The Economic and Social Review, Vol. 24, No. 3, April 1993, pp. 247-273

Reiersøl, Geary and the Idea of Instrumental Variables

JOHN ALDRICH*
The University, Southampton

Abstract: In the 1940s the method of instrumental variables was introduced for use in the Errors in Variables Model. According to Reiersøl, the "idea of using instrumental variables" was introduced independently by himself in 1941 and by Geary in 1943. For the modern reader the "idea" signifies the use of a particular estimator which, under certain assumptions, is consistent for the structural parameters. Geary presents something resembling this but Reiersøl does not. So what was the "idea of using instrumental variables"? The paper tries to answer this question. It sets the work of Reiersøl and Geary in the larger context of work on the EVM in the 1930s and 1940s.

I INTRODUCTION

The method of instrumental variables has an important place in modern econometrics. The method's rise to prominence can be divided into three phases. In the 1940s the method was introduced for use with errors in variables model. In the 1950s it was related to methods devised for the errors in equations model. More recently it has been transformed into an organising principle underlying many apparently distinct models.[1]

This paper presents a detailed study of the work of the 1940s. Two authors, Reiersøl and Geary, are associated with this phase. The aim is to understand their contribution. It is easy to begin — at the end of the story. For the state of the art at the end of the 1940s is depicted in the treatises of the 1960s, in

*I would like to thank Janne Rayner and Alan Ingham for discussion and John Spencer for comments on an earlier draft.

1. For a survey, see Bowden and Turkington (1984). Works expounding the method usually have some information on the history of instrumental variables. However, the information is not always reliable as it often relies on the principals' subsequent reconstructions of their discoveries.

Kendall and Stuart (1967) or Malinvaud (1966).[2] The serious gap is between Geary (1949) or Reiersøl (1950) and the authors' earlier writings.

According to Reiersøl (1950, p. 378), the "idea of using instrumental variables" was introduced independently by himself in 1941 and by Geary in 1943. Neither introduction was very imposing. It was not only the absence of a name for the new idea. As Morgan (1990, p. 226) remarks, Reiersøl's contribution was "hidden in amongst a number of extensions to Frisch's confluence analysis."[3] While Geary subordinated the method to another that he had introduced earlier in 1942.

There is a genuine task in determining what this undramatic multiple discovery consisted of. The modern reader is likely to equate the "idea of using instrumental variables" in the errors in variables model with the use of a particular estimator which, under a particular set of assumptions, is consistent for the structural parameters.[4] Something approximating this specification can be found in Geary's 1943 paper but there is not much of it in Reiersøl's 1941 paper. The idea of using instrumental variables to study the relationship between investigational variables in realised in other ways.

Understanding the work of Reiersøl and Geary on instrumental variables requires setting it in the larger context of work on the errors in variables model in the 1940s. In the case of Reiersøl, the problem situation is unfamiliar for Frisch's confluence analysis and Reiersøl's extensions to it have disappeared from mainstream econometrics.[5] In the case of Geary, the problem situation is more familiar. He was interested in estimating the errors in variables model, though the methods he devised — and which long overshadowed instrumental variables — are no longer familiar.

The paper has three main sections. The first treats Frisch's *Statistical Confluence Analysis*, to discover what Reiersøl found worth imitating and extending. In a more general sense, Geary's work also draws on Frisch. The contributions of Reiersøl and Geary can be covered in separate sections for, though Reiersøl was aware of Geary's work by 1947 and Geary of Reiersøl's by 1948, their influence on one another seems to have been slight.[6] The

2. For a recent survey see Fuller (1987).

3. Morgan (1990, Chapter 7) provides essential background. The present paper picks up the story where she leaves off. See also Hendry and Morgan (1989).

4. Thus the instrumental variables estimator of α_{12} in (2.2) on p. 252 below is given by $-\Sigma x_3 x_2 / \Sigma_3 x_1$, where x_i is ξ_i measured with error. When the measurement errors are independent the estimator is consistent.

5. However, Kalman (1982) and Klepper and Leamer (1984) carry on certain parts of the confluence analysis programme.

6. The only evidence from their publications of any interaction beyond the citing of each other's papers is the acknowledgement in Geary (1949, p. 57): "The author wishes particularly to thank Mr. Olav Reiersøl for constructive criticism which led to an extension in the paper as originally drafted."

section on Reiersøl marks his passage from confluence analysis to identification theory. The section on Geary marks the movement of instrumental variables from the periphery to the centre of his attention. The 1949 paper was his only one to be wholly devoted to the method. That paper, rather than any earlier contribution, marks the real arrival of the method. Our final section has some reflections on the elusive "idea of using instrumental variables."

II FRISCH'S CONFLUENCE ANALYSIS[7]

We begin with Frisch's *Confluence Analysis* (1934), or rather with the work as it appeared to later theorists.[8] Frisch held that a combination of measurement error and multicollinearity makes economic data difficult to analyse; multicollinearity is the condition of several relations holding between the variables. Later authors may have put different emphases on these two factors but the big point of his book, that the methods of least squares regression could not be applied safely to such data, was well taken.

Frisch posed a number of problems and provided influential discussions of them but he did not solve them. His general influence on work on errors in variables was strong; his influence on Reiersøl's work of 1941 and 1945 was overwhelming. There was even an indirect influence on Reiersøl's later work on identification: Frisch's analysis of multicollinearity influenced work on identification in the simultaneous equations model which in turn influenced Reiersøl. Frisch's influence went beyond his students; in Section IV we see how Tintner and Geary tackled the problem of multicollinearity in 1945-1948.

Wald (1940, p. 285) remarks that Frisch's theory was not "based on probability concepts." Reiersøl (1950, p. 376) argued that Frisch (and Gini before him) "do not consider a hypothetical universe but a sample; however, their considerations can be immediately translated into statements about a universe." The authors considered here, Koopmans, Wald, Reiersøl, Geary and Tintner, all worked with a probabilised version of Frisch's model.

Our formulation of "Frisch's model" is actually Reiersøl's (1941, p. 2) reformulation in notation based on Hendry and Morgan (1989).[9] The kx1 observable vector, x_t, is made up of a stationary systematic part, ξ_t, and a

7. Morgan (1990, pp. 207-211) and Hendry and Morgan (1989) treat the technicalities of Frisch's confluence analysis.

8. Researchers were aware that the search for an improvement on least squares extended back into the nineteenth century — see e.g. Wald (1940). But the earlier work had no independent influence on the work we consider here.

9. A uniform notation has been adopted to make it easy to compare different contributions. The drawback is that the reader must translate from Reiersøl (say) as presented here to Reiersøl in his own symbols. This policy on notation is pursued ruthlessly down to changing notation in passages quoted.

serially independent measurement error, υ_t

$$x_t = \xi_t + \upsilon_t.$$

ξ_t and υ_t are random vectors with zero mean and

$$E\xi_t\xi_t' = (\varphi_{ij}) = \phi$$
$$E\upsilon_t\upsilon_t' = (\sigma_{ij}) = \Sigma \text{ where } \Sigma \text{ is diagonal}$$
$$E\xi_t\upsilon_t' = 0.$$

Reiersøl refers to these as Frisch's *Conditions 1.* It follows that:

$$Ex_tx_t' = \phi + \Sigma = (\mu_{ij}) = M, \text{ say.}$$

There are $r \geq 1$ linear relationships between the systematic parts,

$$A\xi_t = 0.$$

These are the structural relationships. When $r > 1$, there is multicollinearity.

