SOAS University of London Department of Economics

Working Paper Series

ISSN 1753 - 5816

Please cite this paper as:

Qin, D. and Zhang, Y-Q. (2013), "A History of Polyvalent Structural Parameters: the Case of Instrument Variable Estimators" SOAS Department of Economics Working Paper Series, No. 183, The School of Oriental and African Studies.

No. 183

A History of Polyvalent Structural Parameters: the Case of Instrument Variable Estimators

by

Duo QIN and Yanqun ZHANG

(August, 2013)

Department of Economics
School of Oriental and African Studies
London
WC1H 0XG
Phone: + 44 (0)20 7898 4730
Fax: 020 7898 4759
E-mail: economics@soas.ac.uk
http://www.soas.ac.uk/economics/

The **SOAS Department of Economics Working Paper Series** is published electronically by The School of Oriental and African Studies-University of London.

Design and layout: O. González Dávila

A History of Polyvalent Structural Parameters: the Case of Instrument Variable Estimators

Duo QIN[1]

Department of Economics, SOAS, University of London, UK

Yanqun ZHANG

IQTE, Chinese Academy of Social Sciences, China

Abstract

This paper investigates the rise and fall of the IV method in macro-econometric models and its subsequent revival in micro-econometric models. The key findings are: (i) the IV method implicitly breaks the contemporaneously circular causality postulated in a simultaneous-equation model (SEM) by redefining the conditional variable concerned as a suboptimal conditional expectation of it; (ii) the IV method falls out of favour in macro-econometrics mainly because of lack of empirical validations for such redefinitions; (iii) the IV method wins its popularity in micro-econometrics by its capacity to produce multiple suboptimal conditional expectations of the latent conditional variables of interest under the disguise of an SEM consistent estimator; nevertheless, (iv) such suboptimal conditional expectations give rise to the insurmountable difficulty of credibly interpreting the IV-based parameter estimates, especially in the case of prognosticated omitted variable bias. The findings highlight the methodological drawback of the estimator-centric strategy of textbook econometrics.

JEL classification: B23, C13, C18, C50

Key words: Instrumental variables, simultaneity, omitted variable bias, collinearity

[1] We thank Olav Bjerkholt, Chris Gilbert, David Hendry and Risa Morimoto for helpful comments.
Correspondence to the contacting author: dq1@soas.ac.uk .

1. Introduction

The instrumental variable (IV) method is taught as an essential estimation tool in all econometrics textbooks. In recent years, it has risen to a particularly prominent place in micro-econometrics mainly through the empirical popularity of models for measuring treatment effects, e.g. see Cameron (2009). In fact, the popularity has already grown well beyond micro-econometrics and spread into economic development studies, e.g. see Duflo *et al* (2008) and Banerjee and Duflo (2009). Meanwhile, the causal validity of the IV-based programme evaluation models has been critically disputed, e.g. see Angrist *et al* (1996), Heckman (1996; 2010), Deaton (2010) and Imbens (2010).

In the comments to Angrist *et al* (1996), the following statement has caught our attention: ‘There are many unfortunate barriers to effective communication between statisticians and economists. The method of instrumental variables (IV) and associated methods for simultaneous equations and for “structural” estimation constitute one of the greatest. These methods are in the toolkit of virtually every economist and are among the most widely used techniques in the field. … Yet it is scarcely used or discussed by statisticians, who often do not see the point of it all’ (Moffitt, 1996, p. 462). It intrigues us as why there is such a difference in attitude between the two professions.[2]

In Wooldridge’s *Econometric Analysis of Cross Section and Panel Data*, one of the most widely used microeconometrics textbooks, the IV method is considered as ‘probably second only to ordinary least squares in terms of methods used in empirical economic research’ (2010, p. 89). In another popularly used companion textbook, the method is attributed to as the ‘most powerful weapon’ for estimating simultaneous-equation models (Angrist and Pischke, 2009, p.

[2] In a popular book by Pearl, a prominent computer scientist and statistician, ‘instrumental variables’ are categorised as a ‘causal concept’ but without much explanation (2009, p. 40).

114). That tribute is further backed by the following statement: ‘Simultaneous equations models (SEMs) have been enormously important in the history of econometric thought. At the same time, few of today’s most influential applied papers rely on an orthodox SEM framework, though the technical language used to discuss IV methods still comes from this framework. Today, we are more likely to find IV methods used to address measurement error problems than to estimate the parameters of an SEM. Undoubtedly, however, the most important contemporary use of IV methods is to solve the problem of omitted variables bias (OVB)’ (ibid, pp. 114-5). The statement has triggered the present investigation – to track down what has led to such a shift of course, with the hope that the history will help us fathom some logical explanations to the changing fortunes of the IV method and illuminate a path through the methodological labyrinth over the valid use of IVs for applied modellers.

Our historical investigation is presented in two sections, one on the rise and fall of the IV method in macro-econometric studies from the 1940s to the 1970s-1980s (Section 2), and the other on the revitalisation of the IV method in micro-econometric models from the late 1970s up to the 1990s (Section 4). Each of the historical sections is followed by a review section dissecting the basic logical ideas of the history concerned (see Sections 3 and 5). The mathematical illustrations in these review sections are kept at an as elementary as possible level so as to make the logical arguments easily comprehensible for applied economists whose routine econometric toolkits are built around simple regression models.

Our historical investigation reveals that the research strategy of treating specific empirical model design problems as a general problem of parameter estimation often leads into sidetracks and creates methodological confusions. In fact, the drawback of such an estimator-centric strategy was recognised by macro modellers decades ago, e.g. see Qin (2013a). Unfortunately, the micro-econometric community remains largely indifferent to what happened in macro-econometrics during its reformative period of the late twentieth century, as shown from the revival of the IV method there. Specifically, our investigation yields the following key findings concerning the IV method: (i) the method implicitly breaks the contemporaneously circular causality of a postulated simultaneous relationship by redefining the conditional expectation of the modelled variable – substituting the contemporaneous explanatory variable which is assumed *endogenous* by a suboptimal conditional expectation of it, thus rejecting that variable as a correctly postulated conditional variable; (ii) the method falls out of favour among macro-econometric modellers because of lack of evidence which would falsify the validity of those conditional variables, whereas there is relatively abundant evidence showing that dynamically inadequate specification forms the key weakness of the traditional SEMs; (iii) the method wins its popularity in micro-econometrics by its capacity to produce multiple suboptimal conditional expectations of the latent explanatory variables of interest under the disguise of an SEM consistent estimator; however, (iv) those suboptimal conditional expectations give rise to an insurmountable difficulty – finding credible interpretation for those IV-based parameter estimates, especially in the case of prognosticated OVB, the main concern of applied modellers in micro and development studies. These findings are further elaborated in the final section.

2. The Rise and Fall of IVs in Macro-econometric Modelling

The term 'instrumental variables' is commonly acknowledged to come from O. Reiersøl's thesis (1945) 'Conference Analysis by Means of Instrumental Sets of Variables', see Morgan (1990, section 7.3). The ideas of using IVs in econometrics were introduced independently by Reiersøl and Geary in the early 1940s, see Aldrich (1993), although the IV method is now known to have been invented much earlier by P.G. Wright (1928), e.g. see Stock and Trebbi (2003). Nevertheless, it should be noted that Reiersøl's IV method was devised for tackling the measurement error problem in the context of error-in-variable models, while, at the same time, it was the SEM in the error-in-equation form which formed the key model of interest in the formalisation of econometrics by the Cowles Commission (CC), e.g. see (Bowden and Turkington, 1990, Section 1.3) and also (Qin, 1993, Chapter 3).

