Identification

Jean-Marie Dufour
Université de Montréal

Cheng Hsiao
University of Souhern California

June 30, 2007

In economic analysis, we often assume that there exists an underlying structure which has generated the observations of real-world data. However, statistical inference can relate only to characteristics of the distribution of the observed variables. Statistical models which are used to explain the behaviour of observed data typically involve parameters and statistical inference aims at making statements about these parameters. For that purpose, it is important that different values of a parameter of interest can be characterized in terms of the data distribution. Otherwise, the problem of drawing inference about this parameter is plagued by a fundamental indeterminacy and can be viewed as "ill-posed".

To illustrate, consider X as being normally distributed with mean $E(X) = \mu_1 - \mu_2$. Then $\mu_1 - \mu_2$ can be estimated using observed X. But the parameters μ_1 and μ_2 are not uniquely estimable. In fact, one can think of an infinite number of pairs (μ_i, μ_j), $i, j = 1, 2, \ldots$ $(i \neq j)$ such that $\mu_i - \mu_j = \mu_1 - \mu_2$. In order to determine μ_1 and μ_2 uniquely, we need additional prior information, such as $\mu_2 = 3\mu_1$ or some other assumption. Note, however, that inference about the variance of X remains feasible without extra assumptions.

More generally, *identification failures* – or situations that are close to it – complicate considerably the statistical analysis of models, so that tracking such failures and formulating restrictions to avoid them is an important problem of econometric modelling.

The problem of whether it is possible to draw inferences from the probability distribution of the observed variables to an underlying theoretical structure is the concern of econometric literature on identification. The first economists to raise this issue were Working (1925, 1927) and Wright (1915, 1928). The general formulations of the identification problems were made by Frisch (1934), Marschak (1942), Haavelmo (1944), Hurwicz (1950), Koopmans and Reiersøl (1950), Koopmans, Rubin and Leipnik (1950), Wald (1950), and many others. An extensive treatment of the theory of identification in simultaneous equation systems was provided by Fisher (1976). Surveys of the subject can be found in Hsiao (1983), Prakasa Rao (1992),, Bekker and Wansbeek (2001), Manski (2003), and Matzkin

(2007); see also Morgan (1990) and Stock and Trebbi (2003) on the early development of the subject.

1 Definition of parametric identification

It is generally assumed in econometrics that economic variables whose formation an economic theory is designed to explain have the characteristics of random variables. Let y be a set of such observations. A structure S is a complete specification of the probability distribution function of y. The set of all a priori possible structures, T, is called a model. In most applications, y is assumed to be generated by a parametric probability distribution function $F(y, \theta)$, where the probability distribution function F is assumed known, but the $q \times 1$ parameter vector θ is unknown. Hence, a structure is described by a parametric point θ, and a model is a set of points $A \subseteq \mathbb{R}^q$.

Definition 1 *Two structures, $S^0 = F(y, \theta^0)$ and $S^* = F(y, \theta^*)$ are said to be* observationally equivalent *if $F(y, \theta^0) = F(y, \theta^*)$ for ('almost') all possible y. A model is* identifiable *if A contains no two distinct structures which are observationally equivalent. A function of θ, $g(\theta)$, is identifiable if all observationally equivalent structures have the same value for $g(\theta)$.*

Sometimes a weaker concept of identifiability is useful.

Definition 2 *A structure with parameter value θ^0 is said to be locally identified if there exists an open neighborhood of θ^0, W, such that no other θ in W is observationally equivalent to θ^0.*

2 General results for identification in parametric models

Lack of identification reflects the fact that a random variable has the same distribution for some if not all values of the parameter. R. A. Fisher's information matrix provides a sensitivity measure of the distribution of a random variable due to small changes in the value of the parameter point (Rao, 1962). It can therefore be shown that, subject to regularity conditions, θ^0 is locally identified if and only if the information matrix evaluated at θ^0 is nonsingular (Rothenberg, 1971).

It is clear that unidentified parameters cannot be consistently estimated. There are also pathological cases where identified models fail to possess consistent estimators (*e.g.*, Gabrielson, 1978). However, in most practical cases, we may treat identifiability and the existence of a consistent estimator as equivalent; for precise conditions, see Le Cam (1956) and Deistler and Seifert (1978).

3 Some specific parametric models

The choice of model structure is one of the basic ingredients in the formulation of the identification problem. In this section we briefly discuss some identification conditions for different types of models in order to demonstrate the kind of prior restrictions required.

3.1 Linear regression with collinearity

One of the most common models where an identification problem does occur is the linear regression model:

$$\mathbf{y} = X\beta + \mathbf{u} \tag{1}$$

where $\mathbf{y}$ is an $n \times 1$ vector of dependent observable variables, X is an $n \times k$ fixed matrix of observable variables, β a $k \times 1$ unknown coefficient vector, and $\mathbf{u}$ is an $n \times 1$ vector of disturbances whose components are (say) independent and identically distributed according to a normal distribution $N(0, \sigma^2)$ with unknown positive variance σ^2.

