Econometric Theory, **19**, 2003, 481–494. Printed in the United States of America.
DOI: 10.1017/S0266466603193073

DENIS SARGAN: SOME PERSPECTIVES

P.M. ROBINSON
London School of Economics

I attempt to present Denis Sargan's work in some kind of historical perspective, in two ways. First, I discuss some previous members of the Tooke Chair of Economic Science and Statistics, which was founded in 1859 and which Sargan held. Second, I discuss one of his articles, "Asymptotic Theory and Large Models," in relation to modern preoccupations with semiparametric econometrics.

1. INTRODUCTION

Unlike other contributors to this special issue, I was not a research student or colleague of Denis Sargan's, and little of my own research has been very close to his. I am not qualified, therefore, to offer a detailed appraisal of his contributions or a fund of personal anecdotes. However, I am very much aware not only of the striking originality of his research but also of his role, along with T.W. Anderson, E.J. Hannan, and others, in creating from the 1950s and 1960s today's rigorous discipline of econometric theory. I would like to try to place his work in some sort of historical context, but I am not a historian of econometrics, and so I will try to do so in a rather "individual" way.

From 1982 until his retirement in 1984, Denis Sargan held the Tooke Chair of Economic Science and Statistics at the London School of Economics (LSE). This chair dates from 1859—by way of comparison, the first university department of statistics, at University College London, was founded in 1911, whereas the Econometric Society was founded in 1930. I hope that a brief survey, in Section 2, of the previous incumbents of the Tooke Chair may give a (very casual) picture of how the interface between statistics and econometrics developed prior to Sargan, whose own treatment of econometric problems was in the highest traditions of mathematical statistics. Such a review seems especially apt in that one of the previous incumbents of the chair was F.Y. Edgeworth, who gave his name to Edgeworth expansions, one of Sargan's major contributions to econometrics being his development of higher order asymptotic theory.

This research was supported by a Leverhulme Trust Personal Research Professorship and ESRC grant R000239252. I thank Peter Phillips for helpful comments and Javier Hualde for obtaining material employed in Section 2. Section 2 is based partly on my *Doctor Honoris Causa* investiture address at Carlos III Madrid University, in October 2000. Address correspondence to: Peter M. Robinson, Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, UK.

© 2003 Cambridge University Press 0266-4666/03 $12.00

https://doi.org/10.1017/S0266466603193073 Published online by Cambridge University Press

Nowadays, the boundary between the sort of theoretical and empirical research pursued by econometricians on the one hand, and many statisticians on the other, has become very blurred. The influence has not been all in one direction, from statistics to econometrics. In particular, "identifiability" is a general issue with statistical models, but it was studied early on in the context of linear simultaneous equations systems in econometrics, and Sargan's own contributions (e.g., Sargan, 1983) to this subject were of particular depth. Also, increasingly many statisticians have become interested in problems arising in economics or finance. On the other hand, in recent years econometricians have drawn on an ever wider range of statistical models and techniques. A particular topic that has attracted great econometric attention, at both the theoretical and empirical levels, and which has largely developed since Sargan's retirement, is semiparametric inference, involving smoothed nonparametric estimation. (Indeed it has been much studied at the LSE since the mid-1980s.) The interest in semiparametric methods reflects the incomplete information typically available in econometric model building, in addition to increasing availability of large data sets and improved computing power. Sargan's own work was predominantly of a "parametric" character, but one of his papers, "Asymptotic Theory and Large Models" (Sargan, 1975), can be related to subsequent research on semiparametric estimation, and I will discuss this paper in Section 3, providing also some, minor, development of it. Section 4 contains some brief final comments.

2. THE TOOKE CHAIR

Denis Sargan was the ninth Tooke Professor of Economic Science and Statistics. The Tooke Chair was founded in 1859 at Kings College London, by public subscription and in memory of Thomas Tooke, who had died the previous year. Born in Cronstadt, Russia, the son of a chaplain of a British factory there, Tooke established himself first as a merchant and as an early supporter of the free trade movement and then founded, with Ricardo, Malthus, James Mill, and others, the Political Economy Club. As an economist, he devoted himself to systematically collecting and analyzing statistical and historical information concerning price changes from 1793 onward, concluding that the fluctuations over the following 45 years were due to circumstances affecting the supply of commodities rather than to changes in the system of currency, the popular view then being that the suspension of convertibility from 1793, and its resumption in 1819, led to a depreciation, and then an increase, in the value of currency. Tooke's investigations also led him to oppose the quantity theory of money. The data he collected were subsequently used by other researchers, including Jevons.

Both the first and the third holders of the Tooke Chair, at Kings College, had been ordained as Church of England clergymen and were pioneers of economic history. James Thorold Rogers held the chair from 1859 to 1890. He espoused controversial political opinions and was briefly a Member of Parliament, but

https://doi.org/10.1017/S0266466603193073 Published online by Cambridge University Press

he also conducted minute historical and statistical investigations leading to many publications, the best known being a history of agriculture and prices from 1759 to 1793. William Cunningham held the chair from 1891 to 1897 and wrote the first textbook on economic history, which went to seven editions; Cunningham's efforts had much to do with the establishment of economic history as a separate discipline.