One of the most influential technical advances by the CC was arguably the limited-information maximum likelihood (LIML) estimator (see Anderson and Rubin, 1949). The LIML was put forward as a computationally more convenient method than the full-information maximum likelihood (FIML) estimator, the optimal method for SEMs, following Haavelmo's (1943) demonstration that the ordinary least squares (OLS) estimator was inconsistent with the SEM specification. The equivalence of the LIML to the IV method was subsequently recognised by Durbin (1954). Around the same time, Theil (1953) proposed the two-stage least squares (2SLS) as another convenient estimator for an SEM. Theil's 2SLS was soon interpreted as an IV estimator by Klein (1955).

Formal extension of the IV method for SEMs was subsequently explored by Sargan (1958; 1959). Sargan was apparently so attracted to the versatile capacity of the IV method that he spent a long time developing the computer programme for his IV estimator for an extended SEM with autocorrelated residual terms, see (Gilbert, 1989). The first trial experiment was carried out on the wage and price models built by Klein and Ball (1959), using the UK quarterly time-series data, see Sargan (1964). The trial IV estimates of the model, however, turned out to be so poorly determined that Sargan decided 'there seemed little point in trying to find a better set of instrumental variables' (1964, p. 39). Sargan therefore abandoned his own IV invention half way through and moved on to conduct an extensive dynamic model specification search to try and improve the Klein-Ball model, aimed mainly by the OLS. Sargan's search resulted in an error-correction model, which was to become one of the most popular model forms in macro-econometrics over two decades later. It was also mainly due to his dynamic model search that Sargan's 1964 paper has been regarded as the ground-breaking work for the

LSE (London School of Economics) modelling approach, e.g. see Hendry (2003) and also Qin (2013a, Chapter 4).

In fact, Sargan's (1964) empirical modelling success with the OLS had been anticipated by H. Wold's 'proximity theorem' over a decade before, see Wold and Juréen (1953, pp. 37-38) and also Wold and Faxér (1957). According to Reiersøl, interestingly, Wold was 'the first opponent' of his IV method (Willassen and Reiersøl, 2000, p.118). Being a staunch proponent of the OLS method, Wold was highly critical of the SEM approach developed by Haavelmo and the CC group. In a series of papers (see Wold, 1954; 1956; 1960; 1961; 1965), Wold criticised the CC's SEM specification as fundamentally flawed due to its inadequately formulated causal structure. He pointed out that '*conditional expectation'* was a 'key notion' to provide 'the rationale for the operative use of the relation in theoretical and applied work, and … for estimating its parameters by the classic method of least squares regression' (1961). Wold also emphasised that the choice between a causal chain model and an SEM was 'not a matter of estimation technique' (1965). Nevertheless, his 'proximity theorem' demonstrated that the inconsistency of the OLS in an SEM should remain practically small as long as the model was approximately of the 'recursive' or 'causal chain' type with serially uncorrelated residual terms, and that the magnitude of the inconsistency would dwindle with the size of the variance of the residuals.

Similar to Sargan's 1964 work, Wold's viewpoints were largely overshadowed by the Haavelmo-CC SEM approach during its consolidation period, see Qin (2013a, Chapter 1). Ironically, Wold's causal ordering principle was adopted as the key rule to guide the *a priori* choice of eligible IVs from all the exogenous and lagged variables of SEMs in practice, e.g. see Fisher (1965). The adoption implied at least two important messages. First, the appropriate choice of IVs entailed 'using information on the dynamic and causal structure' of *a priori* postulated SEMs (p. 633, ibid). Second, the choice 'is best done through continual application of the a priori structural information which governs the formulation of the entire model in the first place, rather than through relatively arbitrary statistical devices' (p. 590, ibid). Unfortunately at the time, these messages were somehow ignored in most of the empirical studies which adopted the IV method, or in those experimental studies which tried to rank various estimators by means of Monte Carlo simulations, e.g. see Christ (1966, Chapter 9). Nevertheless, although the empirical studies which used the IV-based estimators were on the increase, no clear verdict was reached, either from those empirical studies or from various Monte Carlo experiments, as whether the SEMs estimated by the IV-based methods were definitely superior to the ones by the OLS.

The situation altered drastically during the 1970s and the 1980s when dynamic specification and formulation caught the focal attention of macro-econometric modelling led by the reformative movement of the VAR and the LSE approaches, e.g. see Qin (2013a). Noticeably, a key drive for the movement was the failures of conventionally built macro-econometric models in forecasting the turbulent economic recessions in the wake of the 1973 oil crisis. The exigency to improve forecasting precision helped undoubtedly to secure the status of VAR models in macroeconomics. Simultaneous relationships were absent in the initial VAR specification, whereas simultaneity became implicit in the covariance matrix of the VAR residual terms. The general dynamic setting of VARs resulted in a significant reduction of the standard errors of the error terms as compared to those of the error terms in the traditional SEMs irrespective what estimators were used. The clear and abundant evidence of VARs outperforming traditional SEMs greatly dispelled concerns over the OLS inconsistency with SEMs among macro modellers. They have learnt to attach much more importance to having as small as possible white-noise residuals, i.e. innovation error terms, rather than to circumventing possible correlations between regressors and the associate error term in *a priori* tightly parameterised structural equations.

Even in the case of structural VARs, which were developed mainly to appease the strong SEM conviction in the profession, the FIML estimator is usually applied rather than IV estimators, indicating a common faith in both the symmetric model formulation and the choice of conditional variables in individual equations. In fact, most empirical VAR modellers pay far less attention on individual parameter estimates than the overall model performance in the form of shock-based impulse analyses. However, such analyses are predicated on imposing certain restrictions, e.g. orthogonal restrictions on the covariance matrix of the error terms, and the imposition followed effectively the principle of Wold's causal ordering, e.g. see Sims (1980).

The LSE approach put further emphasis on the importance of forecasting precision by advocating the use of innovation error terms as a fundamental model evaluation criterion for designing specifically robust conditional models, which effectively extended Wold's ideas of causal chain models with white-noise residuals, e.g. see Hendry (1995). As mentioned earlier, the LSE approach grew largely from Sargan's 1964 paper. One important turning point of its initial growth was a shift of focal attention from estimation issues to model specification issues. Interestingly, a computer programme developed by Hendry in the early 1970s to facilitate the shift was named GIVE – Generalised IV Estimator. The programme was soon dubbed the 'model destruction programme' at the LSE because of the high rate of model rejections it generated through comparison of estimated results by various estimators under different model specifications, e.g. see Ericsson and Hendry (2004). GIVE, and its subsequent versions known as Pc-GIVE, have certainly helped reinforce Sargan's 1964 choice to abandon the IV estimator in applied model specification searches. In fact, a switch from the estimator-centric strategy to dynamic model specification research marked the rise of the LSE approach, e.g. see Qin (2013a, Chapter 4). The works by proponents of the LSE approach as well as the VAR approach have now won over the majority of applied macroeconomists to use relatively minimum white-noise residuals as a primary empirical model evaluation criterion. Inevitably, the IV method has lost most of its appeal.[3]

