Ordinary least squares as a method of measurement

Balzer, Wolfgang; Haendler, E.-W.

W. BALZER AND E.-W. HAENDLER

ORDINARY LEAST SQUARES AS A METHOD OF MEASUREMENT*

Statistical estimation plays a decisive role within the empirically oriented branches of economics and various other social sciences. Statistical methods are also applied in the natural sciences; the empiricity of the natural sciences is, however, anchored in a more direct mode in "repeatable" measurement, which assigns statistical estimation only a secondary status.

On a narrow conception, measurement consists in "comparison with a unit", a conception which, together with the paradigm of fundamental measurement, has been widely accepted in psychology and the social sciences, too. Yet typical examples of "measurement" in the natural sciences can be subsumed under this concept only at the expense of severe biasedness. Typically, what we encounter here are situations in which theoretical equations, "theories", are employed in order to "calculate" the desired "measured" values. We will adopt here without argument a broad conception of measurement according to which even such theory-dependent methods of determination are termed measurement: the structuralist view of measurement.[1] According to this view the calculation of parameters from a set of equations, for instance, constitutes a method of measurement for these parameters, provided there is a unique solution.

The latter procedure is of course analogous to an estimation of parameters in the social sciences. So, is estimation a method of measurement (in the broad, structuralist sense)? This is not a mere question of terminology. Rather, the attempt of answering this question reveals precise distinctions between related procedures (which we call measurement) in "statistical" social science on the one hand and in the natural sciences on the other. Besides, our investigation will shed light on some fundamental methodological differences between the natural and the social sciences.

We will restrict our analysis to the simple case of applying the method of ordinary least squares (OLS) in order to estimate the two parameters of a linear demand function, a case typical for a wide range of similar elementary applications of OLS. Our aim is to

Erkenntnis **30** (1989) 129–146.

subsume OLS under the general structuralist concept of a method of measurement. By this we do not want to adopt once again the imperalist strategy of recommending methodological ideas from physical science to the social scientist. It will turn out (in Section V) that the notion of theory-dependent measurement which was developed to cover the respective phenomena within physical science has to be generalized to be applicable to OLS. We introduce the notion of a regression method of measurement which covers methods like OLS as well as "ordinary" theory-dependent measurement in the natural sciences (and, of course, fundamental measurement as well). This enables us to work out features common to both the natural and the social sciences as well as to illustrate fundamental differences between the natural and the social sciences.

Furthermore, we discuss the question of how to justify OLS as a method of measurement (Section IV). In this context, again, we encounter strong similarities but at the same time clear differences between the natural and the social sciences.

I. MEASUREMENT

The structuralist view of measurement starts from considering actual procedures of measurement and focuses on the structure of single, isolated processes of measurement, in the course of which one value, the *measured value*, is produced. Such procedures involve a real system, the measuring apparatus, and in most cases some theoretical equations which come from one single or from several established theories and which are used in order to calculate the measured value. The real system thus is represented as a model or a chain of models of one or several theories which *govern* the process of measurement. Such models we call measuring models. A measuring model therefore represents one single process of measurement (or even only a part of it) as far as it can be subsumed under some given theory. As a borderline case the theory may be a mere theory of measurement, like in fundamental measurement the "theory" of extensive systems.[2]

Measuring models have at least four general features in common. First, they are characterized by a law-like proposition (the laws of the theory governing the process of measurement, see (D1–3) below). Second, the measured value in each measuring model is uniquely determined by other parts of the model, and by the law characterizing the model (D1–5). Third, the measured value is effectively computable from other "parts" of the model (D1–6), and fourth, the measured value is a continuous function of those other parts from which it can be computed (D1–7).[3]

The law-like propositions may be of various kinds. They may be proper axioms of established theories, or simpler equations "derived" from more complicated such axioms. In fundamental measurement they are the axioms put forward by measurement theorists. Continuity is required in order to exclude contrived cases, like setting the measured value equal to 1 by definition. In such cases there is no point in measuring since the "measured value" is fixed purely conceptually. Usually, the "same" measuring model will be used to measure not just one value but a set of such values (by repeated application). Accordingly, uniqueness may be required not just for one value but for a set of values of the function F to be measured. For reasons of simplicity we even require that F be uniquely determined for all of its arguments.[4]

These conditions may be integrated into a general definition of a measuring model as follows.

(D1) x is a *measuring model for function F characterized by B, Σ*, τ and $\approx$ iff there exist $D_1, \ldots, D_k$, $A_1, \ldots, A_m$, $R_1, \ldots, R_n$ such that

(1) τ is a type and $x = \langle D_1, \ldots, D_k, A_1, \ldots, A_m, R_1, \ldots, R_n, F\rangle$ is a set theoretic structure of type τ
(2) F is a function
(3) B is a law-like statement, and valid in x
(4) $\approx$ is an equivalence relation on the class of all functions of the type of F
(5) F is uniquely determined by B in x (up to $\approx$)
(6) each function value $F(a)$ can be computed from a finite substructure of $\langle D_1, \ldots, R_n\rangle$ (up to $\approx$, and after appropriate encoding)
(7) $\Sigma = \langle \Sigma_1, \Sigma_2\rangle$ is a pair of topologies such that B defines a function $\langle D_1, \ldots, R_n\rangle \to F$ which is piecewise continuous w.r.t. Σ_1 and Σ_2.

