Social Interactions - Is There Really an Identification Problem?

Ulf von Kalckreuth, Deutsche Bundesbank

This version: November 17, 1999

Abstract:

Social interactions are central to economic theory. This essay examines the position of Charles F. Manski with respect to the inferential problems posed by social effects. He maintains that, as a rule, models with social effects are not identified. Only under very favourable conditions, he claims, can they be distinguished from other reasons for correlations within social groups, such as selectivity. It is shown here that this argument depends critically on a special assumption. In Manski's econometric model, social effects do not flow from the *outcomes* realized within the group, but from their *conditional mathematical expectations*. By relaxing this assumption, a fully identified model is obtained. Furthermore, FIML estimators for all parameters are explicitly calculated and statistically evaluated.

Key Words:

Social Effects, Identification, Panel Data, FIML Estimation

JEL Classification: C 30, C 12, D 62

Correspondence: Deutsche Bundesbank, P.O. Box 10 06 02, 60006 Frankfurt a. M., Germany. Tel.: +49 69 9566 2217, Fax: +49 69 9566 4317, E-mail: Ulf.von-Kalckreuth@bundesbank.de.

Social Interactions - Is There Really an Identification Problem?[1]

1. Introduction

Social interactions are central to modern economic theory as represented by works such as Durlauf (1996), Bénabou (1996a, 1996b) or Borjas (1992, 1995), that explain growth and income distribution jointly. This essay examines the position of Charles F. Manski concerning "endogenous social effects", as published in Manski (1993a), Manski (1993b) and Manski (1995). Endogenous social effects are given when

> the propensity of an individual to behave in some way varies with the prevalence of that behaviour in some reference group containing the individual.[2]

It is an everyday experience that the behaviour of individuals belonging to the same social group tends to be correlated. In his seminal work, Manski differentiates two basic types of feedback between group and individual and he maintains that it is not possible to discriminate between the two by mere observation. What is more: Only under very favourable conditions can social effects be distinguished from other reasons for correlations within social groups, such as selectivity.

Manski's forceful critique challenges not only the numerous empirical efforts to understand the nature of social interactions. In the light of his arguments many theoretical disputes in the social sciences suddenly appear to be rather futile. Thus, a further analysis of his position seems well justified.

The result is encouraging. Identification can quite generally be reached by relaxing one of Manski's key assumptions. In his econometric model, social effects do not flow from the *outcomes* realized within the group, but from their respective *conditional mathematical expectations*. By substituting this assumption by a less demanding one, a model is obtained that is fully identified, if one is willing to make use of restrictions on the structural error terms. Within the new framework, Manski's analysis rightly characterizes the case of infinitely sized reference groups. In the second part of the essay, FIML estimators of all parameters are explicitly derived for the modified model. This enables us to analyse finite sample properties of the new estimator. In general, the new estimator allows one to differentiate clearly between endogenous social effects, exogenous social effects and correlated effects.

---

[1] The opinions expressed in this essay are those of the author alone and do not necessarily represent the view of the Deutsche Bundesbank. I owe a great debt to Jürgen Schröder and Martin Hellwig for encouragement and many helpful comments, and even more so to Klaus Winckler. At a later stage, I got important feedback from Axel Börsch-Supan and Reinhold Schnabel. I appreciate especially the thorough comment by Charles F. Manski that led to a major revision. Bob Chirinko gave me encouragement and some final hints. I also thank participants of two seminars in Mannheim and my audience at the Annual Congress of the *Verein für Socialpolitik* in Mainz, here especially Dieter Bös. Stefanie von Kalckreuth and Angela Lahee proof-read the English version.

[2] Manski (1993b), p. 531.

2. Identifying Social Effects

2.1. Endogenous, Exogenous and Correlated Effects

In order to discuss the inferential problems posed by social effects, Manski constructs a meta-model that embraces many phenomena as special cases. Let every individual in a population be characterized by a vector of jointly distributed variables $(y \quad \mathbf{x} \quad \mathbf{z} \quad v)$. The scalar y is a variable that may partly depend on social effects. An example of such a variable might be the school records of a pupil or his or her occupational aspirations. The J-dimensional vector $\mathbf{x}$ contains all the relevant exogenous characteristics of an individual's reference group. This is the group within which mutual influence seems possible, such as the pupil's class or his or her neighbourhood. A reference group might also be characterized by general attributes such as ethnicity or sex.[3] The K-dimensional vector $\mathbf{z}$ stands for individual qualities with relevance for the dependent variable, such as socio-economic background or health. Vectors $(y \quad \mathbf{x} \quad \mathbf{z})$ can be observed. The random variable v is not observable. Manski offers three general hypotheses that might explain why the behaviour of individuals belonging to the same group often shows a high degree of correlation.

To begin with, the variable y might be directly influenced by the mean of that same variable within a reference group. This is what he calls an *endogenous social effect.* Examples are peer effects among pupils, herding behaviour during financial crises or interdependencies in consumer demand.[4] In Manski's theoretical exposition, the endogenous social effect does not originate from the *outcome* of other individuals in the same group. Instead, the *conditional expectation* $\mathrm{E}(y|\mathbf{x})$ of the variable is deemed relevant, given the general characteristics $\mathbf{x}$ of the reference group. The analytical consequences of this assumption will be analysed in the next sections.

Closely related are the possible effects of the exogenous characteristics of the actors in the social context. In the above example, an *exogenous social effect* is present if not the academic performance of the classmates, but their socio-economic status or national composition act upon the achievements of a pupil. As before, Manski assumes that exogenous effects operate via a conditional expectation, in this case $\mathrm{E}(\mathbf{z}|\mathbf{x})$. The distinction between endogenous and exogenous effects is of great practical importance with respect to the effect of discretionary interventions. Tutoring weaker pupils, for example, will have a beneficial effect on their classmates only in cases of endogenous social effects.

Completely different conclusions are reached by assuming that the variable y might directly depend on the characteristics $\mathbf{x}$ of the reference group, whether a social interaction takes place or not. On average, children of foreign parents in Germany show – depending upon the coun-

---

3 The concept goes back to Herbert Hyman (1942), see also Hyman (1968). In the original formulation, the term is not limited to groups that contain the individual. For Manski's problem, group membership is constitutive.

4 For an example see the empirical study by Case and Katz (1991).

try of origin – a much weaker performance at school than ethnically German children.[5] It is conceivable that this is the result of endogenous or exogenous social effects. The socio-economic status of many foreign families in Germany is relatively low. If their status affects the performance of their children, and if these children's reference group comprises mainly classmates of their own nationality, then even the performance of foreign children with average exogenous characteristics will be substandard.[6]

Alternatively, this empirical regularity can be explained by invoking the language problems and cultural interferences associated with the group characteristic "foreign pupil of nationality $\mathbf{x}$" without social effects playing any role whatsoever. Any such direct causal relationship is labelled a *correlated effect* by Manski. Correlated effects can also be a consequence of *institutional influences,* if foreign pupils of certain nationalities are systematically discriminated against in German schools, or if the pupils of a given school are all exposed to the same bad teachers.[7] A further important source of correlated effects is *self-selection.* This phenomenon is of special importance in the study of social effects within neighbourhoods. Persons with unfavourable but unobserved characteristics might concentrate in low-cost neighbourhoods, which causes a spurious correlation of income and other variables.[8]

2.2. The Reflection Problem

Manski characterizes the inductive task posed by endogenous social effects as

> ...the problem that arises when a researcher observing the distribution of behaviour in a population tries to infer whether the average behaviour in some group influences the behaviour of the individuals that comprise the group.[9]

He introduces the term "reflection problem". The problem is

> similar to that of interpreting the almost simultaneous movements of a person and his reflection in a mirror. Does the mirror image cause the person's movement or reflect them? An observer who does not understand something of optics and human behaviour would not be able to tell.

The basic idea shall be developed using a simplified model. Let the outcome y of a person be determined solely by correlated effects, endogenous social effects, and a disturbance term. The structural equation is:

$$y = \alpha + \beta \mathrm{E}(y|\mathbf{x}) + \mathbf{x}' \mathrm{d} + v \quad . \qquad (1)$$

Here, $\mathbf{x}$ is a vector of K characteristics of the person's reference group. The conditional expectation $\mathrm{E}(v|\mathbf{x})$ is zero. Manski assumes that $\mathrm{E}(y|\mathbf{x})$ can be estimated consistently and he treats the regressor as known. If $\beta \neq 0$, the linear regression expresses an endogenous social

---

5 Alba, Handl and Müller (1994).

6 This thesis is maintained by Borjas in his studies on "ethnic capital" with regard to the relative economic performance of immigrants in the USA. See Borjas (1992) and Borjas (1995).

7 Jencks and Mayer (1990), p. 115.

8 The correlation in the behaviour of adolescents mentioned above can by explained in this manner, as well as the influence of the social composition of the neighbourhood on the academic performance of pupils. The problem is analysed in Rauch (1993) and Corcoran et al. (1992).

9 Manski (1993b), p. 532.

effect. The term $\mathbf{x}'\mathrm{d}$ allows for correlated effects. As an example we can take the correlation between performance at school and ethnicity of pupils in Germany. Here the reference group characteristics $\mathbf{x}$ have a dual function. In addition to their *direct* influence on y – the consequences of the inability to speak German properly and possibly discrimination – it *conditions the expectation* $\mathrm{E}(y|\mathbf{x})$ *that plays the role of another regressor variable*. It is impossible to distinguish between these two aspects of belonging to a certain social group. Solving for the conditional expectation, one obtains:

$$\mathrm{E}(y|\mathbf{x}) = \frac{1}{1-\beta}(\alpha + \mathbf{x}'\mathrm{d}) \ .$$

There is perfect collinearity between the regressor variables $\mathrm{E}(y|\mathbf{x})$ and $\mathbf{x}$ in (1). Elimination of the mathematical expectation from (1) leads to the reduced form:

$$y = c_0 + \mathbf{x}'\mathbf{c}_1 + v \qquad \text{with} \qquad c_0 = \frac{1}{1-\beta}\alpha \ ; \qquad \mathbf{c}_1 = \frac{1}{1-\beta}\mathrm{d}.$$

Under appropriate circumstances, this equation may be consistently estimated. However, such an estimate does not contribute to answering the question of whether or not there are endogenous social effects in the system. For any hypothetical $\beta^* \neq 1$ we can state a vector $(\alpha^* \quad \mathrm{d}^*)$ such that

$$\frac{1}{1-\beta^*}\alpha^* = c_0 \qquad \text{and} \qquad \frac{1}{1-\beta^*}\mathrm{d}^* = \mathbf{c}_1 \ .$$

A linear space of bogus parameters $(\alpha^* \quad \beta^* \quad \mathrm{d}^*)$ leads to the same reduced form as the true parameters $(\alpha \quad \beta \quad \mathrm{d})$: *They are observationally equivalent.* Whatever the size of the data set, it will not be sufficient to decide whether the data were generated by a system with parameters $(\alpha \quad \beta \quad \mathrm{d})$ or by one of the systems with parameters $(\alpha^* \quad \beta^* \quad \mathrm{d}^*)$.

2.3. The Complete Linear Model

In addition to endogenous and correlated effects, Manski features exogenous social effects as well as the action of individual characteristics. His complete specification is:

$$y = \alpha + \beta\mathrm{E}(y|\mathbf{x}) + \mathrm{E}(\mathbf{z}|\mathbf{x})'\mathrm{g} + \mathbf{x}'\mathrm{d} + \mathbf{z}'\mathrm{h} + v \quad . \tag{2}$$

The K-dimensional vector h represents the effect of individual characteristics $\mathbf{z}$. Exogenous social effects are present if the K-dimensional vector g is not zero: y then varies with the mathematical expectation of the exogenous variable $\mathbf{z}$ in the reference group.

