (105) and (106). This is the
expected value of the treatment for an individual with characteristics $X_{i,g}$ and $R_{i,g}$ in school $g$ and represents the object which we wish to estimate. A distinct quantity of interest is $E(\Delta_{i,g} \mid X_{i,g}, R_{i,g}, \xi_{i,g}, \varepsilon_{i,g} = 1)$ which Heckman (1997) refers to as the effect of the “treatment on the treated for persons with characteristics” $X_{i,g}$ and $R_{i,g}$. Notice that the selection problem holds because there is information about $\varepsilon_{i,g}(0)$ and $\varepsilon_{i,g}(1)$ when the treatment, i.e. group membership, is a choice variable.

In order to identify $E(\Delta_{i,g} \mid X_{i,g}, R_{i,g}, m_g)$, one proceeds as follows. First, assume that the effect of group membership is additive, so that

$$\omega_{i,g,1} - \omega_{i,g,0} = k(X_{i,g}, m_g)$$  \hspace{1cm} (108)

for some function $k(\cdot, \cdot)$ which means that

$$k(X_{i,g}, m_g) = E(\Delta_{i,g} \mid X_{i,g}, m_g)$$  \hspace{1cm} (109)

Eq. (108) is often referred to as a shifted outcome assumption. Notice that this is a minor generalization of Heckman (1997), although not Heckman and Robb (1985), in that we allow the $k$'s to vary with respect $X_{i,g}$ and $m_g$, which is natural if one thinks the treatment effect varies across individuals.

Next, consider what is estimable from the data. The group means $m_g$ are of course observable. Further, one can estimate the conditional expectations of behavior for individual given their group memberships, i.e.

$$E(\omega_{i,g,0} \mid X_{i,g}, R_{i,g}, m_g, \xi_{i,g} = 0)$$  \hspace{1cm} (110)

and

$$E(\omega_{i,g,1} \mid X_{i,g}, R_{i,g}, m_g, \xi_{i,g} = 1)$$  \hspace{1cm} (111)
The identification of an endogenous interaction effect can be thought of as requiring that one can move from these conditional expectations to 
\[ E(\omega_{i,g,0} | \tilde{X}_{i,g},\tilde{R}_{i,g},\tilde{m}_{g}) \] and 
\[ E(\omega_{i,g,1} | \tilde{X}_{i,g},\tilde{R}_{i,g},\tilde{m}_{g}) \]. To do this, it is necessary to be somewhat more careful about the process of group formation. We therefore assume that individuals join groups at least partially on the basis of the expected behavior in the groups, and that these expected behaviors are rational. This is nothing more than the self-consistency idea we have used throughout.

We now can consider the estimation of \( k(X_{i,g},\tilde{m}_{g}) \). Letting 
\[ \mu(\xi_{i,g} | X_{i,g},\tilde{R}_{i,g},\tilde{m}_{g}) \] denote the conditional probability of group membership, it is immediate that

\[
E(\omega_{i,g,0} | \tilde{X}_{i,g},\tilde{R}_{i,g},\tilde{m}_{g}) =
\]
\[
E(\omega_{i,g,0} | \tilde{X}_{i,g},\tilde{R}_{i,g},\tilde{m}_{g},\xi_{i,g} = 0)\mu(\xi_{i,g} = 0 | \tilde{X}_{i,g},\tilde{R}_{i,g},\tilde{m}_{g}) +
\]
\[
E(\omega_{i,g,0} | \tilde{X}_{i,g},\tilde{R}_{i,g},\tilde{m}_{g},\xi_{i,g} = 1)\mu(\xi_{i,g} = 1 | \tilde{X}_{i,g},\tilde{R}_{i,g},\tilde{m}_{g})
\] (112)

and

\[
E(\omega_{i,g,1} | \tilde{X}_{i,g},\tilde{R}_{i,g},\tilde{m}_{g}) =
\]
\[
E(\omega_{i,g,1} | \tilde{X}_{i,g},\tilde{R}_{i,g},\tilde{m}_{g},\xi_{i,g} = 0)\mu(\xi_{i,g} = 0 | \tilde{X}_{i,g},\tilde{R}_{i,g},\tilde{m}_{g}) +
\]
\[
E(\omega_{i,g,1} | \tilde{X}_{i,g},\tilde{R}_{i,g},\tilde{m}_{g},\xi_{i,g} = 1)\mu(\xi_{i,g} = 1 | \tilde{X}_{i,g},\tilde{R}_{i,g},\tilde{m}_{g})
\] (113)