The values of F we call *measured values*. The special case in which an explicit definition D for F in terms of the other components $D_1, \ldots, R_n$ is "contained" in statement B will be of particular importance in the following. In this case we speak of measuring models for the defined term F, by which label we refer to structures from which F has been removed ((D2-2) below).

(D2) x is a *measuring model for the defined term F characterized by* B, Σ, τ, $\approx$ and D iff there exist $D_1, \ldots, D_k$, $A_1, \ldots, A_m$, $R_1, \ldots, R_n$ such that
(1) $\langle D_1, \ldots, R_n, F\rangle$ is a measuring model for function F characterized by B, Σ, τ and $\approx$,
(2) $x = \langle D_1, \ldots, R_n\rangle$,
(3) D is an explicit definition of F in terms of x,
(4) there is a law-like statement B^* such that $B(D_1, \ldots, R_n, F)$ is equivalent with $B^*(x) \wedge D(x, F)$.

The law-like proposition B (or B^*) of course can be extracted only from the investigation of a whole class of many similar measuring models. Such a class we call a method of measurement. The connection to the usual use of the word "method" is established by observing that each method determines the class of all systems in which it can be successfully applied, and conversely.

(D3) M_m is a *method of measurement for function* $\{F\}$ iff there exist B, Σ, τ and $\approx$ such that M_m is the class of all measuring models for function F characterized by B, Σ, τ and $\approx$, and $M_m \neq \emptyset$.

(D4) M_m is a *method of measurement for the defined term* $\{F\}$ iff there exist B, Σ, τ, $\approx$ and D such that M_m is the class of all measuring models for the defined term F characterized by B, Σ, τ, $\approx$ and D, and $M_m \neq \emptyset$.

In the natural sciences there are many cases of measurement in the course of which a given theory is used and presupposed in order to calculate the measured values. Any method of measurement with this feature we call theory-dependent.

(D5) If T is a theory[5] with class M of models then M_m is a *T-dependent* method of measurement iff, for some $\{F\}$, M_m is a method of measurement for $\{F\}$ or M_m is a method of measurement for the defined term $\{F\}$, and $M_m \subseteq M$.

As a paradigm for theory-dependent measurement let us consider the measurement of mass by collisions as governed by classical collision mechanics (CCM). The models (M(CCM)) are defined as follows. $x \in M$(CCM) iff x has the form $\langle P, \{b, a\}, \mathbf{R}, v, m\rangle$ and (1) P is a non-empty, finite set (of "particles"), (2) $\{b, a\}$ is a two-element set (of "instants": "*b*efore" and "*a*fter" the collision), (3) $m: P \to \mathbf{R}^+$ ("mass-function"), (4) $v: P \times \{b, a\} \to \mathbf{R}^3$ ("velocity-function"), (5) $\Sigma_{p \in P}\, m(p)v(b, p) = \Sigma_{p \in P}\, m(p)v(a, p)$ (law of conservation of total momentum).

The models are intended to describe collisions of two or more particles. Let $\bar{m}$ denote the term "mass". A method of measurement M_m(CCM) for $\bar{m}$ is defined as follows. $x \in M_m$(CCM) iff (1) $x = \langle P, \{b, a\}, \mathbf{R}, v, m\rangle \in M$(CCM), (2) P is a two-element set ($P = \{p, p'\}$), (3) v is such that all $v(p^*, t)$, $p^* \in P$, $t \in \{b, a\}$ are on a straight line, (4) for $t \in \{b, a\}$, $v(p, t)$ and $v(p', t)$ have opposite direction, (5) $v(p, b) \neq v(p', a)$.

The topologies required in (D1–7) are generated by neighbourhoods defined as follows. For $x = \langle P, \{b, a\}, \mathbf{R}, v\rangle$, $y \in U_x^\epsilon$ iff $y = \langle P, \{b, a\}, \mathbf{R}, v'\rangle$ and for all $p \in P$ and $t \in \{b, a\}$, $|v(p, t) - v'(p, t)| < \epsilon$. For $m: P \to \mathbf{R}^+$, $m' \in U_m^\epsilon$ iff $m': P \to \mathbf{R}^+$ and for all $p \in P$: $|m(p) - m'(p)| < \epsilon$. $\approx$ is given by: $m \approx m'$ iff $\mathrm{Dom}(m) = \mathrm{Dom}(m')$ and there exists $\alpha \in \mathbf{R}^+$ such that for all $p \in \mathrm{Dom}(m)$: $m(p) = \alpha \cdot m'(p)$. The value $m(p')$ in $x \in M_m$(CCM) is computable from v up to some $\alpha \in \mathbf{R}^+$, and, for given such α, it varies smoothly with variation of v. It is not difficult to prove that M_m(CCM), in fact, is a method of measurement for $\{m\}$.

As an example of a method of measurement for a defined term F think of the method of determining the mean velocity of a uniformly moving particle. If $s_p(t)$ indicates the position of particle p at time t, and if p moves uniformly, then its mean velocity v_p (which incidentally is also its actual velocity) is defined as

$$v_p = \frac{s_p(t) - s_p(t')}{t - t'}$$

where t, t' are different instants (which we treat as real numbers for the sake of simplicity) and $t' < t$. A corresponding measuring model for v_p has the form $\langle\{p\}, T, \mathbf{R}^3, s_p\rangle$ where $T \subseteq \mathbf{R}$ is an open interval, s_p: $T \to \mathbf{R}^3$ is such that its image is contained in a straight line and Ds_p is a constant function, and v_p is defined as just indicated. It is easy to define the corresponding method of measurement for the defined term $\{v_p\}$.