Frisch avoided probability in two respects: first, by not specifying the model in probabilistic terms and (consequently) by not considering the properties in repeated sampling of the inference procedures he proposed. Koopmans's (1937) maximum likelihood analysis of the errors in variables model was based on probability concepts in both senses, as was the work of Wald and Geary. Of course, the necessity of probability in both respects was the point emphasised in Haavelmo's *Probability Approach in Econometrics* (1944).[10]

Reiersøl's work is probabilistic in the first sense only: "We shall study only the hypothetical universe and not give any sampling-theoretical analysis." However, he wanted to use his conclusions about the relationships between population moments and structural coefficients to analyse data. In his work there is an implicit appeal to a method of moments principle.[11]

Reiersøl *never* gave any sample-theoretical analysis though he saw in such analysis the essential complement of his own work. It was a matter of first things first, or perhaps possible things first. In 1934 Frisch (p. 7) wrote:

10. See Morgan's Chapter 8 on "Haavelmo's Probability Model." With the fervour of the convert, Haavelmo was hard on Frisch's confluence analysis:

> The purely geometric properties of a set of points in the sample space are insufficient as a basis for statistical inference. In fact, a sample of observations is just a set of cold, uninteresting numbers unless we have a theory concerning the stochastic mechanism that has produced them. (1950, p. 265.)

11. The principle formulated by Goldberger (1968) seems to underlie much of the work discussed here: "the *analogy principle* of estimation ... proposes that population parameters be estimated by sample statistics which have the same property in the sample as the parameters do in the population." We would extend the scope of the principle to include work on limits.

> ... if the sampling aspect should be studied from a sufficiently general set of assumptions, I found that it would lead to such complicated mathematics that I doubted whether anything useful would come out of it.

This assessment of the possible was much less sound after 1940. The change is well reflected in the work of Geary and Tintner, discussed in Section IV.

Frisch and Reiersøl's work on the simultaneous errors in variables model was technically cruder than the work of Haavelmo and Koopmans on the errors in equations model. But beyond this technical difference was a difference in aim; the aim was *not* to consistently estimate the A matrix, the form of which is known a priori.[12] Frisch and Reiersøl were particularly concerned with specification uncertainty, e.g. Reiersøl (1941, p. 12) points out, "In the statistical analysis of economic time series we often do not know in advance the structural lags." Their work was closer to exploratory multivariate analysis than to the economic model-based analysis of Haavelmo and Koopmans.[13] According to Reiersøl (1945, p. 1), the aims of confluence analysis are "partly to determine which variables are to be included in the analysis and partly to find limits to the structural equations ... in the sets of variables which are found to be admissible."

Frisch (1934) had insight into the difficulties of data analysis in a noisy multicollinear world. One difficulty was multicollinearity itself. Suppose, the following pair of relationships hold between four variables:

$$\begin{aligned} \alpha_{11}\xi_1 + \alpha_{12}\xi_2 + \alpha_{14}\xi_4 &= 0 \\ \alpha_{21}\xi_1 + \alpha_{22}\xi_2 + \alpha_{23}\xi_3 &= 0 \end{aligned} \tag{2.1}$$

Any nonsingular transformation yields a pair of equations that hold just as well as the original pair. In particular, there are infinitely many equations between all four variables — just multiply the first equation by p and the second by q and add. As the variables are measured with error, the statistician investigating the equation with all four variables may not realise that it is not well-defined. This made multicollinearity a *problem*.[14]

12. Geraci (1977) analyses a model that combines errors in equations and errors in variables but in the manner of Haavelmo and Koopmans. On the other hand, Fuller (1987, Chapter 4) treats a pure errors in variables model written in a form analogous to the classical multivariate regression model.

13. Klepper and Leamer (1984) were cited above for continuing the work of confluence analysis. Perhaps Leamer with his emphasis on specification uncertainty provides the closest parallel in modern econometrics to Frisch and Reiersøl.

14. Aldrich (1992) shows how Frisch's multicollinearity analysis was transformed into Haavelmo's analysis of "arbitrary parameters", the first identification theory for the simultaneous equations model.

Even without multicollinearity there is a problem with variables measured subject to error. Least squares does not give satisfactory estimates. When there is a single relationship between two variables:

$$\xi_1 + \alpha_{12}\xi_2 = 0 \tag{2.2}$$

Frisch (and Gini) found the following relationships between the structural coefficient and the elements of M, the variance matrix of the observables:

$$\frac{\mu_{12}}{\mu_{22}} \leq \frac{\varphi_{12}}{\varphi_{22}} = -\alpha_{12} = \frac{\varphi_{11}}{\varphi_{12}} \leq \frac{\mu_{11}}{\mu_{12}} \tag{2.3}$$

The inequalities are strict unless at least one of ξ_1 and ξ_2 is measured without error. The tightness of the bounds depends on the error variances.

There were two distinct reactions to this theorem — represented below by Reiersøl and Geary respectively. The inequalities imply that there is more information in the "observable" variance matrix, M, than is extracted by a single regression of x_1 on x_2, i.e. μ_{12}/μ_{22}. But they also imply that the structural coefficients cannot be determined from the second moments of the data: they are not "identifiable" as Reiersøl (1947) put it much later.[15]

Confluence Analysis by means of *complete regression systems* is about extracting information from M using all possible regressions in the variables under study. Bunch map analysis is a graphical representation of the results. Examining the bunch maps, particularly their tightness, is a way of discovering whether a particular variable is superfluous to a relation.[16] As Reiersøl (1941, p. 4) says, "Frisch uses the theorem [on two variables] as a heuristic argument for the main idea of the bunch analysis of any number of variates." For Reiersøl, Frisch's theorem was the paradigm result of confluence analysis, or at least of that part concerned with finding "limits to the structural equations." He reproduced it in new settings and generalised it — the result was confluence analysis by means of instrumental variables.

Others wanted conditions that would "determine" the parameter values; in modern terms — and the modern age begins in the 1940s — they wanted consistent estimates, not estimates converging to some point in an interval bracketing the unknown parameter value. One line, followed in Koopmans's (1937) maximum likelihood analysis based on normal errors was to assume the variance matrix of the errors, Σ, known up to a factor of proportionality.

15. In the normal case when there are only second moments the structural parameters are not identified. Koopmans (1937, p. 70) discussed this case.

16. Malinvaud (1966, pp. 32-36) has an exposition of bunch maps analysis. Hendry and Morgan (1989) make a detailed study of the method and give an evaluation of it.

The other line was to "by-pass" Σ, as Morgan (1990, p. 226) puts it. Geary (1942) escaped from dependence on second moments and exploited relationships between the structural parameters and higher order cumulants. Geary (1948) found another way of by-passing Σ. The use of instrumental variables offers yet another way. Indeed, the title of Geary's instrumental variables paper of 1949 seemed to make this the point of the method — "determination of linear relations between systematic parts of variables with errors of observation *the variance of which are unknown.*"

There was nothing contradictory in doing research both into limits and into determination, but only Frisch and Reiersøl did so on any extensive scale.[17]

III CONFLUENCE ANALYSIS WITH INSTRUMENTAL VARIABLES

Reiersøl's publications between 1941 and 1950 fall into two groups: there are two papers on confluence analysis, the "preliminary publication" of 1941 and the "much larger work" of 1945, and three papers on identification, a general paper with Koopmans (1950), one on the errors in variables model (1950) and one on the factor analysis model (1950a). Between the two groups is an unpublished study of 1947 relating confluence analysis to work on the errors in equations model. Here Reiersøl uses Koopmans's concept of identification to interpret confluence analysis as inference for an under-identified model.

