Sargan's Instrumental Variables Estimation and the Generalized Method of Moments

Manuel ARELLANO
CEMFI, 28014 Madrid, Spain (*arellano@cemfi.es*)

This article surveys J. D. Sargan's work on instrumental variables (IV) estimation and its connections with the generalized method of moments (GMM). First the modeling context in which Sargan motivated IV estimation is presented. Then the theory of IV estimation as developed by Sargan is discussed. His approach to efficiency, his minimax estimator, tests of overidentification and underidentification, and his later work on the finite-sample properties of IV estimators are reviewed. Next, his approach to modeling IV equations with serial correlation is discussed and compared with the GMM approach. Finally, Sargan's results for nonlinear-in-parameters IV models are described.

KEY WORDS: Errors in variables; Finite-sample properties; Minimax estimation; Nonlinear models; Serial correlation.

I pretty early realized that the Geary method was very close to LIML except he was using arbitrary functions of time as the instrumental variables, ... One could easily generalize the idea to the case, for example, of using lagged endogenous variables to generate the instrumental variables. That is really where my instrumental variable estimation started from. I was actually using it to estimate macroeconomic models fairly early, but the models didn't turn out very interesting to my way of thinking. I developed various ideas on time lags at that stage, and I actually had an early version of the Phillips curve in my model. I spent a lot of those years when I had spare time using an electric Marchand calculating machine in a Leeds University basement and getting out estimates which I didn't get published myself.

— J. D. Sargan, interviewed by Phillips (1985).

1. INTRODUCTION

This article surveys Denis Sargan's work on instrumental variables (IV) estimation and its connections with the generalized method of moments (GMM). Sargan pursued his interests in IV estimation all through his career. He focused in particular on asymptotic expansions of the distributions of estimators and test statistics. However, his two articles of 1958 and 1959 provided a fully developed theory of IV estimation that directly connects with the generalized method of moments (GMM) perspective of the 1980s. Thus it seems natural to concentrate on those articles for the purpose of this survey. (See Maasoumi 1988 for an account of Sargan's contributions to econometrics.)

There is a modeling connection between GMM and Sargan's focus on moment conditions that do not necessarily provide a complete description of the probability distribution under consideration. There is also a statistical connection, because Hansen's (1982) treatment of GMM estimation built on Sargan's results. Specifically, Hansen's GMM class of estimators generalized Sargan's class of linear and nonlinear IV estimators, and Hansen's analysis of efficiency followed Sargan's approach based on an optimal selection matrix for the moment conditions.

There is, however, an important aspect of Hansen's GMM approach other than greater generality that was not present in Sargan's work. Hansen proposed using GMM estimators based on moment conditions that exhibited dependence over time, constructed to be robust to the unmodeled components. To that end, he suggested using a weighting matrix that took into account temporal dependence. This suggestion had both modeling and statistical implications. I examine the modeling implications in connection with Sargan's work on IV models with serial correlation. A statistical implication was that consistent estimators of the covariance matrix of the sample moment conditions were required. Consistent estimators under various forms of time series dependence were suggested by Hansen (1982) and other authors.

In the early 1950s, errors-in-variables problems and simultaneous equations were pursued in two different literatures. The IV technique was associated with an ad hoc cure for measurement error. The limited-information maximum likelihood (LIML) method was developed for estimation of a single structural equation of a simultaneous system by Anderson and Rubin (1949, 1950). The mathematical analogy between the IV and LIML methods was first noted by Durbin (1954), and two-stage least squares (2SLS) was introduced by Basmann (1957) and Theil (1961). In fact, it was also introduced by Sargan (1958) as an IV estimator in a more general context. Sargan was not only providing a definitive analysis of the IV method, he was also using it to put together simultaneity and errors in variables, providing an IV analog of LIML as a minimax estimator, and (in his 1959 article) developing nonlinear-in-parameters IV estimation and its properties.

Sargan was thinking in terms of moment conditions, overidentifying restrictions, and partially specified models. He was also considering issues of choice of instruments, finite-sample biases, and underidentification. Many of the themes that appeared with renewed impetus in the econometrics literature of the 1980s and 1990s were present in a surprisingly mature way in Sargan's 1958 and 1959 articles. Yet relatively little attention was paid to this way of thinking about econometric estimation for the next 20 years. Most of the textbooks of the 1960s routinely surveyed simultaneous-equations estimators [2SLS, three-stage least squares (3SLS), LIML, full-information maximum likelihood (FIML)] as distinct from IV methods, which were alluded to in only a casual way in the context of discussions of measurement error,

Journal of Business & Economic Statistics
October 2002, Vol. 20, No. 4
DOI 10.1198/073500102288618595

although Hausman (1975) and Hendry (1976), building on Durbin (1963), both linked IV to FIML in each direction. Moreover, when in the mid-1970s the literature on nonlinear structural models began to develop, the connection with Sargan's 1959 article was often overlooked.

The article is organized as follows. Section 2 presents the modeling context in which Sargan motivated IV estimation. Section 3 reviews the theory of IV estimation and inference as developed by Sargan (1958). His approach to efficiency, his minimax estimator, tests of overidentification and underidentification, and his later work on the finite-sample properties of IV estimators are discussed. Section 4 discusses Sargan's approach to modeling IV equations with serial correlation and compares it with the GMM approach. Section 5 describes Sargan's (1959) results for nonlinear-in-parameters IV models. Finally, Section 6 contains some concluding remarks.

2. MODELS AND INSTRUMENTS

2.1 The Model

Sargan (1958) considered a structural equation of the form

$$\alpha_0' w_t^* = \varepsilon_t, \tag{1}$$

where the vector w_t^* (of order $q+1$) contained both endogenous and predetermined variables and ε_t was a structural random shock assumed to be independent of all of the predetermined variables in the complete system.

The variables were observed with error

$$w_t = w_t^* + v_t, \tag{2}$$

where v_t is a vector of measurement errors. In the case of variables measured without error, the corresponding elements of v_t are 0. Sargan interpreted the variables w_t^* as the "actual variables to which the economic agents react." Thus he argued that it was not necessarily true that "the determined variable is also an ideal economic variable in the sense that it is exactly equal to the variable to which some other economic agent later reacts, or that if an economic variable appears as a cause in two different equations the appropriate values of the ideal economic variable are the same" (1958, p. 395).