II. OLS: A RECONSTRUCTION OF LINEAR STOCHASTIC DEMAND SYSTEMS

We now turn to our example from applied econometrics. A linear demand system is a system in which the purchases, d_t, for a commodity (expressed in quantitative, numerical form) are supposed to be a linear function of the price, p_t, of this commodity, where both prices and purchases may change over time, t.

It must be emphasized that the existence of a linear relationship between observed purchases and observed prices does not provide much justification for speaking of "demands" and "demand systems". Our focus in this paper is, however, not on the criteria of identity for the function to be measured. We investigate parameter estimation by means of the method of ordinary least squares on the basis of the very elementary example of a linear relationship between observed purchases and observed prices. The result of our inquiry applies likewise to the more intricate demand theories which actually represent the state of the art.

While theoretical as well as practical considerations may suggest a linear relationship between prices and purchases, the observed time-series $\langle p_t, d_t\rangle$, $t = 1, \ldots, n$ usually will not satisfy a linear equation

$$d_t = \beta_1 + \beta_2 p_t \qquad \text{for all } t \leq n.$$

In order to accomodate for deviations, p_t and d_t are regarded as values of random variables $\tilde{p}_t$, $\tilde{d}_t$, both defined on some suitable set W of events (which ordinarily is not made explicit), and a disturbance variable $\tilde{u}_t$ is introduced which accounts for all those influences on the purchases which are not allowed for by the prices (e.g., the influence of the temperature on the purchases of sodas). These hypothetical random variables are assumed to form the following linear relationship:

$$(1) \qquad \tilde{d}_t(w) = \beta_1 + \beta_2 \tilde{p}_t(w) + \tilde{u}_t(w), \text{ for all } t \leq n \text{ and } w \in W.$$

(See (D7–3) below). p_t and d_t are taken as particular realizations of $\tilde{p}_t$, $\tilde{d}_t$: $\tilde{p}_t(w_o) = p_t$ and $\tilde{d}_t(w_0) = d_t$ for some $w_o \in W$ where w_o is the event which actually has occurred (D7–2).

(D6) x is a *potential linear stochastic demand system* ($x \in M_p$(LSD)) iff there exist $n, p, d, W, \mathfrak{A}, \mu, \tilde{p}, \tilde{d}, \tilde{u}$ such that $x = \langle n, p, d, W, \mathfrak{A}, \mu, \tilde{p}, \tilde{d}, \tilde{u}, \beta\rangle$ and

(1) $n \in \mathbb{N}$, $n > 0$,
(2) $p, d \in \mathbb{R}^n_+$,
(3) $\langle W, \mathfrak{A}, \mu\rangle$ is a probability space,
(4) $\tilde{p}, \tilde{d}, \tilde{u}$: $W \to \mathbb{R}^n$ are random variables,
(5) $\beta \in \mathbb{R}^2$.

For $\tilde{p}, \tilde{d}, \tilde{u}$ and $t \leq n$ we define functions $\tilde{p}_t$: $W \to \mathbb{R}$ by $\tilde{p}_t(w) = (\tilde{p}(w))_t$ etc.

(D7) x is a *linear stochastic demand system* ($x \in M$(LSD)) iff there exist $n, p, d, W, \mathfrak{A}, \mu, \tilde{p}, \tilde{d}, \tilde{u}, \beta$ such that $x = \langle n, p, d, W, \mathfrak{A}, \mu, \tilde{p}, \tilde{d}, \tilde{u}, \beta\rangle$ and

(1) $x \in M_p$(LSD),
(2) there exist $V \in \mathfrak{A}$ and $w_o \in V$ such that $\mu(V) > 0$, $\tilde{p}(w_o) = p$ and $\tilde{d}(w_o) = d$,
(3) for all $w \in W$ and $t \leq n$: $\tilde{d}_t(w) = \beta_1 + \beta_2\tilde{p}_t(w) + \tilde{u}_t(w)$,
(4) for all $t \leq n$ and all $a \in \mathrm{Rge}(\tilde{p})$: the conditional expectation of $\tilde{u}_t$ under condition a, $E(\tilde{u}_t \mid a)$, is zero.

Some auxiliary definitions are needed for the following.

(D8) (a) If $a = \langle a_1, \ldots, a_n\rangle$ and $b = \langle b_1, \ldots, b_n\rangle \in \mathbb{R}^n$, and $f = \langle f_1, \ldots, f_n\rangle$ and $g = \langle g_1, \ldots, g_n\rangle$ are functions, f, g: $W \to \mathbb{R}^n$, then

$$\bar{a} = \frac{1}{n}\sum_{j \leq n} a_j, \qquad s^2(a) = \frac{1}{n}\sum_{j \leq n}(a_j - \bar{a})^2,$$

$$\mathrm{cov}(a, b) = \frac{1}{n}\sum_{j \leq n}(a_j - \bar{a})(b_j - \bar{b}).$$

$\bar{f}, s^2(f), \mathrm{cov}(f, g)$: $W \to \mathbb{R}^n$ are defined by $\bar{f}(w) = \overline{f(w)}$, $s^2(f)(w) = s^2(f(w))$, $\mathrm{cov}(f, g)(w) = \mathrm{cov}(f(w), g(w))$