It is assumed that $\mathrm{E}(v|\mathbf{x},\mathbf{z}) = 0$, and there is no further distributional information. Calculating the mathematical expectation $\mathrm{E}(y|\mathbf{x})$ by integrating (2) with respect to $\mathbf{z}$ and v, one obtains:

$$\mathrm{E}(y|\mathbf{x}) = \frac{1}{1-b}\left[a + \mathrm{E}(\mathbf{z}|\mathbf{x})'(\mathrm{g} + \mathrm{h}) + \mathbf{x}'\mathrm{d}\right] \ .$$

For a given $\mathbf{x}$, the conditional expectation $\mathrm{E}(y|\mathbf{x})$ is a constant. It is a linear function of the regressors $(1 \quad \mathrm{E}(\mathbf{z}|\mathbf{x}) \quad \mathbf{x})$. The reduced form is calculated in the familiar way:

$$y = c_0 + \mathrm{E}(\mathbf{z}|\mathbf{x})'\mathbf{c}_1 + \mathbf{x}'\mathbf{c}_2 + \mathbf{z}'\mathbf{c}_3 + v \quad , \tag{3}$$

with $$c_0 = \frac{1}{1-\beta}\alpha \, ; \qquad \mathbf{c}_1 = \frac{1}{1-\beta}(\mathrm{g} + \beta\mathrm{h}) \, ; \qquad \mathbf{c}_2 = \frac{1}{1-\beta}\mathrm{d} \, ; \qquad \mathbf{c}_3 = \mathrm{h} \, . \tag{4}$$

The parameters of the reduced form do not allow one to deduce the structural parameters α, β, g, and d. Still, estimating this equation does yield information on the structural parameters. The effect h of the individual characteristics can be inferred. Moreover, it is possible to decide whether there are any social effects at all. If $\mathbf{c}_1 \neq \mathbf{0}$, then $\mathrm{g} \neq 0 \vee \beta \neq 0$. If the outcome of y happens to depend on the *expected* outcome $\mathrm{E}(\mathbf{z}|\mathbf{x})$, the presence of endogenous and/or exogenous social effects can be concluded. This is by no means unimportant, as in scientific practice it proves to be quite difficult to establish social effects of any kind.[10]

As a necessary precondition, $\mathrm{E}(\mathbf{z}|\mathbf{x})$ must supply independent information. If $\mathrm{E}(\mathbf{z}|\mathbf{x})$ can be written as a linear combination of the other regressors $(1 \quad \mathbf{z} \quad \mathbf{x})$, even this limited identification is lost. This situation pertains, if, for example:

(a) $\mathbf{z}$ is a (mathematical) function of $\mathbf{x}$. For any $\mathbf{x}$, we have $\mathrm{E}(\mathbf{z}|\mathbf{x}) = \mathbf{z}(\mathbf{x})$;

(b) $\mathrm{E}(\mathbf{z}|\mathbf{x})$ does not vary with $\mathbf{x}$. $\mathrm{E}(\mathbf{z}|\mathbf{x})$ is then a constant and collinear with 1;

(c) $\mathrm{E}(\mathbf{z}|\mathbf{x})$ is a linear function of $\mathbf{x}$.

All in all, Manski concludes, making statements on the presence of social effects is possible only if the variables $\mathbf{x}$ defining reference groups and the exogenous variables $\mathbf{z}$ are related in the population by a moderately strong, but non-linear statistical dependence. The distinction between endogenous and exogenous effects, important as it may be with regard to the results of discretionary changes, is empirically not feasible, nor is the distinction between endogenous and correlated effects. This general identification problem may be "solved" by discriminating in advance in favour of one of the competing hypotheses. If only exogenous social effects and correlated effects are permitted, then $\beta = 0$ by definition and the model is fully identified. Limiting the analysis to endogenous social effects, such that $\mathrm{g} = \mathrm{d} = \mathbf{0}$, yields the same results. Neither Manski nor the author of this essay knows of any empirical work that permits both types of social effects.

2.4. Is There Really an Identification Problem?

Manski uses the mathematical expectations $\mathrm{E}(y|\mathbf{x})$ and $\mathrm{E}(\mathbf{z}|\mathbf{x})$ as regressor variables in order to model social effects. These magnitudes are mathematical functions of the characteristics $\mathbf{x}$. This technique highlights the problem of differentiating between social effects and other consequences of belonging to a certain social group.

However, when one considers social interactions in real life, this procedure seems to idealize matters slightly. In a social group of finite magnitude (a family, a class of pupils, a neighbourhood), the group mean is as stochastic as the individual outcomes and it is difficult to find a

10 As an example see the survey of Jencks and Mayer (1990).

substantive interpretation of why, for example, the mathematical *expectation* of the classmates' performance, but not their actual performance should act as an externality on an individual pupil. In general, a social effect operating via a mathematical expectation can result if the agents hold rational expectations in the sense of Muth, or in strategic situations of a game-theoretic nature. Yet in these two cases, there is no identification problem of the kind described above, because the conditioning variables for the relevant mathematical expectations would also include the individual characteristics of the group members. Alternatively, the structural equation (2) might deal with the limiting case of a social group with infinite size. In this case, the law of large numbers makes the group means converge to the conditional expectation. From the empirical literature on social interactions, however, the author is not aware of any such formulation. Typically, the social environment is supposed to act on the individual via the arithmetic mean or another linear function of values realized by the group members, as is clearly shown by Manski's own characterization of the inferential problem cited above.[11] We will later return to the problem of infinitely sized reference groups.

In order to analyse the significance of this key assumption, Manski's model will here be modified by assuming that the source of the social effects is the *average* outcome within the group. This eliminates the multicollinearity problem, as the group average varies and is not a linear function of the other exogenous variables. Instead, a new, but "classical" identification problem arises: The endogenous variables are determined by a system of interdependent linear equations.[12] However, with the help of restrictions on the error term, this identification problem can be solved, and the road to this solution yields new and interesting insights into the nature of social interactions.

2.5. Social Effects as Group Interactions

The model is modified and specified by the following assumptions:

A1) Endogenous or exogenous social effects derive from group *averages*;

A2) The groups are of finite size;

A3) The error terms in the equations for the individuals are i.i.d. with variance $\sigma^2 > 0$;

A4) β is less than 1 in absolute value.

Assumptions A1 and A2 together remove the multicollinearity. The original model would result if group sizes approach infinity. A3 is a restriction concerning the covariance matrix of the error term, and A4 is a stability condition. When a specific estimation procedure is considered, some additional assumptions concerning the distribution of the error term will prove convenient. Now the modified model is analysed in some detail.

[11] This also holds for the paper by Alessie and Kapteyn (1991) on demand interdependencies cited by Manski.

[12] Theil (1971), pp. 447-8, illuminates the close affinity between these two types of identification problem.

An individual I_{ji} belongs to the group $G_j = \{I_{j1}, I_{j2}, ..., I_{jM_j}\}$ of size M_j. The data set includes N complete groups, i.e. $\sum_{j=1}^{N} M_j = M$ individuals. Group G_j is described by a $J \times 1$-dimensional vector $\mathbf{x}_j$ of characteristics. This vector is a distinctive feature of every member of the group. As several groups may be characterized by a common vector $\mathbf{x}_j$, it can be interpreted as a "type". Furthermore, the individual I_{ji} is described by a $K \times 1$-vector $\mathbf{z}_{ji}$ of exogenous characteristics and the outcome y_{ji} of a scalar endogenous variable.

Within the group, endogenous, exogenous, and correlated effects are permitted. Thus, both the mean of the exogenous variable, $\bar{\mathbf{z}}_j = \frac{1}{M_j} \sum_{i=1}^{M_j} \mathbf{z}_{ji}$, and the mean of the endogenous variable, $\bar{y}_j = \frac{1}{M_j} \sum_{i=1}^{M_j} y_{ji}$ may have a systematic influence. The structural equation for individual I_{ji} reads:

$$y_{ji} = \beta \bar{y}_j + \bar{\mathbf{z}}_j' \gamma + \mathbf{x}_j' \delta + \mathbf{z}_{ji}' \eta + v_{ji} \quad , \tag{5}$$

where the v_{ji} are i.i.d. with zero expectation and variance σ^2. A constant is contained in the vector $\mathbf{x}_j$.

The OLS estimator of this equation is biased and inconsistent, as the v_{ji} and the $\bar{y}_i$ are correlated. The reason is that the dependent variable y_{ji} enters on the right-hand side of the equations for all the remaining group members. Obviously, the problem will not disappear by taking averages on all group members but individual j. The group G_j must be described as a *system of interdependent regression equations.*[13]

Let $\mathbf{y}_j = (y_{j1} \dots y_{jM_j})'$ be the $M_j \times 1$-vector of endogenous variables and $\mathbf{Z}_j = (\mathbf{z}_{j1} \dots \mathbf{z}_{jM_j})'$ the $M_j \times K$-matrix of exogenous variables. Furthermore, let $\underset{(M_j)}{\mathbf{1}}$ be an $M_j \times 1$-vector with elements all identically equal to 1. Then

$$\mathbf{D}_j = \frac{1}{M_j} \underset{(M_j)}{\mathbf{1}} \underset{(M_j)}{\mathbf{1}}' = \frac{1}{M_j} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}_{(M_j \times M_j)}$$

is a matrix that, if postmultiplied by $\mathbf{y}_j$, generates a vector with elements identically equal to $\bar{y}_j$. An analogous result holds for postmultiplication by $\mathbf{Z}_j$. Finally, let $\mathbf{v}_j$ be an $M_j \times 1$-dimen-

[13] The simultaneity established by social effects and the resulting identification problem were first addressed by Duncan, Haller, and Portes (1968) and Duncan (1970). In the literature, the interdependence is not always treated properly. Investigating the determinants of educational expectations and occupational aspiration of high school sophomores, Alexander, Eckland and Griffin (1975) use the best friend's college plans and the proportion of peers planning to attend college as regular exogenous variables. They even happen to observe a high correlation between the social background of a pupil and the college plans of his best friend, but explain this correlation by invoking "status homophily". Sewell and Hauser (1972, 1975) use a similar procedure in their work on the ambitious "Wisconsin model of status attainment". Bank et al. (1990) circumvent the problem by using as exogenous variables the behaviour of close friends *in the past*.

sional random vector with covariance matrix $\sigma^2 \mathbf{I}_{(M_j)}$. With this notation, the system of equations for group G_j can be succinctly stated as

$$\mathbf{y}_j = \beta \mathbf{D}_j \mathbf{y}_j + \mathbf{D}_j \mathbf{Z}_j g + \mathbf{1x}_j' d + \mathbf{Z}_j h + \mathbf{v}_j \quad . \tag{6}$$

By rearranging,

$$(\mathbf{I} - \beta \mathbf{D}_j)\mathbf{y}_j = \mathbf{D}_j \mathbf{Z}_j g + \mathbf{1x}_j' d + \mathbf{Z}_j h + \mathbf{v}_j$$

and premultiplying by the inverse

$$(\mathbf{I} - \beta \mathbf{D}_j)^{-1} = \mathbf{I} + \frac{\beta}{1-\beta} \mathbf{D}_j$$

one obtains as a reduced form the vector equation:

$$\mathbf{y}_j = \mathbf{D}_j \mathbf{Z}_j \mathbf{c}_1 + \mathbf{1x}_j' \mathbf{c}_2 + \mathbf{Z}_j \mathbf{c}_3 + \mathbf{w}_j \quad , \tag{7}$$

with $\mathbf{c}_1 = (g + \beta h)\frac{1}{1-\beta} \quad ; \quad \mathbf{c}_2 = d\frac{1}{1-\beta} \quad ; \quad \mathbf{c}_3 = h \quad ; \quad \mathbf{w}_j = \left(\mathbf{I} + \frac{\beta}{1-\beta}\mathbf{D}_j\right)\mathbf{v}_j \ .$ (8)

These equations correspond to the reduced form (3) and (4) of Manski's model. The coefficients $\mathbf{c}_1$, $\mathbf{c}_2$, and $\mathbf{c}_3$ are equal to their counterparts. Again, *by the coefficients* only h is identified, but not β, g, or d. A closer inspection shows that the reduced form derived here contains more information. The structure of the error term $\mathbf{w}_j$ is determined in a characteristic way by β. The covariance matrix of $\mathbf{w}_j$ is defined by the $(M_j \times M_j)$-matrix Y_j:

$$Y_j = E\,\mathbf{w}_j \mathbf{w}_j' = \sigma^2 \left[\mathbf{I} + \frac{\beta}{1-\beta}\mathbf{D}_j\right]^2 = \sigma^2 \left[\mathbf{I} + \left(\frac{1}{(1-\beta)^2} - 1\right)\mathbf{D}_j\right] \quad . \tag{9}$$

By expanding the above matrix, it can easily be verified that

$$\frac{\operatorname{var} w_{ji} - \operatorname{cov}(w_{ji}, w_{jk})}{M_j \operatorname{cov}(w_{ji}, w_{jk}) + \operatorname{var} w_{ji} - \operatorname{cov}(w_{ji}, w_{jk})} = (1-\beta)^2 \quad ,$$

an equation by which β is determined uniquely because of A4. The remaining parameters of the structural form can be recovered from the equations in (8). A consistent estimation of the reduced form thus allows us to infer all the parameters of the structural form.