Combining the foregoing two equations, a relationship among the observed time series was obtained,

$$\alpha_0' w_t = u_t, \tag{3}$$

in which the equation's error u_t contains both a structural shock and measurement errors,

$$u_t = \varepsilon_t + \alpha_0' v_t. \tag{4}$$

2.2 The Instruments

Sargan assumed the availability of some predetermined variables (a vector of instrumental variables of order r, denoted z_t here) whose measurement errors were independent of v_t and ε_t. This requirement excluded the predetermined variables in the relationship (unless measured without error), lags of variables in the equation (unless—according to Sargan—"one makes the unrealistic assumption that measurement errors are not autocorrelated"), and any predetermined variable constructed from the same data as one of the variables in the equation. Sargan concluded that it was necessary that the *sources of data* used for constructing the instrumental variables should be largely independent of those used to construct the variables in the equation. Thus the only variables in the equation that could be used as instrumental variables were predetermined variables with zero measurement errors, such as the constant term, trends, and seasonal components.

Sargan suggested following Reiersøl (1945) and base estimation on the sample orthogonality conditions,

$$\frac{1}{T}\sum_{t=1}^{T} z_t u_t \equiv \left(\frac{1}{T}\sum_{t=1}^{T} z_t w_t'\right)\alpha = 0,$$

which provide r equations for the q ratios of the coefficients. Thus if $r = q$, then these conditions give a unique set of estimates of the coefficients α_0.

Sargan expressed concern about the real possibilities of this set of assumptions for identifying parameters of interest. He wrote: "It is not easy to justify the basic assumption concerning these errors, namely, that they are independent of the instrumental variables. It seems likely that they will vary with a trend and with a trade cycle. In so far as this is true, the method discussed here will lead to biased estimates of the coefficients. Nothing can be done about this since presumably, if anything were known about this type of error, better estimates of the variables could be produced" (1958, p. 396).

From the macrodata at his disposal and following Stone (1947), Sargan argued that effectively no more than three factors could be used as instrumental variables: a linear trend, the 10-year business cycle, and the rate of change of the 10-year business cycle. If there were large random events (of the same order of magnitude as the cyclical movements), such as strikes and wars, and if the structural errors could be regarded as independent of these events, then this might allow the identification of further coefficients. However, he felt that this possibility was "very rarely realistic." He concluded: "In practice, when data covering less than twenty years are used, it seems appropriate to use three instrumental variables: a linear trend, a lagged variable that leads in the trade cycle, and a lagged variable that lags with reference to the trade cycle. Analyses of single economic time series indicate that if longer periods of time were studied, a factor analysis might disclose more general factors" (1958, p. 415).

At the time, Sargan was particularly interested in wage-price inflation models. He was skeptical about the existence of a stable trade-off between inflation and unemployment (see Desai, Hendry, and Mizon 1997). In modeling wages, Sargan emphasized the role of union behavior and real wage resistance in wage bargaining: "I couldn't quite believe in the Phillips curve and some of the conclusions that were drawn about the ease with which a small increase in unemployment would cure inflation" (Phillips 1985, p. 124). Reading Sargan's 1964 Colston article, one senses that Sargan, who devoted much time to estimating macro models in those years, was probably left with the feeling that the IV method delivered less than he had initially hoped.

2.3 Errors in the Variables Versus Errors in the Equations

The basic motivation for the IV approach adopted by Sargan (1958) was to deal with equations that exhibited both simultaneity and measurement errors in exogenous (or predetermined) variables. In contrast with early econometric practice, the Cowles commission approach to econometrics stressed errors in the equations and simultaneity biases, as opposed to errors in variables and measurement error biases. This switch of emphasis took place even though the relative importance of simultaneity and measurement error had not been clearly established in many empirical areas (see Goldberger 1972, p. 993; Heckman 2000, pp. 71–72). In this respect, by giving symmetric consideration to the two types of errors, Sargan's 1958 article was an exception.

A particularly attractive feature of Sargan's IV framework that has become commonplace in modern econometric practice is to make operational the notion that measurement error in predetermined variables need not result in lack of identification, provided that the structural model contains sufficient overidentifying restrictions (Goldberger 1972, p. 996). Linearity in variables is an essential ingredient to the possibility of adding up errors in variables and errors in equations. Extending his 1958 results, Sargan (1959) considered IV models that were nonlinear in parameters but linear in variables. Beginning with the work of Amemiya (1974) and Jorgenson and Laffont (1974), Hansen (1982) and a sizable part of the GMM literature emphasized fully nonlinear relationships, but the measurement-error perspective was lost in this process.

3. THE INSTRUMENTAL VARIABLES ESTIMATION METHOD

Early contributions to the IV method were made by Wald (1940), Reiersøl (1945), and Geary (1948, 1949). Before these authors, the IV method can also be associated with the work of Working (1927), P. G. Wright (1928), and S. Wright (1934). (See Goldberger 1972 for a survey of the work of S. Wright.) The articles by Chernoff and Rubin (1953) and Durbin (1954) made contributions more closely related to subsequent work by Sargan. These articles were concerned with the connection between simultaneous equations and measurement error. Durbin (1954) discussed alternative processes generating measurement errors and pointed out the analogy between the overidentified IV case and Anderson and Rubin's method for a single structural equation. Chernoff and Rubin's chapter in Cowles Monograph 14 suggested a modification of LIML in which predetermined variables with measurement error are treated as endogenous variables.

The contributions of Sargan's 1958 article were to provide a definitive treatment of the IV method and to establish its asymptotic properties. A review of these contributions is given in this section. Sargan, like Hansen (1982), considered the asymptotic properties of a class of econometric estimators that are defined in terms of orthogonality conditions. This was a bold partial-information approach that focused estimation on objects of economic interest, abstracting from other features of the probability distribution of the variables under consideration. This approach also made it possible to define optimality in a well-defined sense, even if the resulting optimal estimators were not necessarily asymptotically efficient by comparison with full-information methods.