(b) If $x = \langle n, p, d, W, \mathfrak{A}, \mu, \tilde{p}, \tilde{d}, \tilde{u}, \beta\rangle \in M_p$(LSD) we write

$$\hat{\beta}_2(p, d) = \frac{\mathrm{cov}(p, d)}{s^2(p)}, \quad \hat{\beta}_1(p, d) = \bar{d} - \hat{\beta}_2(p, d) \cdot \bar{p}$$

and, for $w \in W$:

$$\tilde{\beta}_2(w) = \frac{\text{cov}(\tilde{p}, \tilde{d})(w)}{s^2(\tilde{p})(w)}, \quad \tilde{\beta}_1(w) = \bar{\tilde{d}}(w) - \tilde{\beta}_2(w) \cdot \bar{\tilde{p}}(w).$$

The method of measurement we want to consider is the following. By minimizing the squared deviations of the observed data from a hypothetical straight line, OLS recommends the values $\hat{\beta}_1(p, d)$, $\hat{\beta}_2(p, d)$ as defined in (D8–b) above as estimations for the unknown parameters β_1, β_2 which occur in the models (see (D7–3)). It is common to call these unknown, hypothetical parameters β_1, β_2 the *true* parameters. "True parameter" is thus defined as "parameter which occur(s) in (one of) the model(s) which (are) is assumed to capture the real system under study." We adopt this usage without inquiring into its possible philosophical interpretations.

The choice of $\hat{\beta}_i$ as candidates for the true values β_i is justified by the fact that the suggested values $\hat{\beta}_i$ are instances of *estimators* $\tilde{\beta}_i$ as defined in (D8–b) above, which exhibit some desirable statistical properties. In the first place $\tilde{\beta}_i$ are *unbiased*. This means that $\hat{\beta}_i(p, d)$ are realizations of $\tilde{\beta}_i$ (for some $w_o \in W$, $\tilde{\beta}_i(w_o) = \hat{\beta}_i(p, d)$), and the mean value of $\tilde{\beta}_i$, i.e., the integral of $\tilde{\beta}_i$ over W with respect to μ, is identical with β_i, i.e., with the "true" but unknown value hypothesized to make equations (1) true.

Unbiasedness cannot, of course, be proved right away. Further idealizing mathematical assumptions have to be postulated. A classical set of idealizing assumptions is given in the following definition.[6]

(D9) x is a *classical* linear stochastic demand system ($x \in$ CLSD) iff there exist $n, p, d, W, \mathfrak{A}, \mu, \tilde{p}, \tilde{d}, \tilde{u}, \beta$ such that $x = \langle n, p, d, W, \mathfrak{A}, \mu, \tilde{p}, \tilde{d}, \tilde{u}, \beta\rangle$ and

(1) $x \in M(\text{LSD})$,

(2) for all $t, t' \leq n$, $t \neq t'$ and all $a \in \text{Range}(\tilde{p})$:

(2.1) *the conditional expectation of* $u_t u_{t'}$ under the condition that $\tilde{p}$ takes value a is zero, i.e., $E(\tilde{u}_t\tilde{u}_{t'} \mid a) = 0$,

(2.2) the conditional variance of $\tilde{u}_t$ under the condition that $\tilde{p}$ takes value a exists and does not vary with t, i.e., there exists σ^2 such that $V(\tilde{u}_t \mid a) = \sigma^2$,

(2.3) $\tilde{u}_t$ is normally distributed,

(2.4) $V^* := \{w \in W / s^2(\tilde{p})(w) > 0\} \in \mathfrak{A}$ and $\mu(V^*) = 1$.

(T1) If $x = \langle n, p, d, \ldots, \beta\rangle$ is a CLSD then
(a) for $i = 1, 2$: β_i equals the expectation of $\tilde{\beta}_i$, i.e.,

$$\beta_i = E(\tilde{\beta}_i) = \int_W \tilde{\beta}_i \, \mathrm{d}\mu,$$

(b) for each $w_o \in W$ such that $\tilde{p}(w_o) = p$ and $\tilde{d}(w_o) = d$, and for $i = 1, 2$:

$$\tilde{\beta}_i(w_o) = \hat{\beta}_i(p, d),$$

Proof. For (a) see: e.g., Schneeweiß (1971) p. 59. (b) follows rectly from (D8).

the following we will concentrate on the property of unbiasedness d leave other desirable properties for statistical estimators like nsistency or sufficiency out of consideration. Our investigation plies to these statistical properties as well, but at the cost of formal mplication. We only note that the generalization expressed in (D12) low may be strengthened (and thereby specialized again) such that rther criteria for estimators can be taken into account.

III. CLASSICAL LINEAR STOCHASTIC DEMAND SYSTEMS AS A METHOD OF MEASUREMENT

ıe actual procedure for the estimation of the parameters of linear ›chastic demand systems is this. Observe time-series $\langle p_t, d_t\rangle$ and lculate $\hat{\beta}_1(p, d)$, $\hat{\beta}_2(p, d)$. These are the best values for β_1, β_2 you n obtain. Full stop.

The rest of the story is justification. If for a moment we do not care out the justification this procedure gives rise to a method of easurement as defined in Section I.

(D10) x is a *measuring model for* the defined term $\hat{\beta}$ *by regression* ($x \in M_m$(CLSD)) iff x has the form $\langle n, p, d, W, \mathfrak{A}, \mu, \tilde{p}, \tilde{d}, \tilde{u}, \beta\rangle$ where $\langle n, \ldots, \beta\rangle \in$ CLSD and $\hat{\beta} = \langle \hat{\beta}_1(p, d), \hat{\beta}_2(p, d)\rangle$

ıe measured values in such measuring models are the parameters (p, d), and a corresponding method of measurement is easily fined.