In this model, the value of β must be determined from the expected value of $\mathbf{y} : E(\mathbf{y}) = X\beta$. If the latter equation has a solution for β (i.e., if the model is correct), the solution is unique if and only the regressor matrix X has rank k. If X has rank zero (which entails $X = 0$), all values of β are equivalent (β is completely *unidentifiable*). If $1 \leq \mathrm{rank}(X) < k$, then not all the components can be determined, but some linear combinations of the components of β (say $c'\beta$) can be determined (i.e., they are *identifiable*). A necessary and sufficient condition for $c'\beta$ to be estimable (identifiable) is that $c = (X'X)d$ for some some vector d. Linear combinations that do not satisfy this condition are not identifiable. The typical way out of such collinearity problems consists in imposing restrictions on β (identifying restrictions) which set the values of the unidentifiable linear combinations (or components) of β.

Correspondingly, when X does bot have full rank, the equation $(X'X)\hat{\beta} = X'\mathbf{y}$ which defines the least squares estimator $\hat{\beta}$ does not have a unique solution. But all solutions of the least squares problem can be determined by considering $\hat{\beta} = (X'X)^{-}X'\mathbf{y}$ where $(X'X)^{-}$ is any generalized inverse of $(X'X)$. Different generalized inverses then correspond to different identifying restrictions on β. For further discussion, see Rao (1973, Chap. 4).

3.2 Linear simultaneous equations models

Consider a theory which predicts a relationship among the variables as

$$B\mathbf{y}_t + \Gamma\mathbf{x}_t = \mathbf{u}_t\,, \quad t = 1, \ldots, n, \tag{2}$$

where $\mathbf{y}_t$ and $\mathbf{u}_t$ are $G \times 1$ vectors of observed and unobserved random variables, respectively, $\mathbf{x}_t$ is a $K \times 1$ vector of observed non-stochastic variables, B and Γ are $G \times G$ and

$G \times K$ matrices of coefficients, with B being nonsingular. We assume that the $\mathbf{u}_t$ are independently normally distributed with mean 0 and variance-covariance matrix Σ. Equations (2) are called structural equations. Solving the endogenous variables, $\mathbf{y}$, as a function of the exogenous variables, $\mathbf{x}$, and the disturbance $\mathbf{u}$, we obtain:

$$\begin{aligned} \mathbf{y}_t &= -B^{-1}\Gamma\mathbf{x}_t + B^{-1}\mathbf{u}_t \\ &= \Pi\mathbf{x}_t + \mathbf{v}_t, \end{aligned} \tag{3}$$

where $\Pi = -B^{-1}\Gamma$, $E\mathbf{v}_t = 0$, $E\mathbf{v}_t\mathbf{v}_t' = V = B^{-1}\Sigma(B^{-1})'$. Equations (3) are called the *reduced form* equations derived from (2) and give the conditional likelihood of $\mathbf{y}_t$ for given $\mathbf{x}_t$ that summaries the information provided by the observed $(\mathbf{y}_t, \mathbf{x}_t)$. The variables in $\mathbf{x}_t$ are often also called "instruments".

From (3), we see that the simultaneous equations model can be viewed as a special case of a multivariate regression model (MLR), such that the regression coefficient matrix Π satisfies the equation:

$$B\Pi = -\Gamma. \tag{4}$$

Provided the matrix $X = [\mathbf{x}_1, \ldots, \mathbf{x}_n]'$ has full rank K (no collinearity), the regression coefficient matrix Π is uniquely determined by the distribution of $Y = [\mathbf{y}_1, \ldots, \mathbf{y}_n]'$ (it is *identifiable*). The problem is then whether B and Γ can be uniquely derived from equation (4). Premultiplying (2) by a $G \times G$ nonsingular matrix D, we get a second structural equation:

$$B^*\mathbf{y}_t + \Gamma^*\mathbf{x}_t = \mathbf{u}_t^*, \tag{5}$$

where $B^* = DB$, $\Gamma^* = D\Gamma$, and $\mathbf{u}_t^* = D\mathbf{u}$. It is readily seen that the reduced form of (5) is also (3). So equation (4) cannot be uniquely solved for B and Γ, given Π. Therefore, the two structures are observationally equivalent and the model is *non-identifiable*.

To make the model identifiable, additional prior restrictions have to be imposed on the matrices B, Γ and/or Σ. Consider the problem of estimating the parameters of the first equation in (2), out of a system of G equations. If the parameters cannot be estimated, the first equation is called *unidentified* or *underidentified*. If given the prior information, there is a unique way of estimating the unknown parameters, the equation is called *just identified*. If the prior information allows the parameters to be estimated in two or more linearly independent ways, it is called *overidentified*. A necessary condition for the first equation to be identified is that the number of restrictions on this equation be no less than $G - 1$ (order condition). A necessary and sufficient condition is that a specified submatrix of B, Γ and Σ be of rank $G - 1$ (rank condition); see Fisher (1976) and Hausman and Taylor (1983). For instance, suppose the restrictions on the first equation are in the form that certain variables do not appear. Then this rank condition says that the first equation is identified if and only if the submatrix obtained by taking the columns of B and Γ with prescribed zeros in the first row is of rank $G - 1$ (Koopmans and Reiersol, 1950).