3. An Anatomy of the Fall of IVs in Macro SEMs

Let us now look into the prescription of 'curing' the OLS inconsistency in an SEM by the IV method. A simplest identifiable two-equation SEM is as follows:

$$(1)\qquad \begin{aligned} y_t &= \alpha_1 x_t + \beta_1 z_{1t} + u_{1t} \\ x_t &= \alpha_2 y_t + \beta_2 z_{2t} + u_{2t} \end{aligned}$$

where z_{it} are assumed exogenous variables. It is shown by Haavelmo (1943) that single-equation OLS estimates, $\hat{\alpha}_i$, are inconsistent with (1) because of the correlations of $cov\left(x_t, u_{1t}\right) \neq 0$ and $cov\left(y_t, u_{2t}\right) \neq 0$ due to the assumed simultaneous relationship of $y_t \leftrightarrow x_t$. In other words, both x_t and y_t are assumed endogenous. The IV method is prescribed as a single-equation estimator for restoring the consistency. Since it coincides with the 2SLS, let us represent the IV method by the 2SLS with respect to the first equation in (1):

$$(2)\qquad \begin{aligned} & x_t = \lambda \Psi_t + e_t \quad \Rightarrow \quad x_t^* = \hat{\lambda}\Psi_t \\ & y_t = \alpha_1 x_t^* + \beta_1 z_{1t} + u_{1t}^* \end{aligned}$$

where Ψ denotes a set of IVs which are significantly correlated with x_t, but exogenous with respect to x_t, and not directly part of the explanatory variables to y_t in its structural equation, e.g. the second equation in (2).

It should be noted from Hausman (1978) specification test that the IV equation in (2) must *not* give a very good fit of x_t, definitely not the best fit in the sense that x_t^* must not be the optimal predictor of x_t in order to enable the IV estimate, $\tilde{\alpha}_1$, from (2) differs significantly from the OLS estimate, $\hat{\alpha}_1$, of (1).[4] In other words, the IV-based conditional expectation,

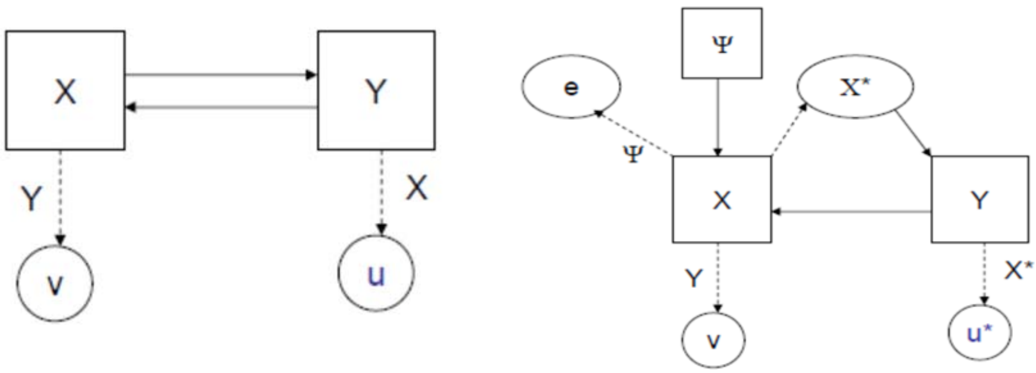
[3] Theoretical research on IVs evolved into the generalised moment method (GMM) since the 1980s, but the method has not been widely applied in macro-econometric models.

[4] We now use 'hat' to denote the OLS and 'tilter' to denote IV estimator.

$x_t^* = E(x_t|\Psi)$, must generate a substantial error term e_t such that x_t^* does not resemble x_t, and that the dissimilarity should holds beyond the sample data. This, however, implicitly alters the postulated causal relationship from $x_t \to y_t$ to $x_t^* \to y_t$. Recognition of the alteration is logically vital because not only of the suboptimal and non-unique nature of the expected variable, $x_t^* = E(x_t|\Psi)$, but also of the consequent break of the contemporaneously circular causality of $y_t \leftrightarrow x_t$, by Ψ. Unfortunately, this alteration remains virtually unrecognised so far, due probably to the prevailed negligence in econometrics of the probabilistic foundation of regression models being conditional expectations, a point stressed repeatedly by Wold decades ago.

Figure 1 illustrates the change of causal conditioning by the IV method in simple path diagrams. The left panel depicts the simultaneity part of a simple SEM, where squares indicate observed variables and circles denote error terms assumed or desired to satisfy the innovation property. The dotted arrows are used to indicate products which are model-derived rather than independently observed prior to the model. The use of an IV estimator for y_t effectively breaks the circle of its symmetrically causal setup with x_t and modifies it as an asymmetric causal chain, as illustrated in the right panel. In particular, it decomposes x_t into two parts at the first stage of the 2SLS to generate a 'latent' expected variable x_t^*. As pointed out earlier, e_t, must not be innovative by intention and that is indicated by the oval shape, so as to secure a suboptimal predictor of x_t. Hence, substantive justifications are needed for the replacement of $x_t \to y_t$ by $x_t^* \to y_t$. One obvious justification is that x_t contains sizeable measurement errors, the original justification for the IV invention. But that is a bit too far from the OLS 'problem' with an SEM. An alternative is to regard x_t as the instrument or medium of Ψ in transmitting its causal effect to y_t. This justification effectively endorses Wold's causal chain specification and also approves of his argument that the problem of SEMs is not in the choice of consistent estimators but the inadequately formulated causal relationships for empirical purposes. However, a logic problem with this justification is that the IVs ought NOT to be selected for substantively causal purposes and hence lacks in general the capacity of shouldering the task of being the ultimate cause.

Figure 1. Path diagrams



It is now easy to see why the IV method has not been successful in the short history of macro-econometric models. There is ample empirical evidence against (1) as a dynamically adequately built model rather than x_t as the correctly postulated conditional variable for y_t in the first equation of (1), or y_t for x_t in the second equation of (1). The move towards VARs retains the spirit of the mutual causality as well as the symmetry of x_t and y_t but simultaneity is explicitly absent in the model formulation. Consider a simple bivariate VAR:

(3) $$\begin{pmatrix} y \\ x \end{pmatrix}_t = A_0 + \sum_{i=1}^{s} A_i \begin{pmatrix} y \\ x \end{pmatrix}_{t-i} + \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}_t \qquad \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}_t = \mathbf{v}_t \sim IN(0, \Sigma)$$

Simultaneity is only implied in the covariance matrix Σ of the error terms. Nevertheless, the 'dynamically' targeted innovation error terms of (3) imply that they are usually much smaller than those of (1) estimated by whatever consistent methods. If we view the estimation of (3) from the IV stance, the conditional expectation, $x_t^* = E(x_t|\Psi)$, implied by (3) is now so similar to x_t that the resulting estimates stand little chance to distinguish themselves from the OLS estimates. In other words, the lagged terms in (3) are 'over-qualified' to serve as IVs in Ψ because they make too precise a prediction of x_t. The rise of VARs thus brings the IV's demise and verifies forcefully Wold's proximity theorem.