The preliminary publication is difficult reading. It is mathematically self-contained but the results are presented with little comment on their significance. Some conclusions are collected into theorems, a treatment that may reflect their importance. There is a small discussion of how some of the results may be used to guide practical data analysis and again some measure of importance may be read into this discussion.

Reiersøl gives results of both the types considered in the last section; there are conditions under which the structural coefficients obey certain limits and conditions under which the coefficients are "exactly determined". The emphasis is on results of the first type. The paper presents fifteen theorems but only one, on "Pfaffian regression", treats the determination of the structural coefficients. The remainder are concerned with developing limits on the structural coefficients either for a single relationship or for a set of relationships — the multicollinearity case.

The theorems relate the structural parameters to the population moments of the observed variables — usually the second moments. The results are clearly intended to support a method of moments like inference procedure.

17. Koopmans (1937, pp. 99-101) contributed to the "limits" stream, by generalising the Gini-Frisch result to the case of several variables.

Later the theorems were taken in other ways: Geary (1949) takes them as the basis for results about the consistency of such estimators; Reiersøl (1950) takes them as results about identification.

In *Confluence Analysis* there was no discussion of dynamic models but later Frisch took to formulating macrodynamic models as systems of linear difference equations, as in the famous memorandum of 1938, "Statistical versus theoretical relations in economic macrodynamics".[18] Reiersøl presents results for both dynamic and non-dynamic cases.

Reiersøl (p. 4) states that "the idea of using lag moments for determination of structural coefficients is due to Frisch." The model for Reiersøl's work on "determination" was Frisch's "method of autouncorrelated residuals" which was first publicised in Reiersøl's article. By using in the analysis variables that are not in the equation, the method could be said to contain the seminal idea for instrumental variable estimation.

Frisch's method applied to a single relationship. Using the notation introduced on p. 4 above but dropping the equation subscript, we define the scalar "residual"

$$P_t = \Sigma\alpha_i x_{it} = \alpha' x_t.$$

It follows that on leading P_t and taking expectations

$$E(P_t P_{t+\tau}) = \alpha' M_\tau \alpha$$

where M_τ is the lag moment matrix of x, i.e.

$$M_\tau = E(x_t x'_{t+\tau}).$$

If P_t is "autouncorrelated" we have a system of determining equations

$$E(P_t P_{t+\tau}) = \alpha' M_\tau \alpha = 0.$$

Reiersøl (p. 4) observes, "If P_t is exactly autouncorrelated this method gives a unique determination of the structural coefficients when we use a sufficient number of the equations."

To see what is involved in the method, return to the structural equation

$$\alpha'\xi_t = 0$$

with its implication that P_t can be written in terms of the errors

$$P_t = \alpha' \upsilon_t.$$

Reiersøl points out that the condition that P_t is autouncorrelated holds in particular when the errors are autouncorrelated and there is no correlation

18. This research programme is discussed by Aldrich (1989, pp. 21-25).

between any error and any other error which is lagged relatively to the former, i.e. when

$$\sigma_{ij.\tau} = E(\upsilon_{it}\upsilon_{j,\,t+\tau}) = 0 \text{ for all i and j and } \tau \neq 0.$$

What *we* would call the instrumental variable method with the lag τ values as instruments is almost immediate. For under these special conditions, the simpler linear system of equations, the lag regression equations,

$$\alpha' M_\tau = 0$$

also holds. However, though Reiersøl considers these equations, he does not assume these special conditions; in particular he does not assume that the errors are autouncorrelated. As a result, his lag regressions generate not exact determination but limits on the structural parameters — variations on the least squares results rather than a new "determination" result.

For his own results Reiersøl (p. 4) specifies the following conditions:

> Condition 3a: There is no correlation between an error and another error lagged τ time units.
> Condition 3b: There is no correlation between the error of one variate and the systematic part of another variate lagged τ time units.

The lag moment matrix of the systematic variables and the lag moment matrix of the total (observable) variables differ in their diagonal elements,

$$M_\tau = \phi_\tau + \Sigma_\tau \quad \text{where } \Sigma_\tau \text{ is diagonal.}$$

Take the equation

$$\alpha' \xi_t = 0 \tag{3.1}$$

multiply by the vector $\xi_{t+\tau}$ and take expectations to obtain

$$\alpha' \phi_\tau = 0.$$

Hence we have for the lag regression equations,

$$\alpha' M_\tau = \alpha'(\phi_\tau + \Sigma_\tau) = \alpha' \Sigma_\tau. \tag{3.2}$$

These equations can generate limits for the structural coefficients, i.e. inequalities like (2.3) — indeed (2.3) corresponds to the case $\tau = 0$.[19]

19. The condition $\sigma_{1i.\tau} \geq 0$ does not hold automatically as in the $\tau = 0$ case. It is one of the extra assumptions made in the limits analysis for lag regressions.

However, there is a method of "determination" available for this set-up, the "Pfaffian regression". It is a way of eliminating the unknown error lag moments, the elements of Σ_τ. Return to Equation (3.1) but this time multiply by the vector $\xi_{t-\tau}$ to obtain in parallel to (3.2):

$$\alpha' M_{-\tau} = \alpha' M'_\tau = \alpha' \Sigma_\tau.$$

Subtract this equation from (3.2) to obtain

$$\alpha'(M_\tau - M'_\tau) = 0.$$

Solving these equations gives the Pfaffian regression.[20]

The Pfaffian regression method is a method of determination yet Reiersøl spends a good deal of time discussing the limits associated with the lag regressions. His (p. 6) justification is that:

> If we are to apply this theorem [on Pfaffian regression] to empirical data the lag-moment matrix must be *significantly* skew. Even if [it] is significantly skew the determination of the structural coefficients by the Pfaffian regression may be uncertain. We shall therefore need also other methods of determining the structural vector or limits to it.

Morgan (1990, p. 226) — see also p. 248 above — writes that "two suggestions for by-passing Σ were hidden in amongst a number of extensions to Frisch's confluence analysis model". One suggestion we have seen already — Frisch's method of autouncorrelated residuals. The other suggestion is contained — or seems to be contained — in Section 11, the second of two sections dealing with multicollinearity. Here Reiersøl treats the limits to structural coefficients implied by the existence of multicollinearity.

So we turn to multicollinearity, which Reiersøl calls double collinearity or r-tuple collinearity, as appropriate. In this part of the paper lag relationships are not used. Multicollinearity implies a rank deficiency of $r>1$ on the variance matrix of the ξ, i.e. on

$$\phi = M - \Sigma.$$

Reiersøl investigates the properties of the determinant of this matrix considered as a function of the diagonal elements of Σ which are assumed to be unknown. In Section IV we consider Tintner's (1945) treatment of the issue.

Reiersøl discusses the nature of the limits when Conditions 1 hold and there is r-tuple collinearity, i.e. r independent linear equations between the

20. The name "Pfaffian" derives from Pfaffian meaning the determinant of a skew symmetric matrix. The matrix $(M_\tau - M'_\tau)$ is skew symmetric.

systematic parts of the variables. He uses (r–1) of the equations to eliminate $\xi_1, \ldots, \xi_{r-1}$ and considers an equation in the set $\xi_r, \ldots, \xi_k$ with coefficients $\alpha_r, \ldots, \alpha_k$.[21] Using Frisch's Conditions 1 we have

$$\sum_{r}^{k} \mu_{ij}\alpha_j = 0 \qquad i = 1, 2, \ldots, r-1 \tag{3.3}$$

After presenting the accompanying limit analysis, Reiersøl points out that the assumption of r-tuple collinearity and Conditions 1 imply that:

$$E(\upsilon_{it}\upsilon_{jt}) = 0, \text{ for } i = r, r+1, \ldots, k; j = 1,2,\ldots, r-1; \text{ all } t. \tag{3.4}$$

He then observes that "Let us now drop the assumption of r-tuple collinearity in the total set. Instead of that we shall assume simple collinearity in the set (ξ_p, ... ξr) and the conditions (3.4). In this case the equations (3.3) will still apply, but not [the limited relations]."