Proposition 1: *Taking into account the structure of the reduced form error, all the parameters of the modified model are exactly identified.*

In other words: Manski's assumption concerning the working of social effects is absolutely critical for his key result.

2.6. The Statistical Fingerprint of Social Effects

Identification succeeds with the help of restrictions concerning the covariance matrix of the error terms in the structural equations.[14] These restrictions permit to trace the key parameter β

---

[14] In many econometric textbooks, the identification problem is reduced to the question of whether the structural equations are identified by the *coefficients* of the reduced form. This question is then answered by the

back from the errors in the reduced form. The seminal works of Goldberger (1972), Griliches (1974), and Chamberlain and Griliches (1975) have explored the special significance of the covariance matrix for the identification of structural models in the social sciences.

Identification of the present model rests on a fundamental property of social effects: For given values of the exogenous variables they lead to covariances in the outcomes of group members that are non-existent with regard to individuals in other groups. These positive or negative covariances serve to amplify or dampen random differences between groups of otherwise identical external attributes. The ratio between the dispersion *within* a group and the dispersion *between* groups is biased in a characteristic way. This will now be made precise.

The reduced form (7) and the structure of its error term (9) bear great resemblance to the standard *random coefficient model* for the econometric analysis of panel data.[15] The reduced form disturbance,

$$w_{ji} = v_{ji} + \frac{\beta}{1-\beta}\bar{v}_j \ ,$$

is composed additively of an individual error term v_{ji} and an error term $\frac{\beta}{1-\beta}\bar{v}_j$ specific for group G_j. The variance of this second error term clearly depends on the intensity of the social interaction. Looking at the *average* outcome in group G_j,

$$\bar{y}_j = \bar{\mathbf{z}}_j'(\mathbf{c}_1 + \mathbf{c}_3) + \mathbf{x}_j'\mathbf{c}_2 + \bar{w}_j \ , \qquad \text{with} \qquad \bar{w}_j = \frac{1}{1-\beta}\bar{v}_j \ , \tag{10}$$

we see that because of

$$\operatorname{var}\bar{w}_j = \frac{1}{M_j}\left(\frac{1}{1-\beta}\right)^2 \sigma^2 \ ,$$

the variance of the error term increases with the strength β of the interdependence between group members. A positive $\beta < 1$ acts as an *amplifier of random disturbances*. A high outcome of $\bar{v}_j$ is translated into a still higher $\bar{w}_j$ in absolute terms. However, looking at the *deviations* of the individual from the average of its reference group,

$$y_{ji} - \bar{y}_j = (\mathbf{z}_{ji} - \bar{\mathbf{z}}_j)\mathbf{c}_3 + w_{ji} - \bar{w}_j \ , \tag{11}$$

the variance of the residual does not depend on β, since by definition

$$w_{ji} - \bar{w}_j = v_{ji} - \bar{v}_j \ .$$

Actually we have:

$$\operatorname{var}(w_{ji} - \bar{w}_j) = \sigma^2\left(1 - \frac{1}{M_j}\right) \ .$$

---

rank and order conditions. Yet the classic treatment of Fisher (1966) already dedicates a whole chapter to the role of the covariance matrix.

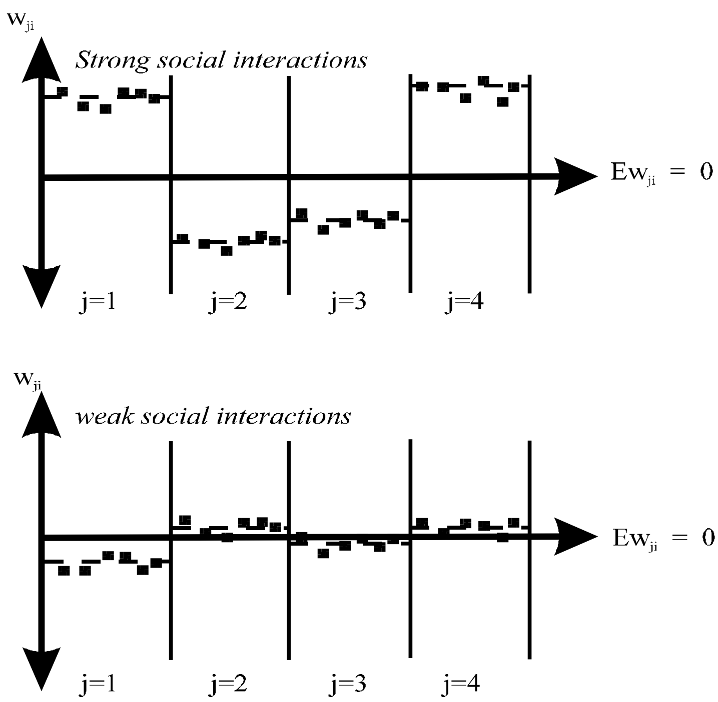
15 See Hsiao (1986), Chaps. 2 and 3. In the case at hand, the disturbances related to group and individual are correlated, as distinct from the formulation in the standard model.

This is the statistical "fingerprint" of endogenous social effects within groups of finite size: Relative to the variability *within groups*, the differences *between* groups are conspicuously large or small. This central feature is suppressed by Manski's quasi-deterministic modelling technique.

As an illustration, the graph below depicts the residuals of four reference groups, for the case of strong and of weak positive social feedbacks. The ratio between the dispersion of the group means on the one hand, and the dispersion around the group means on the other, indicates the social effect:

$$\frac{1}{M_j - 1} \frac{\mathrm{var}(w_{ji} - \overline{w}_j)}{\mathrm{var}\,\overline{w}_j} = (1-\beta)^2 \quad . \tag{12}$$

This equation concisely summarises the information on β contained in the reduced form errors. A consistent estimate for the numerator and the denominator is obtained from the squared residuals of the OLS estimates of the equations (10) and (11).



Dispersion of the reduced form residuals for endogenous effects of varying strength

Beyond the inferential problem, the graph makes clear that endogenous social effects can be very important for the analysis of economic inequality. Differences in the starting positions of social groups such as families or cliques, e.g. with regard to human capital, are reinforced if social interactions cause a positive feedback between the outcome of group members.[16] An endogenous social effect with $0 < \beta < 1$ in (6) is akin to an income multiplier in a simple Keynesian macromodel. This concerns not only the average $\overline{v}_j$ of the residuals, but also the net

[16] For a fascinating example in the same vein see Griliches' (1996) analysis of the "F-connection".

effect of the vectors $\mathbf{x}_j$ and $\bar{\mathbf{z}}_j$. Corresponding to this formal similarity, there is an analogy in the underlying logical structure.

There are two limiting cases in which this road to identification breaks down. The first is the case of infinitely large reference groups, such as "White Anglo-Saxon Protestants" or "My Compatriots". This seems to be the case addressed by Manski. With growing group size, averages converge to their respective conditional expectations, and the variance of average residuals will disappear. Another extreme case is equally devastating: If every reference group is *unique*, i.e. each group merits a dummy of its own, there will be no degrees of freedom left to estimate the variance of the average residuals. Identification works if we have moderate sized groups (families, school classes, clusters of nations related to each other by trade or migration streams) that are comparable in the sense that not every single group forms a type of its own.

One might object that identification depends on imposing a very simple structure for the error process. Actually, it is necessary to possess at least some information on the structural errors. Let $\mathbf{v}_j$ be distributed such that $E\mathbf{v}_j\mathbf{v}_j' = S_j$. The covariance matrix of the reduced form errors will then be $Y_j = E\mathbf{w}_j\mathbf{w}_j' = \left[\mathbf{I}+\frac{\beta}{1-\beta}\mathbf{D}_j\right] S_j \left[\mathbf{I}+\frac{\beta}{1-\beta}\mathbf{D}_j\right]'$. As long as there are not too many unknown parameters in S_j – if the matrix is diagonal for instance – it will still be possible to infer ß from Y_j. If, on the other hand, there is no information on the error term at all, the parameter β remains unidentified. Incidentally, this is exactly what Manski assumes in his own exposition. One might say that for his result of non-identification, this assumption is as important as the way of modelling social interactions.

2.7. Network Analysis

Simultaneous systems of social interactions of the type depicted in (5) were introduced by Erbring and Young (1978).[17] Their approach is labelled *network analysis* or *model of spatial correlation.* With respect to the modelling of social interaction, this approach is more general than the structure explored here, but the literature does not explicitly consider either exogenous effects or correlated effects. The structural equation of Erbring and Young reads:[18]

$$\mathbf{y} = \beta\mathbf{W}\mathbf{y} + \mathbf{Z}\eta + \mathbf{v} \ . \qquad (13)$$

Here, as above, $\mathbf{y}$ is a vector of observations of an endogenous variable, $\mathbf{Z}$ is a matrix of exogenous variables and $\mathbf{v}$ is a random vector. $\mathbf{W}$ is a matrix that describes the structure of the social interactions between the individuals. $\mathbf{W}$ defines an autoregressive relationship between the endogenous variables. Erbring and Young do not assume a number of separate groups, but in principle each individual may interact with every other. They interpret their structural equation as a dynamic equilibrium of an iterative social process:

---

[17] In geographic and biological applications, similar systems were explored even earlier, see for example Ord (1975) and Cliff and Ord (1981), Chap. 9 and the literature cited there. Burt (1980), Friedkin (1990), Friedkin and Johnson (1990), and Friedkin and Cook (1990), further explore this model and concentrate on substantive aspects.

[18] This notation deviates from Erbring and Young (1978) and is adapted to the equation (5).

$$\mathbf{y}_{(t+1)} = \beta \mathbf{W} \mathbf{y}_{(t)} + \mathbf{Z}h + \mathbf{v} \quad , \quad \text{with} \quad \mathbf{y}_{(0)} = \mathbf{0} \quad . \tag{14}$$

The parameter β is dubbed *feedback rate.* The magnitudes $\mathbf{v}$ and $\mathbf{Z}h$ are regarded as being given for the whole "duration" of the process. In the first stage they directly determine $\mathbf{y}_{(1)}$. According to $\mathbf{W}$, the vector of state variables is then transmitted to the interaction partners. The result, $\mathbf{y}_{(2)}$, serves as starting point for the third iterate, and so on. Dynamic equilibrium is given for $\mathbf{y}_{(t+1)} = \mathbf{y}_{(t)} = \mathbf{y}$, i.e., if (13) holds. Again the analogy to the dynamical interpretation of the Keynesian expenditure multiplier is very close. The social "multiplier process" cannot do without restrictions on the parameters. By direct substitution it follows that

$$\mathbf{y}_{(t+1)} = \left(\mathbf{I} + \beta\mathbf{W} + (\beta\mathbf{W})^2 + \ldots + (\beta\mathbf{W})^t\right)(\mathbf{Z}h + \mathbf{v}) \ .$$

The series in brackets must converge if dynamic equilibrium is to be reached for given $\mathbf{v}$ and $\mathbf{Z}h$, i.e., if the difference equation (14) is to be stable. In this case we have

$$\mathbf{I} + \beta\mathbf{W} + (\beta\mathbf{W})^2 + \ldots = (\mathbf{I} - \beta\mathbf{W})^{-1} \quad ,$$

and the system converges to $\mathbf{y} = (\mathbf{I} - \beta\mathbf{W})^{-1}(\mathbf{Z}h + \mathbf{v})$.