3.3 Dynamic models

When both lagged endogenous variables and serial correlation in the disturbance term appear, we need to impose additional conditions to identify a model. For instance, consider the following two equation system (Koopmans, Rubin and Leipnik, 1950):

$$
\begin{aligned}
y_{1t} + \beta_{11} y_{1,t-1} + \beta_{12} y_{2,t-1} &= u_{1t}\,, \\
\beta_{12} y_{1t} + y_{2t} &= u_{2t}\,.
\end{aligned} \tag{6}
$$

If $(u_{1t},\, u_{2t})$ are serially uncorrelated, (6) is identified. If serial correlation in $(u_{1t},\, u_{2t})$ is allowed, then

$$
\begin{aligned}
y_{1t} + \beta^*_{11} y_{1,t-1} + \beta^*_{12} y_{2,t-1} &= u^*_{1t}\,, \\
\beta_{12} y_{1t} + y_{2t} &= u_{2t}\,,
\end{aligned} \tag{7}
$$

is observationally equivalent to (6), where $\beta^*_{11} = \beta_{11} + d\beta_{21}$, $\beta^*_{12} = \beta_{12} + d$, and $u^*_{1t} = u_{1t} + du_{2t}$.

Hannan (1971) derives generalized rank conditions for the identification of this type of model by first assuming that the maximum orders of lagged endogenous and exogenous variables are known, then imposing restrictions to eliminate redundancy in the specification and to exclude transformations of the equations that involve shifts in time. Hatanaka (1975), on the other hand, assumes that the prior information takes only the form of excluding certain variables from an equation, and derives a rank condition which allows common roots to appear in each equation.

3.4 Nonlinear models

For linear models, we have either global identification or else an infinite number of observationally equivalent structures. For models that are linear in parameters, but nonlinear in variables, there is a broad class of models whose members can commonly achieve identification [Brown (1983), McManus (1992)]. For models linear in the variables but nonlinear in the parameters, the state of the mathematical art is such that we only talk about local properties. That is, we cannot tell the true structure from any other substitute; however, we may be able to distinguish it from other structures which are close to it. A sufficient condition for local identification is that the Jacobian matrix formed by taking the first partial derivatives of

$$
\begin{aligned}
\omega_i &= \Psi i(\theta),\ i = 1, \ldots, n, \\
0 &= \varphi_j(\theta),\ j = 1, \ldots, R,
\end{aligned} \tag{8}
$$

with respect to θ be of full column rank, where the ω_i are n population moments of y and the φ_j are the R a priori restrictions on θ (Fisher, 1976).

When the Jacobian matrix of (8) has less than full column rank, the model may still be locally identifiable via conditions implied by the higher-order derivatives. However, the estimator of a model suffering from first-order lack of identification will in finite samples behave in a way which is difficult to distinguish from the behaviour of an unidentified model (Sargan, 1983).

3.5 Bayesian analysis

In Bayesian analysis all quantities, including the parameters, are random variables. Thus, a model is said to be identified in probability if the posterior distribution for θ is proper. When the prior distribution for θ is proper, so is the posterior, regardless of the likelihood function of y. In this sense unidentifiability causes no real difficulty in the Bayesian approach. However, basic to the Bayesian argument is that all probability statements are conditional, that is, they consist essentially in revising the probability of a fixed event in the light of various conditioning events, the revision being accomplished by Bayes' theorem. Therefore, in order for an experiment to be informative with regard to unknown parameters (i.e., for the posterior to be different from the prior), the parameter must be identified or estimable in the classical sense and identification remains as a property of the likelihood function (Kadane, 1975).

Drèze (1975) has commented that exact restrictions are unlikely to hold with probability 1 and has suggested using probabilistic prior information. In order to incorporate a stochastic prior, he has derived necessary rank conditions for the identification of a linear simultaneous equation model.

4 Definition of identification in nonparametric models

When the restrictions of an economic model specify all functions and distributions up to the value of a finite dimensional vector, the model is said to be parametric. When some functions or distributions are left parametrically unspecified, the model is said to be semi-parametric. The model is nonparametric if none of the functions and distributions are specified parametrically. The previous discussion is based on parametric specification. We now turn to the issue of whether economic restrictions such as concavity, continuity and monotonicity of functions, equilibrium conditions, the implications of optimization, and so on, may be used to guarantee the identification of some nonparametric models and the consistency of some nonparametric estimators; see Matzkin (1994).

Formally, an econometric model is specified by a vector of observable dependent and independent variables, a vector of unobservable variables, and a set of known functional relationships among the variables. When such functional relationships are unspecified, the nonparametric identification studies what functions or features of function can be recovered from the joint distribution of the observable variables.