Neither Rogers nor Cunningham was theoretically inclined, and in fact they held strongly critical attitudes toward economic theorizing, but their tenures of the Tooke Chair were separated by that of Francis Ysidro Edgeworth, between 1890 and 1891, who made contributions of lasting value to both mathematical economics and statistics. Born of an Irish landowning father and Spanish refugee mother, Edgeworth had no early formal training in advanced mathematics and started working in the moral sciences, but he went on to father the mathematical approach to economic theory. Edgeworth wrote extensively on such topics as taxation, monopoly and duopoly, international trade, and index numbers. He was a pioneer in using the indifference curve, and his invention of the core was not appreciated by the economics profession for many years. At the same time, he offered numerous important statistical insights. One, which we take for granted today, is the representation of the multivariate normal distribution in terms of pairwise correlation coefficients. Edgeworth's work on correlation greatly influenced Karl Pearson. He derived the *t*-distribution as the posterior distribution for the sample mean. Some of his papers contain some element of a proof of asymptotic efficiency of maximum likelihood estimates, prior to Fisher. To many statisticians and economists he is most associated with the "Edgeworth expansion," however. Recognizing that much real data are not normally distributed, he introduced a series expansion that, as a rival to the Pearson family, is able to flexibly describe skewness, kurtosis, and other phenomena. The "Edgeworth expansion" we refer to nowadays is employed to improve on the central limit theorem in approximating sampling distributions, and it has led to refinements in statistical inference in many areas of statistics and econometrics; more recently, strong theoretical support for the bootstrap has been provided by its ability to achieve an Edgeworth correction. Edgeworth was the first editor of the *Economic Journal*, serving in that role for 21 years.

Edgeworth was the epitome of the absent-minded professor and was an incoherent lecturer. The political economist and historian William Albert Samuel Hewins, Tooke Professor between 1897 and 1904, was more successful as a lecturer, in his early career speaking to working class audiences in the north of England on trade unionism and factory legislation. Hewins became the first director (i.e., chief executive) of the LSE. He was succeeded as Tooke Professor by Charles Stewart Loch, later knighted, and then in 1908 by Edmund John Urwick; both Loch and Urwick wrote on social questions.

In 1919 the Tooke Chair was transferred from Kings College to the LSE. By then, both were colleges of London University, and the aim was to avoid overlapping by concentrating economic study in one of the two colleges—Kings

https://doi.org/10.1017/S0266466603193073 Published online by Cambridge University Press

College has focused on the sciences and humanities, LSE on social science. However, the Tooke Chair was then in abeyance between 1920 and 1931, until it was taken up by Friedrich August von Hayek, who arrived from Vienna. Hayek's scholarly output spreads over six decades and beyond economics. He is widely known for his influence on the monetarist properties later pursued by Western governments, for his critique of Keynesianism, and for his integration of price theory, capital theory, and monetary theory, providing a base for reinterpreting 19th and 20th century economic history and promoting institutional reform. Hayek was awarded the Nobel Prize in economics in 1974.

Hayek left the LSE in 1950, for Chicago, and the Tooke Chair was then vacant until 1958, when it was filled by the promotion of Albert William Housego ("Bill") Phillips. Phillips had a background in electrical engineering, which provided him with a distinctive perspective on economic problems. He pioneered the application of optimal control theory to economics and even built a hydraulic perspex model of a dynamic Keynesian-type economy, the "Phillips machine." Phillips's single most widely known contribution is the "Phillips curve," relating wage inflation and unemployment. Phillips's research on continuous time dynamic models may also be mentioned; it particularly influenced work of fellow–New Zealanders at the LSE and the University of Essex.

Phillips left the LSE in 1967, and the Tooke Chair was vacant until, in 1982, it was transferred to Denis Sargan, then professor of econometrics. None of the previous incumbents of the Tooke Chair comes close to fitting the modern description of an econometrician (though one could imagine a synthesis of Edgeworth's ideas leading to an earlier development of econometrics). But in many of his contributions—his early work in economic theory (Sargan, 1955), his influential study of wages and prices in the UK (Sargan, 1964), his deep theoretical development of Edgeworth expansions (Sargan, 1976), and even his brief interest in continuous time models (Sargan, 1974)—Denis Sargan echoed concerns and achievements of a number of his predecessors.

The preparation of this section has relied very heavily on the references Blaug (1986), Blyth (1987), Dahrendorf (1995), Davis and Weaver (1927), Garrison and Kirzner (1987), Lee (1897, 1899), Mai (1975), Morgan (1990), Newman (1987), Pavetti (1987), Stigler (1987), Weaver (1937), Wickham Legg (1949), and *Who Was Who 1929–1940* and *1941–1950*.