3.1 Asymptotic Properties

Sargan (1958) assumed that the $r \times 1$ vector of sample orthogonality conditions

$$g_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_t u_t$$

had an asymptotic joint normal distribution of the form

$$g_T \xrightarrow{d} \mathcal{N}(0, \sigma^2 \overline{M}_{zz}), \tag{5}$$

where $\sigma^2 = E(u_t^2)$ and $\overline{M}_{zz}$ denotes the stationary limit of $E(z_t z_t')$ given by

$$\overline{M}_{zz} = \lim_{T\to\infty} E\left(\frac{1}{T}\sum_{t=1}^{T} z_t z_t'\right) = p \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} z_t z_t'.$$

This was based on the assumption that the variables were stationary and that u_t was independent of z_s for $t \geq s$ and of u_s for $t \neq s$, for in this case

$$E(g_T g_T') = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} E(u_t u_s z_t z_s') = \sigma^2 E\left(\frac{1}{T}\sum_{t=1}^{T} z_t z_t'\right). \tag{6}$$

Some normalization is introduced so that $u_t = w_t'\alpha_0 = y_t - x_t'\beta_0$, where β_0 is a $q \times 1$ vector of parameters and the usual matrix notation $Z'X = \sum_{t=1}^{T} z_t x_t'$, and so on. Moreover, $\overline{M}_{zx}$ and $\overline{M}_{zw}$ denote the probability limits of $Z'X/T$ and $Z'W/T$. In the just-identified case (i.e., when $r = q$ and $\overline{M}_{zx}$ has full rank), Sargan (1958) showed that the simple IV estimator

$$\hat{\beta} = (Z'X)^{-1} Z'y \tag{7}$$

has limiting distribution

$$\sqrt{T}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, \sigma^2 \overline{M}_{zx}^{-1} \overline{M}_{zz} \overline{M}_{xz}^{-1}). \tag{8}$$

When there are more instruments than unknown parameters (i.e., when $r > q$), Sargan considered q linear combinations of the instruments

$$z_t^* = \Theta z_t,$$

where Θ is a $q \times r$ matrix of coefficients. Now, from the preceding discussion, the asymptotic variance of the IV estimator based on z_t^* is

$$\sigma^2 (\Theta \overline{M}_{zx})^{-1} (\Theta \overline{M}_{zz} \Theta') (\overline{M}_{xz} \Theta')^{-1}.$$

Sargan showed that the optimal choice for Θ that minimizes the asymptotic variance can be taken as

$$\Theta = \overline{M}_{xz} \overline{M}_{zz}^{-1}, \tag{9}$$

in which case

$$\sigma^2 (\overline{M}_{xz} \overline{M}_{zz}^{-1} \overline{M}_{zx})^{-1}. \tag{10}$$

Because Θ is unknown, Sargan argued, the previous result suggests consideration of $\widehat{\Theta} = X'Z(Z'Z)^{-1}$, giving rise to the estimator that solves

$$(X'Z(Z'Z)^{-1}Z'X)\hat{\beta} = X'Z(Z'Z)^{-1}Z'y, \tag{11}$$

which has the optimal asymptotic variance. Sargan suggested using the estimated variance matrix

$$\hat{\sigma}^2 T(X'Z(Z'Z)^{-1}Z'X)^{-1}, \tag{12}$$

where $\hat{\sigma}^2 = \hat{u}'\hat{u}/T$ and $\hat{u} = y - X\hat{\beta}$.

Sargan also noted that a different set of estimates is obtained for each normalization. He showed that these estimates differ asymptotically by quantities of order $1/T$ provided that $\overline{M}_{wz}\overline{M}_{zz}^{-1}\overline{M}_{zw}$ is of rank q. The estimator (11) is, of course, the two-stage least squares estimator, which at the time was being independently introduced from a different perspective by Basmann and Theil.

3.2 Sargan's Minimax Instrumental Variables Estimator

The motivation for the minimax method was to obtain a symmetrically normalized IV estimator for the overidentified case, after noting the lack of invariance to normalization of estimates of the form (11) obtained by deleting an arbitrary equation from the set of optimal sample moment conditions. Because the model posits zero correlation between errors and instruments, the idea was to choose as an estimator of β_0 the value that minimizes the maximum possible sample correlation between the errors and a linear combination of the instruments.

Consider arbitrary linear combinations of the instruments,

$$m_t(\gamma) = z_t'\gamma,$$

and of the variables appearing in the econometric model,

$$u_t(\alpha) = w_t'\alpha.$$

The squared sample correlation between $m_t(\gamma)$ and $u_t(\alpha)$ is given by

$$\rho^2(\alpha, \gamma) = \frac{(\alpha'W'Z\gamma)^2}{(\alpha'W'W\alpha)(\gamma'Z'Z\gamma)}. \tag{13}$$

For given α, the maximum correlation is given by

$$\lambda(\alpha) = \max_{\gamma} \rho^2(\alpha, \gamma) = \frac{\alpha'W'Z(Z'Z)^{-1}Z'W\alpha}{\alpha'W'W\alpha}. \tag{14}$$

Thus the resulting estimator is given by solving the minimax problem

$$\min_{\alpha}\left[\max_{\gamma} \rho^2(\alpha, \gamma)\right] = \min_{\alpha} \lambda(\alpha) = \hat{\lambda}_1, \tag{15}$$

subject to a normalization restriction. The statistic $\hat{\lambda}_1$ is the smallest eigenvalue of the matrix $W'Z(Z'Z)^{-1}Z'W$ in the metric of $W'W$.

The minimax estimator provided an IV analog and a generalization of Anderson and Rubin's LIML method. Sargan (1958, sec. 13) noted that "the LIML method is equivalent to using the instrumental variables method with all the predetermined variables in the model used as instrumental variables. This procedure is reasonable since an essential assumption of the LIML method is that there are no measurement errors."

Thus Sargan's conclusion was that in the context of a structural equation subject to measurement error, one can still use symmetrically normalized IV estimators, but these lack the LIML interpretation. Because the validity of instruments depends on the measurement error properties of the model, it is no longer true that all of the predetermined variables in the system are necessarily valid instruments.

Sargan's minimax estimator can also be regarded as minimizing the largest standardized linear combination of the sample moments,

$$\min_{\alpha}\left[\max_{\gamma} \frac{(u(\alpha)'Z\gamma)^2}{\gamma'V(\alpha)\gamma}\right] = \min_{\alpha}[u(\alpha)'Z]V(\alpha)^{-1}[Z'u(\alpha)], \tag{16}$$

where $V(\alpha) = \mathrm{var}[Z'u(\alpha)]$ and the previous equality follows from application of the Cauchy–Schwarz inequality. For $V(\alpha) = u(\alpha)'u(\alpha)(Z'Z)$, Sargan's estimator is obtained, but for choices of $V(\alpha)$ based on alternative assumptions about the form of the variance matrix of the moments, other generalized minimax or "continuously updated" GMM estimators of the type considered by Hansen, Heaton, and Yaron (1996) may be obtained.

3.3 Inference

3.3.1. Testing Overidentifying Restrictions. Sargan (1958) proposed a specification test of the existence of a relationship that satisfied all moment conditions. He showed that as long as $\overline{M}_{wz}$ has reduced rank q,

$$T\hat{\lambda}_1 \xrightarrow{d} \chi^2_{r-q}. \tag{17}$$

This provided an IV analog to one of the criteria derived by Anderson and Rubin (1949, 1950) for testing overidentification in a single equation from a system of simultaneous equations.