The right hand terms \( E(\omega_{i,g,0} | \tilde{X}_{i,g},\tilde{R}_{i,g},\tilde{m}_{g},\xi_{i,g} = 1) \) and 
\( E(\omega_{i,g,1} | \tilde{X}_{i,g},\tilde{R}_{i,g},\tilde{m}_{g},\xi_{i,g} = 0) \) are not observed since they refer to conditional expectations of behavior for individuals were they members of groups which they did not select into. Hence identification will only occur if some additional assumption overcomes this.
One such assumption is an exclusion restriction with respect to the variables which affect selection into groups versus variables which affect behavior once one is a member of a given the group. Formally, we need the following. For every set of pairs \( X_{i, g} \) and \( R_{i, g} \)

\[
E(\omega_{i, g, 0} | X_{i, g}, R_{i, g}, m_{g}) = E(\omega_{i, g, 0} | X'_{i, g}, R'_{i, g}, m_{g})
\]  

(114)

and

\[
E(\omega_{i, g, 1} | X_{i, g}, R_{i, g}, m_{g}) = E(\omega_{i, g, 1} | X'_{i, g}, R'_{i, g}, m_{g})
\]  

(115)

What this means is that there is a variable which affects selection but not expected behavior for each individual once that person is a group member.

Following Heckman (1997) and Manski (1995) p. 144, the shifted outcome restriction (108) and the exclusion restriction described by eqs. (114) and (115) can be combined to conclude that

\[
E(\omega_{i, g, 1} | X_{i, g}, R_{i, g}, m_{g}, \xi_{i, g} = 1) \mu(\xi_{i, g} = 1 | X_{i, g}, R_{i, g}, m_{g}) +
\]

\[
E(\omega_{i, g, 0} | X_{i, g}, R_{i, g}, m_{g}, \xi_{i, g} = 0) \mu(\xi_{i, g} = 0 | X_{i, g}, R_{i, g}, m_{g}) +
\]

\[
k(X_{i, g}, m_{g}) \mu(\xi_{i, g} = 0 | X_{i, g}, R_{i, g}, m_{g}) =
\]

\[
E(\omega_{i, g, 1} | X_{i, g}, R'_{i, g}, m_{g}, \xi_{i, g} = 1) \mu(\xi_{i, g} = 1 | X_{i, g}, R'_{i, g}, m_{g}) +
\]

\[
E(\omega_{i, g, 0} | X_{i, g}, R'_{i, g}, m_{g}, \xi_{i, g} = 0) \mu(\xi_{i, g} = 0 | X_{i, g}, R'_{i, g}, m_{g}) +
\]

\[
k(X_{i, g}, m_{g}) \mu(\xi_{i, g} = 0 | X_{i, g}, R'_{i, g}, m_{g}).
\]  

(116)

Other than \( k(X_{i, g}, m_{g}) \), each of the terms in this expression can be estimated.
nonparametrically, and so \( k(\tilde{X}_{i,g}, m_g) \) is identified.

**Theorem 8. Nonparametric identification of endogenous interaction effect**

In the presence of self-consistent expectations, shifted outcomes of the form eq. (108), and an exclusion restriction of the form eqs. (114) and (115), the interaction effect is identified.

Two features of this result are worth noting. First, under Theorem 8, one can estimate \( \mathbb{E}(k(\tilde{X}_{i,g}, m_g) | m_g) \) and test for the average interaction effect in a population. Further, if one is willing to assume that

\[
k(\tilde{X}_{i,g}, m_g) = J_g(m_{g,1} - m_{g,0})
\]

then the interaction parameter \( J_g \) for each school may be identified. In principle, cross-school variation in \( J_g \) could be employed to study the determinants of the strength of interactions.