3. SEMIPARAMETRIC ESTIMATION

So much of Sargan's work, on instrumental variables, simultaneous equations, specification testing, and Edgeworth expansions, can be seen as fundamental to, and influential in, the development of econometrics. "Asymptotic Theory and Large Models" (Sargan, 1975) can be seen as something of a one-off contribution, which turns out to contain ideas related to a more modern econometric preoccupation, semiparametric estimation; I referred to Sargan (1975) in my review article on semiparametric econometrics (Robinson, 1988) but did

https://doi.org/10.1017/S0266466603193073 Published online by Cambridge University Press

not discuss it in any depth there. The paper is not included among the two volumes of selections of Sargan's papers (Sargan, 1988) and has indeed rather escaped attention, perhaps because interest in large econometric models was already on the wane by the time the paper was published, in an issue of *International Economic Review* devoted to the proceedings of a symposium in econometric model performance. Though the paper is connected to some of Sargan's other work, including instrumental variables estimation of linear simultaneous equations systems (see Sargan, 1958), the main innovation in Sargan (1975) is the allowance for an increase, with sample size, of the number of variables, and equations, in the model, and the justification of parametric convergence rates with respect to the part being estimated, and asymptotic efficiency, relative to an underlying model involving infinitely many variables and equations. It thus deserves to be associated with the statistical and econometric literature on "adaptive" estimation in the context of a semiparametric model: the finite-dimensional parameter vector is estimated asymptotically as efficiently (in either the Cramér–Rao or Gauss–Markov senses) as if the infinite-dimensional nuisance function were of finite dimension. Hannan (1963) seems to have obtained the first results of this kind, showing adaptivity with respect to disturbance autocorrelation of unknown, nonparametric form, in time series regression models; Hannan followed this up with work on related semiparametric models, including distributed lag and linear simultaneous equations models. (Sargan's own work on frequency domain semiparametric estimation [Espasa and Sargan, 1977] and his early paper on the periodogram [Sargan, 1953] also warrants mention here.) The model in Sargan (1975) is perhaps not really semiparametric in the sense that, even though he was concerned with estimation of only a finite subset of equations, whose number remains constant as sample size increases, all equations in his underlying infinite system are assumed to be identified, each involving variables whose number also remains fixed as sample size increases. I will point out later that the same statistical achievement is possible without this kind of "full-information" requirement, presenting a version of the problem more in keeping with the modern semiparametric literature. First, however, I would like to describe in more detail the basic setup and accomplishments of Sargan (1975).

Define arrays $\{y_{st};\ s,t \geq 1\}$, $\{z_{st};\ s,t \geq 1\}$, $\{u_{st};\ s,t \geq 1\}$, $\{b_{st};\ s,t \geq 1\}$, $\{c_{st};\ s,t \geq 1\}$ and consider the model

$$\sum_{j=1}^{\infty} \{b_{ij}\, y_{jt} + c_{ij}\, z_{jt}\} = u_{it}, \qquad i,t \geq 1. \tag{3.1}$$

In (3.1), i indexes structural equations and t indexes observations (I will mostly try to follow Sargan's notation, though he employed instead an infinite-dimensional matrix presentation here). We will refer to the y_{jt} as "endogenous" variables and the z_{jt} as "exogenous" variables. The u_{it} are unobservable disturbances, satisfying at least

https://doi.org/10.1017/S0266466603193073 Published online by Cambridge University Press

$$E(u_{it}) = 0, \tag{3.2}$$

$$E(u_{it}u_{jt}) = \omega_{ij}, \tag{3.3}$$

$$E(u_{is}u_{jt}) = 0, \qquad s \neq t \tag{3.4}$$

for all i, j, s, t, and the z_{jt} are nonstochastic and uniformly bounded. The ith equation in (3.1) is well defined, for all t, by the requirements that for some $p < \infty$, $q < \infty$,

$$b_{ij} \neq 0 \quad \text{for at most } p \text{ values of } j, \tag{3.5}$$

$$c_{ij} \neq 0 \quad \text{for at most } q \text{ values of } j, \tag{3.6}$$

and is normalized by the requirement

$$b_{ii} = 1. \tag{3.7}$$

It is supposed that (3.5)–(3.7) hold for all $i \geq 1$. Further, Sargan assumed that for all i

$$b_{ij} = 0, \qquad j > i + P, \quad \text{for some } P > 0, \tag{3.8}$$

$$c_{ij} = 0, \qquad j > qi, \tag{3.9}$$

$$\sum_{j:j\neq i}^{\infty} |b_{ij}| < 1, \tag{3.10}$$

and

$$\max_{i,j} |c_{ij}| < \infty. \tag{3.11}$$

The b_{ij}, c_{ij} are otherwise unknown. Condition (3.8) is a kind of near-recursivity requirement, and along with (3.10), which is clearly very strong, and (3.11), it ensures the existence of a stable solution of (3.1), whereas, given (3.6), condition (3.9) is always satisfied by a suitable reordering.

The simultaneous equations system (3.1) involves infinitely many equations and infinitely many endogenous variables y_{jt}, $j \geq 1$, and exogenous variables z_{jt}, $j \geq 1$, observed at time points t. In practice we have observations at $t = 1, \ldots, T$, for T finite, and construct a working model of $n(T)$ equations in y_{jt}, $1 \leq j \leq n(T)$ and z_{jt}, $1 \leq j \leq m(T)$, namely,

$$\sum_{j=1}^{n(T)} b_{ij} y_{jt} + \sum_{j=1}^{m(T)} c_{ij} z_{jt} = u_{it}, \qquad 1 \leq i \leq n(T), \qquad 1 \leq t \leq T, \tag{3.12}$$

with $m(T) + n(T) < T$. In discussing asymptotic statistical theory for estimation of the unknown parameters explaining the b_{ij}, c_{ij} in (3.12), traditionally (see, e.g., Johnston, 1984) $m(T)$ and $n(T)$ are fixed as $T \to \infty$, so $m(T) = m$, $n(T) = n$, say. In (3.12), however, $m(T)$ and $n(T)$ are regarded as increasing slowly with T, so that (3.12) "tends to" (3.1), to reflect an attitude often adopted