As a sketch of the argument, note that factoring $(Z'Z/T)^{-1} = CC'$, in view of (5) and the consistency of $\hat{\sigma}^2$, $C'g_T/\hat{\sigma} \xrightarrow{d} \mathcal{N}(0, I_r)$. Moreover, letting $G = C'(Z'X/T)$,

$$h = \frac{T^{-1/2}C'Z'\hat{u}}{\hat{\sigma}} = [I_r - G(G'G)^{-1}G']\frac{C'g_T}{\hat{\sigma}}.$$

Because the limit of $[I_r - G(G'G)^{-1}G']$ is idempotent and has rank $r-q$, it follows that

$$h'h = T\frac{\hat{u}'Z(Z'Z)^{-1}Z'\hat{u}}{\hat{u}'\hat{u}} \xrightarrow{d} \chi^2_{r-q}, \tag{18}$$

and hence also (17), because, due to their asymptotic equivalence, the same result holds if $\hat{u}$ is replaced by the minimax residual.

A statistic like $T\hat{\lambda}_1$ or $h'h$ has become known as a "Sargan test" and has become a standard complement when reporting IV estimates. Sargan argued that this "provides a significance test for the hypothesis that there is a relationship between the suggested variables with a residual independent of all the instrumental variables" and added that this "is a suitable test even when $\overline{M}_{wz}$ is of rank less than q since it can be shown that in this case the probability of rejecting the hypothesis will be less than in the other case" (1958, p. 404).

Sargan (1959) considered a generalization to nonlinear-in-parameters IV models, and Hansen (1982) extended this type of specification test to a general nonlinear GMM environment with dependent observations (often called a "J test").

3.3.2. Testing for Underidentification. Next, Sargan considered a test of underidentification, that is, a test of the hypothesis of the existence of a multiplicity of relationships that satisfy all moment conditions. This is a test of the null that the rank of $\overline{M}_{wz}$ is less than q against the alternative that the rank is equal to q. This type of test was also considered by Koopmans and Hood (1953) in the context of a single equation from a simultaneous system.

Sargan and Koopmans and Hood stated that when $\overline{M}_{wz}$ has rank $q-1$, letting $\hat{\lambda}_2$ be the second smallest characteristic root,

$$T(\hat{\lambda}_1+\hat{\lambda}_2) \xrightarrow{d} \chi^2_{2(r-q)+2}. \tag{19}$$

Thus this result could be used as a test of the hypothesis that the equation is underidentified and that any admissible equation has a homoscedastic and nonautocorrelated error. Sargan (1958) pointed out that "this hypothesis is not very likely to be true a priori since even if there is a relationship between the suggested variables with a nonautocorrelated residual, it is unlikely that there would be a second combination of these variables not only independent of all instrumental variables but nonautocorrelated as well." However, he regarded the use of the test as "a useful qualitative answer as to whether the estimates are reasonably well identified" (p. 405).

If $\overline{M}_{wz}$ has rank $q-1$, then there is another solution α_0^* not proportional to α_0 that satisfies the original moment equations,

$$E(z_t w_t')(\alpha_0, \alpha_0^*) = 0. \tag{20}$$

So a test of underidentification can be regarded as a test of the overidentifying restrictions in (20) subject to an extended normalization of (α_0, α_0^*). In fact, letting $A=(\alpha, \alpha^*)$, it turns out that $T(\hat{\lambda}_1+\hat{\lambda}_2)$ coincides with the minimizer of

$$T(\alpha' W'Z, \alpha^{*\prime} W'Z)(A'W'WA \otimes Z'Z)^{-1}\begin{pmatrix} Z'W\alpha \\ Z'W\alpha^* \end{pmatrix} \tag{21}$$

subject to $A'W'WA = I_2$. This has been shown by Arellano, Hansen, and Sentana (1999), who used the idea to consider tests of underidentification from a GMM perspective in a wider context.

3.4 Approximating the Distribution of Instrumental Variable Estimates

Sargan (1958) stressed the distinction between the purely theoretical asymptotic results and the accuracy of the asymptotic approximations for finite T. He discussed determinants of the quality of the asymptotic approximation and offered specific recommendations for practitioners. On the basis of unpublished calculations, he asserted that the biases of IV estimates and $T\hat{\lambda}_1$ were of order $r/T\lambda_2$, where λ_2 is the population counterpart of $\hat{\lambda}_2$. The implication was that the asymptotic approximation was poor when the relationship was almost unidentified and when the number of instruments was large relative to the sample size.

Sargan discussed the issue of instrument choice, pointing out a finite-sample trade-off between bias and efficiency. He argued that although the addition of a new instrumental variable will not worsen the asymptotic variance matrix, "the improvements are usually small after the first three or four instrumental variables have been added. Thus there may be no great advantage in increasing the number of instrumental variables, and ... it emerges that the estimates have large biases if the number of instrumental variables becomes too large" (Sargan 1958, p. 400). His practical suggestion was to require that $r \leq T/20$. In a similar vein, he also suggested a crude finite-sample adjustment to the chi-squared statistic. Finally, he compared IV with ordinary least squares (OLS), arguing that although theoretically IV is better than OLS because of consistency, for finite T the advantage of IV is less certain because "the instrumental variable estimates may have large biases especially in the almost unidentified case and in the event the number of instrumental variables is large" (pp. 412–413).

3.4.1. Improved Asymptotic Approximations. The use of asymptotic expansions to obtain improved approximations to the distributions of econometric estimators was pioneered by Sargan. The first work published in this area was that of Sargan and Mikhail (1971), who developed an Edgeworth (or Gram–Charlier) approximation to the distribution of IV estimates and evaluated the accuracy of the approximation for a model with two endogenous variables (the abstract of an earlier version of this article was published in *Econometrica* in 1964). In subsequent work, Sargan (1975a) approximated the distribution of t ratios of IV estimators and proved a general theorem on the validity of Edgeworth expansions for statistics that are defined as functions of a vector of more primitive statistics. This 1975 article also contained an early discussion of empirical Edgeworth approximations. The results of these two articles relied on "classical assumptions" in the sense that excluded lagged endogenous variables and nonnormal errors. The general conclusions were that "the asymptotic approximation will be particularly poor if r/T is not small, and if the variance of the reduced form errors is relatively large. In addition, increased correlation between the reduced form and equation errors worsen the asymptotic approximation" (Sargan 1975a, p. 340).