Second, while Theorem 8 makes some progress in terms of relaxing the parametric assumptions of the interactions model, it is still strong in terms of the underlying behavioral assumptions. As stated by Heckman (1997),

"Any valid application of the method of instrumental variables for estimating these treatment effects in the case where the response to treatment varies among persons requires a behavioral assumption about how persons make their decisions about program participation. This issue cannot be settled by a statistical analysis." (pg. 449)

In our context, the treatment is the membership in group 1 rather than group 0 and the instrument is characterized by eqs. (114) and (115).

One approach to weakening the exclusion restriction on instruments is due to Manski and Pepper (1998). Following their analysis, we replace our previous

69
assumptions with inequality restrictions. First, we replace our assumption of a shifted outcome variable, eq. (108) with

$$\omega_{i,g} \leq k(X_{i,g}, m_g). \quad (118)$$

This assumption means that the effect of shifting a person with individual characteristics $X_{i,g}$ from group 0 to group 1 is bounded from above by $k(X_{i,g}, m_g)$. Second, we assume that a monotonic increase in the selection variables $R_{i,g}$ never decreases the expected outcome for an individual within a given group. Formally, if $R_{i,g} \geq R_{i,g}'$, then

$$E(\omega_{i,g}, \xi | X_{i,g}, R_{i,g}, m_g) \leq E(\omega_{i,g}, \xi | X_{i,g}, R_{i,g}', m_g), \xi = 0, 1 \quad (119)$$

This assumption relaxes eqs. (114) and (115) in that a monotonic increase from $R_{i,g}$ to $R_{i,g}'$ may have an effect on the conditional expectation of $\omega_{i,g}, \xi$, but this effect’s sign must not be negative.

Under these assumptions, we have

$$E(\omega_{i,g}, 1 | X_{i,g}, R_{i,g}, m_g, \xi_{i,g} = 1) \mu(\xi_{i,g} = 1 | X_{i,g}, R_{i,g}, m_g) +$$

$$E(\omega_{i,g}, 0 | X_{i,g}, R_{i,g}, m_g, \xi_{i,g} = 0) \mu(\xi_{i,g} = 0 | X_{i,g}, R_{i,g}, m_g) +$$

$$k(X_{i,g}, m_g) \mu(\xi_{i,g} = 0 | X_{i,g}, R_{i,g}, m_g) \leq$$

$$E(\omega_{i,g}, 1 | X_{i,g}, R_{i,g}', m_g, \xi_{i,g} = 1) \mu(\xi_{i,g} = 1 | X_{i,g}, R_{i,g}', m_g) +$$

$$E(\omega_{i,g}, 0 | X_{i,g}, R_{i,g}', m_g, \xi_{i,g} = 0) \mu(\xi_{i,g} = 0 | X_{i,g}, R_{i,g}', m_g) +$$

$$k(X_{i,g}, m_g) \mu(\xi_{i,g} = 0 | X_{i,g}, R_{i,g}', m_g). \quad (120)$$
We may now consider the quantity, \( Q(X_{i,g}, R_{i,g}, m_g) \) defined as

\[
Q(X_{i,g}, R_{i,g}, m_g) = 
\]

\[
E(\omega_{i,g} | X_{i,g}, R_{i,g}, m_g, \xi_{i,g} = 1) \mu(\xi_{i,g} = 1 | X_{i,g}, R_{i,g}, m_g) + 
\]

\[
E(\omega_{i,g} | X_{i,g}, R_{i,g}, m_g, \xi_{i,g} = 0) \mu(\xi_{i,g} = 0 | X_{i,g}, R_{i,g}, m_g)
\] (121)