https://doi.org/10.1017/S0266466603193073 Published online by Cambridge University Press

by practical modelers that the more data we have, the more unknown parameters one can hope to estimate with reasonable precision. Even with the restrictions (3.5) and (3.6), when $n(T) \to \infty$ the number of unrestricted, unknown b_{ij}, c_{ij} in (3.12) also tends to infinity, so that (3.12) has something of a nonparametric, infinite-dimensional character. However, Sargan (1975) was concerned with estimating only the first N equations of (3.12), where N stays fixed as $T \to \infty$, so that a finite number of no more than $N(p + q - 1)$ unknowns are being estimated. The situation is then more like that in various semiparametric models, where it is desired to estimate a finite-dimensional vector as well as possible, or at least with convergence rate $T^{1/2}$, in the presence of an unknown nuisance function. In fact, Sargan (1975) was not concerned with developing new estimates to deal with this situation but rather with justifying established ones, developed for the traditional case of $m(T), n(T)$ fixed, in the context of the "increasing" model (3.12) nested in the infinite-dimensional system (3.1). Specifically, he showed that estimates that are known to achieve a certain asymptotic efficiency bound with respect to (3.12) with $m(T), n(T)$ fixed achieve the same bound with respect to (3.1).

To significantly simplify the presentation I will take $N = 1$ throughout; because N is in any case fixed, there is no essential loss of generality, the only substantive difference arising in the case $N > 1$ being that an estimate of the covariance matrix of $(u_{1t}, \ldots, u_N)$ is involved to achieve efficiency when those u_{it} are contemporaneously correlated; in the asymptotic theory Sargan developed, and that which I will later develop, this is handled relatively straightforwardly, given the other techniques employed to address the problem, and the allowance for this kind of extension to the theory I will subsequently give is likewise straightforward. We write the first equation of (3.1) or (3.12) as

$$y_{1t} + \sum_{j=2}^{p_1} b_{1j} y_{jt} + \sum_{j=1}^{q_1} c_{1j} z_{jt} = u_{1t} \tag{3.13}$$

for $p_1 \leq p$, $q_1 \leq q$, where $b_{12}, \ldots, b_{1p_1}$, $c_{11}, \ldots, c_{1q_1}$ are unknown and we have employed a convenient ordering of the y_{jt}, z_{jt}. Write

$$y_j = (y_{j1}, \ldots, y_{jT})', \qquad Y_1 = (y_1, \ldots, y_{p_1}), \tag{3.14}$$

$$z_j = (z_{j1}, \ldots, z_{jT})', \qquad Z_1 = (z_1, \ldots, z_{q_1}). \tag{3.15}$$

The estimates considered by Sargan are all of instrumental variables type. For a $T \times (p_1 - 1)$ matrix W we consider the instrumental variables estimate

$$\hat{a}_{IV} = \begin{bmatrix} W'Y_1 & W'Z_1 \\ Z_1'Y_1 & Z_1'Z_1 \end{bmatrix}^{-1} \begin{bmatrix} W'y_1 \\ Z_1'y_1 \end{bmatrix} \tag{3.16}$$

of $a = (b_{12}, \ldots, b_{1p_1}, c_{11}, \ldots, c_{1q_1})'$, assuming the inverse exists. (Our notation here differs somewhat from Sargan's.)

https://doi.org/10.1017/S0266466603193073 Published online by Cambridge University Press

An efficient choice of W involves the reduced form for $\tilde{y}_t = (y_{2t},\ldots,y_{p_1 t})'$. Sargan deduced that

$$E(\tilde{y}_t) = \sum_{j=1}^{\infty} \pi_j z_{jt}, \tag{3.17}$$

where the π_j are $(p_1 - 1) \times 1$ vectors, the π_j indeed converging exponentially to zero as $j \to \infty$ under the preceding conditions (due especially to (3.10)), so that (3.17) is bounded. Introduce the $T \times (p_1 - 1)$ matrix

$$\tilde{Y}_1 = \sum_{j=1}^{\infty} z_j \pi_j'. \tag{3.18}$$

Then take $W = \tilde{Y}$ in (3.16) and thus consider the estimate

$$\hat{a}_{INF} = \begin{bmatrix} \tilde{Y}_1' Y_1 & \tilde{Y}_1' Z_1 \\ Z_1' Y_1 & Z_1' Z_1 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{Y}_1' y_1 \\ Z_1' y_1 \end{bmatrix}, \tag{3.19}$$

where *INF* stands for "infeasible," both because the π_j are unknown and because the z_{jt} are observable only for $j \le n(T)$ (see (3.12)). Sargan argued that $W = \tilde{Y}$ is an efficient choice of instruments, in the sense that under suitable additional conditions the limiting variance matrix of $T^{1/2}(\hat{a}_{IV} - a)$ exceeds that of $T^{1/2}(\hat{a}_{INF} - a)$ by a nonnegative definite matrix; for this and his other results he assumed that the limits

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} z_{it} z_{jt} = m_{ij} \tag{3.20}$$

exist for all i and j and the $m(T) \times m(T)$ matrix (m_{ij}) has eigenvalues that are uniformly bounded away from zero and infinity as $T \to \infty$ (implying $m(T) \to \infty$).