This is the discussion to which Reiersøl referred in 1950 (see above) — *all* of it. There may be instrumental variables, viz. $x_1, \ldots, x_{r-1}$, but there is not an instrumental variable method of estimation. In 1950 Reiersøl (p. 378) states that "it is evident that the vector α is identifiable" when the rank of the moment matrix is k–r–1. In 1941 there was no statement of the "evident". The insight was not developed into an estimation method. There is no result to celebrate as a theorem. The result first appears in 1947.

Hendry and Morgan (p. 45) comment on this 1941 argument, "[the] work is particularly interesting for it was a development of confluence analysis in which the confluent variables were utilised (as instrumental variables) to provide a better estimation method, whereas previously these variables had been a bar to accurate estimation." The point could be turned round; Reiersøl did not expect to find a key to improved estimation in multicollinearity and so did not find it.

Reiersøl's 1945 dissertation, "Confluence Analysis by Means of Instrumental Sets of Variates", is an elaborate study in eight chapters including an introductory chapter on the nature of regression analysis with economic data reminiscent of the early chapters of Haavelmo's *Probability Approach*, a chapter on linear algebra and a chapter of applications to real data. However there are no fundamental new ideas or results.

The text teems with neologisms. In the preface Reiersøl writes, "Almost all new words and phrases have been discussed with Frisch and the final forms

21. This reduction is a way of obtaining an estimable equation and it is the basis of the treatment of multivariate models in Fuller (1987). The equation in the reduced set of variables will not usually be a structural equation in the strong sense of possessing a high degree of autonomy. The development of "autonomy" and "structure" are treated in Aldrich (1989).

D

of the terms are perhaps in most cases due to Frisch." The happiest of the innovations is the term "instrumental set" of variables as distinct from the "investigational set" which appear in the equation of interest. Morgan (1990, p. 228) reports Reiersøl recalling that Frisch had coined the term.

For *each* variable in the "investigational set" there is a corresponding variable in the "instrumental set". Reiersøl comments as follows on the eligibility of instruments:

> In all examples the variables of the instrumental set has some connection with the variables of the investigational set. This is a necessary condition for the instrumental set to be of any use in the following analysis. Most of the methods of this study supposes also that there is a natural correspondence between pairs of variables. (p. 32)

He does not take the discussion any further.

Although the instrumental variables are like the Section 11 "instruments" in not being restricted to lag or lead variables, the 1945 analysis is not otherwise a development of Section 11. Instead it generalises the treatment of the variables themselves in 1934 and the treatment of lag variables in 1941, down to the results on limits and Pfaffian regression. For underlying the analysis is a condition (p. 33) analogous to Condition 1 on p. 250 above and Condition 3 on p. 255, viz.:

> Any error of the investigational set, υ_i, is uncorrelated with any non-corresponding variable of the instrumental set, z_j where $i \neq j$.

Proceeding as before we take the equation

$$\alpha'\xi_t = 0 \tag{3.1}$$

but now multiply by the vector of instruments z_t and take expectations:

$$\alpha'\phi_z = 0.$$

Then, using the fundamental condition, we have analogous to (3.2):

$$\alpha' M_z = \alpha'(\phi_z + \Sigma_z) = \alpha'\Sigma_z$$

where $M_z = (Ez_i x_j)$ and Σ_z is the diagonal matrix $(Ez_i \upsilon_i)$.

The Pfaffian regression of 1941 uses both lead and lagged variables as instruments. Reiersøl generalises this case by specifying a class of cases with two linear relationships with *identical* coefficients, one involving the investigational variables and one the instrumental variables. When each variable is paired with its own lagged value, as in Pfaffian regression, this requirement is natural. Reiersøl (pp. 69-71) develops the necessary formalism but does not give any interpretation for this specification in other cases.

Reiersøl's "Confluence Analysis When the Model Contains Both Shocks and Errors" (1947) assimilates the work that was eventually published in Cowles monograph 10.[22] Reiersøl was particularly influenced by two papers, Koopmans, Rubin and Leipnik (1950) and Koopmans (1950). The first paper had introduced the concept of identifiability for the simultaneous equations model — a matter of the relationship between (reduced form) moments and structural parameters. Reiersøl (p. 4) adapts the notion for the "pure error model" and uses it to classify past work: as "partly been concerned with models leading to identifiability ... and partly been concerned with finding bounds for the parameter point in cases where there is no identifiability." Reiersøl recognises Geary's 1942 and 1943 articles among a number that deal with the identified case. Reiersøl (p. 5) also mentions how the use of instruments — in the manner of his own Section 11 and of Geary (1943) — can achieve identification. This marks Reiersøl's recognition of the instrumental variable *estimator* as it is known in the modern literature.

The set-up in the paper follows the block recursive scheme of Koopmans (1950) but there are errors in variables as well as errors in equations. Reiersøl (p. 4) tells us that "The main purpose of this paper will be to find bounds for the parameter point in the case of non-identifiability." Such cases are "important in practice." (p. 5) Given his purpose and the whole background in confluence analysis, it is not surprising that Reiersøl does not propose the instrumental variable estimator for the structural coefficients, developed in the later literature. In fact, Reiersøl aims for limits on the reduced form coefficients, stating that it would be possible to translate them into limits on the structural coefficients.

In Reiersøl (1950) he took the opportunity to review his own earlier work on errors in variables; he re-presented the works on limits as results on partial identification: "This analysis gives methods which can be used whether α is identifiable or not, but if σ is identifiable we can, of course, do something better than use the limits mentioned." (p. 377). The paper also encompassed results in Koopmans (1937). The paper dealt extensively with the rôle of non-normality in the identifiability of the model — an issue that made retrospective sense of Geary's failure to handle normality with his cumulant based method, the failure that led him to introduce instrumental variables.[23]

22. This paper is referred to by Qin (1989/1993). Reiersøl does not refer back to it.

23. Koopmans (1937, p. 70) had already discussed the problem associated with normality. The proofs in Reiersøl (1950) and Koopmans and Reiersøl (1950) on the rôle of non-normality in the identifiability of the structural coefficients used the same cumulant argument that Geary (1942) had used in developing his estimation method.

The errors in variables paper was restricted to the case of a single relationship. However, multiple relationships were of the essence of factor analysis which Reiersøl discussed in a second paper in 1950. The work on the factor analysis model should not be surprising. The model is a variation on the multicollinearity model of p. 4 above with normality and independence imposed and the structural equations $A\xi_t = 0$ replaced by

$$x_t = B\xi_t + \upsilon_t.$$

Already in his 1941 paper Reiersøl had referred to results in Ledermann (1937) which treated the identification of this model.

The sequence of identification papers is rounded off with the well-known paper with Koopmans. In a technical sense the collaboration with Koopmans on identification was natural. Relating the structural parameters to the population moments of the observed variables — the identification problem, as it was presented in the paper — was what Reiersøl had been doing all along. Koopmans's work on the errors in equations model complemented Reiersøl's work on the errors in variables and the factor analysis models to make a varied collection of examples, that illustrate the main message of the paper. Yet the authors do not refer to the possibility of analysing unidentified models. Confluence analysis was on its way to extinction.[24]

IV GEARY AND ESTIMATION OF THE ERRORS IN VARIABLES MODEL[25]

Moving from Reiersøl to Geary takes us to a different statistical culture. Reiersøl's *statistical* technique in his confluence analysis papers amounts to little more than taking expectations; Geary uses much more sophisticated tools.[26] But more importantly, Geary was immersed in the theory of statistical inference — his (1942a) and (1944) are contributions to the basic theory — while Reiersøl's papers reflect no knowledge of the subject.