This term is an observable analog of the expected outcome of an individual with observed characteristics \( X_{i,g}, R_{i,g}, \) and \( m_g \). Inequality (118) implies that

\[
Q(X_{i,g}, R_{i,g}, m_g) + k(X_{i,g}, m_g) \mu(\xi_{i,g} = 0 | X_{i,g}, R_{i,g}, m_g) \geq 
\]

\[
Q(X_{i,g}, R_{i,g}, m_g) + k(X_{i,g}, m_g) \mu(\xi_{i,g} = 0 | X_{i,g}, R_{i,g}, m_g)
\] (122)

which may be rewritten as

\[
Q(X_{i,g}, R_{i,g}, m_g) - Q(X_{i,g}, R_{i,g}, m_g) \geq 
\]

\[
k(X_{i,g}, m_g)(\mu(\xi_{i,g} = 0 | X_{i,g}, R_{i,g}, m_g) - \mu(\xi_{i,g} = 0 | X_{i,g}, R_{i,g}, m_g))
\] (123)

So long as

\[
Q(X_{i,g}, R_{i,g}, m_g) - Q(X_{i,g}, R_{i,g}, m_g) > 0
\] (124)

and

\[
\mu(\xi_{i,g} = 0 | X_{i,g}, R_{i,g}, m_g) - \mu(\xi_{i,g} = 0 | X_{i,g}, R_{i,g}, m_g) > 0
\] (125)
one can construct an upper bound on $k(X_{i,g}^m g)$. Formulating the bound using the fact that

$$\mu(\xi_i, g = 0 \mid X_{i,g} R_{i,g}^m g) - \mu(\xi_i, g = 0 \mid X_{i,g} R_{i,g}^m g) =$$

$$\mu(\xi_i, g = 1 \mid X_{i,g} R_{i,g}^m g) - \mu(\xi_i, g = 1 \mid X_{i,g} R_{i,g}^m g)$$

we have Theorem 9.

**Theorem 9. Construction of upper bound on the endogenous interaction effect**

Assume that eqs. (118), (119), (124), and (125) hold. Then

$$k(X_{i,g}^m g) \leq$$

$$\frac{Q(X_{i,g} R_{i,g}^m g) - Q(X_{i,g} R_{i,g}^m g)}{\mu(\xi_i, g = 1 \mid X_{i,g} R_{i,g}^m g) - \mu(\xi_i, g = 1 \mid X_{i,g} R_{i,g}^m g)} \quad (126)$$

A weakness of this result is that when it comes to interaction effects, it is probably more interesting to obtain a lower bound, since the presence of such effects are controversial. Notice, however, that the Manski and Pepper approach does suggest a way of constructing such a lower bound. In order to do so, one would need to find a variable which possesses the features that an increase (decrease) in its level would 1) both increase (decrease) the probability of selection into group 1 and 2) decrease (increase) the expected outcome for an individual conditional on the variable. Introspection suggest that it may be difficult to find such a variable, although there may be contexts where it holds.

In contrast, the assumption that one can find a variable in which both effects move in the same direction seems relatively plausible. For example, Manski and Pepper consider the question of how to bound the effect of the returns
to schooling, using SAT as an instrument. It seems natural in this case to assume both that higher SAT’s make additional schooling more likely and higher SAT’s do not reduce the benefit of additional schooling if chosen.

The self-consistency conditions

\[ m_{\xi,g} = \int E(\omega_{i,g} | \xi, X, R, m_g) dF_X, R, g, \xi = 0, 1 \]  

(127)

(where following our previous convention, \(dF_X, R, g\) denotes the joint distribution of \(X\) and \(R\) in school \(g\)), play an essential role in determining the quality of the bound in (126). To see this, consider the extreme case where all individuals within a school \(g\) have identical values of \(X_{i,g}\) and \(R_{i,g}\), then the bound is undefined, since the numerator and denominator of (125) will each equal 0. Alternatively, suppose that the distribution functions of individual characteristics are identical across schools, so \(dF_X, R, g = dF_X, R, g\). If the solutions to the self-consistency conditions (127) are unique, then this implies expected average outcomes must also be identical, i.e. \(m_g = m_{g'}\). If \(k(X_{i,g}, m_g) = k(m_g)\), so that the bound does not depend on individual characteristics, then one can use cross-school information in that \(k(m_g)\) must be bounded by the \(inf\) of the upper bounds computed for each school in isolation. This type of argumentation seems a valuable area for future research.