Sargan then considered feasible efficient estimates, taking $W = \hat{Y}$, where

$$\hat{Y}_1 = \sum_{j=1}^{m(T)} z_j \hat{\pi}_p' = Z\tilde{\pi}_1', \tag{3.21}$$

writing

$$\hat{\Pi}_1 = (\hat{\pi}_1,\ldots,\hat{\pi}_{m(T)}), \qquad Z = (z_1,\ldots,z_{m(T)}) \tag{3.22}$$

for $(p_1 - 1) \times 1$ vector estimates $\hat{\pi}_j$ of the π_j. We then have the feasible estimate

$$\hat{a}_{FEA} = \begin{bmatrix} \hat{Y}_1' Y_1 & \hat{Y}_1' Z_1 \\ Z_1' Y_1 & Z_1' Z_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}_1' y_1 \\ Z_1' y_1 \end{bmatrix}. \tag{3.23}$$

For $\hat{\Pi}_1$ Sargan used preliminary instrumental variables estimates computed from all equations of (3.12). Given estimates $\hat{B}$ and $\hat{C}$ of the $n(T) \times n(T)$ and $n(T) \times m(T)$ matrices (b_{ij}) and (c_{ij}), take

https://doi.org/10.1017/S0266466603193073 Published online by Cambridge University Press

$$\hat{\Pi}_1 = [I_{p_1-1}, 0]\hat{\Pi}, \tag{3.24}$$

where $\hat{\Pi} = -\hat{B}^{-1}\hat{C}$, I_r is the $r \times r$ identity matrix, and 0 is here a $(p_1 - 1) \times (m(T) - p_1 + 1)$ matrix of zeros. In Sargan (1975) $\hat{B}$ and $\hat{C}$ are formed from preliminary single-equation instrumental variables estimates, as indicated by Brundy and Jorgensen (1971). These authors argued that using the least squares estimate $Y_1'Z(Z'Z)^{-1}$ in place of $\hat{\Pi}$ in (3.24) is computationally unattractive when $m(T)$ is large, as a result of the need to invert $Z'Z$; with p_j unexcluded y_{jt} and q_j unexcluded z_{jt} in the jth equation, $p_j \leq p$, $q_j' \leq q$, single equation instrumental variables estimation of the jth equation involves inversion of only a $(p_j - 1 + q_j)$-dimensional square matrix, and Brundy and Jorgensen (1971) preferred the prospect of $n(T)$ such inversions over least squares. Imposing some unprimitive conditions on $\hat{B}$ and $\hat{C}$ (though indicating circumstances under which they would be met), Sargan showed that $\hat{a}_{FEA} - \hat{a}_{INF} = o_p(T^{1/2})$, so that $T^{1/2}(\hat{a}_{FEA} - a)$ has the same limit distribution as $T^{1/2}(\hat{a}_{INF} - a)$, and is thus asymptotically as efficient.

From a technical standpoint, Sargan (1975) solved highly challenging problems by both deriving the reduced form (3.17) and justifying its estimation by the preliminary structural form estimates of B and C. However, computing power has developed considerably over the past 30 years, so that least squares estimation of the reduced form in the presence of a very large number of exogenous variables seems a less daunting prospect nowadays than it did to Brundy and Jorgenson (1971). Thus, the two-stage least squares estimate of a,

$$\hat{a}_{2SLS} = \begin{bmatrix} Y_1'Z(Z'Z)^{-1}Z'Y_1 & Y_1'Z_1 \\ Z_1'Y_1 & Z_1'Z_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1'Z(Z'Z)^{-1}Z'y_1 \\ Z_1'y_1 \end{bmatrix}, \tag{3.25}$$

seems worth considering. This is (3.16) with $W = Y_1'Z(Z'Z)^{-1}$. The term $\hat{a}_{2SLS}$ has a further, minor, advantage over $\hat{a}_{FEA}$ in that the first factor on the right of (3.25) is proportional to a consistent estimate of the limiting covariance matrix of the estimate, whereas the first factor on the right of (3.23), with Sargan's choice of $\hat{Y}_1$, is not symmetric and so will not be satisfactory for covariance matrix estimation. The main distinction from Sargan's work is theoretical, however. Though exploitation of overidentifying restrictions on all structural equations of (3.12) may produce desirable finite-sample statistical properties, in addition to avoiding inversion of a very large matrix, there is a cost: some misspecification of the restrictions could lead to inconsistent estimation of the reduced form and thence at least to a large-sample efficiency loss in the estimates of the first equation. In fact, the large-sample efficiency achievement of Sargan is possible without identifying other structural equations, as two-stage least squares allows in the traditional fixed-$m(T)$ context. We demonstrate this by a representation of the problem that is more in keeping with the modern semiparametric literature than Sargan's. A further, minor, observation is that Sargan did not actually present a central limit theorem for his $\hat{a}_{FEA}$, whereas we will

https://doi.org/10.1017/S0266466603193073 Published online by Cambridge University Press

do so, requiring some additional conditions, though in other respects our conditions are weaker.