However, most empirical VAR modellers care far more for the shock-based impulse analyses than individual parameter estimates in $\hat{A}_i$, as mentioned earlier. In contrast, the robustness of individual parameter estimates occupies a central place in the LSE approach. To achieve that, a dynamic model selection procedure with a set of evaluation criteria is established under the heading of 'the theory of reduction' (see Hendry, 1995, Chapter 9). Conceptually, the procedure is built on the idea of how to search for a data-congruent conditional model through a valid marginalisation of a joint probability distribution of all the variables concerned, e.g. see Hendry and Richard (1982). The search is mainly assisted by a series of exogeneity tests. Suppose that (3) is a model corresponding to such a joint distribution, the LSE approach tries to transform it into:

(4a) $$y_t = a_{10} + \sum_{i=1}(a_{11})_i y_{t-i} + \sum_{j=0}(a_{12})_j x_{t-j} + \varepsilon_{1t}$$

(4b) $$x_t = a_{20} + \sum_{i=1}(a_{22})_i x_{t-i} + \varepsilon_{2t}$$

according to the *a priori* theoretical interest in the conditional relationship $x_t \rightarrow y_t$, which can often be postulated as a simple equilibrium condition:

(5) $$E(y|x) = kx$$

The empirical validity of (4b) can be tested by the Granger causality test, an important test for strong exogeneity. A key prerequisite of the transformation is that ε_{1t} in (4a) passes all the diagnostic tests to fulfil the criterion of being an innovative process. The prerequisite shares virtually the same ground as Wold's proximity theorem. The conditional equation (4a) is then reparameterised into an error-correction model:

$$(6) \qquad \Delta y_t = a_{10} + \sum_{i=1} (\alpha_{11})_i \Delta y_{t-i} + \sum_{j=0} (\alpha_{12})_j \Delta x_{t-j} + \lambda (y - kx)_{t-1} + \varepsilon_{1t}$$

to circumvent the collinear problem among the parameters in (4a) so as to facilitate its economic interpretation at the level of individual parameters. Noticeably, the static theoretical postulate of (5) is embedded in (6) as the long-run equilibrium state. Clearly, the IV method is totally dispensed with here. In fact, the OLS is used most frequently to assist the whole process or model selection and evaluation since the process is essentially built on Wold's proximity theorem, i.e. to secure first a dynamic regression model with the smallest possible white-noise error term.

4. Revitalization of IVs in Microeconometrics

Just as the IV method was submerging in macro-econometrics in the 1970s, there came a wave of interest in using IVs for estimating SEMs which involved endogenous limited dependent variables, or endogenous explanatory variables of the truncated or dichotomous type. Such models arose mainly from micro-econometric studies using household survey data. One of the leading research fields at the time was labour economics. A well-known example is to conduct survey-based analyses of female job participation decisions. Such decisions were conceived of involving models for the interdependence of the choice to work and/or the number of work hours vis-à-vis wage rate. J.J. Heckman was one of the pioneering econometricians who delved into generalising the early labour supply models and devising consistent estimators following closely the Haavelmo-CC tradition. Soon after devising a two-step estimation procedure to circumvent the possible selection-bias issue for a Tobin type truncated regression model (e.g. see Heckman, 1974; 1976), Heckman extended the procedure to an SEM involving an endogenous dummy variable (1978).[5] In particular, a probit model conditional on a set of IVs was used as the first-stage of the 2SLS procedure to filter out the suspected endogenous trait from the dummy variable (see the next section for a more detailed description). Heckman's exploration was strengthened, around the same time, by similar studies with respect to SEMs involving limited dependent variables, e.g. see Nelson and Olson (1978), Lee (1978) and Lee *et al* (1980), and also Maddala (1983, Chapter 7). These works played an important role in the development of selection models or latent-index models using censored regressions in micro-econometrics.

Empirical findings of the subsequent decade or so, however, presented thin and inconclusive supports to the IV (or 2SLS-based) method in tackling the 'endogenous' issue in SEMs involving limited dependent variables. A citation search in JSTOR of the relevant applied studies which cite those works mentioned in the previous paragraph and were published during the 1980s up to the early 1990s has resulted in below five cases. In comparison, there are far more cases which resort to the Heckman's two-step procedure for the purpose of tackling the issue of selection bias but not simultaneity 'bias'.[6] Of the few which have used the method for *a priori* postulated SEMs, the results show either that the evidence of endogeneity is rather weak, e.g. see Stern (1989), or that the difference between the OLS estimates and the IV estimates is mostly statistically insignificant because of the relatively large standard errors of the latter estimates, e.g. see Addison and Portugal (1989). The findings actually verify what has already been found by macro modellers.

It was not until the early 1990s when another wave of promoting the IV method came and brought about a real boom of using IVs in empirical studies. The boom was associated with

[5] Note that Amemiya (1974; 1976) was among the first to extend a Tobin model into an SEM and derive consistent estimators for the model, as acknowledged by Heckman and others who worked on the topic.
[6] There was a certain conceptual confusion in the literature mixing 'selection bias' with 'endogeneity', e.g. (Duncan and Leigh, 1985). What we discuss here is endogeneity due to SEM specification rather than possible selection bias, which was effectively an OVB, see Heckman (1979).

models designed for evaluating the effects of social programmes. The objective of those models oriented the econometricians' attention towards measurement issues with particular respect to the single parameter of a specific social programme of interest, which is commonly represented by a dummy explanatory variable. Since it was seldom possible to conduct controlled experiments with social programmes, the research was inevitably focused on how best one could isolate and measure the sole impact of the programme. Heckman's SEM with endogenous dummy variable was regarded as an immediately applicable model, since the observed dummy variable normally covered a mixture of participants and non-participants of the programme and therefore was susceptive to selection bias or measurement errors, e.g. see Heckman and Hotz (1989). Moreover, the observed outcome could easily be due to a mixture of the programme together with other related factors. Inadequate consideration of these factors would result in the omitted variable bias (OVB) in the estimated programme treatment effect. However, some of those related factors might not be directly observable in practice, and a two-step modelling procedure was naturally thought of as an expedient solution. The idea led to the IV method being chosen as the ideal and general remedy for tackling the 'endogenous' treatment variable problem compounded with selection bias, latent OVB and/or measurement errors.

Much of the early promotions of the IV approach in programme evaluation models stemmed from Angrist's empirical studies published in the very early 1990s. One of his early studies was to measure the effect of military service on subsequent earnings (Angrist, 1990). In spite of the fact that the military veteran status obviously preceded the earning information, the status variable was nevertheless considered as 'endogenous' because it might be 'correlated with the unobserved components of the earnings equation' included in the error term (p. 318). Therefore, the draft lottery data was used as the IV to randomise the sample of veterans so as to secure a consistent estimate for the military service effect. A similar exercise was carried out in a joint study jointly with Krueger. The primary aim of the study was to measure the effect of education attainment on earning. Again, the 2SLS procedure was adopted out of the fear that the education variable explaining wage might be correlated with the error term of the wage equation due possibly to OVB. The seasonal information of birthdays was used as the IVs. Since the resulting IV estimates were consistent and therefore argued to be the credible estimates of the education effect (Angrist and Krueger, 1991). These IV experiments led Angrist to further extending, in collaboration with Imbens, a general model framework for measuring 'local average treatment effects' (LATE). In particular, Angrist and Imbens (1991) argued for the sole reliance on the IV method to filter out the potential selection bias from the assumed endogenous dummy variable representing the programme of concern, as against the approach of building a separate latent index of the bias as an additional explanatory variable. In their eyes, correlation of the programme dummy variable with the error term of the regression model formed potentially the fundamental threat, and thus the IV-based estimators offered a simple and general solution. Moreover, they emphasised that the IVs should be selected in such a way that they were uncorrelated with those potentially relevant omitted variables to make the IV-based prescription free of any latent OVB worries. They subsequently justified their LATE approach by interpreting the model result as measuring the causal effect of '*potential outcomes* or *counterfactuals*' of the programme in concern, see Imbens and Angrist (1994).