In this view, the reduced form describes the stationary state of a dynamical system. The model can be usefully employed for quite diverse purposes. Doreian (1981) uses it to analyse spatial interdependencies: in the military activities of a rebel formation, for example, or in voting decisions in Louisiana. Burt and Doreian (1982) investigate interdependencies in the evaluation of leading scientific journals by the scientific community. Burt (1987) undertakes a network-theoretic analysis of the diffusion of a novel antibiotic among physicians in the American Midwest and Case (1991) explores spatial interdependencies in the demand of consumer goods.

The network model can be estimated by a maximum likelihood routine. The technique proposed by Erbring and Young (1978) and Doreian (1981) actually goes back to Ord (1975).[19] The procedure described in these publications has one serious drawback: in general, the likelihood equations cannot be solved explicitly and the likelihood function must be maximized numerically. Fortunately, for the case at hand we can give an exact solution. This not only greatly facilitates the *interpretation* of the resulting estimators, it even enables us to make detailed statements about their *finite sample properties*.

3. FIML-Estimation of the System

3.1. The Structural Equations

Collecting together the data, the structural equations of the modified model can be written as

$$\mathbf{y} = \beta\mathbf{D}\mathbf{y} + \mathbf{D}\mathbf{Z}\xi + \mathbf{X}d + \mathbf{Z}h + \mathbf{v} \quad , \qquad \text{where} \tag{15}$$

---

[19] See also Doreian (1981), Cliff and Ord (1981), Chaps. 6, and 9 and Anselin (1988), Chaps. 6 and 12.

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}_{(M\times 1))} \quad \mathbf{D} = \begin{pmatrix} \mathbf{D}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}_N \end{pmatrix}_{(M\times M)} \quad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_N \end{bmatrix}_{(M\times K)} \quad \mathbf{X} = \begin{bmatrix} \underset{(M_1)}{\mathbf{1}} & \mathbf{x}_1' \\ \underset{(M_2)}{\mathbf{1}} & \mathbf{x}_2' \\ \vdots & \\ \underset{(M_N)}{\mathbf{1}} & \mathbf{x}_N' \end{bmatrix}_{(M\times J)} \quad \mathbf{v} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_N \end{bmatrix}_{(M\times 1)} .$$

A further simplification lies in the notation

$$(\mathbf{I} - \beta\mathbf{D})\mathbf{y} = \mathbf{Q}\,\mathrm{f} + \mathbf{v} \;, \text{ with } \quad \mathbf{Q} = [\mathbf{DZ} \quad \mathbf{X} \quad \mathbf{Z}] \quad \text{ and } \quad \mathrm{f}' = [\,\mathrm{g}' \quad \mathrm{d}' \quad \mathrm{h}'\,] \,. \qquad (16)$$

Note that $\mathbf{y}$ is formed from the *vectors* $\mathbf{y}_j$ of group G_j. The random vector $\mathbf{v}$ and the matrices $\mathbf{Z}$ and $\mathbf{X}$ are constructed in an analogous way. The matrix $\mathbf{D}$ is block-diagonal, with submatrices $\mathbf{D}_j = \frac{1}{M_j} \underset{(M_j)}{\mathbf{1}} \underset{(M_j)}{\mathbf{1}}'$. It is idempotent and postmultiplication by $\mathbf{y}$ generates a vector that yields for every individual the average of the endogenous variable in his or her group. Postmultiplying by $\mathbf{Z}$ yields an analogous result for the exogenous variables. The block-diagonal matrix $\mathbf{D}$ plays the role of $\mathbf{W}$ in the model of Erbring and Young. The series

$$\mathbf{I} + \beta\mathbf{D} + (\beta\mathbf{D})^2 + (\beta\mathbf{D})^3 \ldots = \mathbf{I} + (\beta + \beta^2 + \beta^3 \ldots)\mathbf{D}$$

converges for $|\beta| < 1$. This is presupposed by the assumption A4. Furthermore, we shall now specify:

A5) $\mathbf{X}$ and $\mathbf{Z}$ – and thus $\mathbf{Q}$ – are fixed real matrices and $\mathbf{Q}'\mathbf{Q}$ is of full rank;

A6) The elements of $\mathbf{v}$ are jointly normal with expectation zero and covariance $\sigma^2\mathbf{I}$.

3.2. The Likelihood Function

Our point of departure is the structural equation for the simultaneous system (16). The random vector $\mathbf{v}$ is normal, with density:

$$f_v(\mathbf{v}) = (2\pi\sigma^2)^{-\frac{M}{2}} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{v}'\mathbf{v}\right\} .$$

The observed variable $\mathbf{y}$ is generated by a linear transformation from the non-observed random variable $\mathbf{v}$. Thus, $\mathbf{y}$ is also normal, with density:[20]

$$f_y(\mathbf{y}) = (2\pi\sigma^2)^{-\frac{M}{2}} \exp\left\{-\frac{1}{2\sigma^2}[(\mathbf{I}-\beta\mathbf{D})\mathbf{y} - \mathbf{Q}\,\mathrm{f}]'[(\mathbf{I}-\beta\mathbf{D})\mathbf{y} - \mathbf{Q}\,\mathrm{f}]\right\} \mathrm{abs}|\mathbf{I}-\beta\mathbf{D}| \;.$$

Since the matrix $\mathbf{I} - \beta\mathbf{D}_j$ possesses the $(M_j - 1)$-fold eigenvalue 1 plus the simple eigenvalue $1-\beta$, it follows that $\mathrm{abs}|\mathbf{I} - \beta\mathbf{D}| = (1-\beta)^N$. Hence the log-likelihood function is:

$$\mathrm{l}(\mathrm{f}, \sigma^2, \beta|\mathbf{y}) = -\frac{M}{2}\ln 2\pi - \frac{M}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}A + N\ln(1-\beta) \;,$$

with $\quad A = [(\mathbf{I}-\beta\mathbf{D})\mathbf{y} - \mathbf{Q}\,\mathrm{f}]'\,[(\mathbf{I}-\beta\mathbf{D})\mathbf{y} - \mathbf{Q}\,\mathrm{f}]$

---

20 $|\mathbf{I} - \beta\mathbf{D}|$ is the Jacobian for the transformation $\mathbf{y} = (\mathbf{I} - \beta\mathbf{D})^{-1}(\mathbf{Q}\,\mathrm{f} + \mathbf{v})$, see, e.g., Fisz (1963), p. 56.

$$= \mathbf{y}'\left(\mathbf{I} - 2\beta\mathbf{D} + \beta^2\mathbf{D}\right)\mathbf{y} - 2\mathbf{y}'\left(\mathbf{I} - \beta\mathbf{D}\right)\mathbf{Q}\,\mathrm{f} + \mathrm{f}'\mathbf{Q}'\mathbf{Q}\,\mathrm{f} \ .$$

3.3. The Likelihood Equations and Their Solutions

We need the combination $\mathrm{f}, \sigma^2, \beta$ that maximizes the likelihood function for given $\mathbf{y}$ and $\mathbf{Q}$. As necessary conditions for an interior solution, the following likelihood equations must hold:

$$\frac{\partial l}{\partial \mathrm{f}} = \frac{1}{\sigma^2}\left[\mathbf{Q}'\left(\mathbf{I} - \beta\mathbf{D}\right)\mathbf{y} - \mathbf{Q}'\mathbf{Q}\,\mathrm{f}\right] \overset{!}{=} \mathbf{0} \quad ; \tag{17}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{M}{2\sigma^2} + \frac{A}{2\sigma^4} \overset{!}{=} 0 \quad ; \tag{18}$$

$$\frac{\partial l}{\partial \beta} = \frac{1}{\sigma^2}\mathbf{y}'\mathbf{D}\left[\left(\mathbf{I} - \beta\mathbf{D}\right)\mathbf{y} - \mathbf{Q}\,\mathrm{f}\right] - N\frac{1}{1-\beta} \overset{!}{=} 0 \quad . \tag{19}$$

The first two equations yield

$$\mathrm{f}_{\mathrm{ML}} = \left(\mathbf{Q}'\mathbf{Q}\right)^{-1}\mathbf{Q}'\left(\mathbf{I} - \beta_{\mathrm{ML}}\mathbf{D}\right)\mathbf{y} \ ; \tag{20}$$

$$\sigma^2_{\mathrm{ML}} = \frac{1}{M}\left[\mathbf{Q}'\left(\mathbf{I} - \beta_{\mathrm{ML}}\mathbf{D}\right)\mathbf{y} - \mathbf{Q}\,\mathrm{f}_{\mathrm{ML}}\right]'\left[\mathbf{Q}'\left(\mathbf{I} - \beta_{\mathrm{ML}}D\right)\mathbf{y} - \mathbf{Q}\,\mathrm{f}_{\mathrm{ML}}\right]$$

$$= \frac{1}{M}\mathbf{y}'\left(\mathbf{I} - \beta_{\mathrm{ML}}\mathbf{D}\right)\mathbf{B}\left(\mathbf{I} - \beta_{\mathrm{ML}}\mathbf{D}\right)\mathbf{y} \ , \tag{21}$$

with $\mathbf{B} = \mathbf{I} - \mathbf{Q}\left(\mathbf{Q}'\mathbf{Q}\right)^{-1}\mathbf{Q}'$. (22)

$\mathbf{B}$ is a symmetric and idempotent $M \times M$-matrix. From the third likelihood equation, after substitution of

$$\left(\mathbf{I} - \beta_{\mathrm{ML}}\mathbf{D}\right)\mathbf{y} - \mathbf{Q}\,\mathrm{f}_{\mathrm{ML}} = \mathbf{B}\left(\mathbf{I} - \beta_{\mathrm{ML}}\mathbf{D}\right)\mathbf{y}$$

and $\sigma^2_{\mathrm{ML}} = \frac{1}{M}\mathbf{y}'\left(\mathbf{I} - \mathbf{D} + \left(1 - \beta_{\mathrm{ML}}\right)\mathbf{D}\right)\mathbf{B}\left(\mathbf{I} - \mathbf{D} + \left(1 - \beta_{\mathrm{ML}}\right)\mathbf{D}\right)\mathbf{y}$,

we obtain:

$$\left(1 - \beta_{\mathrm{ML}}\right)\mathbf{y}'\mathbf{DB}\left(\mathbf{I} - \mathbf{D}\right)\mathbf{y} + \left(1 - \beta_{\mathrm{ML}}\right)^2\mathbf{y}'\mathbf{DBDy} - \frac{N}{M}\mathbf{y}'\left(\mathbf{I} - \mathbf{D}\right)\mathbf{B}\left(\mathbf{I} - \mathbf{D}\right)\mathbf{y}$$

$$- 2\frac{N}{M}\left(1 - \beta_{\mathrm{ML}}\right)\mathbf{y}'\mathbf{DB}\left(\mathbf{I} - \mathbf{D}\right)\mathbf{y} - \frac{N}{M}\left(1 - \beta_{\mathrm{ML}}\right)^2\mathbf{y}'\mathbf{DBDy} = 0 \ . \tag{23}$$