Geary's work on errors in variables is oriented towards "exact determination", or estimation, rather than to "limits". Listing his papers on errors in variables shows not only a continuing interest in the topic but an interest in considering a variety of methods: two papers (1942, 1943) on his own cumulant-based method, a paper (1948) focusing on Tintner's maximum likelihood method and a paper (1949) on instrumental variable estimation.

24. Hendry and Morgan (1989) discuss another aspect of this extinction: the disappearance of bunch map analysis as a form of practical data analysis.

25. See the very useful reviews of Geary's work by Spencer (1976 and 1983).

26. By the time of the work we consider here Geary was a well established author. His first paper was published in 1925.

Subsequently he published another errors in variables paper (1953) and a general piece reflecting on the nature of regression (1963).

This list of papers with their "headlines" seems to indicate a late entrant to the field of instrumental variables. Such an impression would be wrong for the method is actually proposed in the 1943 paper and some large sample properties are contained in the 1948 paper. However these findings were contained in papers concerned with other, seemingly bigger, business. These contributions are rendered more invisible as Geary scarcely refers back to them.

There is a basic similarity between the analysis of Geary (1942) and Reiersøl (1941), though the techniques are very different; Geary uses sophisticated distribution theory instead of Reiersøl's geometric linear algebra. Reiersøl sets out relationships between the structural vector and the moments of the observables while Geary sets out relationships between the structural vector and the cumulants or semi-invariants of the observables. Using cumulants may not be particularly easy or obvious but, as Spencer (1976, p. 235) points out, Geary's expertness in the use of cumulants had been apparent from his earliest papers.[27]

Geary (1942) gets down to expounding his method straight away though he refers to Frisch (1934) and Wald (1940) in the course of the article. The failure of the classical regression analysis is taken for granted. Geary makes Frisch's assumption that the disturbances are independent of one another and of the systematic parts — see Conditions 1 on p. 250 above. There is a single exact linear relationship between the systematic parts of the k variables

$$\alpha'\xi_t = 0.$$

Geary obtains the following series of equations, "in general infinite in number":

$$\Sigma\alpha_i L(c_1, c_2, \ldots, c_i + 1, \ldots, c_k) = 0$$

where the L are the cumulants or semi-invariants of powers indicated in the variates x_i for which two at least of the numbers c_j are different from zero. This requirement implies that the semi-invariants are of total power 3 or over.

The coefficients α_i, subject to normalisation, are determinable from any (k–1) of these equations. Like Reiersøl, Geary envisaged using sample quantities in place of the population quantities in such equations. The method found a place in the literature and is outlined in Kendall and Stuart (p. 395)

27. As Spencer (1976, p. 235) says, the papers treating this method "are ingenious and technically highly instructive".

and Malinvaud (pp. 356-358), to which reference can be made for further details.

Yet from Geary's point of view, this first paper tackles only half the job. "Relations between random variables" was followed the next year by "relations between statistics". Besides the missing sampling theory, there was another item of unfinished business: what to do when the joint distribution of the x_i is normal. In the normal case all the higher cumulants are zero and, though the equations hold, they are useless for determining the α_i.[28] Geary (1943, p. 184) argued, "It seems desirable to consider a particular set of circumstances in which a modification of the theory will apply when the joint distribution of the variates ... is normal, provided that their variances and covariances exist." On this first outing, instrumental variables was an improvisation to accommodate the awkward normal case.

Geary divides the k variables into two groups $\xi_t = (\xi_{1t}, \xi_{2t})$ with k_1 and k_2 elements. There are conformable partitions of α and ϕ. The second group of variables does not enter into the equation, i.e. $\alpha_2 = 0$. So

$$\alpha'_1\xi_{1t} = 0.$$

Multiplying through by ξ_{2t} and taking expectations

$$\alpha'_1\phi_{12} = \alpha'_1 M_{12} = 0$$

Geary presents these results without any motivating argument. Besides obtaining them from the cumulant equations, they could be reached from the least squares "normal equations" for α setting $\alpha_2 = 0$.

About presenting the equations, Geary makes some observations. If $k_2 = k_1 - 1$ the equations will be sufficient in general to determine the coefficients up to a factor of proportionality. If $k_2 > k_1 - 1$, it "is necessary that certain determinants should vanish" so that the equations have a non-zero solution. This problem of superfluous instruments, corresponding to overidentification in the errors inequations model was picked up in his 1949 paper. There may be more than one set of solutions: "Iif there are many variates in the investigation it is quite likely that such linear relations will be found." Geary returned to the issue of multicollinearity in 1948.

Geary's assessment of the method was that:

> The method may be quite effective when all the inherent linear relations in the data are in fewer than half the variates measured. The disadvantage is, of course, that it cannot determine relations in more than half the variates. (p. 185)

28. Of course it was not an idiosyncrasy of Geary's method that it did not work in this case; the model is not identified. See p. 259 and footnote 23 above.

Geary discusses the sampling properties of the estimator in conjunction with the sampling properties of his cumulant-based estimator. Fiducial intervals are constructed but the results are more or less buried under complex notation designed for the more complicated estimator. There is no reference to Reiersøl (1941), presumably because Geary was not yet aware of it.

Geary's next contribution (1948) was the outcome of a stay at the Cambridge Department of Applied Economics. The Cambridge connection is very apparent. Stone (1947) had applied principal components analysis to some macroeconomic data. Geary analyses the same data and discusses the connection between Tintner's analysis of the simultaneous errors in variables model and principal components analysis. T.W. Anderson, a later visitor to the Department, considered the multivariate analysis underlying Geary's work in greater detail.

The paper begins with a brief account of the position: three "angles of attack" have been tried, the confluence analysis of Frisch and Reiersøl, the work of Koopmans and Tintner and Geary's own semi-invariants work. These attacks are "complementary and not antagonistic." As often in Geary's papers, the tentativeness of conclusions and the need for experimentation are stressed. The paper focuses on Tintner's method; Geary elucidates its relationship with principal components analysis and suggests some modifications to the method. Once again instrumental variables is overshadowed; the author's "new method" is an extension of Tintner's method.

Geary (1943, pp. 181-182) had a brief discussion of how to apply his semi-invariant method to the multicollinearity case — the case of a system of relationships — but this case is much more prominent here. Tintner (1945) extended Koopmans's (1937) maximum likelihood method to this case. Like Koopmans, Tintner assumed that the variance of each measurement error was known, though he had a proposal for estimating these variances. However Tintner did not confine himself to estimation in the context of a *known* number of relationships. An important objective of the analysis was to find the number of independent relationships in the data, to estimate the degree of multicollinearity.