In an important article, Sargan (1976) obtained explicit formulas for the second-order Edgeworth expansion of a statistic defined as a smooth function of sample second moments of the data. These formulas covered models containing lagged dependent variables (a case also independently considered in Phillips 1977a, b). To summarize the setting of Sargan's results, let $\phi(p)$ represent a scalar estimator or test statistic as a function of a vector p of sample second moments that have been generated by some stationary stochastic process. If a central limit theorem is available for p, $\sqrt{T}(p-\mu) \xrightarrow{d} \mathcal{N}(0, \Sigma)$, using the delta method, a first-order approximation to the cdf of $b_T = \sqrt{T}[\phi(p) - \phi(\mu)]$ is given by

$$\Pr(b_T \leq s) = \Phi\left(\frac{s}{\sigma_\phi}\right) + O(T^{-1/2}),$$

where $\sigma_\phi^2 = d'\Sigma d$, $d = \partial\phi(\mu)/\partial p$, and $\Phi(\cdot)$ is the standard normal cdf. Sargan (1976) provided formulas for the coefficients of a refined approximation of the form

$$\Pr(b_T \leq s) = \Phi\left(\frac{s}{\sigma_\phi} + \frac{h_0}{\sqrt{T}} + \frac{h_1}{T}\left(\frac{s}{\sigma_\phi}\right) + \frac{h_2}{\sqrt{T}}\left(\frac{s}{\sigma_\phi}\right)^2 + \frac{h_3}{T}\left(\frac{s}{\sigma_\phi}\right)^3\right) + O(T^{-3/2}),$$

where the coefficients h_j were expressed as functions of the cumulants of p and the derivatives of $\phi(p)$. In this way, Sargan gave an explicit method for finding Edgeworth approximations for a large class of statistics from linear-in-variables time series econometric models. Successive corrections to these formulas appeared in Sargan's (1977) erratum, in Tse's (1981) Ph.D. thesis, and in appendix B of Arellano and Sargan (1990); the first correct general formulas, however, were published using a slightly different notation by Phillips (1977b).

3.4.2. Resampling Methods. Sargan's 1976 article was a remarkable work in many other respects. The abundance of technical results were coupled with a genuine concern with the problems of applying the theoretical refinements to improving the accuracy of asymptotic significance tests in models of a realistic size.

As an alternative to analytical Edgeworth expansions, Sargan (1976) suggested a resampling method that he called a "Barnard approximation" in implicit reference to Barnard's (1963) Monte Carlo testing. He argued that "an alternative approach, which has only rarely been used in econometrics, is to estimate the probability associated with a given significance test by simulating the model, using the estimated parameters, and observing the resulting proportion of simulated criteria falling within the asymptotic confidence interval" (1976, p. 428). Sargan pointed out that at first sight the method suffered from depending on the use of estimated parameters, but he noted that "the criteria have the property that their asymptotic distributions are independent of the parameters of the model. If we are considering a symmetric confidence interval for a t ratio, the difference between the asymptotic probability and the finite sample probability is usually a differentiable function of the parameters in the neighborhood of the true value uniformly of order $1/T$. It follows that the error in the estimated probability induced by using estimated values for these parameters is of order $T^{-3/2}$" (p. 429). He nevertheless was concerned that the asymptotic result might overstate the finite-sample properties of the resampling method, and suggested investigating the matter further through Monte Carlo experimentation.

These matters were pursued by Sargan (1981) (and later briefly summarized in Sargan 1993), who studied the asymptotic properties of a parametric bootstrap procedure. He also undertook a Monte Carlo exercise using a two-equation overidentified dynamic model estimated by 3SLS to study the properties of his bootstrap method with $T = 20$ and 50. He found that the bootstrap procedure gave some very poor estimates of the true size of the various confidence intervals considered.

3.4.3. Approximations When the Number of Instruments Is Large. Sargan (1975b) argued that conventional asymptotic theory, in which $T \to \infty$ and the model remains constant, was quite irrelevant for large models where the total number of variables was large relative to the sample size. As an alternative, he considered an asymptotic framework in which not only T, but also the number of instruments and the number of equations, were tending to infinity. He focused on estimation of a fixed number of parameters occurring in a subsystem of equations and studied the asymptotic properties of an iterated IV estimator that used the overidentifying restrictions in forming the instruments (of the type considered in Brundy and Jorgenson 1971). Sargan's theorem 3 established that the feasible and unfeasible IV estimators were asymptotically equivalent and concluded that for these type of estimators, the results of asymptotic theory were still a good approximation even in large models.

This article pioneered a literature that focused on the properties of IV and GMM estimates when the number of moment conditions and the sample size tended to infinity (cf. Kunitomo 1980; Morimune 1983; Bekker 1994).

4. SERIAL CORRELATION

The theory for Sargan's 1958 article was developed under the assumption that the error u_t was not autocorrelated. Sargan nevertheless regarded the assumption of lack of autocorrelation in measurement errors as unrealistic, and pointed out that because his IV estimates were based on minimal assumptions, they were still consistent even if the errors were autocorrelated. Because of this, he suggested that "it is probably not wise to use lagged values of a variable appearing in the relationship as instrumental variables" (1958, p. 413). The idea was to rely on instruments that would not lose their validity in the event of serial correlation in measurement errors.

A contribution of Hansen's GMM perspective has been the following reaction to this problem: If the economic problem suggests that $E(g_T) = 0$, but $\text{var}(g_T)$ differs from (6) because of serial dependence, then we obtain a consistent estimate $\widehat{\text{var}}(g_T)$ to perform optimal inference relative to the original moments g_T. A similar perspective was present in the IV estimators proposed independently by Chamberlain (1982) and White (1982) for cross-sectional and panel data linear models with heteroscedasticity and by Cumby, Huizinga, and Obstfeld (1983) for linear rational expectations models. To do this, one could specify a parametric process for u_t or try to obtain a robust estimate of $\text{var}(g_T)$ under more general assumptions (as in Hansen and Hodrick 1980; Hansen 1982). If the sample size is small, then the former may be a better idea than the latter, but even if the process for u_t is misspecified, the IV estimates will still be consistent. In any event, the suggested estimates will minimize

$$g_T'\left[\widehat{\text{var}}(g_T)\right]^{-1} g_T. \tag{22}$$

The motivation for Sargan's 1959 article was IV estimation of models with autocorrelated errors, but he did not follow the

GMM route. Sargan specified a reduced-form process for the errors and used this specification to change the orthogonality conditions. He then considered joint estimation of the structural parameters and those in the error process. By doing so, he gained efficiency but at the cost of a more fragile estimator. (An early discussion of the pros and cons of these two approaches was given in Griliches 1967, pp. 40–41.)