The 'counterfactual' causal interpretation plus the operational ease of the LATE model has apparently worked wonders finally for popularising the IV method among applied modellers. The method has now become almost routinely applied in programme evaluation models not only in microeconomic studies, but also in development studies with particular reference to measuring the effects of foreign aid projects, e.g. see Angrist and Pischke (2009). Meanwhile, the increasing popularity of treatment models has stimulated more technical research, such as extension of the IV method to panel data models or to a system of equations, e.g. see Wooldridge (1995; 1996), and various tests for weak IVs, e.g. see Stock and Yogo (2005) and Andrews *et al* (2007). These empirical and theoretical developments have been further enlivened by methodological discussions and debates over the 'identification' capacity of the IV-based programme evaluation models, e.g. see Angrist *et al* (1996) and the comments following that paper, Deaton (2010), Heckman (2010) and Imbens (2010), as well as over much more extended issues concerning the capacity of causal inference using micro-econometric models, e.g. see Heckman (2005; 2008) and Chen and Pearl (2012). However, most of this literature is too recent to fit under a historical lens.

5. An anatomy of the IV revival in Micro-econometrics

Let us starts from a regression model of a truncated variable, y_i (assuming the truncation occurs at 0 for simplicity) explained by a set of variables, X_i, in a cross-section data setting:

$$\text{(7)} \quad \begin{array}{ll} y_i = \alpha X_i + u_i & \text{if} \quad y_i > 0 \\ y_i = 0 & \text{otherwise} \end{array}$$

Heckman's two-step procedure is to extend the second equation in (7) via representing the truncation by a binary variable, d_i:

$$\text{(8)} \quad \begin{array}{ll} y_i = \alpha X_i + u_i & y_i > 0 \\ d_i = \gamma Z_i + e_i & d_i \in \{0,1\} \end{array}$$

When the two error terms are assumed to be correlated, $\rho_{ue} \neq 0$, the selection decision from the second equation would affect the outcome, y_i, of the first equation, because:

$$\text{(9)} \quad \begin{array}{l} E(y_i | d_i > 0) = \alpha X_i + E(u_i | -e_i < \gamma Z_i) = \alpha X_i + \beta[f(\gamma Z_i)] \\ \Rightarrow \quad y_i = \alpha X_i + \beta[f(\gamma Z_i)] + \varepsilon_i \end{array}$$

where function $f(\gamma Z_i)$ relates to the conditional expectation of the second equation of (8), e.g. the inverse Mill's ratio derived from a probit regression in the Heckman procedure. Equation (9) shows how a simple truncated regression model (7) can involve an extra regressor due to 'selection bias'. Furthermore, the extra regressor may be correlated with the existing variable set, X_i, especially when $Z_i \supset X_i$. Under such a circumstance, obviously, the OLS estimate of α using the first equation of (8) would suffer from OVB when β turns out to be statistically significant.[7]

Model (8) together with (9) is further developed into an SEM involving a dummy endogenous variable, d_i (see Heckman, 1978):

$$\text{(10)} \qquad \begin{aligned} y_i &= \alpha X_i + \beta d_i + \varepsilon_i \\ d_i &= \gamma \Psi_i + v_i \qquad d_i \in \{0,1\} \end{aligned}$$

where the modelled variable y_i in the first equation is no longer limited to the truncated type. It should be noted, however, that (10) is a pseudo SEM if compared to SEMs in the CC tradition, such as (1), since y_i is not assumed to simultaneously explain d_i in the second equation. Moreover, the second equation is not treated in equal substantive importance as the first one. It is mainly an 'instrumental' equation to justify, by endogenising d_i, the rejection of the OLS estimator for being inconsistent. In other words, the second equation serves effectively as the first stage of the 2SLS procedure and hence Ψ_i is regarded as an IV set. At the second stage, the first equation of (10) becomes:

$$\text{(11)} \qquad y_i = \alpha X_i + \beta d_i^* + \varepsilon_i^* = \alpha X_i + \beta\left(\hat{\gamma}\Psi_i\right) + \varepsilon_i^*$$

Notice, $d_i^* = E\left(d_i \middle| \Psi_i\right)$ must not be the optimal conditional expectation of d_i and therefore is not uniquely determinable, as shown in Section 3. Nevertheless, it implicitly revises the

[7] The substantial linear feature of the inverse Mill's ratio is clearly demonstrated in Puhani (2002). The demonstration shows how much the statistical significance of the ratio derives from the collinearity of this probit regression generated variable with other explanatory variables.

originally postulated equation by conditioning y_i indirectly on Ψ_i via a dummy filter. In other words, the use of the IV method designates (10) as a single-equation structural model involving a latent regressor, just as $f(\gamma Z_i)$ in (9), rather than a structurally postulated SEM. It is thus unnecessary to bring in Haavelmo's argument of the OLS simultaneity 'bias' here. On the other hand, the very assignment of Ψ_i being IVs delimitates the second equation of (10) to be an inadequately, if not incorrectly, specified selection equation from a substantive viewpoint. In other words, the use of IVs undermines fundamentally the credibility of the resulting fitted d_i^* being *the* right representation of sample selection bias in reality. But this logic problem is somehow camouflaged by the substantive-matter devoid choice of IVs in the sense that it is not difficult to find some d_i^* which would produce statistically significant $\tilde{\beta}$ given the prevalence of high inter-correlation among many economic variables. Using Frisch's terminology, what IVs live on are 'confluent relations' instead of 'structural relations', see Qin (2013b). That is why the significance of $\tilde{\beta}$ in practice frequently results from choosing $\Psi_i \supset X_i$, albeit the choice will inevitably give rise to collinearity or multicollinearity between X_i and d_i^*. Ironically, such collinear artefacts have been used commonly as empirical evidence for selection bias, because the bias is *a priori* demonstrated as a special form of OVB and also because it is almost impossible to measure the bias directly from available survey data. In fact, OVB is taken to be equivalent to selection bias in some subsequent literature on programme evaluation modelling.

In models for programme evaluation purposes, the dummy regressor, d_i, is widely used to represent the programme of concern and hence, β in (10) becomes the focal parameter of interest. The possibility of d_i being endogenous grows into such a major concern now that the assumption of d_i as an 'endogenous regressor' forms a hallmark of the programme evaluation models, e.g. see Angrist *et al* (1996). One particular line of its justification is potential OVB because the impact of a social programme is seldom isolated. Suppose that the hypothetical programme evaluation model is:

(12) $y_i = \beta d_i + \lambda q_i + \varepsilon_i$

where q_i represents the other potential factor correlated with d_i. Now, assume that q_i is unobservable directly, as in the case of selection bias. (12) collapses into:

(13) $y_i = bd_i + u_i$

Obviously, the OLS estimator will result in the well-known OVB: $\hat{\beta} - \hat{b} = -\delta_{dq}\lambda$, where $\delta_{dq} = cov(d_i, q_i) / var(q_i)$. But since q_i is unobservable, a direct estimation of (12) is impossible. On the other hand, suppose $b = \beta$ in (13), we have $u_i = \lambda q_i + \varepsilon_i$ according to (12), the correlation problem between d_i and u_i is diagnosed as the culprit, e.g. see (Angrist, 1990). The IV method is thus proposed as the remedy. In particular, a set of IVs, Ψ_i, is chosen which is correlated with d_i but uncorrelated with q_i, e.g. see Imbens and Angrist (1994). (13) now becomes:

(14) $y_i = bE(d_i | \Psi_i) + \upsilon_i \quad \Rightarrow \quad \hat{y}_i = \tilde{b} d_i^*$

Obviously, the IV estimator, $\tilde{b} \neq \hat{b}$, which is argued to be the consistent estimator for β of the hypothetical model (12). In fact, the IV estimator is seen as a universal remedy to orthogonalise d_i against any potential correlation problem of with the error term, u_i.