At a minimum, this discussion illustrates two points. First, semi- and non-parametric approaches to inference can be adapted to achieve either point or interval identification of interaction effects. Second, the conditions required for identification require careful consideration of the underlying socioeconomic theories under analysis in order to identify appropriate instruments.

b. duration data
Identification can also be considered for nonparametric approaches to duration data. As compellingly demonstrated by Heckman and Singer (1984b), errors in the assumed form of the hazard function and in the "mixing distribution" (by which they mean the distribution of unobservables) can lead to wildly misleading estimates. These problems of course are also relevant when interaction effects may be present. As far as we know, the extension of the methods studied by Heckman and Singer and subsequent authors to models with interactions has yet to be studied and it is beyond the scope of this paper to do so. However, we do sketch a slight extension to one approach to nonparametric identification in duration models, due to Elbers and Ridder (1982), in order to illustrate how such argumentation can in principle proceed; the reader is advised to see Heckman and Singer (1984c) for an evaluation of alternative conditions for nonparametric identification in this context.

Following Elbers and Ridder, suppose that the hazard function may be written as

\[ \lambda_i(t) = a(t)h(M_i)\nu_i \]  \hspace{1cm} (128)

Relative to our original treatment of hazards, this incorporates two additional terms: \( a(t) \) which allows for duration dependence, and \( \nu_i \) which allows for unobserved heterogeneity. The \( \nu_i \)'s are assumed to be drawn from a common distribution \( F_\nu(\cdot) \) with associated density \( f_\nu(\cdot) \). Elbers and Ridder show that subject to appropriate regularity conditions, if \( F(t \mid M_i) \) is nondefective (which means that all spells are completed), then it is possible to identify \( a(\cdot), h(\cdot), \) and \( f_\nu(\cdot) \). They do this as follows.

Define the conditional survivor function for individual \( i \) as

\[ S(t, M_i, \nu_i) = \exp\left( \int_0^t a(r)h(M_i)\nu_i dr \right) = \exp(A(t)h(M_i)\nu_i) \]  \hspace{1cm} (129)

where \( A(t) = \int_0^t a(r)dr \) and the conditional survivor function

74
\[
S(t, M_i) = \int S(t, M_i; \nu) dF_\nu = \int \exp(\tau \nu) dF_\nu
\]

where \( \tau = A(t)h(M_i) \). The last term in (130) indicates how the conditional survivor function is the LaPlace transform of \( dF_\nu \). The analyst is assumed to observe a family of nondefective distribution functions \( G(t, M_i) = 1 - S(t, M_i) \) from which he wishes to recover \( a(\cdot), h(\cdot), \) and \( f_\nu(\cdot) \). Elbers and Ridder assume that 1) \( \nu \) is nonnegative with mean 1, 2) \( M_i \) lies in an open set in the \( k \)-dimensional reals for some \( k \), and 3) \( h(\cdot) \) is defined on this open set and is nonnegative, differentiable, and nonconstant on the set.

Working through the proof that these conditions allow for identification, reveals the following. First, if one differentiates \( G(t, M_i) \) with respect to \( t \), \( h(M_i) \) may be recovered regardless of whether a self-consistency condition like eq. (99) holds when \( m_{n(i)} \) is a component of \( M_i \). Second, Elbers and Ridder exploit the LaPlace transform relationship to obtain a differential equation by assuming, without loss of generality, that \( M_i \) is one-dimensional. This argument requires differentiability and nonconstancy of \( h(M_i) \).