In order to proceed, it is necessary to show the identity $\mathbf{DB}\left(\mathbf{I} - \mathbf{D}\right)\mathbf{y} = \mathbf{0}$. Because the equation for f_{ML} in (20) has, for given β_{ML}, the form of an OLS estimator, the equality

$$\left(\mathbf{I} - \beta_{\mathrm{ML}}\mathbf{D}\right)\mathbf{y} = \mathbf{Q}\,\mathrm{f}_{\mathrm{ML}} + \hat{\mathbf{v}}$$

holds, with

$$\hat{\mathbf{v}} = \mathbf{B}\left(\mathbf{I} - \beta_{\mathrm{ML}}\mathbf{D}\right)\mathbf{y} \qquad \text{and} \qquad \mathbf{Q}'\hat{\mathbf{v}} = \mathbf{0}.$$

It immediately follows that

$$\left(\mathbf{I} - \mathbf{D}\right)\mathbf{y} = \left(\mathbf{I} - \mathbf{D}\right)\mathbf{Q}\,\mathrm{f}_{\mathrm{ML}} + \left(\mathbf{I} - \mathbf{D}\right)\hat{\mathbf{v}}.$$

Because $\mathbf{BQ}=\mathbf{0}$ and $\mathbf{BDQ}=\mathbf{B}(\mathbf{DZ} \quad \mathbf{X} \quad \mathbf{DZ})=\mathbf{0}$, it follows that $\mathbf{B}(\mathbf{I}-\mathbf{D})\mathbf{Q}=\mathbf{0}$. Furthermore, with $\mathbf{Q}'\hat{\mathbf{v}}=\mathbf{0}$, we also have $\mathbf{Q}'\mathbf{D}\hat{\mathbf{v}}=\mathbf{0}$. This yields:

$$\mathbf{DB}(\mathbf{I}-\mathbf{D})\mathbf{y}=\mathbf{DB}((\mathbf{I}-\mathbf{D})\mathbf{Q}\,\mathrm{f}_{\mathrm{ML}}+(\mathbf{I}-\mathbf{D})\hat{\mathbf{v}})=\mathbf{DB}(\mathbf{I}-\mathbf{D})\hat{\mathbf{v}}=\mathbf{D}(\mathbf{I}-\mathbf{D})\hat{\mathbf{v}}=\mathbf{0} \ .$$

Thus equation (23) simplifies to

$$(1-\beta_{\mathrm{ML}})^2=\frac{N}{M-N}\frac{\mathbf{y}'(\mathbf{I}-\mathbf{D})\mathbf{B}(\mathbf{I}-\mathbf{D})\mathbf{y}}{\mathbf{y}'\mathbf{DBDy}} \ .$$

Since $|\beta|<1$, the estimator β_{ML} is uniquely determined. The other estimators are obtained by substituting in equations (20) to (22). In the appendix it is shown that the second order conditions for a local maximum hold. The results are summarized as follows:

Proposition 2: *The maximum likelihood estimators* $\beta_{\mathrm{ML}},\sigma^2_{\mathrm{ML}},\mathrm{g}_{\mathrm{ML}},\mathrm{d}_{\mathrm{ML}}$ *and* h_{ML} *for system (15) are given by*

$$\beta_{\mathrm{ML}}=1-\sqrt{e} \ , \ \textit{with} \qquad e=\frac{N}{M-N}\frac{\mathbf{y}'(\mathbf{I}-\mathbf{D})\mathbf{B}(\mathbf{I}-\mathbf{D})\mathbf{y}}{\mathbf{y}'\mathbf{DBDy}} \ ;$$

$$\sigma^2{}_{\mathrm{ML}}=\frac{1}{M-N}\mathbf{y}'(\mathbf{I}-\mathbf{D})\mathbf{B}(\mathbf{I}-\mathbf{D})\mathbf{y} \ ;$$

$$\mathrm{f}_{\mathrm{ML}}=(\mathrm{g}'_{\mathrm{ML}} \quad \mathrm{d}'_{\mathrm{ML}} \quad \mathrm{h}'_{\mathrm{ML}})'=(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'(\mathbf{I}-\beta_{\mathrm{ML}}\mathbf{D})\mathbf{y} \ .$$

3.4. Calculating the Estimators

It is readily seen that e can be interpreted as the ratio of two sums of squared residuals from OLS estimates. The ML estimation of the key parameter β can be carried out in three steps:

a) First, an OLS estimate is run with $\mathbf{Q}$ as regressor and the deviations $(\mathbf{I}-\mathbf{D})\mathbf{y}$ from the group average as regressand. The sum of squared residuals for this auxiliary regression,

$$\hat{\mathbf{v}}_{(1)}'\hat{\mathbf{v}}_{(1)}=\mathbf{y}'(\mathbf{I}-\mathbf{D})\mathbf{B}(\mathbf{I}-\mathbf{D})\mathbf{y} \ ,$$

yields the ML-estimator for the variance σ^2, after correcting with $1/(M-N)$. The expression can also be generated by means of a regression on the $M\times K$-matrix $(\mathbf{I}-\mathbf{D})\mathbf{Z}$.

b) A second auxiliary regression uses $\mathbf{Dy}$, the group averages of the endogenous variables, as regressand vector and again $\mathbf{Q}$ as regressor matrix. This yields the SSR

$$\hat{\mathbf{v}}_{(2)}'\hat{\mathbf{v}}_{(2)}=\mathbf{y}'\mathbf{DBDy} \ .$$

It is possible to generate the same expression using the residuals of a regression on the $(K+J)\times M$ -matrix $[\mathbf{DZ} \quad \mathbf{X}]$.

c) Finally, the ratio of the SSRs in a) and b) is calculated and multiplied by $N/(M-N)$:

$$e=\frac{N}{M-N}\frac{\hat{\mathbf{v}}_{(1)}'\hat{\mathbf{v}}_{(1)}}{\hat{\mathbf{v}}_{(2)}'\hat{\mathbf{v}}_{(2)}} \ .$$

The square root of this expression is equal to $1-\beta_{\mathrm{ML}}$.

This account shows *that the structure of the ML-estimator corresponds exactly to the solution of the identification problem worked out in Section 2.6 above*, as shown in equation (12). In order to calculate $\mathrm{f}_{\mathrm{ML}} = (\mathbf{Q'Q})^{-1}\mathbf{Q'}(\mathbf{I}-\beta_{\mathrm{ML}}\mathbf{D})\mathbf{y}$, it is necessary to generate $(\mathbf{I}-\beta_{\mathrm{ML}}\mathbf{D})\mathbf{y}$ first. This is done by subtracting from each observation y_{ji} the amount $\beta_{\mathrm{ML}}\bar{y}_{\mathrm{j}}$. Then a regression of this transformed variable on the exogenous variables in $\mathbf{Q}$ is performed.

3.5. Characterizing the Estimators

We now turn our attention to the finite sample distributions of β_{ML} and σ^2_{ML}. Consider first the SSR $\hat{\mathbf{v}}_{(1)}'\hat{\mathbf{v}}_{(1)}$. It is $(\mathbf{I}-\mathbf{D})\mathbf{y} = (\mathbf{I}-\mathbf{D})(\mathbf{Q}\mathrm{f}+\mathbf{v})$, and as $\mathbf{B}(\mathbf{I}-\mathbf{D})\mathbf{Q} = \mathbf{0}$, this leads to

$$\hat{\mathbf{v}}_{(1)}'\hat{\mathbf{v}}_{(1)} = \mathbf{v}'\mathbf{B}_{(1)}\mathbf{v}, \text{ with } \mathbf{B}_{(1)} = (\mathbf{I}-\mathbf{D})\mathbf{B}(\mathbf{I}-\mathbf{D}) \quad .$$

The matrix $\mathbf{B}_{(1)}$ is symmetric and idempotent. Its rank is equal to its trace:

$$\begin{aligned}\mathrm{Rank}\,\mathbf{B}_{(1)} &= \mathrm{tr}\left\{(\mathbf{I}-\mathbf{D}) - (\mathbf{I}-\mathbf{D})\mathbf{Q}(\mathbf{Q'Q})^{-1}\mathbf{Q'}(\mathbf{I}-\mathbf{D})\right\} \\ &= \mathrm{tr}(\mathbf{I}-\mathbf{D}) - \mathrm{tr}(\mathbf{Q'Q})^{-1}\mathbf{Q'Q} + \mathrm{tr}(\mathbf{Q'Q})^{-1}\mathbf{Q'DQ} \;.\end{aligned}$$

Now $\mathrm{tr}(\mathbf{I}-\mathbf{D}) = \sum_{\mathrm{j}=1}^{N} \mathrm{tr}\left(\mathbf{I}_{(M_{\mathrm{j}})} - \frac{1}{M_{\mathrm{j}}}\mathbf{1}_{(M_{\mathrm{j}})}\mathbf{1}_{(M_{\mathrm{j}})}'\right) = M - N$ and $\mathrm{tr}(\mathbf{Q'Q})^{-1}\mathbf{Q'Q} = 2K + J$. Furthermore, $\mathbf{DX} \equiv \mathbf{X}$ and $\mathbf{DQ} = [\mathbf{DZ} \quad \mathbf{X} \quad \mathbf{DZ}]$ lead to:

$$(\mathbf{Q'Q})^{-1}\mathbf{Q'DQ} = \begin{pmatrix} \mathbf{Z'DZ} & \mathbf{Z'X} & \mathbf{Z'DZ} \\ \mathbf{X'Z} & \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'DZ} & \mathbf{Z'X} & \mathbf{Z'Z} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Z'DZ} & \mathbf{Z'X} & \mathbf{Z'DZ} \\ \mathbf{X'Z} & \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'DZ} & \mathbf{Z'X} & \mathbf{Z'DZ} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{(\mathrm{K})} & \mathbf{0} & \mathbf{I}_{(\mathrm{K})} \\ \mathbf{0} & \mathbf{I}_{(\mathrm{J})} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{24}$$

so that $\mathrm{tr}(\mathbf{Q'Q})^{-1}\mathbf{Q'DQ} = K + J$. Therefore one obtains:

$$\mathrm{tr}\,\mathbf{B}_{(1)} = M - N - K \quad .$$

For the distribution of the sum of squared residuals, it follows that:

<u>Proposition 3:</u> *The expression*

$$\frac{1}{\sigma^2}\hat{\mathbf{v}}_{(1)}'\hat{\mathbf{v}}_{(1)} \tag{25}$$

is distributed χ^2 *with* $M - N - K$ *degrees of freedom. The statistic*

$$s^2 = \frac{1}{M-N-K}\mathbf{v}_{(1)}'\mathbf{v}_{(1)} = \frac{M-N}{M-N-K}\sigma^2_{\mathrm{ML}}$$

is an unbiased estimator for σ^2 *with variance* $\frac{2\sigma^4}{M-N-K}$. *The estimator* σ^2_{ML} *is consistent for* $N \to \infty, \frac{N}{M} \to \rho$ *and* $\frac{1}{N}\mathbf{Q'Q} \to \mathbf{C}$ *and* $\mathrm{Rank}\,\mathbf{C} = 2K + J$.