Let us follow McFadden (1999) and multiple the IVs to (12):

(15) $\Psi'_i y_i = \beta \Psi'_i d_i + \lambda \Psi'_i q_i + \Psi'_i \varepsilon_i$

Because of $cov(\Psi_i, q_i) = 0$, and also $cov(\varepsilon_i, \Psi_i) = 0$, $\tilde{\beta} = (\Psi' d)^{-1} \Psi' y$ is cured of the ‘endogenous regressor’ problem as well as the collinear problem with q_i.

However, a basic logic problem arises: The original OLS $\hat{\beta}$ of (12) should be the desired and unbiased estimate if q_i were observable. Now that $\tilde{\beta} = \tilde{b} \neq \hat{\beta}$, implying that the IV method can never correct the OVB: $\tilde{b} - \hat{b} \neq \hat{\beta} - \hat{b} = -\delta_{dq}\lambda$. In other words, the restrictive condition that $cov(\Psi_i, q_i) = 0$ destines Ψ_i not to contain adequate information to recover the product, $\delta_{dq}\lambda$, i.e. the 'collinear bias' caused by q_i. But the confusion is covered up by the unobservable nature of q_i and further camouflaged by the fear for OLS inconsistency with 'endogenous regressors'.

Meanwhile, the IV-based treatment for selection bias and latent OVB leads to another interpretation via the 'counterfactual' justification. Since the model fitted result is seen as the 'potential' outcome, the corresponding treatment variable should also be potential or intentional rather than the observed. The interpretation amounts to treating the observed d_i as containing substantive measurement errors.[8] Remarkably, the interpretation returns to the very origin of the IV method. It also justifies the first-stage filtering, $d_i^* = E(d_i|\Psi_i)$, not being the optimal, e.g. the 'compliers' only variable in the LATE literature. However, the resulting parameter deviates substantively from the one originally postulated, e.g. $\tilde{\beta} = \tilde{b} \neq \hat{\beta}$ as shown above. That may explain much of the controversies concerning what IV-based models can really deliver in practice, e.g. see Heckman and Urzua (2010). Moreover, the overt recognition of the programme treatment variable being *latent* opens up multiple possibilities in defining this latent variable and thus aggravates the problem of none unique 'identification', albeit offering a fertile ground for IV-based empirical model results. Nevertheless, the idea of using

[8] Note that some IV methods targeting at the heteroscedasticity problem in micro modelling can be regarded as treatment of a type of measurement errors.

the no correlation restriction in the IV selection to isolate the parameter of interest from any potential OVB may lead to the difficulty of having only weak IVs.

6. Retrospective Comments

We have traced the history of how the IV method has gone out of favour among applied macro economists but has subsequently restored its popularity in micro-econometric studies and further extended its territory into development economics. The historical examination has enabled us to unravel much of the conceptual confusion over the application of the IV method and draw the following lessons.

Primarily, the IV estimator achieves the 'consistency' criterion by essentially redefining the originally postulated conditional variable of concern – substituting it by a conditional expectation of the variable. Moreover, this conditional expectation must NOT be the statistically optimal predictor and hence not required to be based on a substantively causal relationship. As such, there exists a multitude of such suboptimal conditional expectations. Application of the IV method to an SEM effectively breaks the contemporaneously circular causal relationship between the explained variable and the conditional variable in the model and forms an asymmetric causal chain with the IVs as the initial 'exogenous' drivers, which cast their impact on the explained variable solely via the suboptimal conditional expectation of the original conditional variable. It is thus not so surprising for us to find, from the history, that no serious cases of empirical successes with IV-based macro-econometric SEMs, because they are ultimately driven by sets of 'instruments' without substantively serious 'causal' grounds.

From the viewpoint of the parameter of interest concerned, the consistency of its IV estimate comes at a price of changing its 'master', i.e. the variable from which the parameter derives its interest. The estimate no longer measures the impact of the conditional expectation of the original conditional variable, but that of a suboptimal conditional expectation of that variable. However, this inadvertent swap of masters has been ignored by the profession at large since the basic statistical theory that the probability foundation of a regression is conditional expectation is far from a unanimously understood concept. That is the most evident from ‘endogenous regressor’, a term widely used in textbooks and lecture notes. Note how the ‘endogenous’ attribute rejects the direct use of a regression model as the valid conditional expectation! When there is serious doubt over the empirical appropriateness of an *a priori* postulated model, it makes little sense to go first and foremost for estimators *consistent* with that model. Now, we can better understand why statisticians are far less interested in the IV method than economists and econometricians, and also why Wold’s repeated arguments over half a century ago have been ignored and almost forgotten.

Once we explicitly equate regression models with conditional expectations, much of the conceptual confusion over IVs can be deciphered. The very suboptimal characteristic of the IV-based conditional expectations determines the method applicable only for situations where the observed explanatory variables are measured with non-negligible errors, since the essence of the method is to reject the valid conditioning on those observed variables. Modellers should therefore avoid using the IVs when they have no substantive ground or evidence to doubt the valid conditioning of their selected explanatory variables. We can now see why the IV prescription is virtually abandoned in macro-econometric modelling research, since accruing evidence on model weaknesses there points decidedly towards inadequately specified dynamic conditional expectations rather than incorrectly selected conditional variables. Here, it is important to note that a crucial drive for the development of explicitly formulated dynamic econometric models is to raise their forecasting accuracy and that the drive mirrors into explicit specification searches for multiple regression models whose error terms should exhibit the innovation properties. In other words, the error terms are treated explicitly as model-derived residuals without any a priori assumed autonomous status, e.g. see Qin and Gilbert (2001). As a result, the essential task of the modellers falls on the choice of data-congruent regression models, since there is no need for going for estimators consistent to initially untested model specifications and the OLS estimator works thanks to Wold's proximity theorem.

In the case of micro-econometric studies using mostly cross-section survey data, forecasting is seldom on the agenda or considered relevant. The substantive matters are mainly related to measuring the effects of one or a few postulated conditioning variables pertinent to certain policy issues on a particular variable, such as wage. Although wage is known to be affected by a multitude of factors other than those policy related ones, the modeller has no substantive interest in estimating the effects of those other factors. Hence, such studies are inevitably biased towards a very partial use of the data evidence on the modelled variable, especially when the survey samples are large and designed for multiple purposes, such as the US 'Panel Study of Income Dynamics'. Consequently, the innovation properties of the error terms are not required, let alone being used as an essential prerequisite for model selection. In fact, the essential model selection criterion commonly used in micro-econometrics is simply that the estimated parameters of interest are both statistically significant and substantively interpretable. That explains why the choice of estimators consistent with the *a priori* model specification remains a key task of micro modellers. Since micro models are mostly composed of static structural relations, concerns over Haavelmo's OLS simultaneity 'bias' have almost grown into a widely spread paranoia, as reflected from the popular IV prescription to guard against 'endogenous regressors'. Furthermore, microeconometric evidence has mostly been used for the purpose of policy debates. Given the reality that almost no policies can have their full and detailed causal effects identified, let alone measured accurately, there is relatively little incentive for modellers to try and raise the precision of their estimates or examine the invariance property of the estimates beyond data samples. Most of these samples are not regularly updated anyway. Elegantly constructed impressionist stories based on internally consistent theoretical derivations are far more powerful in political persuasion than painstaking realist reports on the statistical findings pertaining to particular samples. In other words, the practical value of microeconometric evidence lies mainly in statistical illustrations and hence its role is firmly subordinate to the maintenance of *a priori* theoretical postulates. Here, the non-uniqueness of IVs comes in handy for enhancing such maintenance.