In order to generalize this step to allow for endogenous interactions with a self-consistency condition such as eq. (99), the reference group for each individual \( i \) must be broad enough to allow differentiability with respect to a nontrivial subvector of \( M_i \). Specifically, one needs to be able to vary \( X_i \) and \( X_{n(i)} \) without \( m_{n(i)} \) varying. For example, suppose that within each neighborhood, all \( X_i \)'s are identical. In this case, the self-consistent average choice level is

\[
m_{n(i)} = E(t \mid X_i; Y_{n(i)}; m_{n(i)}) = \int tdG(t, X_i; X_{n(i)}; m_{n(i)}).
\]

Generically, eq. (131) will have only a finite number of self-consistent solutions (if such solutions exist). Therefore, in this case, \( (X_i; X_{n(i)}) \) cannot be varied independently of \( m_{n(i)} \) and it is not obvious how to adapt the Elbers and Ridder (1982) proof (pg. 405) to this case. Hence for the case of arbitrarily fine-grained
reference groups, identification is currently problematic.

In contrast, suppose that individuals reference on a coarse group $G$. In this case the self-consistency condition is

$$m_{n(i)} = \int_X E(t \mid X, X_{n(i)}; m_{n(i)}) dF_X = H(m_{n(i)}) \quad (132)$$

where $dF_X$ is the distribution of $X$'s within $n(i)$. In this case, it is possible to locate a nontrivial set of sufficient conditions on $dG(t, X_i; X_{n(i)}, m_{n(i)})$ such that the hazard function is differentiable with respect to $(X_i, X_{n(i)})$ on the self-consistent solutions defined by eq. (132). This appears to be sufficient to extend Elbers and Ridder's identification argument to the case of interactions.

5. Sampling properties

In this section, we develop some asymptotics for the parameter estimates for interactions-based models and consider the effects on such estimates of omitted variables. The sampling properties for data generated by interactions-based models are no different from that associated with standard discrete choice and linear regression models. The critical property which one needs to verify is that the behavioral data obey the standard limits theorem necessary for asymptotics when there is sufficient dependence across observations to induce multiple equilibria. Similarly, the effects of omitted variables mirror results found in other contexts. We therefore focus on the binary choice case.

i. laws of large numbers

Despite the dependence introduced by interactions, the data generated by the noncooperative version of the binary choice model with interactions generates
a law of large numbers. Brock and Durlauf (1995) showed this for the special case where \(h_i = h \ \forall \ i\); it is straightforward (cf. Ash (1972) pg. 234) to extend this result to non-identically distributed choices.

**Theorem 10. Law of large numbers for realized average choice levels in noncooperative version of the binary choice model with interactions**

Suppose that a population of agents holds a common belief that the expected value of the average population choice is \(m^*\), where \(m^*\) is a solution to eq. (22). Then a weak law of large numbers will apply as \(I\) becomes arbitrarily large such that

\[
\lim_{\omega I \to \omega} m^*. \tag{133}
\]

**ii. naive estimator**

The “naive” estimator whose identification properties we have analyzed does not introduce any new econometric issues with respect to asymptotic normality. Theorem 11 is standard; necessary conditions for it are given, for example, in McFadden (1984), p. 1399. The specific conditions we cite are found in Amemiya (1985) pg. 270.5

**Theorem 11. Consistency and asymptotic normality of naive estimates in the binary choice model with global interactions**

Let \(b = (k,c,d,J)^t\) and \(M_i = (1,X_i,Y_n(i),m^e_n(i))^t\). If the binary choice model with

\[5\] In Amemiya (1985), it is also assumed that the \(M_i\) elements are uniformly bounded when asymptotic normality is proved, which contradicts our identification assumption that the \(Y_n(i)\)'s are unbounded. However, as Amemiya points out, the boundedness assumption can be dispensed with.
interactions is globally identified, and if

i. $b$ lies in an open, bounded subset of $R^r + s + 2$.

ii. $\lim_{I \to \infty} I^{-1} \sum_{i \in I} M_i M'_i$ is a finite nonsingular matrix.

iii. The empirical distribution function of $M_i$ converges to a distribution function.