The set of restrictions on the unknown functions and distributions in an econometric model defines the set of functions and distributions to which these belong. Let the model T denote the set of all a priori possible unknown functions and distributions. Let m denote a vector of the unknown functions and distributions in T and $P(m)$ denote the joint distribution of the observable variables under m. Then the identification of m can be defined as follows.

Definition 3 *The vector of functions m is identified in T if for any other vector, $m* \in T$ such that $m \neq m*$, $P(m) \neq P(m^*)$.*

Let $C(m)$ denote some feature of m, such as the sign of some coordinate of m.

Definition 4 *The feature $C(m)$ of m is identified if $C(m) = C(m^*)$ for all m, $m* \in T$ such that $P(m) = P(m^*)$.*

5 Examples of nonparametric identification

Contrary to the parametric model, there is no general result for nonparametric identification. We shall therefore give some examples of how restrictions can be used to identify nonparametric functions.

5.1 Generalized regression models

Economists often consider a model of the form

$$y = g(\mathbf{x}) + u. \tag{9}$$

When $E(u|\mathbf{x}) = 0$ and $g(\cdot)$ is a continuous function $g : \mathbf{x} \to \mathbb{R}$, then $g(\cdot)$ can be recovered from the joint distribution of $(y, \mathbf{x})$ because $E(y|\mathbf{x}) = g(\mathbf{x})$.

In some cases, the object of interest is not a conditional mean function $g(\cdot)$, but some "deeper" function, such as a utility function generating the distribution of demand for commodities by a consumer. For example, $\mathbf{x}$ in (9) can be a price vector for K commodities and the income of a consumer. Mas-Colell (1977) has shown that we can recover the underlying utility function from the distribution of demand if we restrict $g(\cdot)$ to be monotone increasing, continuous, concave and strictly quasi-concave functions.

5.2 Simultaneous equations models

Suppose $(\mathbf{y}, \mathbf{x})$ satisfies by the structural equations

$$\mathbf{r}(\mathbf{x}, \mathbf{y}) = \mathbf{u}, \tag{10}$$

where $\mathbf{y}$ and $\mathbf{u}$ denote $G \times 1$ vectors of observable endogenous and unobservable variables, respectively, $\mathbf{x}$ is a $K \times 1$ vector of observable exogenous variables, $\mathbf{r}$ denotes the G unknown functions, and let $p(\mathbf{r})$ and $p(\mathbf{r}^*)$ represent the joint distributions of the observables under $\mathbf{r}$ and $\mathbf{r}^*$ respectively. Assume also that: (i) $\forall(\mathbf{x}, \mathbf{y})$, $\partial\mathbf{r}/\partial\mathbf{y}$ has full rank, (ii) there exists a function $\pi(\cdot)$ such that $y = \pi(\mathbf{x}, \mathbf{u})$ [for conditions ensuring this, see Benkard and Berry (2006)], and (iii) $\mathbf{u}$ is distributed independently of $\mathbf{x}$. Then a necessary and sufficient condition guaranteeing that $p(\mathbf{r}^*) = p(\mathbf{r})$ is that

$$\operatorname{rank}\begin{pmatrix} \frac{\partial \mathbf{r}_i^*}{\partial(\mathbf{x},\mathbf{y})} \\ \frac{\partial \mathbf{r}}{\partial(\mathbf{x},\mathbf{y})} \end{pmatrix} < G + 1, \tag{11}$$

for all $(\mathbf{x}, \mathbf{y})$ and $i = 1, \ldots, G$, and all , where $\mathbf{r}_i^*$ denotes the i-th coordinate function of $\mathbf{r}^* \in T$; see Roehrig (1988) and Matzkin (2007).

5.3 Latent variable models and the measurement of treatment effects

For each person i, let (y_{0i}^*, y_{1i}^*) denote the potential outcomes in the untreated and treated states, respectively. Then the treatment effect for individual i is

$$\mathit{\Delta}_i = y_{1i}^* - y_{0i}^*$$

and the average treatment effect (ATE) is defined as

$$E(\mathit{\Delta}_i) = E(y_{1i}^* - y_{0i}^*)\,; \tag{12}$$

see Heckman and Vytlacil (2001).

Let the treatment status be denoted by the dummy variable d_i where $d_i = 1$ denotes the receipt of treatment and $d_i = 0$ denotes nonreceipt. The observed data are often in the form