<u>Proof:</u> The elements of $\mathbf{v}$ are stochastically independent and distributed as $\mathrm{N}(0,\sigma^2)$. Since matrix $\mathbf{B}_{(1)}$ is idempotent with rank $M - N - K$, the expression $\frac{1}{\sigma^2}\mathbf{v}'\mathbf{B}_{(1)}\mathbf{v}$ is χ^2 distributed with the same number of degrees of freedom. The second part of the proposition follows from taking into account the expectation and variance associated with the χ^2-distribution. ■

With regard to β_{ML}, the distribution of the SSR $\hat{\mathbf{v}}_{(2)}'\hat{\mathbf{v}}_{(2)}$ must be determined in an analogous fashion. By definition it is

$$\mathbf{Dy} = \frac{1}{1-\beta}\left(\mathbf{DQ}\,\mathrm{f} + \mathbf{Dv}\right) \quad , \tag{26}$$

and because of $\mathbf{BDQ} = \mathbf{0}$ we obtain:

$$\hat{\mathbf{v}}_{(2)}'\hat{\mathbf{v}}_{(2)} = \frac{1}{(1-\beta)^2}\mathbf{v}'\mathbf{B}_{(2)}\mathbf{v}, \text{ with } \mathbf{B}_{(2)} = \mathbf{DBD} \quad .$$

The matrix $\mathbf{B}_{(2)}$, too, is symmetric and idempotent. Its rank can be calculated as

$$\text{Rank}\,\mathbf{B}_{(2)} = \text{tr}\,\mathbf{D} - \text{tr}\left(\mathbf{Q}'\mathbf{Q}\right)^{-1}\mathbf{Q}'\mathbf{DQ} = N - K - J \quad .$$

Consequently the expression

$$\frac{(1-\beta)^2}{\sigma^2}\hat{\mathbf{v}}_{(2)}'\hat{\mathbf{v}}_{(2)} = \frac{1}{\sigma^2}\mathbf{v}'\mathbf{B}_{(2)}\mathbf{v} \tag{27}$$

follows a χ^2-distribution with $N - K - J$ degrees of freedom. Finally we have

$$\mathbf{B}_{(1)} \cdot \mathbf{B}_{(2)} = (\mathbf{I} - \mathbf{D})\mathbf{B}\underbrace{(\mathbf{I} - \mathbf{D}) \cdot \mathbf{D}}_{\mathbf{0}}\mathbf{BD} = \mathbf{0} \quad .$$

The quadratic forms $\mathbf{v}'\mathbf{B}_{(1)}\mathbf{v}$ and $\mathbf{v}'\mathbf{B}_{(2)}\mathbf{v}$ are thus stochastically independent and the ratio of the two expressions (25) and (27), corrected by their respective degrees of freedom, follows an F-distribution:

$$(1-\beta)^2\frac{M-N-K}{N-K-J}\frac{\hat{\mathbf{v}}_{(2)}'\hat{\mathbf{v}}_{(2)}}{\hat{\mathbf{v}}_{(1)}'\hat{\mathbf{v}}_{(1)}} \sim \mathrm{F}_{N\text{-}K\text{-}J,M\text{-}N\text{-}K} \quad .$$

For $N \to \infty$ and $\frac{N}{M} \to \rho$, this expression converges stochastically to unity. Therefore setting

$$q = \frac{N-K-J}{M-N-K} \cdot \frac{\hat{\mathbf{v}}_{(1)}'\hat{\mathbf{v}}_{(1)}}{\hat{\mathbf{v}}_{(2)}'\hat{\mathbf{v}}_{(2)}} = \frac{(N-K-J)/N}{(M-N-K)/(M-N)} \cdot e$$

implies immediately:

$$\operatorname*{plim}_{\substack{N\to\infty \\ \frac{N}{M}\to\rho}} q = \operatorname*{plim}_{\substack{N\to\infty \\ \frac{N}{M}\to\rho}} e = (1-\beta)^2 \quad .$$

Let $\mathrm{F}_{n,m,\alpha}$ be the value that with probability α is exceeded by a random variable following an F-distribution with *n* and *m* degrees of freedom. Then we can state

Proposition 4: *For* $N \to \infty$, $\frac{N}{M} \to \rho$ *and* $\frac{1}{N}\mathbf{Q'Q} \to \mathbf{C}$, *with* $\text{Rank}\,\mathbf{C} = 2K + J$, *the estimator* β_{ML} *is consistent. The equation*

$$W\left\{1 - \sqrt{q \cdot F_{N-K-J,M-N-K,\frac{\alpha}{2}}} \le \beta \le 1 - \sqrt{q/F_{M-N-K,N-K-J,\frac{\alpha}{2}}}\right\} = 1 - \alpha$$

defines a confidence interval for β.

Proof: The second part of the proposition follows from

$$W\left\{F_{N-K-J,M-N-K,1-\frac{\alpha}{2}} \le \frac{(1-\beta)^2}{q} \le F_{N-K-J,M-N-K,\frac{\alpha}{2}}\right\} = 1 - \alpha \quad ,$$

taking into account the identity $F_{N-K-J,M-N-K,1-\frac{\alpha}{2}} = \frac{1}{F_{M-N-K,N-K-J,\frac{\alpha}{2}}}$. ■

Ultimately, for the ML estimator $f_{\text{ML}} = (g_{\text{ML}} \quad d'_{\text{ML}} \quad h'_{\text{ML}})'$, the following holds:

Proposition 5: *For* $N \to \infty$, $\frac{N}{M} \to \rho$ *and* $\frac{1}{N}\mathbf{Q'Q} \to \mathbf{C}$, *with* $\text{Rank}\,\mathbf{C} = 2K + J$, *the estimator* f'_{ML} *is consistent for* $f = (g \quad d' \quad h')'$.

Proof: Under the given conditions, the OLS estimator $(\mathbf{Q'Q})^{-1}\mathbf{Q'}(\mathbf{I} - \beta\mathbf{D})\mathbf{y}$ is consistent for f in equation (16). The parameter β is not known, but β_{ML} is available. The difference vector,

$$f_{ML} - (\mathbf{Q'Q})^{-1}\mathbf{Q'}(\mathbf{I} - \beta\mathbf{D})\mathbf{y} = (\beta - \beta_{ML})(\mathbf{Q'Q})^{-1}\mathbf{Q'Dy},$$

is the product of the estimation error $\beta - \beta_{\text{ML}}$ and an OLS estimator for equation (26), with $\mathbf{Q}$ as design matrix. Because $\mathbf{DQ} = [\mathbf{DZ} \quad \mathbf{X} \quad \mathbf{DZ}]$, the magnitude $(\mathbf{Q'Q})^{-1}\mathbf{Q'Dy}$ converges in probability to the fixed vector $\frac{1}{1-\beta}[g + h' \quad d' \quad 0']'$. Finally, $\beta - \beta_{\text{ML}}$ is stochastically convergent to zero and the proposition follows. ■

3.6. Asymptotic Distribution of the Estimators

The information on the finite sample properties of β_{ML} is very valuable in the given context: The parameter β stands for endogenous social effects and the distribution of its estimator allows us to distinguish between different hypotheses concerning the correlation of behaviour in social groups. Yet some questions remain. The estimator f_{ML} is the product of a normally distributed variate and one that follows an F distribution. We know that it is consistent, but this is not enough for hypothesis testing. Besides, we would like to have the covariance matrix of the estimators.

It is well known that, under weak conditions, ML estimators are asymptotically normal with the inverse of the information matrix as covariance matrix, if the observations are independent and identically distributed.[21] A similar convergence result can also be derived in the present

---

[21] See Cramér (1946). For explanations and proofs, see, e.g., Theil (1971).

case. It is convenient to refine the notation and slightly strengthen the assumptions concerning the parameter space:

A7) Let $\theta' = \begin{pmatrix} \phi' & \sigma^2 & \beta \end{pmatrix}$. The true parameters $\theta_0' = \begin{pmatrix} \phi_0' & \sigma_0^2 & \beta_0 \end{pmatrix}$ are in the interior of the parameter space Θ. The latter is a closed and convex interval of the $\mathbb{R}^{2K+J+2}$, with $\sigma^2 > 0$ and $|\beta| < 1$ for all $\theta \in \Theta$.

This notation explicitly differentiates between *permitted* parameters, $\theta \in \Theta$, and the *true* parameters, θ_0. For a formal statement on the asymptotic distribution of the ML-estimators we need the following lemma:

Lemma 1: *Let the elements* q_{tk}, $t = 1,\ldots,M$; $k = 1,\ldots,2K+J$ *of matrix* $\mathbf{Q}$ *satisfy*

$$|q_{tk}| < \bar{q} < \infty \quad \text{, and} \quad \lim_{\substack{M\to\infty \\ \frac{N}{M}\to\rho}} \frac{1}{N}\mathbf{Q'Q} = \mathbf{C}, \text{ with } \operatorname{rank}\mathbf{C} = 2K+J \ .$$

Then, for $N \to \infty$, $\frac{N}{M} \to \rho$, *the following holds:*

$$\frac{1}{\sqrt{N}} \left.\frac{\partial l}{\partial \theta}\right|_{\theta_0} \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \overline{\mathbf{R}}(\theta_0)\right) ,$$

where

$$\overline{\mathbf{R}}(\theta) = \lim_{\substack{N\to\infty \\ \frac{N}{M}\to\rho}} \frac{1}{N}\mathbf{R}_N(\theta) \qquad \text{and} \qquad \mathbf{R}_N(\theta) = -\underset{\theta_0}{\mathrm{E}} \frac{\partial^2 l(\cdot)}{\partial\theta\partial\theta'} \ .$$

Specifically, one obtains:

$$\mathbf{R}_N(\theta_0) = \frac{1}{\sigma_0^2}\begin{pmatrix} \mathbf{Q'Q} & \mathbf{0} & \frac{1}{1-\beta_0}\mathbf{Q'DQ}\phi_0 \\ \mathbf{0'} & \frac{M}{2\sigma_0^2} & \frac{N}{1-\beta_0} \\ \frac{1}{1-\beta}\phi_0'\mathbf{Q'DQ} & \frac{N}{1-\beta_0} & \frac{1}{(1-\beta_0)^2}\left(\phi_0'\mathbf{Q'DQ}\phi_0 + 2N\sigma_0^2\right) \end{pmatrix} . \tag{28}$$

Proof: First, calculating the negative of the expected Hesse matrix, $-\underset{\theta_0}{\mathrm{E}}\frac{\partial^2 l(\theta|\mathbf{y})}{\partial\theta\partial\theta'}$, one obtains a function $\mathbf{R}_N(\theta)$ of θ. Evaluating this function at θ_0 yields $\mathbf{R}_N(\theta_0)$, the so-called *information matrix*. Second, by substituting the system equation (16) into the derivatives of the likelihood function as stated in (17) to (19), one obtains for θ_0, the true parameter values:

$$\left.\frac{\partial l}{\partial \theta}\right|_{\theta_0} = \begin{pmatrix} \partial l/\partial\phi \\ \partial l/\partial\sigma^2 \\ \partial l/\partial\beta \end{pmatrix}_{\theta_0} = \begin{pmatrix} \frac{1}{\sigma_0^2}\mathbf{Q'v} \\ \frac{1}{2\sigma_0^4}\left(\mathbf{v'v} - M\sigma_0^2\right) \\ \frac{1}{(1-\beta_0)\sigma_0^2}\left(\phi_0'\mathbf{Q'Dv} + \mathbf{v'Dv} - N\sigma_0^2\right) \end{pmatrix} . \tag{29}$$

In this expression, $\partial l/\partial f$ is a $(2K+J)$-vector; $\partial l/\partial \sigma^2$ and $\partial l/\partial \beta$ are scalars. The mathematical expectation of (29) is equal to the zero vector. Calculating the covariance matrix leads to the same matrix $\mathbf{R}_N(q_0)$ as given in (28):

$$\underset{q_0}{\mathrm{cov}}\left(\left.\frac{\partial l}{\partial f}\right|_{q_0}\right)=\underset{q_0}{\mathrm{E}}\left(\left.\frac{\partial l}{\partial q}\right|_{q_0}\cdot\left.\frac{\partial l}{\partial q'}\right|_{q_0}\right)=-\underset{q_0}{\mathrm{E}}\frac{\partial^2 l(q|\mathbf{y})}{\partial q\partial q'}=\mathbf{R}_N(q_0)\quad .$$

This identity, of course, holds quite generally; see, e.g., Theil (1971). A random vector is asymptotically normal if the distribution of any nontrivial linear combination of its elements converges to the univariate normal. Consider the components of (34). Every element of the subvector $\partial l/\partial f|_{q_0}$ is normal. In $\partial l/\partial \sigma^2|_{q_0}$, the magnitude $\frac{1}{\sigma_0^2}\mathbf{v}'\mathbf{v}$ follows a χ^2 distribution with M degrees of freedom. Finally, in $\partial l/\partial \sigma^2|_{q_0}$ the expression $f_0'\mathbf{Q}'\mathbf{D}\mathbf{v}$ is normal and $\frac{1}{\sigma_0^2}\mathbf{v}'\mathbf{D}\mathbf{v}$ is χ^2 distributed with N degrees of freedom, because $\mathbf{D}$ is idempotent with rank N. Every linear combination of the elements of vector $\frac{1}{\sqrt{N}}\left.\frac{\partial l}{\partial q}\right|_{q_0}$ is a linear combination of normal or asymptotically normal random variables. Thus, this vector is asymptotically normal with the parameters stated above. ■