We assume (3.13) with

$$\tilde{y}_t = \sum_{j=1}^{\infty} \pi_j z_{jt} + v_t, \tag{3.26}$$

with

$$\sum_{j=1}^{\infty} j^{1/2} \|\pi_j\| < \infty, \tag{3.27}$$

where u_{1t} and the $(p_1 - 1) \times 1$ unobservable vector v_t of reduced form disturbances are such that the (u_{1t}, v_t') are independent and identically distributed with zero mean and finite fourth moment. Like Sargan (1975), we assume the z_{jt}, $j \geq 1$, $t \geq 1$, are nonstochastic and uniformly bounded, and we also impose the condition in the sentence surrounding (3.20). Denoting by e_j the $(p_1 - 1) \times 1$ vector whose jth element is 1 and whose other elements are zero, by 0 the $q_1 \times 1$ vector of zeros, and $\tilde{\pi}_j = (\pi_j', e_j')'$, $1 \leq j \leq q_1$, $\tilde{\pi}_j = (\pi_j', 0')'$, $j > q_1$, we assume that

$$(\tilde{\pi}_1, \tilde{\pi}_2, \ldots) \text{ has full row rank;} \tag{3.28}$$

this is just an identifiability condition. Then with

$$m(T)^{-1} + m(T)^2/T \to 0, \qquad \text{as } T \to \infty, \tag{3.29}$$

we have that

$$T^{1/2}(\hat{a}_{INF} - a) \to_d N(0, \sigma^2 \Omega^{-1}), \tag{3.30}$$

$$T^{1/2}(\hat{a}_{2SLS} - a) \to_d N(0, \sigma^2 \Omega^{-1}), \tag{3.31}$$

where

$$\sigma^2 = V(u_{11}), \qquad \Omega = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \tilde{\pi}_j m_{jk} \tilde{\pi}_k', \tag{3.32}$$

and $\sigma^2 \Omega^{-1}$ can be estimated consistently by $\hat{\sigma}^2 \hat{\Omega}^{-1}$, where

$$\hat{\sigma}^2 = \frac{1}{T} y_1' y_1 - \hat{a}_{2SLS}' \hat{\Omega} \hat{a}_{2SLS}, \tag{3.33}$$

$$\hat{\Omega} = \frac{1}{T} \begin{bmatrix} Y_1' Z (Z'Z)^{-1} Z' Y_1 & Y_1' Z_1 \\ Z_1' Y_1 & Z_1' Z_1 \end{bmatrix}. \tag{3.34}$$

The proof is described in the Appendix.

It follows that $\hat{a}_{2SLS}$ has the same asymptotic efficiency as the infeasible optimal estimate $\hat{a}_{INF}$, and inference can be carried out in the manner presented in elementary textbooks (e.g., Johnston, 1984) for two-stage least squares with a fixed number of explanatory variables. Note that condition (3.27) is much

https://doi.org/10.1017/S0266466603193073 Published online by Cambridge University Press

milder than the exponential decay of the π_j derived by Sargan (1975) from his setup. The prescription (3.26) is rather appealing because it may be felt that whereas we may be prepared to identify a single structural equation of interest, the included endogenous variables y_{jt} are affected by a potentially unlimited number of exogenous variables z_{jt}, only some of which we observe. Moreover, the z_{jt} could be nonlinear functions of a possibly finite number of observables, so that $\sum_{j=1}^{m(T)} \pi_j z_{jt}$ constitutes a series approximation to a nonlinear $E(\tilde{y}_t)$ in (3.17) and $\sum_{j=1}^{m(T)} \tilde{\pi}_j z_{jt}$ a corresponding series estimate, $\tilde{\pi}_j$ being the jth column of $Y_1' Z(Z'Z)^{-1}$. Our setup can then be compared with that of Newey (1990), who considered semiparametric estimation using nearest neighbor (instead of our series) nonparametric estimation and assumed stochastic exogenous variables, unlike our nonstochastic ones. Indeed, Andrews (1991) and Newey (1994) have provided extensive theoretical treatments of the use of series estimation in semiparametric problems, in a wide context, though they did not explicitly consider the present problem and their regularity conditions were motivated by more general and elaborate situations. A paper closer to ours is that of Koenker and Machado (1999), who considered, under different kinds of conditions, generalized method of moments estimation with an increasing number of moment conditions. Our discussion here aims to draw parallels between Sargan's (1975) paper and later econometric developments, rather than offer something distinctively novel relative to the current literature. Extended versions of the problem discussed here might involve the structural equation or equations of interest being semiparametric or nonparametric or the number of equations of interest being thought of as increasing slowly with T.

4. FINAL COMMENTS

The London School of Economics was a sympathetic environment for research in quantitative economics and statistics before Sargan's arrival, and in a more global sense these areas were already quite developed. However, Denis Sargan deserves to be regarded as a founder of high-level econometric theory at the LSE, and in the UK, and also, alongside a few others, internationally. Econometrics has developed considerably since, and interests have changed, but nevertheless much of Sargan's work was ahead of its time and remains relevant and inspiring today.

REFERENCES

Andrews, D.W.K. (1991) Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica* 59, 307–345.

Blaug, M. (1986) *Who's Who in Economics: A Biographical Dictionary of Major Economists 1700–1986*, 2nd ed. Brighton: Wheatsheaf.