Then, the maximum likelihood estimates $\hat{b}_I$ of the binary choice model with global interactions are consistent and asymptotically normal with limiting behavior

$$I^{1/2}(\hat{b}_I - b) \Rightarrow_w N(0, \mathfrak{I}^{-1}) \quad (134)$$

where

$$\mathfrak{I} = \lim_{I \to \infty} I^{-1} \sum_{i} \frac{\exp(\hat{b}'M_i)}{(1 + \exp(\hat{b}'M_i))^2} M_i M'_i \quad (135)$$

($\mathfrak{I}$ is of course the suitably normalized information matrix of the likelihood function and is consistently estimable for this model.)

iii. asymptotics for data generated by social planner

Models which incorporate realized contemporaneous interactions between individuals introduce several mathematical complexities relative to standard econometric models. As noted before, this occurs because of the quadratic terms which appear in the likelihood. The following theorem is proved in the Mathematical Appendix; unlike Theorem 10 it does not apply to the case of
heterogeneous $h_i$'s, although results in Amaro de Matos and Perez (1991) suggest this can be done.

**Theorem 12. Large economy limit for realized average choice levels in social planner's version of the binary choice model with interactions**

Suppose that the vector of choices in a population is determined by a social planner with preferences consistent with eq. (44). The sample mean of these choices will converge weakly such that

$$\lim_{\omega \to \omega} m^*$$

where $m^*$ is the solution to $m^* = \tanh(\beta h + \beta Jm^*)$ with the same sign as $h$.

Unfortunately, maximum likelihood estimation has yet to be developed for data generated by a social planner problem of the type we have studied. While techniques developed in Amaro de Matos and Perez (1991), Brock (1993), and Ellis (1985) all suggest that the development of these asymptotics is feasible, the argument seems sufficiently complicated that we are not comfortable making a conjecture on the asymptotic distribution of the estimator. In the Mathematical Appendix, we provide some initial discussion of these issues to illustrate how such a theory could be developed.

**iv. unobserved variables**

Perhaps the most serious criticism made of efforts to identify interaction effects is the difficulty in identifying interaction effects in the presence of unobserved individual or group characteristics. This is true because the main
groupings for which interactions are conjectured to exist, neighborhoods, schools, firms, etc., the neighborhoods are endogenously determined. Presumably, neighborhood contextual and endogenous characteristics influence individual choices as to neighborhood membership. Hence, it seems very likely that omitted variables which influence individual behavior once that person is a member of a neighborhood will also be correlated with the various group effects which are captured in a statistical model. This point is distinct from the self-selection issues which are discussed above.

In particular, we are interested in determining how omitted variables will affect inferences concerning \( J \). We do this following a maximum likelihood approach due to Cameron and Heckman (1998). This approach is straightforward to describe in sample, rather than population terms which is why we place it here.

In our framework, assume that the binary choices \( \omega_i \) are coded 0,1 and are generated by the probability model

\[
\mu(\omega_i = 1 \mid X_{i,o}, o, X_{i,u}, m_{n(i)}^e) =
F_{\epsilon}(k + c_{o}oX_{i,o} + d_{o}oY_{i,o} + c_{u}uX_{i,u} + d_{u}uY_{i,u} + Jm_{n(i)}^e)
\]

(137)

where subscripts \( o \) and \( u \) refer to observed and unobserved variables respectively. Let \( \Theta' = (k, c_{o}, d_{o}, J), Z_i' = (1, X_{i,o}, o, X_{i,u}, m_{n(i)}^e) \) and \( \eta_i = c_{u}uX_{i,u} + d_{u}uY_{i,u} \). This means the true probability structure can be rewritten as

\[
\mu(\omega_i = 1 \mid X_{i,o}, o, X_{i,u}, u, Y_{i,u}, m_{n(i)}^e) = F_{\epsilon}(\Theta'Z_i + \eta_i)
\]

(138)

which produces a likelihood function of the form

\[
L = I^{-1} \sum_i (\omega_i \log F_{\epsilon}(\Theta'Z_i + \eta_i) + (1 - \omega_i) \log (1 - F_{\epsilon}(\Theta'Z_i + \eta_i))
\]

(139)