Tintner's methodological presuppositions were quite different from those of Haavelmo (and Koopmans) in their contemporary work on the errors in equations model. Tintner (1946, p. 14) describes the difference very clearly:

> The possibility of estimating the "true dimensionality" of our problem in statistical terms (a problem already envisaged by Frisch and contemplated by Fisher and others in "discriminant analysis") distinguishes our approach from the one of Haavelmo and his followers. They simply *assume* R, the number of independent linear relationships between the

variables as given. Hence it seems possible that they may endeavour to accomplish too much, i.e., to determine a greater number of equations than actually exist in the data. This would by necessity lead to non-sensical results. The reason for this distinction between their method and our procedure lies in the different rôle of economic theory and the reliance put into a priori assumed relationships, which seem to be greater with Haavelmo and his school than with us.[29]

As we saw in our discussion of (2.1), the coefficient matrix A is not uniquely determined by the observable variance matrix, M. But, it is possible to estimate the space spanned by the rows of the matrix. Hendry and Morgan (1989) have related modern interest in cointegration to Frisch's work on the errors in variables model. The associated revival of interest in "true dimensionality" has taken the form of identifying the cointegrating space. It is interesting that Johansen (1991) includes a reference to Anderson (1951) which reviewed Tintner and Geary's work on "true dimensionality".

Suppose the system of relationships is known to consist of r equations

$$A\xi_t = 0$$

Tintner finds an orthonormal basis for A by imposing the constraint $A\Sigma A' = I$. He estimates the rows of A as the vectors satisfying

$$a(\Sigma x_t x_t' - \lambda\Sigma) = 0$$

corresponding to the r smallest roots of

$$|\Sigma x_t x_t' - \lambda\Sigma| = 0.$$

When r is unknown, the procedure is to find the characteristic roots and arrange them in order of magnitude. When there are r relations the normalised sum of the roots is asymptotically χ^2 with a specified number of degrees of freedom. When the statistic is large the hypothesis of r relations is rejected in favour of more than r relations. The technique was adapted from discriminant analysis but found use in other branches of multivariate analysis.[30]

29. Morgan (1990, Chapter 8) and Aldrich (1989) discuss the structural econometrics of "Haavelmo and his followers". It is not clear to what extent Geary shared Tintner's views, though Spencer (1976, p. 163) remarks, "A striking feature of much of this work and, indeed, of much of all of his work is [Geary's] reluctance to allow significant intrusion of the a priori, so common in econometrics.

30. Anderson (1948), a companion paper to Geary (1948), discusses the theorems on the distribution of latent roots underlying the analysis of Tintner and Geary. Both Geary and Anderson were at the Department of Applied Economics when their work was done. Unfortunately Geary and his work are not recalled in Anderson's reminiscences (Phillips, 1986).

The need to assume that the error variances are known was the weakness of the Koopmans-Tintner approach. Geary suggested a way round this problem, which also improved on the approach by producing consistent estimates of the systematic parts of the variables (the incidental parameters) which maximum likelihood failed to do. In Tintner's set-up there was no structure on the systematic variables beyond the possibility of one or more linear relationships between them. Geary proposed that these variables be represented as smooth functions of time:

$$\xi_{it} = \gamma_{i1}f_{1t} + \dots + \gamma_{ik}f_{mt}$$

where the functions $f_{1t}, \dots, f_{mt}$ are orthogonal polynomials. For each variable it would be possible to estimate the γs and also the error variance, σ_{ii}, using least squares. Geary expresses the true values of the variables as functions of time and uses the least squares fitted values in the maximum likelihood approach proposed by Tintner, including the method for determining the number of relations.

In this paper Geary refers for the first time to Reiersøl's work. He generously refers to the instrumental variable estimator as Reiersøl's method" — a practice he followed in subsequent publications. Of course, the "method" was the one Geary had himself introduced in 1943. Why does he attribute it to Reiersøl? Reiersøl published in 1941, before Geary, but he did not publish the *method*. As we have seen, with Reiersøl's measurement assumptions instrumental variables does not provide consistent estimates. Yet for someone, who knew what Geary knew and wanted to do what Geary wanted to do, the alteration needed to make instrumental variables "work" was trivial. Geary may have read the method into Reiersøl's work because it was so obvious or he may have regarded Reiersøl as having done almost all that was necessary to make the method work. Either way, attributing the method to Reiersøl would seem reasonable. The "idea" of instrumental variables certainly belonged to Reiersøl. He made much more of a fuss about instrumental variables than Geary did; they were not just plugging a gap left by another more important technique. Reiersøl had also suggested a source of instrumental variables, viz. lagged values.

For the case of a single equation Geary compares several methods, the Koopmans-Tintner maximum likelihood method, Geary's modification that we have discussed plus some other simplifications and lastly the instrumental variable estimator. Geary shows that the instrumental variable estimator is asymptotically unbiased and obtains its limiting variance. In general, the comparisons between estimators make the author's "new method" look good. Geary recognises that there is a non-comparability between some of the

E

estimators because they use different information.

Geary's extension of Tintner is referenced in the later literature on the estimation of sets of relations.[31] The paper's influence on the subsequent development of instrumental variables was curious. Sargan, the most influential writer of the next generation, describes its impact on him:

> The article which started me up was [the] article by Geary ... That took me back to the earlier work by Reiersøl, and I pretty early realised that the Geary method was very close to LIML except he was using arbitrary functions of time as the instrumental variables, particularly polynomials in the time variable. 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. (Phillips, 1986, p. 123)[32]

Sargan's description of "the Geary method" is not Geary's description. Geary does *not* say that his method is instrumental variables with $f_{1t}, \ldots f_{mt}$ as instruments.

With Geary (1949) we reach the definitive paper on instrumental variable estimation for the errors in variables model. Unlike the 1943 paper this one treats instrumental variables as a method in its own right — its own right being as a method for the case when the error variances are unknown. Thus the method is complementary to those treated in the 1948 paper.

This is a much less "social" paper than the 1948 piece. It does not depend in any important way on the contributions of others but draws heavily on earlier results obtained by its author. However, the data set from Stone is again used to illustrate the theory. From the vantage point of this ambitious and comprehensive paper the thinness of the earlier contributions — by both Reiersøl and Geary — is evident. It is not primarily a matter of supplying details or refinements but of fundamental explanations of how the method works.

Geary's work on the distribution of the estimator raises issues that were beyond Reiersøl's horizons. We saw on p. 258 above the sketchiness of Reiersøl's account of what makes instruments useful. Geary could characterise useful instruments and consider the question of the optimal weighting of different instrumental variables. He could escape the pairing of instrumental and investigational variables that had been a feature of his earlier treatments of instrumental variable estimation. This analysis was followed

31. See Gleser (1981) for a modern treatment of this problem, including a discussion of Tintner and Geary's work.

32. LIML was introduced by Anderson and Rubin (1949). It is discussed together with the Tintner and Geary work in Anderson (1951).

up by Klein (1955) in his account of the relationship between instrumental variables and two-stage least squares.

After establishing that the estimator is consistent under fairly general conditions, Geary presents the distribution theory for the instrumental variable estimator in three cases. The first case may be considered the primary one: Geary obtains the most complete results for this case and he relates the other two cases to it. It is the case in which the observations on the two investigational and the one instrumental variable are a random sample from a trivariate normal distribution. Historically it is the primary case too. It is the one for which he first proposed the estimator in 1943 and the analysis draws on a line of research going back to his 1930 paper on the frequency distribution of the quotient of two normal variates. Using characteristic function methods he obtains the exact distribution for the slope coefficient.

The second case is referred to as the "simplest stationary type" but this is misleading; it is not the set-up envisaged by Reiersøl. It is the case when the true values of the investigational variables can be treated as fixed in repeated sampling. A third case preserves this assumption on the investigational variables but uses as the instrumental variable a lagged or advanced investigational variable. With these subordinate cases Geary is concerned with the quality of approximation afforded by the results for the first case. For moderate size samples he argues that the approximation is good.