Sargan considered an equation of the form

$$a_0' w_t^\dagger = \eta_t \tag{23}$$

with autoregressive errors

$$\eta_t = \sum_{j=1}^{J} \psi_j \eta_{t-j} + u_t, \tag{24}$$

leading to the transformed equation

$$a_0' w_t^\dagger - \sum_{j=1}^{J} \psi_j a_0' w_{t-j}^\dagger \equiv \alpha(\theta_0)' w_t = u_t, \tag{25}$$

where $\alpha(\theta_0)' = (a_0', -\psi_1 a_0', \ldots, -\psi_J a_0')$ and

$$w_t = \begin{pmatrix} w_t^\dagger \\ w_{t-1}^\dagger \\ \vdots \\ w_{t-J}^\dagger \end{pmatrix}.$$

He then suggested joint estimation of the structural parameters a_0 and the autoregressive parameters ψ_j from a set of nonlinear moment conditions of the form

$$E(z_t u_t) = 0. \tag{26}$$

(Specification tests of the common-factor bilinear restrictions in $\alpha(\theta_0)$ have been studied in Sargan 1964, 1980.)

To examine the nature of these moments in more detail, suppose that a vector $z_t^\dagger$ provides valid instruments for the errors of the original equation, such that

$$E\big(z_t^\dagger \eta_t\big) = 0. \tag{27}$$

Sargan's 1958 estimator based on these moments would still be consistent even if the errors were serially correlated. If (27) holds and η_t is serially correlated, in general also $E(z_t^\dagger \eta_{t-j}) = 0$ for $j = 1, \ldots, J$ (e.g., if $z_t^\dagger$ were strictly exogenous instruments). If this is so, then it will also be true that

$$E\big(z_t^\dagger u_t\big) \equiv E\big(z_t^\dagger \eta_t\big) - \sum_{j=1}^{J} \psi_j E\big(z_t^\dagger \eta_{t-j}\big) = 0. \tag{28}$$

However, the autoregressive parameters are not identified from the latter set of moments, because they will hold not only for the true value of the ψ_j, but also for any other value. Under the assumption of independence at all lags and leads between all latent variables, u_t will also be orthogonal to $w_{t-1}^\dagger, \ldots, w_{t-J}^\dagger$, which can then be used as additional instruments to secure identification of the ψ_j. The problem is that if the error process is misspecified, then the IV estimates of the structural parameters jointly estimated with the ψ_j will be inconsistent. On the other hand, if the error process is well specified, then the joint estimates will be more asymptotically efficient than those based on $E(z_t^\dagger \eta_t) = 0$ alone.

Sargan was aware that assuming a low-order autoregressive process for η_t was conceptually problematic, given that he was regarding the errors as a combination of structural shocks and measurement errors. He argued that "it would appear logical to assume that each measurement error, and the random component, are being determined by a different mechanism, for example, by an autoregressive equation. However the problem of estimation which this assumption involves is very complicated, and indeed the only treatment which appears promising is that which assumes that the whole residual is determined by an autoregressive equation" (1959, p. 101).

In later work, Sargan and coauthors considered structural dynamic models with serial correlation in various settings. Sargan (1961) studied the properties of ML estimates of dynamic simultaneous systems with vector autoregressive errors; Espasa and Sargan (1977) considered the spectral estimation of simultaneous systems with stationary errors, and Bhargava and Sargan (1983) analyzed models for short panels with unrestricted autocovariance matrix. In the latter two articles, completely exogenous instrumental variables were required to distinguish structural dynamics from the generic patterns of serial correlation used.

5. NONLINEAR INSTRUMENTAL VARIABLES ESTIMATION

A remarkable contribution of Sargan's 1959 article was to develop in a general and rigorous way the IV estimation of nonlinear-in-parameters models. Sargan realized that estimation of the transformed model (25) was a special case of a nonlinear-in-parameters model. Thus he set about to develop the theory for estimation of a general model in which parameter restrictions were expressed as functions of a smaller set of free parameters. In this way, the contribution of Sargan (1959) goes well beyond the resolution of the problem of estimating models with serially correlated errors that originally motivated the analysis.

The setting was

$$\alpha(\theta_0)' w_t = u_t \tag{29}$$

and

$$E(z_t u_t) = 0, \tag{30}$$

where $\alpha(\cdot)$ is a vector of $q+1$ functions of a $k \times 1$ parameter vector θ and z_t is an $r \times 1$ vector of instrumental variables.

Begin by considering the simple case where $r = k$. Sargan first considered the nature of the solutions of the set of r equations

$$\overline{M}_{zw} \alpha(\theta_0) \equiv \left(p \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} z_t w_t' \right) \alpha(\theta_0) = 0. \tag{31}$$

He did so because the behavior of the solutions of the sample moment equations

$$\left(\frac{1}{T} \sum_{t=1}^{T} z_t w_t' \right) \alpha(\theta) = 0 \tag{32}$$

depends on the nature of the solutions to the limiting equations (31).

He distinguished three cases: (a) (31) has a unique solution, (b) (31) has multiple solutions, and (c) (31) has a continuous infinity of solutions so that they just determine a curve in θ space (the "a priori unidentified case"). Moreover, in cases (a) and (b) he also distinguished between singular and nonsingular solutions. The value $\theta = \theta^*$ is a singular solution if $\overline{M}_{zw}\alpha(\theta^*) = 0$ and the $r \times k$ Jacobian matrix

$$\overline{M}_{zw}\frac{\partial\alpha(\theta^*)}{\partial\theta'}$$

does not have full rank.

Next, Sargan considered IV estimators that solve

$$\min_{\theta}\alpha(\theta)'M_{wz}M_{zz}^{-1}M_{zw}\alpha(\theta) \tag{33}$$

and argued that the probability that there is a solution of (32) near each solution of (33) tends to unity as $T \to \infty$, provided that the solution is nonsingular. Then he used a general consistency theorem for extremum estimators to establish consistency of the nonlinear IV estimator. (This type of theorem, given without proof, predated a large literature in nonlinear econometrics; its ideas were further elaborated in Sargan 1975c.)

The final step was to establish asymptotic normality. The result was

$$\sqrt{T}(\hat{\theta}-\theta_0) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\left[\left(\frac{\partial\alpha(\theta_0)}{\partial\theta'}\right)' \times \overline{M}_{wz}\overline{M}_{zz}^{-1}\overline{M}_{zw}\left(\frac{\partial\alpha(\theta_0)}{\partial\theta'}\right)\right]^{-1}\right), \tag{34}$$

where $\hat{\theta}$ is the minimum of $\alpha(\theta)'M_{wz}M_{zz}^{-1}M_{zw}\alpha(\theta)$ within or on the boundary of a small region in θ space surrounding a nonsingular solution θ_0 of the limiting equation (31).