Blyth, C.A. (1987) Albert William Housego Phillips. In J. Eatwell, M. Milgate, & P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*, pp. 857–858. London: MacMillan.

https://doi.org/10.1017/S0266466603193073 Published online by Cambridge University Press

Brundy, J.M. & D.W. Jorgenson (1971) Efficient estimation of simultaneous equations by instrumental variables. *Review of Economics and Statistics* 53, 207–224.
Dahrendorf, R. (1995) *A History of the London School of Economics and Political Science 1895–1995*. Oxford: Oxford University Press.
Davis, H.W.C. and J.R.H. Weaver (1927) *Dictionary of National Biography, 1912–1921*. London: Humphrey Milford.
Espasa, A. and J.D. Sargan (1977) The spectral estimation of sets of simultaneous equations with lagged endogenous variables. *International Economic Review* 18, 583–605.
Garrison, R.W. and I.M. Kirzner (1987) Friedrich August von Hayek. In J. Eatwell, M. Milgate, & P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*, pp. 609–614. London: MacMillan.
Hannan, E.J. (1963) Regression for time series. In M. Rosenblatt (ed.), *Time Series Analysis*, pp. 17–37. New York: Wiley.
Johnston, J. (1984) *Econometric Methods*, 3rd ed. New York: McGraw-Hill.
Koenker, R. and J.A.F. Machado (1999) GMM inference when the number of moment conditions is large. *Journal of Econometrics* 93, 327–344.
Lee, S. (1897) *Dictionary of National Biography,* vol. 49. London: Smith, Elder and Co.
Lee, S. (1899) *Dictionary of National Biography,* vol. 57. London: Smith, Elder and Co.
Mai, L.H. (1975) *Men and Ideas in Economics: A Dictionary of World Economists Past and Present*. Totowa: Rowman and Littlefield.
Morgan, M.S. (1990) *The History of Economic Ideas*. Cambridge: Cambridge University Press.
Newey, W.K. (1990) Efficient instrumental variable estimation of nonlinear models. *Econometrica* 58, 809–837.
Newey, W.K. (1994) Series estimation of regression functionals. *Econometric Theory* 10, 1–28.
Newman, P. (1987) Francis Ysidro Edgeworth. In J. Eatwell, M. Milgate, & P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*, pp. 84–98. London: Macmillan.
Pavetti, M. (1987) Thomas Tooke. In J. Eatwell, M. Milgate, & P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*, pp. 657–659. London: MacMillan.
Robinson, P.M. (1988) Semiparametric econometrics: A survey. *Journal of Applied Econometrics* 3, 35–51.
Sargan, J.D. (1953) An approximate treatment of the properties of correlogram and periodogram. *Journal of the Royal Statistical Society, Series B* 8, 140–152.
Sargan, J.D. (1955) The period of production. *Econometrica* 23, 151–165.
Sargan, J.D. (1958) The estimation of economic relationships using instrumental variables. *Econometrica* 26, 397–415.
Sargan, J.D. (1964) Wages and prices in the U.K.: A study in econometric methodology. In P.E. Hart, G. Mills, & J.K. Whitaker (eds.), *Econometric Analysis for National Planning*, pp. 25–59. London: Butterworth.
Sargan, J.D. (1974) Some discrete approximations to continuous time stochastic models. *Journal of the Royal Statistical Society, Series B* 36, 74–90.
Sargan, J.D. (1975) Asymptotic theory and large models. *International Economic Review* 16, 75–91.
Sargan, J.D. (1976) Econometric estimators and the Edgeworth expansion. *Econometrica* 44, 421–448.
Sargan, J.D. (1983) Identification and lack of identification. *Econometrica* 51, 1605–1633.
Sargan, J.D. (1988) *Contributions to Econometrics*, E. Maasoumi (ed.), vols. 1 and 2. Cambridge: Cambridge University Press.
Stigler, S.M. (1987) Edgeworth as a statistician. In J. Eatwell, M. Milgate, & P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*, pp. 98–99. London: MacMillan.
Weaver, J.R.H. (1937) *Dictionary of National Biography, 1922–1930*. London: Humphrey Milford.
Wickham Legg, L.G. (1949) *Dictionary of National Biography, 1931–1940*. London: Geoffrey Cumberlege.
Who Was Who 1929–1940. London: A&C Black.
Who Was Who 1941–1950. London: A&C Black.

https://doi.org/10.1017/S0266466603193073 Published online by Cambridge University Press

APPENDIX

We describe the main details of the proof of (3.31); the proofs of consistency of $\hat{\sigma}^2$ and $\hat{\Omega}$ for σ^2 and Ω use similar techniques but are easier, whereas the proof of (3.30) involves straightforward use of standard central limit theory for independent nonidentically distributed variables, noting the conditions on the z_{jt} and (u_{1t}, v_t'); it is perhaps only necessary here to observe that Ω is nonsingular as a result of the condition on the m_{jk} and (3.28).