$$y_i = d_i y_{1i}^* + (1 - d_i) y_{0i}^*. \tag{13}$$

Suppose $y_{1i}^* = \mu_1(\mathbf{x}_i, u_{1i})$, $y_{0i}^* = \mu_0(\mathbf{x}_i, u_{0i})$, and $d_i^* = \mu_D(\mathbf{z}_i) - u_{di}$, where $d_i = 1$ if $d_i^* \geq 0$ and 0 otherwise, $\mathbf{x}_i$ and $\mathbf{z}_i$ are vectors of observable exogenous variables and (u_{1i}, u_{0i}, u_{di}) are unobserved random variables. The average treatment effect and the complete structural econometric model can be identified with parametric specifications of $(\mu_1(\cdot), \mu_0(\cdot), \mu_D(\cdot))$ and the joint distributions of (u_{1i}, u_{0i}, u_{di}) even though we do not simultaneously observe y_{1i}^* and y_{0i}^*. In the case that neither $(\mu_1(\cdot), \mu_0(\cdot), \mu_D(\cdot))$ nor the joint distribution of (u_1, u_0, u_d) are specified, certain treatment effects may still be nonparametrically identified under weaker assumptions. For instance, under the assumption that d_i is orthogonal to (y_{1i}^*, y_{0i}^*) conditional on a set of confounders (x, z) (conditional independence or ignorable selection), the ATE is identifiable and estimable by comparing the difference of the average outcomes from the treatment group and from the untreated (control) group (Heckman and Robb, 1985, Rosenbaum and Rubin, 1985). If the focus is on the average treatment effect for someone who would not participate if $p(\mathbf{z}) \leq p(\mathbf{z}_0)$ and would participate if $p(\mathbf{z}) > p(\mathbf{z}_0)$ (the local average treatment effect (LATE)), where $p(\mathbf{z}) = \mathrm{Prob}(d = 1|\mathbf{z})$ (propensity score), Imbens and Angrist (1994) show that under the assumptions of separability of the effects of observable factors and unobservable factors and independence between observed factors and unobserved factors, they can be estimated by the sample analogue of

$$\Delta^{LATE}(\mathbf{x},\, p(\mathbf{z}),\, p(\mathbf{z}_0)) \equiv \frac{E(y|\mathbf{x}, p(\mathbf{z})) - E(y|\mathbf{x}, p(\mathbf{z}_0))}{p(\mathbf{z}) - p(\mathbf{z}_0)} \tag{14}$$

where without loss of generality, we assume $p(\mathbf{z}) > p(\mathbf{z}_0)$. The limit of LATE provides the local instrumental variable (LIV) estimand (Heckman and Vytlacil, 1999):

$$\Delta^{LIV}(\mathbf{x},\, p(\mathbf{z})) \equiv \frac{\partial E(y|\mathbf{x}, p(\mathbf{z}))}{\partial p(\mathbf{z})}\,. \tag{15}$$

Heckman and Vytlacil (2001) give conditions that suitably weighted versions of LIV identify the ATE.

6 Weak instruments and weak identification

The most common way of trying to achieve identification consists in imposing exclusion restrictions on the variables of a structural equation. In model (2), suppose that $\mathbf{y}_t$ and $\mathbf{x}_t$ are partitioned as $\mathbf{y}_t = (y_{1t}, \mathbf{y}_{2t}', \mathbf{y}_{3t}')'$ and $\mathbf{x}_t = (\mathbf{x}_{1t}', \mathbf{x}_{2t}')'$ where y_{1t} is a scalar, $\mathbf{y}_{it}$ has dimension G_i $(i = 2, 3)$ and $\mathbf{x}_{it}$ has dimension K_i $(i = 1, 2)$. If $\mathbf{y}_{3t}$ and $\mathbf{x}_{2t}$ are excluded from the first equation and the coefficient of y_{1t} is normalized to one, this yields an equation of the form:

$$y_{1t} - \mathbf{y}_{2t}'\beta_1 = \mathbf{x}_{1t}'\gamma_1 + u_{1t}\,, \quad t = 1,\, \ldots\,,\, n. \tag{16}$$

Let us also rewrite the reduced equation for $\mathbf{y}_{2t}$ in terms of $\mathbf{x}_{1t}$ and $\mathbf{x}_{2t}$:

$$\mathbf{y}_{2t} = \Pi_{21}\mathbf{x}_{1t} + \Pi_{22}\mathbf{x}_{2t} + \mathbf{v}_{2t}\,. \tag{17}$$

Then, substituting (17) into (16), we see that the reduced form for y_{1t} is:

$$y_{1t} = \Pi_{11}\mathbf{x}_{1t} + \Pi_{12}\mathbf{x}_{2t} + \mathbf{v}_{1t}\,, \tag{18}$$

where $\mathbf{v}_{1t} = u_{1t} + \mathbf{v}_{2t}'\beta_1$, $\Pi_{11} = \gamma_1' + \beta_1'\Pi_{21}$ and

$$\Pi_{12}' = \Pi_{22}'\beta_1\,. \tag{19}$$

Since γ_1 is free, Π_{11} is not restricted, but equation (19) determines the identifiability of β_1, hence also of γ_1. Provided equation (19) has a solution [i.e., if equation (16) is consistent with the data], the solution is unique if and only if the rank of the $G_2 \times K_2$ matrix Π_{22} is equal to G_2, the dimension of β_1 :