Sargan then moved to the overidentified case where the number of instrumental variables is greater than the number of parameters $r > k$. As in his previous article, he considered linear combinations of the z_t,

$$z_t^* = \Phi z_t$$

where Φ is a $k \times r$ transformation matrix, showing that the optimal choice for Φ that minimizes the asymptotic variance is given by

$$\Phi = \left(\frac{\partial\alpha(\theta_0)}{\partial\theta'}\right)'\overline{M}_{wz}\overline{M}_{zz}^{-1}, \tag{35}$$

so that the corresponding (unfeasible) estimates solve

$$\left(\frac{\partial\alpha(\theta_0)}{\partial\theta'}\right)'\overline{M}_{wz}\overline{M}_{zz}^{-1}M_{zw}\alpha(\theta) = 0,$$

and the asymptotic variance matrix is as in (34). Sargan (1959) argued that this suggests that one consider the equations

$$\left(\frac{\partial\alpha(\theta)}{\partial\theta'}\right)'M_{wz}M_{zz}^{-1}M_{zw}\alpha(\theta) \equiv \frac{\partial}{\partial\theta}\left[\alpha(\theta)'M_{wz}M_{zz}^{-1}M_{zw}\alpha(\theta)\right] = 0. \tag{36}$$

Thus he considered estimators that minimize the nonlinear IV criterion $\alpha(\theta)'M_{wz}M_{zz}^{-1}M_{zw}\alpha(\theta)$, establishing consistency and asymptotic normality with variance matrix equal to that in (34).

Next, Sargan briefly considered the singular case, which was further developed in his 1980 World Congress Presidential Address, (Sargan 1983), where

$$rank\left(\overline{M}_{zw}\frac{\partial\alpha(\theta^*)}{\partial\theta'}\right) \leq k-1. \tag{37}$$

He argued that in the case where the rank is $k-1$, "there is asymptotically a probability of 1/2 that there is a single minimum with error of order $T^{-1/2}$ and a probability of 1/2 of two minima with errors of order $T^{-1/4}$. However, this case has a mainly academic interest since it is a priori unlikely that $\overline{M}_{zw}$ will take just those values which makes the solution singular, and if the solution is only almost singular the errors will be large but as in the previous sections" (1959, p. 97).

The Minimax Approach. Sargan (1959) generalized the minimax approach of his 1958 article to the nonlinear case. The idea is the same as before—namely, to minimize the largest squared correlation between the errors and a linear combination of the instruments,

$$\min_{\theta}\left[\max_{\gamma}\frac{(\alpha(\theta)'M_{wz}\gamma)^2}{(\alpha(\theta)'M_{ww}\alpha(\theta))(\gamma'M_{zz}\gamma)}\right] = \min_{\theta}\lambda(\theta), \tag{38}$$

where

$$\lambda(\theta) = \frac{\alpha(\theta)'M_{wz}M_{zz}^{-1}M_{zw}\alpha(\theta)}{\alpha(\theta)'M_{ww}\alpha(\theta)}. \tag{39}$$

Sargan showed that this estimator was asymptotically equivalent to the previous one, and argued that its advantage was that "in the application to the study of autoregressive residuals it gives a set of estimates symmetric between the different variables in the relationship," and also that the minimum $\hat{\lambda}_1$ provided the basis for "a significance test for the existence of a relationship of the proposed type" (1959, p. 99). Next, he derived the asymptotic distribution of $T\hat{\lambda}_1$ (the test statistic of overidentifying restrictions). The general theory concluded with discussions of testing for underidentification and of confidence regions.

6. CONCLUDING REMARKS

Sargan's 1958 and 1959 articles provided lasting foundations for the theory of IV estimation of linear and nonlinear-in-parameters models. His results and his way of thinking about econometric estimation are very much in the agendas of present-day econometricians and empirical economists.

Hansen (1982) extended Sargan's framework by considering fully nonlinear models, but also by abstracting from equations and expressing an estimation problem as a list of moment conditions. Moreover, Sargan's departure from conventional asymptotic efficiency was taken a significant step further by Hansen. Hansen considered orthogonality conditions defined in terms of errors before filtering to remove serial correlation and generalized the properties of errors, allowing for flexible forms of serial dependence and conditional heteroscedasticity.

Sargan's work on instrumental variables had also dispensed with likelihood functions and blended errors in variables with simultaneity problems by relying on a given set of moment conditions. However, something was lost in the process of moving toward fully nonlinear models, because in Sargan's original motivation measurement errors played an important role alongside structural shocks.

ACKNOWLEDGMENTS

This article is dedicated to the memory of Denis Sargan. The author thanks Olympia Bover, Giorgio Calzolari, Eric Ghysels, Alastair Hall, David Hendry, Peter Phillips, and Enrique Sentana for helpful comments and discussions.

[Received July 2001. Revised June 2002.]

REFERENCES

Amemiya, T. (1974), "The Nonlinear Two-Stage Least-Squares Estimator," *Journal of Econometrics*, 2, 105–110.

Anderson, T. W., and Rubin, H. (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

——— (1950), "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 570–582.

Arellano, M., Hansen, L. P., and Sentana, E. (1999), "Underidentification?" unpublished manuscript, CEMFI, Madrid.

Arellano, M., and Sargan, J. D. (1990), "Imhof Approximations to Econometric Estimators," *Review of Economic Studies*, 57, 627–646.

Barnard, G. A. (1963), Discussion of "The Spectral Analysis of Point Processes" by M. S. Bartlett, *Journal of the Royal Statistical Society*, Ser. B, 25, 294.

Basmann, R. L. (1957), "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation," *Econometrica*, 25, 77–83.

Bekker, P. A. (1994), "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 657–681.

Bhargava, A., and Sargan, J. D. (1983), "Estimating Dynamic Random Effects Models From Panel Data Covering Short Time Periods," *Econometrica*, 51, 1635–1659.

Brundy, J. M., and Jorgenson, D. W. (1971), "Efficient Estimation of Simultaneous Equations by Instrumental Variables," *Review of Economics and Statistics*, 53, 207–224.

Chamberlain, G. (1982), "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5–46.

Chernoff, H., and Rubin, H. (1953), "Asymptotic Properties of Limited Information Estimates Under Generalized Conditions," *Cowles Commission Monograph 14*, 200–212.