With this preliminary work done, the asymptotic distribution of the maximum likelihood estimators can now be characterized as follows:

<u>Proposition 6:</u> *The conditions of Lemma 1 hold. Let* $q_{ML}'=(f_{ML}'\quad \sigma^2_{ML}\quad \beta_{ML})$, *and* $x'=(g' \; h' \quad d' \quad 0')_{(2K+J)}$. *Then for* $N\to\infty, \frac{N}{M}\to\rho$ *the vector* $\sqrt{N}(q_{ML}-q_0)$ *is asymptotically normal with expectation zero and covariance matrix* $\overline{\mathbf{R}}(q_0)^{-1}=\lim_{\substack{N\to\infty\\ \frac{N}{M}\to\rho}} N\mathbf{R}_N(q_0)^{-1}$, *where*

$$\mathbf{R}_N(q_0)^{-1}=\sigma_0^2\begin{pmatrix}(\mathbf{Q}'\mathbf{Q})^{-1}\left[\frac{1}{2\sigma_0^2N\left(1-\frac{N}{M}\right)}x_0x_0'\right] & \frac{1}{M-N}x_0 & -\frac{1-\beta_0}{2\sigma_0^2N\left(1-\frac{N}{M}\right)}x_0\\ \frac{1}{M-N}x_0 & 2\frac{\sigma_0^2}{M-N} & -\frac{1-\beta_0}{M-N}\\ -\frac{1-\beta_0}{2\sigma_0^2N\left(1-\frac{N}{M}\right)}x_0 & -\frac{1-\beta_0}{M-N} & \frac{(1-\beta_0)^2}{2\sigma_0^2N\left(1-\frac{N}{M}\right)}\end{pmatrix}.$$

<u>Proof:</u>[22] The gradient of the log likelihood in q_{ML} can be written as a Taylor series:

$$\left.\frac{\partial l}{\partial q}\right|_{q_{ML}}=\left.\frac{\partial l}{\partial q}\right|_{q_0}+(q_{ML}-q_0)\left.\frac{\partial^2 l}{\partial q\partial q'}\right|_{q^*}\quad ,$$

---

22 The proof follows Amemiya (1985), pp. 111-113 and pp. 121-123.

with q^* between q_0 and q_{ML}. As a necessary condition for an interior solution, q_{ML} satisfies the likelihood equations. Therefore, the left-hand side is equal to the zero vector:

$$\frac{1}{\sqrt{N}} \frac{\partial l}{\partial q}\bigg|_{q_0} = -\frac{1}{N} \frac{\partial^2 l}{\partial q \partial q'}\bigg|_{q^*} \sqrt{N}\left(q_{ML} - q_0\right) \ . \tag{30}$$

Now let $l_j\left(q \middle| \mathbf{y}_j\right)$ be the log likelihood as it is computed from the marginal density for the vector of endogenous variables in group G_j. Because the $\mathbf{y}_j$ are independent, we have $l\left(q \middle| \mathbf{y}\right) = \sum_{j=1}^{N} l_j\left(q \middle| \mathbf{y}_j\right)$ and therefore

$$-\frac{1}{N} \frac{\partial^2 l(\mathbf{y}, q)}{\partial q \partial q'} = -\frac{1}{N} \sum_{j=1}^{N} \frac{\partial^2 l_j(\mathbf{y}, q)}{\partial q \partial q'} \ .$$

The elements of the matrix on the left are arithmetic means of N independent variables with bounded variance. The strong law of large numbers makes them converge to the mathematical expectation. The matrix $-\underset{q_0}{E} \frac{\partial^2 l(\cdot)}{\partial \theta \partial \theta'} = \mathbf{R}_N(q)$, on the other hand, converges uniformly to the continuous function $\overline{\mathbf{R}}(q)$, as $\frac{1}{N}\mathbf{Q}'\mathbf{Q}$ converges to a fixed matrix $\mathbf{C}$. Under these conditions it is sufficient[23] for

$$\underset{\substack{N \to \infty \\ \frac{N}{M} \to \rho}}{\text{plim}} -\frac{1}{N} \frac{\partial^2 l}{\partial q \partial q'}\bigg|_{q^*} = \overline{\mathbf{R}}(q_0)$$

that q^* converges in probability towards q_0. As q_{ML} is consistent for q_0, this is indeed the case. In the appendix it is shown that $\overline{\mathbf{R}}(q_0)$ is positive definite. The determinant of a matrix is a continuous function of the elements, so $\overline{\mathbf{R}}(q)$ is non-singular in the neighbourhood of q_0. Now consider the left-hand side of (35). Lemma 1 states that the distribution of this expression converges to $N\left(\mathbf{0}, \overline{\mathbf{R}}(q_0)\right)$. This completes the proof. The appendix shows how to calculate the asymptotic covariance matrix as stated in the proposition. ■

4. Summary and Evaluation

Interactions in social groups can be the reason why the outcome of a variable at the individual level is strongly influenced by the average outcome of the same variable in the environment. This average, conversely, is determined by the same exogenous characteristics as the individual realizations. Social effects can therefore act as an amplifier for systematic differences between persons. The average realizations of groups with different exogenous characteristics do not reveal which part of the observed differences can be attributed to social effects. Manski works out this problem *in nuce*. Actually, it is striking how social scientists routinely make an *a priori* decision in favour of one of the competing hypothesis.

---

23 See Amemiya (1985), p. 113.

Yet Manski's forceful exposition obstructs the view on possible solutions to these problems. It was shown here that not only systematic differences between individuals are magnified, but also random differences in the mean realization of the endogenous variable. The social effects give rise to a *group identity:* The outcome of groups with identical characteristics deviate in a statistically conspicuous way. This allows identification even under very unfavourable circumstances. As demonstrated above, the estimation procedure itself is relatively simple.

The modification of Manski's metamodel can be regarded as a variant of the network model discussed in Section 2.7, with a special matrix **W** and extended to include exogenous social effects and correlated effects. Manski briefly mentions this class of models. His criticism is this: The network model is capable of capturing social interactions in smaller groups of friends, colleagues or members of the same household. For large social groups like neighbourhoods, researchers usually have to use random samples. A literal interpretation of the network analytic approach would then amount to assuming that the members of the random sample know each other and choose their outcome only *after* they have been selected into the sample.

This argument is not complete. On the one hand, random samples can well be used if the structure of the interaction matrix is appropriate. Case (1991) investigates interregional interdependencies in consumer demand and for each region uses a random sample; see also the estimation procedures in Doreian (1981). The model elaborated here is not affected either – the procedure outlined in Section 2.6 will provide a consistent estimator even when there is only a random sample from each group. Furthermore, the objection does not constitute an identification problem, but primarily reflects the difficulties in finding adequate data.[24]

Manski replaces stochastic interactions between individuals by a functional relationship. By relaxing this idealization, the identification problem that arises can generally be solved.

Appendix: Second Order Conditions and Asymptotic Covariance Matrix

The likelihood equations have a unique solution. To make sure that this solution characterizes a local maximum, the Hesse matrix evaluated at the solution $\mathrm{q}'_{\mathrm{ML}} = \left(\mathrm{f}'_{\mathrm{ML}} \quad \sigma^2_{\mathrm{ML}} \quad \beta_{\mathrm{ML}}\right)$ is tested for sign definiteness:

$$\left.\frac{\partial^2 l}{\partial \mathrm{q}\partial \mathrm{q}'}\right|_{\mathrm{ML}} = -\frac{1}{\sigma^2_{\mathrm{ML}}}\mathbf{H}^*, \text{ with } \mathbf{H}^* = \begin{pmatrix} \mathbf{Q}'\mathbf{Q} & \mathbf{0} & \dfrac{\mathbf{Q}'\mathbf{D}\mathbf{Q}\mathrm{f}_{\mathrm{ML}}}{1-\beta_{\mathrm{ML}}} \\ \mathbf{0}' & \dfrac{M}{2\sigma^2_{\mathrm{ML}}} & \dfrac{N}{1-\beta_{\mathrm{ML}}} \\ \dfrac{\mathrm{f}'_{\mathrm{ML}}\mathbf{Q}'\mathbf{D}\mathbf{Q}}{1-\beta_{\mathrm{ML}}} & \dfrac{N}{1-\beta_{\mathrm{ML}}} & \dfrac{\mathrm{f}'_{\mathrm{ML}}\mathbf{Q}'\mathbf{D}\mathbf{Q}\mathrm{f}_{\mathrm{ML}} + 2N\sigma^2_{\mathrm{ML}}}{(1-\beta_{\mathrm{ML}})^2} \end{pmatrix}. \quad (31)$$

Let $\mathbf{p}' = (\mathbf{p}_1' \quad p_2 \quad p_3) \in \mathrm{R}^{2K+J+2}$. Here, $\mathbf{p}_1'$ is a $(2K+J)$-vector and p_2 and p_3 are scalars. Consider the quadratic form

24 See Marsden (1990) on the collection of data for network models.

$$\mathbf{p}'\mathbf{H}^*\mathbf{p} = \mathbf{p_1}'\mathbf{Q}'\mathbf{Q}\mathbf{p_1} + p_2^2\frac{M}{2\sigma_{\mathrm{ML}}^2} + p_3^2\frac{1}{(1-\beta_{\mathrm{ML}})^2}\left(\phi'_{\mathrm{ML}}\,\mathbf{Q}'\mathbf{D}\mathbf{Q}\phi_{\mathrm{ML}} + 2N\sigma_{\mathrm{ML}}^2\right) +$$

$$+\frac{1}{1-\beta_{\mathrm{ML}}}\mathbf{p_1}'\mathbf{Q}'\mathbf{D}\mathbf{Q}\phi_{\mathrm{ML}}\,p_3 + p_3\frac{1}{1-\beta_{\mathrm{ML}}}\phi'_{\mathrm{ML}}\,\mathbf{Q}'\mathbf{D}\mathbf{Q}\mathbf{p_1} + 2p_2p_3\frac{N}{1-\beta_{\mathrm{ML}}}$$

$$=\left[\mathbf{Q}\mathbf{p_1} + \frac{1}{1-\beta_{\mathrm{ML}}}\mathbf{D}\mathbf{Q}\phi_{\mathrm{ML}}\,p_3\right]'\left[\mathbf{Q}\mathbf{p_1} + \frac{1}{1-\beta_{\mathrm{ML}}}\mathbf{D}\mathbf{Q}\phi_{\mathrm{ML}}\,p_3\right]$$

$$+\left[\sqrt{\frac{M}{2\sigma_{\mathrm{ML}}^2}}\,p_2 + \sqrt{\frac{2\sigma_{\mathrm{ML}}^2}{M}}\,\frac{N}{1-\beta_{\mathrm{ML}}}\,p_3\right]^2 + 2\left(1-\frac{N}{M}\right)\frac{N\sigma_{\mathrm{ML}}^2}{(1-\beta_{\mathrm{ML}})^2}\,p_3\ .$$

This sum of squares is positive whenever $\mathbf{p}$ is not the zero vector, so $\mathbf{H}^*$ is positive definite. Then $\left.\frac{\partial^2 l}{\partial\theta\partial\theta'}\right|_{\mathrm{ML}}$ is negative definite, and θ_{ML} specifies a unique local maximizer.