(T2) M_m(CLSD) is a method of measurement for the defined term $\hat{\beta}$ characterized by B, Σ, τ, $\approx$ and D where

(1) D is the definition of $\hat{\beta}$ (compare (D8–b)),

(2) B is the conjunction of the axioms (D6–1) to (D6–5), (D7–2) to (D7–4), (D9–2), and (D),

(3) $\Sigma = \langle \Sigma_1, \Sigma_2 \rangle$ where Σ_1, Σ_2 are generated by neighbourhoods of the form U_z^ϵ, $U_{\hat{\beta}}^\epsilon$ defined by $z' \in U_z^\epsilon$ iff $z = \langle n, p, d, W, \mathfrak{A}, \mu, \tilde{p}, \tilde{d}, \tilde{u}, \beta \rangle$, $z' = \langle n, p', d', W, \mathfrak{A}, \mu, \tilde{p}, \tilde{d}, \tilde{u}, \beta \rangle$, $|p - p'| < \epsilon$ and $|d - d'| < \epsilon$, and $\hat{\beta}' \in U_{\hat{\beta}}^\epsilon$ iff $\hat{\beta} = \langle \hat{\beta}_1, \hat{\beta}_2 \rangle$, $\hat{\beta}' = \langle \hat{\beta}_1', \hat{\beta}_2' \rangle$ and $|\hat{\beta} - \hat{\beta}'| < \epsilon$,

(4) τ is the type of linear stochastic demand systems extended by $\hat{\beta}$,

(5) $\approx$ is identity.

Proof. $\hat{\beta}_i(p, d)$ is a continuous function up to one singularity for $p = 0$. The remaining requirements are trivially satisfied.

Note that the measuring models refer to two sets of parameters: $\langle \beta_1, \beta_2 \rangle$ which are purely hypothetical and occur in the linear equation (1), the "law" of the theory, and $\langle \hat{\beta}_1(p, d), \hat{\beta}_2(p, d) \rangle$ which are defined in terms of the data given by p and d. $\langle \beta_1, \beta_2 \rangle$ are the values one would like to determine (to measure) while $\langle \hat{\beta}_1(p, d), \hat{\beta}_2(p, d) \rangle$ are the values actually obtained. Usually, $\hat{\beta}_i(p, d)$ differs from β_i. So in what sense can we say that the method of measurement described is a method of measurement *for* β (and not only for $\hat{\beta}$, which holds trivially)?

In answering this question we refer to the statistical justification already mentioned. $M_m \in$ (CLSD) is a method of measurement *for* β inasfar as (1) the values actually measured, $\hat{\beta}_i$, are realizations of estimators $\tilde{\beta}_i$, and (2) integration over each $\tilde{\beta}_i$ yields the value β_i, i.e., the "true" value one wants to find out.

It must be emphasized that the provable identity of $\int \tilde{\beta}_i \, d\mu$ and β_i does not help to determine the desired values β_i. For of all the values of $\tilde{\beta}_i$ the only one we know is $\tilde{\beta}_i(w_o)$, where w_o is the event for which $\tilde{p}(w_o) = p$ and $\tilde{d}(w_o) = d$. These are the only data at hand. For all other $w \in W$, $w \neq w_o$, $\tilde{\beta}_i(w)$ is unknown, and therefore the value of the integral cannot be calculated.

Also it has to be noted that neither the theoretical assumption of linearity nor the true parameters β_i occurring in it are used or play any direct role in the determination of the measured values $\hat{\beta}_i(p, d)$.

IV. JUSTIFICATION

In order to discuss the justification of M_m(CLSD) and its differences to methods of measurement in the natural sciences, let us use the phrase:

> measuring model x for function F is intended to measure function G.

The reason for using this phrase is the following. Usually, if we want to measure some function G we already have at hand some, perhaps rather weak, criteria of identity for G, conditions that G has to satisfy independently of the result of the measurement process. In most cases the function we want to measure already has a name, like "mass", "utility", or "price". In these cases we do not simply perform some process of measurement and accept the result as that value we intended to measure. Rather, we have to justify why some measuring model measures price, and not, say, utility. On the other hand the measuring model actually produces *some* value, independently of whether we accept this value as adequate or not. We call these values, the function values of the function F, *measured values*.

Consider two typical cases from the natural sciences.

Case A. The function G one intends to measure is a primitive of the theory T which governs the process of measurement. In this case the criteria of identity for G are given by theory T. The function one intends to measure is the function G as determined by its role in theory T. In order to guarantee that these criteria are met in the course of measurement it is assumed that the measuring model is a proper model of theory T. If the measuring model is a model of T then the values of F actually measured satisfy rather strong conditions as given by the axioms of T: they are in this sense consistent with T. Since the criteria for G consist of these same requirements, both F and G satisfy the same criteria. In many concrete cases this entails identity of F and G.

This is the situation of theory-dependent measurement introduced in Section I. The criteria for G are given by the axioms of T, these same axioms are used and presupposed in the course of the measurement which yields values of F, and therefore the values of F are acceptable, they "are" (i.e., may be accepted as) values of the function

G one intends to measure. The situation may be depicted as in Figure 1. In this case the justification for the assertion that the method of measurement really produces the values one intends to measure is this. Because the process of measurement is assumed to be a model of the theory, the measured values of F satisfy the criteria of identity for G, the function one intends to measure. Therefore, values of F are acceptable as values for G.