$$\operatorname{rank}(\Pi_{22}) = G_2 \,. \tag{20}$$

If $\operatorname{rank}(\Pi_{22}) < G_2$, the vector β_1 is not identifiable. However, it is completely unidentifiable only if $\operatorname{rank}(\Pi_{22}) = 0$, or equivalently if $\Pi_{22} = 0$. If $1 < \operatorname{rank}(\Pi_{22}) < G_2$, some linear combinations $c'\beta_1$ are identifiable, but not all of them. Failure of the identification condition means that the regressors (or the "instruments") $\mathbf{x}_{2t}$ do not move enough to separate the effects of the different variables in $\mathbf{y}_{2t}$. Condition (20) underscores two important things: first, exclusion and normalization restrictions – which are easy to check – are not sufficient to ensure identification; second, identification depends on the way the exogenous variables $\mathbf{x}_{2t}$ *excluded* from the structural equation of interest (16) are related to endogenous variables $\mathbf{y}_{2t}$ included in the equation. The latter feature is determined by the matrix Π_{22} whose rows should be linearly independent. Since Π_{22} is not observable, this may be difficult to determine in practice.

A situation that can lead to identification difficulties is the one where the identification condition (20) indeed holds, but, in some sense, Π_{22} is "close" not to have sufficient rank. In such situations, we say that we have *weak instruments*. In view of the fact that the distributions of most statistics move continuously as functions of Π_{22}, the practical consequences of being close to identification failure are essentially the same. Assessing the closeness to non-identification may be done in various ways, for example by considering the eigenvalues of the matrices which measure the "size" of Π_{22}, such as $\Pi_{22}\Pi_{22}'$, $\Pi_{22}X_2'M(X_1)X_2\Pi_{22}'$ or a *concentration matrix* $\Sigma_{22}^{-1/2}\Pi_{22}X_2'M(X_1)X_2\Pi_{22}'\Sigma_{22}^{-1/2}$, where $X_1 = [\mathbf{x}_{11}, \ldots, \mathbf{x}_{1n}]'$, $X_2 = [\mathbf{x}_{21}, \ldots, \mathbf{x}_{2n}]'$, Σ_{22} is the covariance matrix of $\mathbf{v}_{2t}$, $\Sigma_{22}^{-1/2}$ is its square root, and $M(X_1) = I_n - X_1(X_1'X_1)^{-1}X_1'$. More generally, any situation where a parameter may be difficult to determine because we are close to a case where a parameter ceases to be identifiable may be called *weak identification*. Weak identification was highlighted as a problem of practical interest by Nelson and Startz (1990), Bound, Jaeger and Baker (1995), Dufour (1997), and Staiger and Stock (1997); for reviews, see Stock, Wright and Yogo (2002) and Dufour (2003).

7 Statistical consequences of identification failure

Identification failure has several detrimental consequences for statistical analysis:

1. parameter estimates, tests and confidence sets computed for unidentified parameters have no clear interpretation; this situation may be especially misleading if the statistical instruments used do not reveal the presence of the problem;

2. consistent estimation is not possible unless additional information is supplied;

3. many standard distributional results used for inference on such models are not anymore valid, even with a large sample size [see Phillips (1983, 1989) and Rothenberg (1984)];

4. numerical problems also easily appear, due for example to the need to invert (quasi) singular matrices.

Weak identification problems lead to similar difficulties, but may be more treacherous in the sense that standard asymptotic distributional may remain valid, but they constitute very bad approximations to what happens in finite samples:

1. standard consistent estimators of structural parameters can be heavily biased and follow distributions whose form is far from the limiting Gaussian distribution, such as bimodal distributions, even with fairly large samples [Nelson and Startz (1990), Hillier (1990), Buse (1992)];

2. standard tests and confidence sets, such as Wald-type procedures based on estimated standard errors, become highly unreliable or completely invalid [Dufour (1997)].

A striking illustration of these problems appears in the reconsideration by Bound, Jaeger and Baker (1995) of a study on returns to education by Angrist and Krueger (1991). Using 329000 observations, these authors found that replacing the instruments used by Angrist and Krueger (1991) with randomly generated (totally irrelevant) instruments produced very similar point estimates and standard errors. This result indicates that the original instruments were weak. Recent work in this area is reviewed in Stock, Wright and Yogo (2002) and Dufour (2003).

8 Concluding remarks

The study of identifiability is undertaken in order to explore the limitations of statistical inference (when working with economic data) or to specify what sort of a priori information is needed to make a model estimable. It is a fundamental problem concomitant with the existence of a structure. Logically it precedes all problems of estimation or of testing hypotheses.

An important point that arises in the study of identification is that without a priori restrictions imposed by economic theory it would be almost impossible to estimate economic relationships. In fact, Liu (1960) and Sims (1980) have argued that economic relations are not identifiable because the world is so interdependent as to have almost all variables appearing in every equations, thus violating the necessary condition for identification. However, almost all the models we discuss in econometrics are only approximate. We use convenient formulations which behave in a general way that corresponds to our economic theories and intuitions, and which cannot be rejected by the available data. In this sense, identification is a property of the model but not necessarily of the real world. It is also important to be careful about situations where identification almost does not hold (weak identification), since these are in practice as damaging for statistical analysis as identification failure itself.