Geary's interest is not confined to the properties of estimators; he discusses tests of whether the relationship is "complete", i.e., that no variables are omitted from the relationship. One of these is based on the idea that, for a properly specified equation, estimates based on different instruments should be in close agreement. Another is based on the idea that, when the equation is complete, an additional investigational variable will not contribute to explaining the fitted value: the procedure is to investigate the relationship between the fitted value and the new variable using an additional instrument. However, compared to the attention given to the properties of the estimators, these ideas have the status of tentative suggestions.

The paper ends by drawing a contrast between the proposed methods and Frisch's confluence analysis:

> Classical regression theory on which confluence analysis so largely depends, will only afford consistent estimates in the almost trivial case of no disturbance or in which so many variables have been introduced into the equation that the fit of the plane to the observations is very close ... This is emphatically *not* to say that confluence analysis has not a value for determining just the set of variables constituting a complete set ...

> Actually all the familiar difficulties of collinearity, etc., encountered in confluence analysis arise in the technique here discussed, but they have an entirely different character. They can all be resolved if the variables and observations are numerous enough, whereas in classical regression analysis, no matter how many observations there are, the estimates of the coefficients are biased. (p. 57)

The emphatic disclaimer in the first paragraph is not really convincing for confluence analysis as a means of determining which are the relevant variables also rests on "classical regression theory". Geary's last word on the new method itself was not the resolvability of all the problems but that, "It only remains now to build up a set of applications of the theory to test its practical efficacy!"

V CONCLUSION

The method of instrumental variables for the errors in variables model is very simple. Its history should be just as simple. The papers marking the end of the process by which the method was established present a simple history. Thus Geary (1949) supplies the following context for his work:

> Reiersøl (1941, 1945) and Geary (1942, 1943) have shown how relations, *consistent* in the statistical sense, between variables can be derived. Reiersøl used what he termed the *instrumental sets* of variables in contradistinction to the *investigational set*, which are the variables the systematic parts of which enter into the relations; Reiersøl's method is that used in this communication. Geary showed how relations could be established from knowledge of the investigational set alone by having recourse to semi-invariants of power greater than two and of dimension greater than unity. (p. 30)

The author not only writes himself right out of the instrumental variables story but gives a misleading account of Reiersøl's work.

Geary's paper, with Reiersøl's 1950 paper, summarise the achievement of the decade, in the sense of stating the main findings. But at the same time they gave a rationalised reconstruction of the development of ideas. During the decade perspectives changed radically and a great deal of work was scrapped. Some of this scrap had been the setting for the original work on instrumental variables. Most of the work done under the heading of confluence analysis belongs in this category.

We have seen that the development of ideas between 1941 and 1950 was not simple. In both the early work of Reiersøl and that of Geary the estimator was buried. In the case of Reiersøl, his choice of measurement assumptions barred him from the textbook theorem about the textbook estimator. For Geary, the textbook theorem for the textbook estimator was there to be found but he was not interested in extracting and emphasising it.

The work of Reiersøl and Geary on instrumental variables was motivated by different problems with different views of what constitutes a "good" result. But there was an "idea of using instrumental variables" common to their work; this amounted to the use of variables outside the relationship to gain information about the relationship between the "inside" variables.

The present account should do something to "de-homogenise" Reiersøl and Geary.[33] Homogenisation is usually accompanied by mis-attribution. We have seen a number of cases: Geary's account of Reiersøl and Sargan's account of Geary's (1948) method. Another instance is Anderson and Rubin's (1949, p. 50) description of Reiersøl's (1941) method in the context of estimating a single exactly identified equation from a system. They state that the idea behind Reiersøl's method is "similar" to that behind indirect least squares. For the historian such mis-attributions are fascinating. On the one hand, they make the task of historical reconstruction harder. On the other, mis-attributions can be important historical facts in their own right. The mis-attribution of an idea to somebody may signal misunderstanding — perhaps a creative misunderstanding, by which the misunderstanding improves on the original. Or the mis-attributor may understand very well that the idea is not in the original but consider it such a trivial extension — from his own standpoint — that it is not worth recording as a discovery. In either case it is important that such acts are identified as creative acts.

It is appropriate that instrumental variables should be very hard to pin down for it has itself been one of the great homogenising agents in econometrics. Durbin (1954, p. 26) describes Wald's (1940) grouping method as an instrumental variable estimator and numerous estimation methods have since been given an instrumental variable "interpretation". Most spectacularly — perhaps with tongue in cheek — Farebrother (1986) uses the title "The Early History of Instrumental Variable Estimation" for a discussion that begins around 1750. Numerous estimators from the days before the introduction of least squares can be given an instrumental variable interpretation — as well as least squares itself, of course! But they come from different intellectual universes from that (or those) inhabited by Reiersøl and Geary.

In a recent book Manski (1988, p. ix) writes, "Around 1980 ... I developed the approach I called 'closest empirical distribution' estimation ... For the first time, I felt that I understood the essence of the instrumental variables method introduced by Wright and Reiersøl for the estimation of linear

33. The term is taken from Jaffé (1976).

models."[34] He does not mean he has divined what Reiersøl and Wright were getting at. There are as many "essences" as there are valid points to be made about the method and Manski's "essence" is different from the essence(s) for the writers of the fifties — Durbin (1954), Klein (1953, 1955) and Sargan (1958) — who saw the connection between instrumental variables and the methods devised for the errors in equations model.[35] For the writers of the 1940s the essence changed as they worked with the method: for Geary in 1943 the point of the method was that it could deal with normal variables while in 1949 its point was that it could deal with variables whose error variance was unknown. There was no contradiction between the arguments underlying these points but the emphases were different.

REFERENCES

ALDRICH, J., 1989. "Autonomy", *Oxford Economic Papers*, Vol. 41, pp. 15-34.

ALDRICH, J., 1992. *Haavelmo's Identification Theory*, Southampton University Discussion Paper No. 9218.

ANDERSON, T.W., 1948. "The Asymptotic Distributions of the Roots of Certain Determinantal Equations", *Journal of the Royal Statistical Society, B*, Vol. 10, pp. 132-139.

ANDERSON, T.W., 1951. "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions", *Annals of Mathematical Statistics*, Vol. 22, pp. 327-351.

ANDERSON, T.W., and H. RUBIN, 1949. "Estimation of the Parameters of a Single Equation in a Complete Set of Stochastic Equations", *Annals of Mathematical Statistics*, Vol. 20, pp. 46-65.

BOWDEN, R., and D.A. TURKINGTON, 1984. *Instrumental Variables*, New York: Cambridge University Press.

DURBIN, J., 1954. "Errors in Variables", *Review of the International Statistical Institute*, Vol. 22, pp. 23-32.

FAREBROTHER, R.W., 1986. *The Early History of Instrumental Variable Estimation*, University of Manchester: Department of Econometrics and Social Statistics Discussion Paper.

FRISCH, R., 1934. *Statistical Confluence Analysis by Means of Complete Regression Systems*, Oslo: University Institute of Economics.

FRISCH, R., 1938. "Statistical Versus Theoretical Relations in Economic Macrodynamics", *League of Nations Memorandum*, Geneva: League of Nations.

FULLER, W.A., 1987. *Measurement Error Models*, New York: Wiley.

GEARY, R.C., 1930. "The Frequency Distribution of the Quotient of Two Normal Variates", *Journal of the Royal Statistical Society*, Vol. 93, pp. 442-446.

GEARY, R.C., 1942. *Inherent Relations between Random Variables, Proceedings of the Royal Irish Academy*, Vol. 47A, pp. 63-76.

34. Wright's (1928) work played no part in the developments treated in this paper. His work was neglected by econometricians until Goldberger (1972) drew attention to it.