Cumby, R. E., Huizinga, J., and Obstfeld, M. (1983), "Two-Step Two-Stage Least Squares Estimation in Models With Rational Expectations," *Journal of Econometrics*, 21, 333–355. Errata, 23, 407.

Desai, M. J., Hendry, D. F., and Mizon, G. E. (1997), "John Denis Sargan," *The Economic Journal*, 107, 1121–1125.

Durbin, J. (1954), "Errors in Variables," *Review of the International Statistical Institute*, 22, 23–32.

——— (1963), "Maximum-Likelihood Estimation of the Parameters of a System of Simultaneous Regression Equations," Paper presented at the Copenhagen meeting of the Econometric Society. Published as Durbin, J. (1988), *Econometric Theory*, 4, 159–170.

Espasa, A., and Sargan, J. D. (1977), "The Spectral Estimation of Simultaneous Equation Systems With Lagged Endogenous Variables," *International Economic Review*, 18, 583–605.

Geary, R. C. (1948), "Studies in Relations between Economic Time Series," *Journal of the Royal Statistical Society*, Ser. B, 10, 140–158.

——— (1949), "Determination of Linear Relations Between Systematic Parts of Variables With Errors of Observation the Variances of Which Are Unknown," *Econometrica*, 17, 30–58.

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

Griliches, Z. (1967), "Distributed Lags: A Survey," *Econometrica*, 35, 16–49.

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Hansen, L. P., Heaton, J., and Yaron, A. (1996), "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, 14, 262–280.

Hansen, L. P., and Hodrick, R. J. (1980), "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy*, 88, 829–853.

Hausman, J. A. (1975), "An Instrumental Variable Approach to Full Information Estimators for Linear and Certain Nonlinear Econometric Models," *Econometrica*, 43, 727–738.

Heckman, J. J. (2000), "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective," *Quarterly Journal of Economics*, 115, 45–97.

Hendry, D. F. (1976), "The Structure of Simultaneous Equations Estimators," *Journal of Econometrics*, 4, 51–88.

Jorgenson, D. W., and Laffont, J. (1974), "Efficient Estimation of Nonlinear Simultaneous Equations With Additive Disturbances," *Annals of Economic and Social Measurement*, 3, 615–640.

Koopmans, T. C. and Hood, W. C. (1953), "The Estimation of Simultaneous Linear Economic Relationships," *Cowles Commission Monograph 14*, 112–199.

Kunitomo, N. (1980), "Asymptotic Expansions of the Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations," *Journal of the American Statistical Society*, 75, 693–700.

Maasoumi, E. (1988), "Denis Sargan and His Seminal Contributions to Economic and Econometric Theory," in *Contributions to Econometrics: John Denis Sargan*, Vol. 1, ed. E. Maasoumi, Cambridge, UK: Cambridge University Press.

Morimune, K. (1983), "Approximate Distributions of k-Class Estimators When the Degree of Overidentifiability Is Large Compared With the Sample Size," *Econometrica*, 51, 821–841.

Phillips, P. C. B. (1977a), "A General Theorem in the Theory of Asymptotic Expansions as Approximations to the Finite Sample Distributions of Econometric Estimators," *Econometrica*, 45, 1517–1534.

——— (1977b), "Approximations to Some Finite Sample Distributions Associated With a First Order Stochastic Difference Equation," *Econometrica*, 45, 463–485.

——— (1985), "ET Interviews: Professor J. D. Sargan," *Econometric Theory*, 1, 119–139.

Reiersøl, O. (1945), *Confluence Analysis by Means of Instrumental Sets of Variables*, Uppsala, Sweden: Almquist and Wiksell.

Sargan, J. D. (1958), "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393–415.

——— (1959), "The Estimation of Relationships With Autocorrelated Residuals by the Use of Instrumental Variables," *Journal of the Royal Statistical Society*, Ser. B, 21, 91–105.

——— (1961), "The Maximum Likelihood Estimation of Economic Relationships With Autoregressive Residuals," *Econometrica*, 29, 414–426.

——— (1964), "Wages and Prices in the United Kingdom: A Study in Econometric Methodology," in *Econometric Analysis for National Economic Planning*, eds. P. E. Hart, G. Mills, and J. K. Whitaker, London: Butterworths.

——— (1975a), "Gram–Charlier Approximations Applied to t Ratios of k-Class Estimators," *Econometrica*, 43, 327–346.

——— (1975b), "Asymptotic Theory and Large Models," *International Economic Review*, 16, 75–91.

——— (1975c), "The Identification and Estimation of Sets of Simultaneous Stochastic Equations," Discussion paper A1, LSE. Published in *Contributions to Econometrics: John Denis Sargan*, Vol. 1, ed. E. Maasoumi, Cambridge, UK: Cambridge University Press, 1988.

——— (1976), "Econometric Estimators and the Edgeworth Approximation," *Econometrica*, 44, 421–448. Erratum, 45, 272.

——— (1980), "Some Tests of Dynamic Specification for a Single Equation," *Econometrica*, 48, 879–897.

——— (1981), "Some Experiments With Post Estimation Simulation," LSE Econometrics Research Programme, unpublished manuscript.

——— (1983), "Identification and Lack of Identification," *Econometrica*, 51, 1605–1633.

——— (1993), "Some Alternatives to the Edgeworth Approximation for Econometric Statistics," in *Models, Methods, and Applications of Econometrics, Essays in Honor of A. R. Bergstrom*, ed. P. C. B. Phillips, Blackwell.

Sargan, J. D., and Mikhail, W. M. (1971), "A General Approximation to the Distribution of Instrumental Variables Estimates," *Econometrica*, 39, 131–169.

Stone, R. (1947), "On the Interdependence of Blocks of Transactions" (with discussion), *Journal of the Royal Statistical Society*, 9 (suppl), 1–45.

Theil, H. (1961), *Economic Forecasts and Policy*, (2nd ed.), Amsterdam: North-Holland.

Tse, Y. K. (1981), "Edgeworth Approximation to the Finite Sample Distribution of Econometric Estimators and Test Statistics," unpublished Ph.D. thesis, University of London.

Wald, A. (1940), "The Fitting of Straight Lines if Both Variables Are Subject to Error," *Annals of Mathematical Statistics*, 11, 284–300.

White, H. (1982), "Instrumental Variables Regression With Independent Observations," *Econometrica*, 50, 483–499.

Working, E. J. (1927), "What Do Statistical 'Demand Curves' Show?" *Quarterly Journal of Economics*, 41, 212–235.

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

Wright, S. (1934), "The Method of Path Coefficients," *Annals of Mathematical Statistics*, 5, 161–215.