Given (3.30), we immediately deduce (3.31) on proving

$$\hat{a}_{INF} - \hat{a}_{2SLS} = o_p(T^{-1/2}). \tag{A.1}$$

We may write the left side of (A.1) as

$$D_1^{-1} d_1 - D_2^{-1} d_2 = (D_1^{-1} - D_2^{-1})\, d_1 + D_2^{-1}(d_1 - d_2),$$

where

$$D_1 = \frac{1}{T}\begin{bmatrix} \tilde{Y}_1' Y_1 & \tilde{Y}_1' Z_1 \\ Z_1' Y_1 & Z_1' Z_1 \end{bmatrix}, \qquad d_1 = \frac{1}{T}\begin{bmatrix} \tilde{Y}_1' u_1 \\ Z_1' u_1 \end{bmatrix},$$

$$D_2 = \frac{1}{T}\begin{bmatrix} Y_1' Z(Z'Z)^{-1} Z' Y_1 & Y_1' Z_1 \\ Z_1' Y_1 & Z_1' Z_1 \end{bmatrix}, \quad d_2 = \frac{1}{T}\begin{bmatrix} Y_1' Z(Z'Z)^{-1} Z_1' u_1 \\ Z_1' u_1 \end{bmatrix},$$

with $u_1 = (u_{11}, \ldots, u_{1T})'$. Because $D_1 \to_p \Omega$ is entailed in the proof of (3.30), it suffices to show that $D_1 - D_2 \to_p 0$ and $d_1 - d_2 = o_p(T^{-1/2})$. We give the proof only for the latter, the proof of the former being easier because of the slower convergence that suffices.

Now

$$\tilde{Y}_1' u_1 = \Pi Z' u_1 + c_1,$$

$$Y_1' Z(Z'Z)^{-1} Z' u_1 = \Pi Z' u_1 + c_2 + c_3,$$

where

$$c_1 = \sum_{j=m(T)+1}^{\infty} \pi_j \sum_{t=1}^{T} z_{jt} u_{1t},$$

$$c_2 = \sum_{t=1}^{T} v_t \tilde{z}_t' (Z'Z)^{-1} \sum_{t=1}^{T} \tilde{z}_t u_{1t},$$

$$c_3 = \sum_{j=m(T)+1}^{\infty} \pi_j \sum_{t=1}^{T} z_{jt} \tilde{z}_t' (Z'Z)^{-1} \sum_{t=1}^{T} \tilde{z}_t u_{1t},$$

writing $\tilde{z}_t = (z_{1t}, \ldots, z_{m(T),t})'$. It thus remains to show that

$$c_i = o_p(T^{1/2}), \qquad i = 1,2,3. \tag{A.2}$$

https://doi.org/10.1017/S0266466603193073 Published online by Cambridge University Press

First, c_1 has mean zero and variance matrix

$$\sigma^2 \sum_{j=m(T)+1}^{\infty} \sum_{k=m(T)+1}^{\infty} \pi_j \sum_{t=1}^{T} z_{jt} z_{kt} \pi_k',$$

whose largest eigenvalue is

$$O\left(T \left(\sum_{j=m(T)+1}^{\infty} \|\pi_j\| \right)^2 \right) = o(T), \quad \text{as } T \to \infty,$$

where $\|A\|$ denotes the square root of the largest eigenvalue of $A'A$. This proves (A.2) for $i = 1$.

Next, c_2 has mean

$$E(v_1 u_{11}) \sum_{t=1}^{T} \tilde{z}_t' (Z'Z)^{-1} \tilde{z}_t = E(v_1 u_{11}) \mathrm{tr}((Z'Z)^{-1} Z'Z)$$

$$= O(m(T)) = o(T^{1/2})$$

and variance matrix

$$E\{v_1 u_{11} - E v_1 u_{11}\}\{v_1 u_{11} - E v_1 u_{11}\}' \sum_{t=1}^{T} \{\tilde{z}_t^{-1} (Z'Z)^{-1} \tilde{z}_t\}^2$$

$$+ \{E(v_1 v_1') E(u_{11}^2) + E(v_1 u_{11}) E(v_1' u_{11})\} \sum_{t=1}^{T} \sum_{\substack{s=1 \\ s \neq t}}^{T} \{\tilde{z}_s' (Z'Z)^{-1} \tilde{z}_t\}^2$$

$$= O\left(\max_{1 \le t \le T} \|\tilde{z}_t\|^2 \|(Z'Z)^{-1}\| \right) + O(m(T))$$

$$= O\left(\frac{m(T)^2}{T} + m(T) \right) = o(T)$$

so (A.2) is proved for $i = 2$.

Finally, c_3 has mean zero and variance matrix

$$\sigma^2 \sum_{j=m(T)+1}^{\infty} \sum_{k=m(T)+1}^{\infty} \pi_j \sum_{t=1}^{T} z_{jt} \tilde{z}_t (Z'Z)^{-1} \sum_{t=1}^{T} \tilde{z}_t z_{kt} \pi_k'. \tag{A.3}$$

Now

$$\left\| \sum_{t=1}^{T} z_{jt} \tilde{z}_t \right\| \le \left\{ \sum_{t=1}^{T} z_{jt}^2 \sum_{t=1}^{T} \|\tilde{z}_t\|^2 \right\}^{1/2} = O(Tm(T)^{1/2}),$$

so (A.3) is

$$O\left(Tm(T) \left(\sum_{j=m(T)+1}^{\infty} \|\pi_j\| \right)^2 \right) = O\left(T \left(\sum_{j=m(T)+1}^{\infty} j^{1/2} \|\pi_j\| \right)^2 \right) = o(T),$$

to prove (A.3) for $i = 3$.

https://doi.org/10.1017/S0266466603193073 Published online by Cambridge University Press