Case B. The function F, whose values are the actual result of the process of measurement, is a defined term. Usually the function G, the function one intends to measure, is given by the same definition as F. Take measurement of mean velocity as an example. Mean velocity is defined in terms of positions, the measured values are calculated by using this definition, and the function one intends to measure is velocity as given by the same definition. The criteria for G will reduce

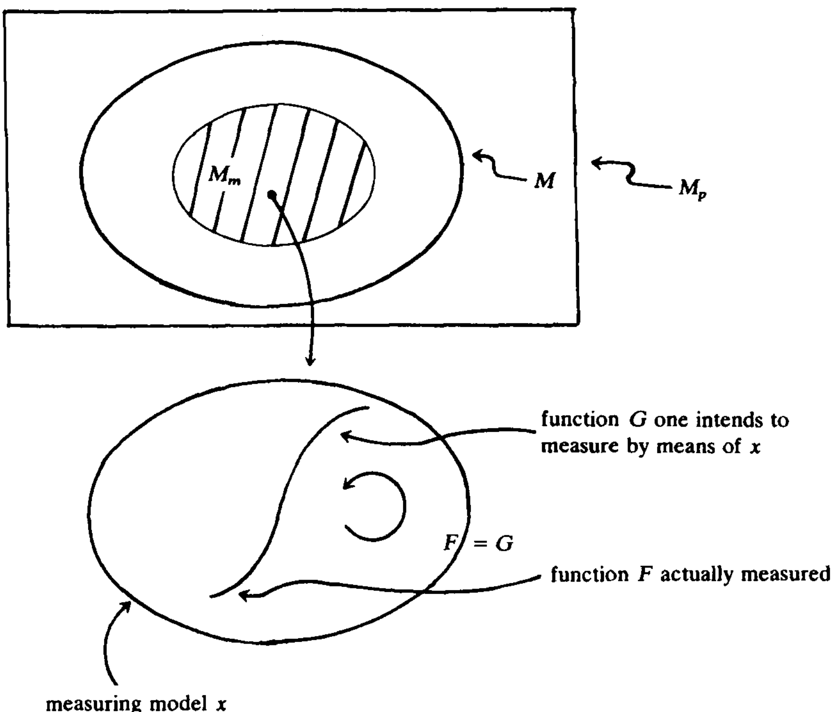

Fig. 1.

to criteria for those functions in terms of which G is defined. This leads to the situation of case A just discussed: the criteria of identity are given by the axioms of the theory governing the process of measurement. The situation is depicted in Figure 2. The important point is the following: in the natural sciences measurement of a defined term means that the function G one wants to measure by means of producing measured values of function F *practically never* is a primitive of the theory which is used in the respective measuring model.

What is the justification for the assertion that measurement of a defined term really produces the value one intends to measure? M_m is

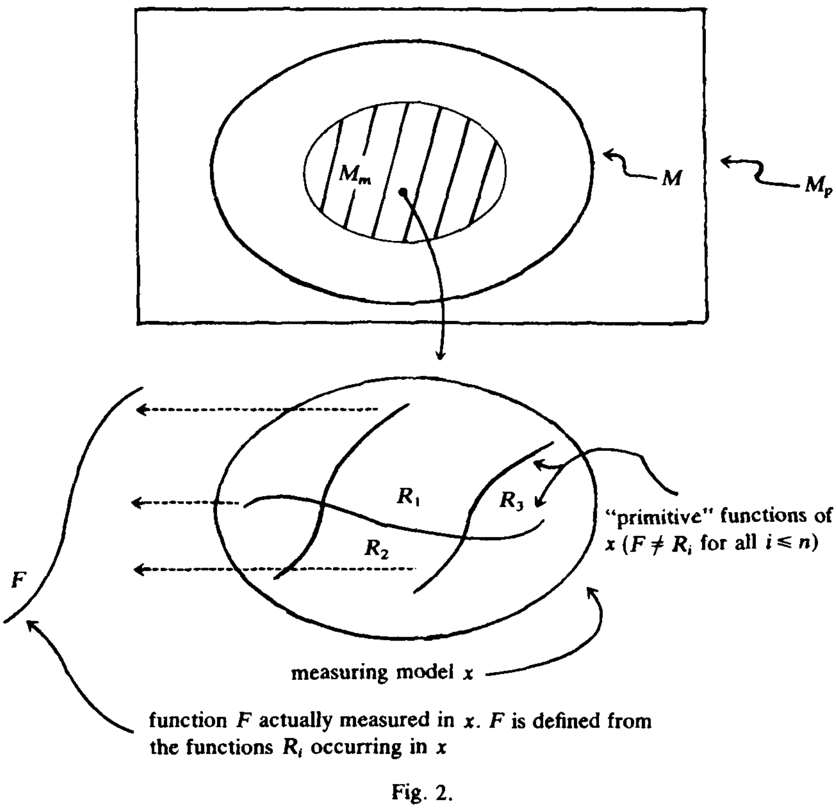

Fig. 2.

accepted as a method of measurement for G because the process of measurement is assumed to be a model of the theory, and because the definitions of F and G are identical. These assumptions entail that the *definiens* of F satisfies the criteria of identity for the *definiens* of G as given by the axioms of the theory. Therefore the *definiens* for F is acceptable as a *definiens* for G, and, since the definitions of F and G are the same, the values of F are acceptable as values for G.