We have shown that most of textbook micro models are not symmetric SEMs on which Haavelmo's 1943 paper was based. It is thus misleading to force a simultaneity interpretation on top of the issues of serious concern for empirical micro modellers, i.e. issues such as measurement errors, latent variables or omitted variables, and the related collinearity due to latent OVB. In the case of OVB, in particular, it is a pure fallacy to prescribe the problem as caused by endogeneity of the un-omitted regressor since its conditional status is never in doubt regardless of whether the regression includes or excludes those prognosticated omitted variables. In other words, correlation of the error term with the regressors of a single-equation based regression is merely a phantom because the error term is derivative, rather than autonomous, of the regression. Uses of the IV method in such a situation frequently lead to logically dubious results. It is no wonder that many applied modellers find it impossible to find 'strong' IVs to generate robust and conclusive results. When the method is used to generate a latent omitted variable, e.g. the inverse Mill's ratio, it is difficult to give much credit to the IV-generated variable since the valid choice of IVs denies similarity between the fitted variable and the intended latent variable. When the method allegedly corrects OVB, all it actually does is to alter both the originally postulated parameter of interest and its 'master' by swapping the variable of interest with a suboptimal conditional expectation of it, in the hope that such an expectation would be immune to the prognosticated OVB. It is not surprising that the method gives rise to difficulties and ambiguities when it comes to causal interpretations. More fundamentally, there cannot be a universal orthogonal treatment by IVs for any un-specified, correlated omitted variables. It is unsurprising that many tried IV treatments are found to be weak when the uncontrolled modelled variable is known to be susceptive to a multitude of unspecified co-varying factors.

Ironically, the very fact that the modelled variable is frequently susceptive to multiple unspecified co-varying factors provides a fertile ground for the production of statistically significant IV estimates. The non-uniqueness of IV choices facilitates modellers to exploit the other facet of OVB – collinearity or multicollinearity. For example, it is not difficult to find an IV-based latent variable which becomes significant in the second stage of the 2SLS as long as the chosen IVs in the first stage are correlated with some of the 'un-omitted' regressors, as shown in the previous section. It was forty years ago when Leamer showed that collinearity is essentially a 'problem of interpreting multidimensional evidence' in a 'parameter-by-parameter fashion' (1973). The problem was subsequently pinned down to one of parameter design and circumvented effectively through explicit model reparameterisation by the LSE approach, see Davidson *et al* (1978) and also (Qin, 2013a, Chapter 7). Sadly, these lessons have been totally neglected in the recent promotion of the IV method as a universal remedy for OVB. The 'schizophrenic' mentality of interpreting OVB separately from collinearity (Farrar and Glauber, 1967) is still prevalent. Nevertheless, our analysis helps to explain why such an IV-based prescription has caused so much controversy concerning programme evaluation models, since it would require a set of miraculous IVs to enable the reduction of any multidimensional evidence into a single and interpretable parameter.

Finally, our investigation shows just how inefficient, if not counterproductive, it is to treat particular and often disparate empirical model design problems as a general problem of estimator choice or to choose estimators before *a priori* postulated models have been rigorously tested. Such a textbook approach often yields little substantive gains but piles of mathematical proofs glossing over dubious assumptions such as 'endogenous regressors', instead of a straightforward '1+1=2' statement (Siegfried, 1970). It should also be noted that conceptual muddles of such a kind are by no means new in the short history of econometrics. Hopefully, the present IV story can encourage more from the profession to take the history much more seriously than quoting whatever comes in handy merely for self-justification or persuasion.

References

Addison, J.T. and P. Portugal (1989) Job displacement, relative wage changes, and duration of unemployment, *Journal of Labor Economics*, **7**, 281-302.

Aldrich, J. (1993) Reiersøl, Geary and the idea of instrumental variables, *The Economic and Social Review*, **24**, 247-73.

Amemiya, T. (1974), Multivariate regression and simultaneous equation models when the dependent variables are truncated normal, *Econometrica*, **42**, 999-1012.

Amemiya, T. (1976) The maximum likelihood, the minimum chi-square, and the nonlinear weighted least-squares estimator in the general qualitative response model, *Journal of the American Statistical Association*, **71**, 347-51.

Anderson T.W. and H. Rubin (1949) Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics* **20**, 46-63.

Andrews, D.W.K., M. J. Moreira, and J.H. Stock (2007) Performance of conditional Wald tests in IV regression with weak instruments, *Journal of Econometrics*, **139**, 116-32.

Angrist, J. (1990) Lifetime earnings and the Vietnam era draft lottery: Evidence from social security administrative records. *American Economic Review*, **80**, 313-36.

Angrist, J. and G. Imbens (1991) Sources of identifying information in evaluation models, *NBER Technical Working Papers* 0117, National Bureau of Economic Research.

Angrist, J., G. Imbens and D. Rubin (1996), Identification of causal effects using instrumental variables, *Journal of the American Statistical Association*, **91**, 444-55.

Angrist, J. and A. Krueger (1991) Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, **106**, 979-1014.

Angrist J. D. and J. Pischke (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Banerjee, A.V. and E. Duflo (2009) The experimental approach to development economics, *Annual Review of Economics*, **1**, 151-78.

Bowden, R and D.A. Turkington (1990) *Instrumental Variables*, Cambridge University Press.

Cameron, A.C. (2009) Microeconometrics: Current methods and some recent developments, in Patterson, K. and Mills, T. C. (eds.) *Palgrave Handbook of Econometrics*, vol. 2. Houndmills: Palgrave MacMillan, pp. 729-74.

Chen, B. and J. Pearl (2012) Regression and Causation: A critical examination of econometrics textbooks, *UCLA Cognitive Systems Laboratory, Technical Report* (R-395).

Christ, C. F. (1966) *Econometric Models and Methods*, New York: John Wiley & Sons.

Davidson, J.E.H., Hendry, D.F., Srba, F. and Yeo, S. (1978) Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom, *Economic Journal*, **88**, 661-92.

Deaton, A. (2010) Instruments, randomization, and learning about development, *Journal of Economic Literature*, **48**, 424-55.

Duflo, E., R. Glennerster and M. Kremer (2008) Using randomization in development economics research: A toolkit, in T. P. Schultz and J. A. Strauss (eds.), *Handbook of Development Economics*, Elsevier, Chapter 61.

Duncan, G. M. and D. E. Leigh (1985) The endogeneity of union status: An empirical test, *Journal of Labor Economics*, **3**, 385-402.

Durbin, J. (1954) Errors in variables, *Review of the International Statistical Institute* **22**, 23-32.

Ericsson, N.R. and D.F. Hendry (2004) The ET Interview: Professor David F. Hendry, *Econometric Theory*, **20**, 745-806.

Farrar, D.E. and R.R. Glauber (1967) Multicollinearity in regression analysis: the problem revisited, *The Review of Economics and Statistics*, **49**, 92-107.