The problem of identification arises in a number of different fields such as automatic control, biomedical engineering, psychology, systems science, etc., where the underlying physical structure may be deterministic [*e.g.*, see Aström and Eykoff (1971)]. It is also aptly linked to the design of experiments (*e.g.*, Kempthorne, 1947; Bailey, Gilchrist and Patterson, 1977). Here, we restrict our discussion to economic applications of statistical identifiability involving random variables.

See also: econometrics; endogeneity and exogeneity; estimation; simultaneous equations models; treatment effects.

References

Angrist, J. D. & Krueger, A. B. (1991), 'Does compulsory school attendance affect schooling and earning?', *Quarterly Journal of Economics* **CVI**, 979–1014.

Aström, K. J. & Eykhoff, P. (1971), 'System identification - a survey', *Automatica* **7**, 123–162.

Bailey, R. A., Gilchrist, F. H. L. & Patterson, H. D. (1977), 'Identification of effects and confounding patterns in factorial designs', *Biometrika* **64**, 347–354.

Bekker, P. & Wansbeek, T. (2001), Identification in parametric models, *in* B. Baltagi, ed., 'Companion to Theoretical Econometrics', Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, U.K., chapter 7, pp. 144–161.

Benkard, C. L. & Berry, S. (2006), 'On the nonparametric identification of nonlinear simultaneous equations models: Comment on Brown (1983) and Roehrig (1988)', *Econometrica* **74**, 1429–1440.

Bound, J., Jaeger, D. A. & Baker, R. M. (1995), 'Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak', *Journal of the American Statistical Association* **90**, 443–450.

Brown, B. W. (1983), 'The identification problem in systems nonlinear in the variables', *Econometrica* **51**, 175–196.

Buse, A. (1992), 'The bias of instrumental variables estimators', *Econometrica* **60**, 173–180.

Deistler, M. & Seifert, H.-G. (1978), 'Identifiability and consistent estimability in econometric models', *Econometrica* **46**, 969–980.

Drèze, J. (1975), Bayesian theory of identification in simultaneous equations models, *in* (Fienberg & Zellner 1975).

Dufour, J.-M. (1997), 'Some impossibility theorems in econometrics, with applications to structural and dynamic models', *Econometrica* **65**, 1365–1389.

Dufour, J.-M. (2003), 'Identification, weak instruments and statistical inference in econometrics', *Canadian Journal of Economics* **36**(4), 767–808.

Fienberg, S. E. & Zellner, A., eds (1975), *Studies in Bayesian Econometrics and Statistics*, North-Holland, Amsterdam.

Fisher, F. M. (1976), *The Identification Problem in Econometrics*, Krieger Publishing Company, Huntington (New York).

Frisch, R. (1934), *Statistical Confluence Analysis by Means of Complete Regression Systems*, Publication No.5, Universitetes Okonomiske Institutt, Oslo, Norway.

Gabrielson, A. (1978), 'Consistency and identifiability', *Journal of Econometrics* **8**, 261–263.

Griliches, Z. & Intriligator, M. D., eds (1983), *Handbook of Econometrics, Volume 1*, North-Holland, Amsterdam.

Haavelmo, T. (1944), 'The probability approach in econometrics', *Econometrica* **12 (supplement)**, 1–115.

Hannan, E. J. (1971), 'The identification problem for multiple equation systems with moving average errors', *Econometrica* **39**, 751–766.

Hatanaka, M. (1975), 'On the global identification of the dynamic simultaneous equations model with stationary disturbances', *International Economic Review* **16**, 545–554.

Hausman, J. A. & Taylor, W. E. (1983), 'Identification, estimation and testing in simultaneous equations models with disturbance covariance restriction', *Econometrica* **51**, 1527–1549.

Heckman, J. J. & Vytlacil, E. (1999), 'Local instrumental variables and latent variables models for identifying and bounding treatment effects', *Proceedings of the National Academy of Sciences* **96**, 4730–4734.

Heckman, J. J. & Vytlacil, E. (2001), Local instrumental variables, *in* C. Hsiao, K. Morimune & J. L. Powell, eds, 'Nonlinear Statistical Modeling Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya', Cambridge University Press, Cambridge, U.K., pp. 1–46.

Heckman, J. & Robb, R. (1985), Alternative methods for evaluating the impact of interventions, *in* J. Heckman & B. Singer, eds, 'Longitudinal Analysis of Labor Market Data', Cambridge University Press, New York.

Hillier, G. H. (1990), 'On the normalization of structural equations: Properties of direction estimators', *Econometrica* **58**, 1181–1194.

Hsiao, C. (1983), Identification, *in* (Griliches & Intriligator 1983), chapter 4, pp. 223–283.

Hurwicz, L. (1950), Generalization of the concept of identification, *in* (Koopmans 1950).

Imbens, G. & Angrist, J. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**, 467–476.

Kadane, J. B. (1975), The role of identification in Bayesian theory, *in* (Fienberg & Zellner 1975).