We now are prepared to turn to the present example of measurement by OLS. In this case the term F actually measured, namely $\hat{\beta}$, is explicitly defined, so we are in the situation of case B just discussed. But there is a decisive difference between measurement by OLS and measurement of a defined term in the natural sciences. Now the function G one intends to measure *is* a primitive of the theory which governs the measuring model. In this case our terminology yields the following. We have measuring models for the *defined* term F which are intended to measure a *primitive* function G ("primitive" with respect to the theory governing the measuring models). The picture is as in Figure 3. In the natural sciences the measuring model is a model of some theory T, the function F actually measured is explicitly defined, and the function G one intends to measure is *not* a primitive of T. In the case of OLS the function G one intends to measure *is* a primitive of the theory.

The criteria for identity of G, i.e., for β in the case of OLS, are given by the theory from which $G(\beta)$ comes from: the "theory" of linear stochastic demand systems. These criteria will, however, usually not be met by the measured values $\hat{\beta}$. The justification for taking OLS as a method of measurement for G was worked out in the previous section: OLS is accepted because $\hat{\beta}$ is a realization of the random variable $\tilde{\beta}$ whose mean value is identical with β, at least under the restrictions of classical LSD's. This may be formalized one step further.

(D11) We say that y is *the mean value of* $x \in M_m(\text{CLSD})$ ($y = E(x)$) iff x has the form $\langle n, p, d, W, \mathfrak{A}, \mu, \tilde{p}, \tilde{d}, \tilde{u}, \beta\rangle$ and $y = \langle n, p, d, W, \mathfrak{A}, \mu, \tilde{p}, \tilde{d}, \tilde{u}, E(\tilde{\beta})\rangle$

(T3) If $x \in M_m(\text{CLSD})$ then $E(x) \in M(\text{LSD})$

Proof. Trivial by (T1).

Prima facie we could omit completely the theoretical apparatus

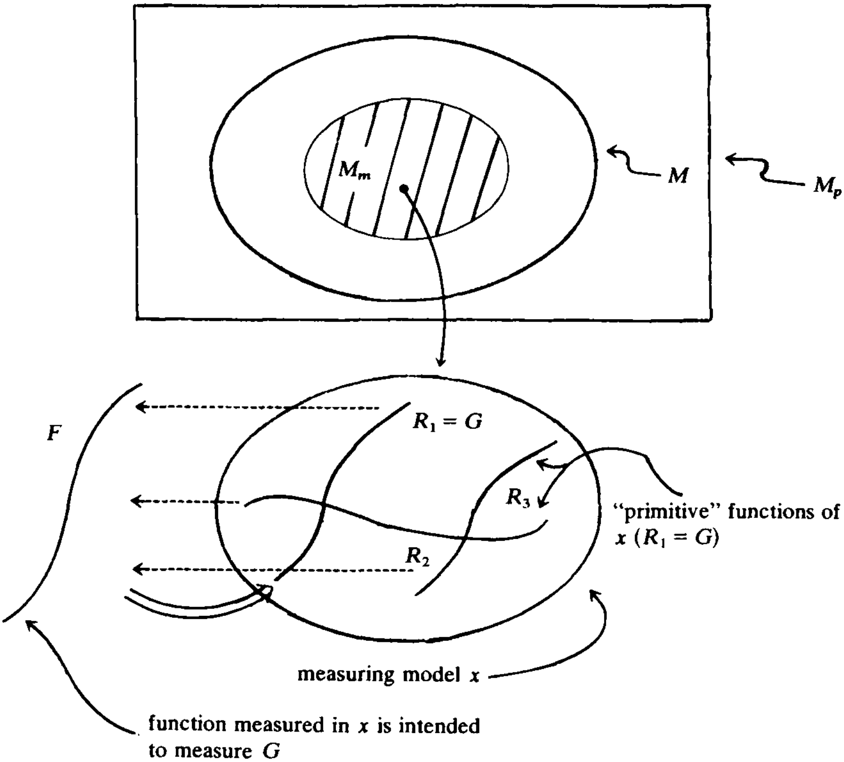

Fig. 3.

together with the probability space in the measuring models which would leave us with structures of the form $\langle n, p, d, \hat{\beta}(p, d)\rangle$. But then the statistical justification for the particular choice of $\hat{\beta}$ could not even be stated for then we lack of the true value β, and the theoretical condition of linearity which is essential for making β "true".

These considerations directly lead us to a central difference between the natural and the social sciences: in the social sciences the models of the respective theories exhibit some distance to observational or experimental results. In order to account for OLS as a method of measurement the criteria for acceptance known from the natural sciences have to be generalized and weakened.

The acceptance and justification of methods like OLS implicitly refers to some counterfactual argument:

(2) If we could collect further data for different, and many events $w \in W$ the mean-values of the corresponding $\tilde{\beta}_i(w)$ in the long run would converge (with probability 1) towards β_i.

In the natural sciences the same argument is *not* counterfactual. There, it simply describes what actually happens. In the natural sciences we are *in fact able to repeat experiments with the* "same" system (philosophical objections notwithstanding), and in this way to realize the antecedent of (2). This difference between the phenomena dealt with by natural and social sciences is of course well known, but it is only in connection with concepts of measurement that it may be formulated sharply.