Fisher, F. M. (1965) Dynamic structure and estimation in economy-wide econometric models, in Duesenberry, J.S., G. Fromm, L.R. Klein and E. Kuh (eds.) *The Brookings Quarterly Econometric Model of the United States*. Chicago: Rand McNally, pp. 589-635.

Gilbert, C.L. (1989) LSE and the British approach to time series econometrics. *Oxford Economic Papers*, **41**, 108-28.

Haavelmo, T. (1943) The statistical implications of a system of simultaneous equations, *Econometrica*, **11**, 1–12.

Hausman, J. (1978) Specification tests in econometrics, *Econometrica*, **46**, 1251-71.

Heckman, J. (1974) Effects of child-care programs on women's work effort, *Journal of Political Economy*, **82**, S136-S163.

Heckman, J. (1976) A life-cycle model of earnings, learning, and consumption, *Journal of Political Economy*, **84**, S11-44.

Heckman, J. (1978) Dummy endogenous variables in a simultaneous equation system, *Econometrica*, **46**, 931-59.

Heckman, J. (1979) Sample selection bias as a specification error, *Econometrica* **47**, 153-61.

Heckman, J. (1996) Comment (to Angrist *et al*, 1996), *Journal of the American Statistical Association*, **91**, 459-62.

Heckman, J. (2005) The scientific model of causality, *Sociological Methodology*, **35**, 1-97.

Heckman, J. (2008) Econometric causality, *International Statistical Review*, **76**, 1-27.

Heckman, J. (2010) Building bridges between structural and program evaluation approaches to evaluating policy, *Journal of Economic Literature*, **48**, 356-98.

Heckman, J. and V.J. Hotz (1989) Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training, *Journal of the American Statistical Association*, **84**, 862-74.

Heckman, J. and S, Urzua (2010) Comparing IV with structural models: What simple IV can and cannot identify, *Journal of Econometrics*, **156**, 27-37.

Hendry, D.F. (1995) *Dynamic Econometrics,* Oxford University Press.

Hendry, D.F. (2003) J. Denis Sargan and the origins of LSE econometric methodology, *Econometric Theory,* **19**, 457-80.

Hendry, D.F. and J-F. Richard (1982) On the formulation of empirical models in dynamic econometrics, *Journal of Econometrics*, **20**, 3-33.

Imbens, G. (2010) Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009), *Journal of Econometrics*, **156**, 399-423.

Imbens, G. and J. Angrist (1994) Identification and estimation of local average treatment effects, *Econometric*a, **62**, 467-475.

Klein, L. R. (1955) On the interpretation of Theil's method of estimation of economic relations, *Metroeconomica,* **7**, 147–53.

Klein, L.R. and R.J. Ball (1959) Some econometric of the determination of absolute wages and prices, *Economic Journal*, **69**, 465-82.

Leamer, E.E. (1973) Multicollinearity: a Bayesian interpretation, *Review of Economics and Statistics*, **55**, 371-80.

Lee, L-F. (1978) Unionism and wage rates: A simultaneous equations model with qualitative and limited dependent variables, *International Economic Review*, **19**, 415-33.

Lee, L-F., G. Maddala and R. Trost (1980) Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity, *Econometrica*, 48, 491-503.

Maddala, G. S. (1983) *Limited-Dependent and Qualitative Variables in Economics*, New York: Cambridge University Press.

McFadden, D. (1999) Chapter 4. Instrument Variables, lecture notes, Econ. 240B, at: http://emlab.berkeley.edu/users/mcfadden/e240b_f01/ch4.pdf .

Moffitt, R.A. (1996) Comment (to Angrist *et al*, 1996), *Journal of the American Statistical Association*, **91**, 462-65.

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

Nelson, F. and L. Olson (1978) Specification and estimation of a simultaneous equation model with limited dependent variables, *International Economic Review*, **19**, 695-709.

Pearl J. (2009) *Causality: Models, Reasoning and Inference*, Cambridge University Press.

Puhani, P.A. (2002) The Heckman correction for sample selection and its critique. *Journal of Economic Survey* **14**, 53-68.

Qin, D. (1993) *Formation of Econometrics: A historical perspective*, Oxford University Press.

Qin, D. (2013a) *A History of Econometrics: The Reformation from the 1970s*, Oxford University Press.

Qin, D. (2013b) Inextricability of confluence and autonomy in econometrics, draft paper for *Oeconomia* forthcoming special issue on the post Cowles Econometrics.

Qin, D. and C.L. Gilbert (2001) The error term in the history of time series econometrics, *Econometric Theory*, **17**, 424-450.

Reiersøl, O. (1945) Confluence analysis by means of instrumental sets of variables, *Arkiv for Matematik, Astronomi och Fysik* , **32**A, 1-119.

Sargan, J.D. (1958) The estimation of economic relationships using instrumental variables, *Econometrica*, **26**, 393-415.

Sargan, J.D. (1959) The estimation of relationships with autocorrelated residuals by the use of the instrumental variables, *Journal of the Royal Statistical Society*, Series B, **21**, 91-105.

Sargan, J.D. (1964) Wages and prices in the United Kingdom: a study in econometric methodology, in R.E. Hart, G. Mills and J.K. Whittaker, eds., *Econometric Analysis for National Economic Planning*, Butterworth, London, 25-63.

Siegfried, J. (1970) A first lesson in econometrics, *Journal of Political Economy* **78**, 1378-79.

Sims, C.A. (1980) Macroeconomics and reality, *Econometrica*, **48**, 1-48.

Stern, S. (1989), Measuring the effect of disability on labor force participation, *The Journal of Human Resources*, **24**, 361-95.

Stock, J. and M. Yogo (2005) Testing for weak instruments in linear IV regression, in D. W. K. Andrews, and J. H. Stock eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge University Press, pp. 80–108.

Stock, J.H. and F. Trebbi (2003) Retrospectives: Who invented instrumental variable regression? *Journal of Economic Perspectives*, **17**, 177-94.

Theil, H. (1953), Estimation and simultaneous correlation in complete equation systems, The Hague: Centraal Planbureau, (Mimeographed).

Willassen, Y. and O. Reiersøl (2000) The ET Interview: Reiersøl Olav, *Econometric Theory*, **16**, 113-25.

Wold, H.O.A. (1954) Causality and econometrics, *Econometrica*, **2**2, 162-77.

Wold, H.O.A. (1956) Causal inference from observational data: A review of ends and means, *Journal of Royal Statistical Society Series* A, **119**, 28-61.

Wold, H.O.A. (1960) A generalization of causal chain models, *Econometrica*, **28**, 443-63.

Wold, H.O.A. (1961) Construction principles of simultaneous equations models in econometrics, *Bulletin of the Institute of International Statistics*, **38**, 111-36.

Wold, H.O.A. (1965) A letter report to Professor P. C. Mahalanobis, in C.R. Rao ed., *Essays on Econometrics and Planning*, Pergamon Press, Oxford, pp. 309-20.

Wold, H. and P. Faxér (1957) On the specification error in regression analysis, *Annals of Mathematical Statistics*, **28**, 265-67.

Wold, H.O.A. and L. Juréen (1953) *Demand Analysis: A Study in Econometrics*, Wiley and Sons, New York.

Wooldridge, J. M. (1995) Selection corrections for panel data models under conditional mean independence assumptions, *Journal of Econometrics*, **68**, 115-32.

Wooldridge, J. M. (1996) Estimating systems of equations with different instruments for different equations, *Journal of Econometrics*, **74**, 387-405.

Wooldridge, J. M. (2010) *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, second edition.

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