Kempthorne, O. (1947), 'A simple approach to confounding and factorial replication in factorial experiments', *Biometrika* **34**, 255–272.

Koopmans, T. C., ed. (1950), *Statistical Inference in Dynamic Economic Models*, number 10 *in* 'Cowles Commission Monographs', John Wiley & Sons, New York.

Koopmans, T. C. & Reiersøl, O. (1950), 'The identification of structural characteristics', *Annals of Mathematical Statistics* **21**, 165–181.

Koopmans, T. C., Rubin, H. & Leipnik, R. B. (1950), Measuring the equation systems of dynamic economics, *in* (Koopmans 1950), chapter II, pp. 53–237.

Le Cam, L. (1956), On the asymptotic theory of estimation and testing hypotheses, *in* 'Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability', The University of California Press, Berkeley, California.

Liu, T. C. (1960), 'Underidentification, structural estimation, and forecasting', *Econometrica* **28**, 855–865.

Manski, C. (2003), *Partial Identification of Probability Distributions*, Springer Series in Statistics, Springer-Verlag, New York.

Marschak, J. (1942), Economic interdependence and statistical analysis, *in* O. Lange, F. McIntyre & T. O. Yntema, eds, 'Studies in Mathematical Economics and Econometrics', University of Chicago Press, Chicaho, Illinois.

Mas-Collel, A. (1977), 'On the recoverability of consumers preferences from market demand behavior', *Econometrica* **45**, 1409–1430.

Matzkin, R. (1994), Restrictions of economic theory in nonparametric methods, *in* R. F. Engle & D. L. McFadden, eds, 'Handbook of Econometrics, Volume 4', North-Holland, Amsterdam, chapter 42, pp. 2523–2558.

Matzkin, R. (2007), Nonparametric identification, *in* J. J. Heckman & E. E. Leamer, eds, 'Handbook of Econometrics, Volume 6', North-Holland, Amsterdam.

McManus, D. A. (1992), 'How common is identification in parametric models?', *Journal of Econometrics* **53**, 5–23.

Morgan, M. S. (1990), *The History of Econometric Ideas*, Cambridge University Press, Cambridge, U.K.

Nelson, C. R. & Startz, R. (1990), 'The distribution of the instrumental variable estimator and its t-ratio when the instrument is a poor one', *Journal of Business* **63**, 125–140.

Phillips, P. C. B. (1983), Exact small sample theory in the simultaneous equations model, *in* (Griliches & Intriligator 1983), chapter 8, pp. 449–516.

Phillips, P. C. B. (1989), 'Partially identified econometric models', *Econometric Theory* **5**, 181–240.

Prakasa Rao, B. L. S. (1992), *Identifiability in Stochastic Models: Characterization of Probability Distributions*, Academic Press, New York.

Rao, C. R. (1962), 'Problems of selection with restriction', *Journal of the Royal Statistical Society, Series B* **24**, 401–405.

Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, second edn, John Wiley & Sons, New York.

Roehrig, C. S. (1988), 'Conditions for identification in nonparametric and parametric models', *Econometrica* **56**, 433–477.

Rosenbaum, P. & Rubin, D. (1985), 'Reducing bias in observational studies using subclassification on the propensity score', *Journal of the American Statistical Association* **79**, 516–524.

Rothenberg, T. J. (1971), 'Identification in parametric models', *Econometrica* **39**, 577–591.

Rothenberg, T. J. (1984), Approximating the distributions of econometric estimators and test statistics, *in* Z. Griliches & M. D. Intriligator, eds, 'Handbook of Econometrics, Volume 2', North-Holland, Amsterdam, chapter 15, pp. 881–935.

Sargan, J. D. (1983), 'Identification and lack of identification', *Econometrica* **51**, 1605–1633.

Sims, C. (1980), 'Macroeconomics and reality', *Econometrica* **48**, 1–48.

Staiger, D. & Stock, J. H. (1997), 'Instrumental variables regression with weak instruments', *Econometrica* **65**(3), 557–586.

Stock, J. H. & Trebbi, F. (2003), 'Who invented IV regression?', *The Journal of Economic Perspectives* **17**(3), 177–194.

Stock, J. H., Wright, J. H. & Yogo, M. (2002), 'A survey of weak instruments and weak identification in generalized method of moments', *Journal of Business and Economic Statistics* **20**(4), 518–529.

Wald, A. (1950), Note on the identification of economic relations, *in* (Koopmans 1950), chapter III, pp. 238–244.

Working, E. J. (1927), 'What do statistical demand curves show?', *The Quarterly Journal of Economics* **41**, 212–235.

Working, H. (1925), 'The statistical determination of demand curves', *The Quarterly Journal of Economics* **39**, 503–543.

Wright, P. G. (1915), 'Moore's economic cycles', *The Quarterly Journal of Economics* **29**, 631–641.

Wright, P. G. (1928), *The Tariff on Animal and Vegetable Oils*, Macmillan, New York.