It is tempting here to introduce the notation of a repeated experiment given by a sequence $w_1, w_2, \ldots$ in W and corresponding data $\tilde{d}(w_i)$, $\tilde{p}(w_i)$, and to refer to corresponding convergent sequences of mean-values. But this account would introduce an element which, in fact, is not operationally accessible and has no real referent in many situations in the social sciences. Such convergent sequences are typical for natural science but in the social sciences they often simply are not feasible. The only way to obtain such sequences in the social sciences is to increase the length of the time-series (n in (D6)). But social systems change quickly over time, and there is no warrant for regarding such a sequence as a repetition of the "same" situation. In the natural sciences, by contrast, the systems are stable enough to regard a "time series" of observations as "repetitions of observations of the same system". And this is of course the basic justification for applying probabilities and statistics.

V. REGRESSION METHODS OF MEASUREMENT

The kind of measurement studied here may easily be generalized. We do not attempt to give the most general formulation. Sticking to cases in which the determination of real parameters is at stake our definition still covers a wide range of examples in which OLS is applied. Roughly, the generalized measuring models which we call regression measuring models are measuring models in the sense of Section I for some parameters $\hat{\beta}$ which are added to the models of a given theory

((D12–a) below). $\hat{\beta}$ is required to be definable in terms of the other components of the measuring model (D12–a–1). Furthermore, the model should contain a probability space $\langle W, \mathfrak{A}, \mu \rangle$ (D12–a–4), and for some estimators $\tilde{\beta}$, also definable in terms of the components of the measuring model (D12–a–5), the value $\hat{\beta}$ actually measured should be a realization of the estimator (D12–a–5.3). A final, central condition is that the mean value of $\tilde{\beta}$ when replaced for the "true" parameters β in the original model $x[\beta]$ should yield a model of the theory (D12–a–5.4). This condition, which we call model replacement condition, is satisfied, for instance, whenever $E(\tilde{\beta}) = \beta$.

(D12) Let T be a theory with classes M_p and M of potential models and models, respectively.
- (a) x is a *regression measuring model* for β *in* T relative to B, Σ, τ, W, s, $\hat{\beta}$, D iff
  - (1) x is a measuring model for the defined term $\hat{\beta}$ characterized by B, Σ, τ, $=$ and D,
  - (2) $s \in \mathbb{N}$, and β, $\hat{\beta} \in \mathbb{R}^s$,
  - (3) x has the form $x[\beta]$ and $x[\beta] \in M$,
  - (4) W is a component of x and among the components of x there are $\mathfrak{A}$, μ such that $\langle W, \mathfrak{A}, \mu \rangle$ is a probability space
  - (5) there is some $\tilde{\beta}$: $W \to \mathbb{R}^s$ such that
    - (5.1) $\tilde{\beta}$ is definable in terms of the components of
    - (5.2) β is integrable with respect to μ,
    - (5.3) there is some $w_o \in W$ such that $\tilde{\beta}(w_o) = \hat{\beta}$
    - (5.4) $x[E(\tilde{\beta})] \in M$.
- (b) M_m is a *regression method of measurement for* $\{\beta\}$ iff there exist B, Σ, τ, s and D such that for all $x: x \in M_m$ there are W and $\hat{\beta}$ such that x is a regression measuring model for β in T relative to B, Σ, τ, W, s, $\hat{\beta}$ and D.

(T4) If M_m is a regression method of measurement for β then M_m is a method of measurement for the defined term $\hat{\beta}$

Proof. Obvious.

NOTES

* We are indebted to M. Kuettner and M. Kuokkanen for helpful remarks on an earlier draft.

[1] Compare (Balzer, 1985) and (Balzer, 1988) for more detailed accounts of this view.
[2] A standard reference on fundamental measurement is (Krantz et al., 1971). There, also various kinds of extensive systems are studied.
[3] Compare (Balzer, 1988) for further details.
[4] Cases in which only "parts" of F are uniquely determined can be treated simply by restricting the process of measurement to those arguments of F for which uniqueness obtains. Note the difference between requiring that F be uniquely determined in terms of B and the components of x different from x, and requiring that the function values of F are uniquely determined by F's arguments and by F. The former requirement amounts to

$$\forall F \forall F'(B(D_1, \ldots, R_n, F) \wedge B(D_1, \ldots, R_n, F') \rightarrow F = F')$$

while the latter means

$$\forall b \forall b'(F(a) = b \wedge F(a) = b' \rightarrow b = b').$$

[5] Compare (Balzer–Moulines–Sneed, 1987) for a detailed account of empirical theories.
[6] Note the close analogy to cases of theory-dependent measurement in the natural sciences. There, the basic laws of a theory usually also are not sufficient to guarantee uniqueness of the function to be measured. Further ad hoc assumptions, which can be drawn from a large stock of possibilities, have to be added in order to obtain measuring models. Compare (Balzer, 1985).

REFERENCES

Balzer, W.: 1985, *Theorie und Messung*, Springer, Berlin.

Balzer, W.: 1988, 'The Structuralist View of Measurement: An Extension of Received Measurement Theories', to appear in the *Minnesota Studies in the Philosophy of Science*.

Balzer, W., C. U. Moulines and J. D. Sneed: 1987, *An Architectonic for Science*, Reidel, Dordrecht.

Krantz, D. H., R. D. Luce, P. Suppes and A. Tversky: 1971, *Foundations of Measurement*, Academic Press, New York.

Schneeweiß, H.: 1971, *Ökonometrie*, Physica-Verlag, Würzburg.

Manuscript received 25 January, 1988

Universität München
Ludwigstr 31, D-800 München 22