35. Qin (1989/1993) discusses the absorption of instrumental variables estimation into simultaneous (errors in) equations analysis.

GEARY, R.C., 1942a. "The Estimation of Many Parameters", *Journal of the Royal Statistical Society*, Vol. 105, pp. 213-217.

GEARY, R.C., 1943. *Relations between Statistics: The General and the Sampling Problem when the Samples are Large, Proceedings of the Royal Irish Academy*, Vol. 49A, pp. 177-196.

GEARY, R.C., 1944. "Comparison of the Concepts of Efficiency and Closeness of Estimate for Consistent Estimates of a Parameter", *Biometrika*, Vol. 33, pp. 123-128.

GEARY, R.C., 1948. "Studies in Relations Between Economic Time Series", *Journal of the Royal Statistical Society, B*, Vol. 10, pp. 140-158.

GEARY, R.C., 1949. "Determination of Linear Relations between Systematic Parts of Variables with Errors of Observation the Variances of which are Unknown", *Econometrica*, Vol. 17, pp. 30-58.

GEARY, R.C., 1953. "Non-linear Functional Relationships Between Two Variables when One Variable is Controlled", *Journal of the American Statistical Association*, Vol. 48, pp. 94-103.

GEARY, R.C., 1963. "Some Remarks about Relations between Stochastic Variables: A Discussion Document", *Review of the International Statistical Institute*, Vol. 22, pp. 23-32.

GLESER, L.J., 1981. "Estimation in a Multivariate 'Errors in Variables' Regression Model: Large Sample Results", *Annals of Statistics*, Vol. 9, pp. 24-44.

GOLDBERGER, A., 1968. *Topics in Regression Analysis*, New York: Macmillan.

GOLDBERGER, A., 1972. "Structural Equations Methods in the Social Sciences", *Econometrica*, Vol. 40, pp. 979-1,001.

HAAVELMO, T., 1943. "The Statistical Implications of a System of Simultaneous Equations", *Econometrica*, Vol. 11, pp. 1-12.

HAAVELMO, T., 1944. "The Probability Approach in Econometrics", Supplement to *Econometrica*, Vol. 12.

HAAVELMO, T., 1950. "Remarks on Frisch's Confluence Analysis and its Use in Econometrics", in T.C. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*, Chapter 5 and pp. 258-265. New York: Wiley.

HENDRY, D.F., and M.S. Morgan, 1989. "A Re-analysis of Confluence Analysis", *Oxford Economic Papers*, Vol. 41, pp. 35-52.

JAFFÉ, W., 1976. "Menger, Jevons and Walras De-homogenised", *Economic Inquiry*, Vol. 14, pp. 511-524.

JOHANSEN, S., 1991. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica*, Vol. 59, pp. 1,551-1,580.

KALMAN, R.E., 1982. "System Identification from Noisy Data", in A.R. Bednarek and L. Cesari (eds.), *Dynamical Systems II*, Lecture 9 and pp. 134-164, San Diego: Academic Press.

KENDALL, M.G., and A. STUART, 1967. *The Advanced Theory of Statistics*, Volume 2, second edition, London: Griffin.

KLEIN, L.R., 1953. *A Textbook of Econometrics*, New York: Row, Peterson.

KLEIN, L.R., 1955. "On the Interpretation of Theil's Method of Estimating Economic Relationships", *Metroeconomica*, Vol. 7, pp. 147-153.

KLEPPER, S., and E. LEAMER, 1984. "Consistent Sets of Estimates for Regressions with Errors in All Variables", *Econometrica*, Vol. 52, pp. 163-184.

KOOPMANS, T.C., 1937. *Linear Regression Analysis of Economic Time Series*, Haarlem: F. Bohn.

KOOPMANS, T.C., 1950. "When is an Equation System Complete for Statistical Purposes?", in T.C. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*, Chapter 17 and pp. 393-409, New York: Wiley.

KOOPMAN, T.C., and O. REIERSØL, 1950. "The Identification of Structural Characteristics", *Annals of Mathematical Statistics*, Vol. 21, pp. 165-181.

KOOPMANS, T.C., H. RUBIN, and R.B. LEIPNIK, 1950. "Measuring the Equation Systems of Dynamic Economics", in T.C. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*, Chapter 2 and pp. 52-237, New York: Wiley.

LEDERMANN, W., 1937. "On the Rank of the Reduced Correlational Matrix in Multiple Factor Analysis", *Psychometrika*, Vol. 2, pp. 85-93.

MANSKI, C.F., 1988. *Analog Estimation Methods in Econometrics*, London: Chapman and Hall.

MALINVAUD, E., 1966. *Statistical Methods of Econometrics*, Amsterdam: North Holland.

MORGAN, M.S., 1990. *The History of Econometric Ideas*, New York: Cambridge University Press.

PHILLIPS, P.C.B., 1985. "The ET Interview: Professor J.D. Sargan", *Econometric Theory*, Vol. 1, pp. 119-139.

PHILLIPS, P.C.B., 1986. "The ET Interview: Professor T.W. Anderson", *Econometric Theory*, Vol. 2, pp. 249-288.

QIN, D., 1989. "Formalisation of Identification Theory", *Oxford Economic Papers*, Vol. 41, pp. 73-93.

QIN, D., 1989/93. *History of Econometric Thought: The Formation of Econometrics (1930-1960)*, Oxford D. Phil. thesis, to be published as *Formation of Econometrics: A Historical Perspective* by Oxford University Press, 1993.

REIERSØL, O., 1941. "Confluence Analysis by Means of Lag Moments and Other Methods of Confluence Analysis", *Econometrica*, Vol. 9, pp. 1-24.

REIERSØL, O., 1945. "Confluence Analysis by Means of Instrumental Sets of Variables", *Arkiv för Matematik, Astronomi och Fysik*, Band 32A, pp. 1-119.

REIERSØL, O., 1947. "Confluence Analysis When the Model Contains Both Shocks and Errors", *Cowles Commission Discussion Papers and Minutes* (Statistics), No. 306.

REIERSØL, O., 1950. "Identifiability of a Linear Relation between Variables which are Subject to Error", *Econometrica*, Vol. 18, pp. 375-389.

REIERSØL, O., 1950a. "On the Identifiability of Parameters in Thurstone's Multiple Factor Analysis", *Psychometrika*, Vol. 15, pp. 121-149.

SARGAN, J.D., 1958. "The Estimation of Economic Relationships Using Instrumental Variables", *Econometrica*, Vol. 26, pp. 393-415.

SPENCER, J.E., 1976. "The Scientific Work of Robert Charles Geary with List of Published Works", *The Economic and Social Review*, Vol. 7, No. 3, pp. 233-247.

SPENCER, J.E., 1983. "Robert Charles Geary — An Appreciation", *The Economic and Social Review*, Vol. 14, No. 3, pp. 161-164.

STONE, J., 1947. "On the Interdependence of Blocks of Transactions (with Discussion)", *Journal of the Royal Statistical Society, (Supplement)*, Vol. 9, pp. 1-45.

TINTNER, G., 1945. "A Note on Rank, Multicollinearity and Multiple Regression", *Annals of Mathematical Statistics*, Vol. 16, pp. 304-308.

TINTNER, G., 1946. "Multiple Regression for Systems of Equations", *Econometrica*, Vol. 14, pp. 5-36.

WALD, A., 1940. "The Fitting of Straight Lines If Both Variables are Subject to Error", *Annals of Mathematical Statistics*, Vol. 11, pp. 284-300.

WRIGHT, S., 1928. Appendix to *The Tariff on Animal and Vegetable Oils*, by P.G. Wright, New York: Macmillan.