For the determination of the asymptotic covariance matrix of θ_{ML} it is necessary to examine the expectation of the Hesse matrix, i.e. the matrix $\underset{\theta_0}{\mathrm{E}}\frac{\partial^2 l(\theta|\mathbf{y})}{\partial\theta\partial\theta'}$. The likelihood is evaluated at $\theta\in\Theta$, with the expectation being based on the distribution of $\mathbf{y}$ according to the true parameters:

$$\mathbf{y} = (\mathbf{I}-\beta_0\mathbf{D})^{-1}(\mathbf{Q}\phi_0+\mathbf{v}) = \left(\mathbf{I}+\frac{\beta_0}{1-\beta_0}\mathbf{D}\right)(\mathbf{Q}\phi_0+\mathbf{v})\ , \text{ with } \mathbf{v}\sim \mathrm{N}(\mathbf{0},\sigma_0^2\mathbf{I})\ .$$

Direct calculation shows that all elements of $\underset{\theta_0}{\mathrm{E}}\frac{\partial^2 l(\theta|\mathbf{y})}{\partial\theta\partial\theta'}$ are continuous functions of ϕ,σ^2 and β in the whole parameter space. If, for $N\to\infty$ and $\frac{N}{M}\to\rho$, the matrix $\frac{1}{N}\mathbf{Q}'\mathbf{Q}$ converges to a fixed matrix $\mathbf{C}$, then $\frac{1}{N}\mathbf{Q}'\mathbf{D}\mathbf{Q}$ also converges, and all the components of

$$\overline{\mathbf{R}}(\theta) = -\lim_{\substack{N\to\infty\\ \frac{N}{M}\to\rho}}\underset{\theta_0}{\mathrm{E}}\frac{1}{N}\frac{\partial^2 l(\theta,\mathbf{y})}{\partial\theta\partial\theta'}$$

are continuous functions of ϕ,σ^2 and β in the whole parameter space. Evaluation of the expected negative Hesse matrix, $-\underset{\theta_0}{\mathrm{E}}\frac{\partial^2 l(\theta,\mathbf{y})}{\partial\theta\partial\theta'}$, in θ_0 yields the information matrix as it was defined in Lemma 1. The estimator θ_{ML} is consistent for θ_0. Thus it comes as no surprise that the matrix $\mathbf{R}_N(\theta_0)$ can be formally derived from $-\left.\frac{\partial^2 l}{\partial\theta\partial\theta'}\right|_{\mathrm{ML}}$ by simply replacing the magnitudes ϕ_{ML}, σ_{ML}^2 and β_{ML} in (31) by the corresponding true parameters. The matrix $-\left.\frac{\partial^2 l}{\partial\theta\partial\theta'}\right|_{\mathrm{ML}}$ is positive definite, and the same applies to $\mathbf{R}_N(\theta_0)$, as well as to the asymptotic matrix $\overline{\mathbf{R}}(\theta_0)$.

The asymptotic covariance matrix of the ML estimator is given by the inverse of the information matrix $\mathbf{R}_N(\theta_0)$. With the help of the Gauss algorithm one obtains

$\mathbf{R}_N(\mathrm{q}_0)^{-1} =$

$$
= \sigma_0^2 \begin{pmatrix}
(\mathbf{Q'Q})^{-1}\left[\mathbf{I} + \dfrac{1}{2\sigma_0^2 N\left(1-\dfrac{N}{M}\right)}\mathbf{Q'DQ}\mathrm{f}_0\mathrm{f}_0'\mathbf{Q'DQ}(\mathbf{Q'Q})^{-1}\right] & \dfrac{1}{M-N}(\mathbf{Q'Q})^{-1}\mathbf{Q'DQ}\mathrm{f}_0 & -\dfrac{1-\beta_0}{2\sigma_0^2 N\left(1-\dfrac{N}{M}\right)}(\mathbf{Q'Q})^{-1}\mathbf{Q'DQ}\mathrm{f}_0 \\
\dfrac{1}{M-N}\mathrm{f}_0'\mathbf{Q'DQ}(\mathbf{Q'Q})^{-1} & 2\dfrac{\sigma_0^2}{M-N} & -\dfrac{1-\beta_0}{M-N} \\
-\dfrac{1-\beta_0}{2\sigma_0^2 N\left(1-\dfrac{N}{M}\right)}\mathrm{f}_0'\mathbf{Q'DQ}(\mathbf{Q'Q})^{-1} & -\dfrac{1-\beta_0}{M-N} & \dfrac{(1-\beta_0)^2}{2\sigma_0^2 N\left(1-\dfrac{N}{M}\right)}
\end{pmatrix}.
$$

Consideration of identity (29) furthermore leads to

$$
(\mathbf{Q'Q})^{-1}\mathbf{Q'DQ}\mathrm{f} = \begin{pmatrix} \mathbf{I}_{(\mathbf{K})} & \mathbf{0} & \mathbf{I}_{(\mathbf{K})} \\ \mathbf{0} & \mathbf{I}_{(\mathbf{J})} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathrm{g} \\ \mathrm{d} \\ \mathrm{h} \end{pmatrix} = \begin{pmatrix} \mathrm{g} + \mathrm{h} \\ \mathrm{d} \\ 0 \end{pmatrix} = \mathrm{x}
$$

which in turn yields $\mathbf{R}_N(\mathrm{q})^{-1}$ as stated in Proposition 6.

References

ALESSIE, R. and KAPTEYN, A. (1991), "Formation, Interdependent Preferences and Demographic Effects in an Almost Ideal Demand System", *The Economic Journal* **101**, 404-419

ALEXANDER, K. L., ECKLAND, B. K. and GRIFFIN, L. J. (1975), "The Wisconsin Model of Socioeconomic Achievement: A Replication", *American Journal of Sociology* **81**, 324-342

ALBA, R., HANDL, J. and MÜLLER, W. (1994), "Ethnic Inequalities in the German School System" (Working Paper AB I, No. 8, Mannheimer Zentrum für Europäische Sozialforschung, Mannheim)

AMEMIYA, T. (1985), *Advanced Econometrics* (Cambridge, Mass.: Harvard University Press)

ANSELIN, L. (1988), *Spatial Econometrics: Methods and Models (*Dordrecht, Boston, London: Kluwer)

BANK, B. J. , SLAVINGS, R. L. and BIDDLE, B. J. (1990), "Effects of Peer, Faculty and Parental Influence on Students' Persistence", *Sociology of Education* **63**, 208-225

BÉNABOU, R. (1996a), "Heterogeneity, Stratification and Growth: Macroeconomic Implications of Community Structures and School Finance", *American Economic Review* **86**, 584-609

BÉNABOU, R. (1996b), "Equity and Efficiency in Human Capital Investment: The Local Connection", *Review of Economic Studies* **63**, 237-264

BORJAS, G. J. (1992), "Ethnic Capital and Intergenerational Mobility", *The Quarterly Journal of Economics* **107**, 123-150

BORJAS, G. J. (1995), "Ethnicity, Neighborhoods and Human Capital Externalities", *The American Economic Review* **85**, 365-390

BURT, R. S. (1980), "Models of Network Structure", *Annual Review of Sociology* **6**, 79-141

BURT, R. S. (1987), "Social Contagion and Innovation: Cohesion versus Structural Equivalence", *American Journal of Sociology* **92**, 1287-1335

BURT, R. S. and DOREIAN, P. (1982) "Testing a Structural Model of Perception: Conformity and Deviance with Respect to Journal Norms in Elite Sociological Methodology", *Quality and Quantity* **16**, 109-150

CASE, A. C. (1991), Spatial Patterns in Household Demand, *Econometrica* **59**, 953-965

CASE, A. C. and KATZ. L. F. (1991), "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youth" (Working Paper No. 3705, National Bureau of Economic Research, Cambridge, Mass.)

CHAMBERLAIN, G. and GRILICHES, Z. (1975), "Unobservables with a Variance-Components Structure: Ability, Schooling, and the Economic Success of Brothers", *International Economic Review* **16**, 422-448

CLIFF, A. and ORD, K. (1981), *Spatial Processes. Models and Applications* (London: Pion)

CORCORAN, M., GORDON, R., LAREN, D. and SOLON, G. (1992), "The Association between Men's Economic Status and Their Family and Community Origins", *The Journal of Human Resources* **27**, 575-601

CRAMÉR, H. (1946), *Mathematical Methods of Statistics* (Princeton: Princeton University Press)

DOREIAN, P. (1981), "Estimating Linear Models with Spatially Distributed Data", in Samuel Leinhardt (editor), *Sociological Methodology 1981* (San Francisco, Washington, London: Jossey-Bass)

DUNCAN, O. D. (1970), "Duncan's Corrections of Published Text of 'Peer Influences on Aspirations: A Reinterpretation'", *American Journal of Sociology* **75**, 1042-6

DUNCAN, O. D., HALLER, A. O. and PORTES, A. (1968), "Peer Influences of Aspirations: A Reinterpretation" *American Journal of Sociology* **74**, 119-137

DURLAUF, S. N. (1996), "A Theory of Persistent Income Inequality" *Journal of Economic Growth* **1**, 75-93

ERBRING, L. and YOUNG, A. A. (1978), "Individuals and Social Structure. Contextual Effects as Endogenous Feedback", *Sociological Methods and Research* **7**, 396-430

FISHER, F. M. (1966), *The Identification Problem in Econometrics* (New York: McGraw-Hill)

FISZ, M. (1963), *Probability Theory and Mathematical Statistics* (3. ed.; New York, London, Sydney: Wiley)

FRIEDKIN, N. (1990), "Social Networks in Structural Equation Models", *Social Psychology Quarterly* **53**, 316-328

FRIEDKIN, N. E. and COOK, K. S. (1990), "Peer Group Influence", *Sociological Methods and Research* **19**, 122-143

FRIEDKIN, N. E. and JOHNSON, E. E. C. (1990), "Social Influence and Opinions", *Journal of Mathematical Sociology* **15**, 193-205

GOLDBERGER, A. S. (1972), "Structural Equation Methods in the Social Sciences", *Econometrica* **40**, 979-1001

GRILICHES, Z. (1974), "Errors in Variables and Other Unobservables", *Econometrica* **42**, 971-998

GRILICHES, Z. (1996), "Education, Human Capital and Growth: A Personal Perspective", (Working Paper 5426, National Bureau of Economic Research, Cambridge, Mass.)

HSIAO, C. (1986), *Analysis of Panel Data* (Cambridge, New York etc.: Cambridge University Press)

HYMAN, H. H. (1942) "The Psychology of Status" *Archives of Psychology* No. 269

HYMAN, H. H. (1968) "Reference Groups", in: David L. Sills (editor), International Encyclopedia of the Social Sciences **13** (New York: MacMillan)

JENCKS, C. and MAYER, S. E. (1990) "The Social Consequences of Growing Up in a Poor Neighborhood" in: Laurence E. Lynn and Michael G.H. McGreary (editors), Inner City Poverty in the United States (Washington, D.C.: National Academy Press)

MANSKI, C. F. (1993a) "Identification Problems in the Social Sciences" *Sociological Methodology* **23**, 1-56

MANSKI, C. F. (1993b) "Identification of Endogenous Social Effects: The Reflection Problem" *Review of Economic Studies* **60**, 531-542

MANSKI, C. F. (1995) *Identification Problems in the Social Sciences* (Cambridge, Mass., London: Harvard University Press)

MARSDEN, P. V. (1990) "Network Data and Measurement" *Annual Review of Sociology* **16**, 435-463

ORD, K. (1975) "Estimation Methods for Models of Spatial Interaction" *Journal of the American Statistical Association* **70**, 120-126

RAUCH, J. E. (1993) "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities" *Journal of Urban Economics* **34**, 380-400

SEWELL, W. H. and HAUSER, R. M. (1972) "Causes and Consequences of Higher Education: Models of the Status Attainment Process" *American Journal of Agricultural Economics* **54**, 851-861

SEWELL, W. H. and HAUSER, R. M. (1975) *Education, Occupation and Earnings. Achievement in the Early Career* (New York, San Francisco, London: Academic Press)

THEIL, H. (1971) *Principles of Econometrics* (Amsterdam, London: North Holland)