

Resurgence of the Endogeneity-Backed Instrumental Variable Methods

Duo Qin

Abstract

This paper investigates the nature of the IV method for tackling endogeneity. By tracing the rise and fall of the method in macroeconometrics and its subsequent revival in microeconometrics, it pins the method down to an implicit model respecification device—breaking the circular causality of simultaneous relations by redefining it as an asymmetric one conditioning on a non-optimal conditional expectation of the assumed endogenous explanatory variable, thus rejecting that variable as a valid conditional variable. The revealed nature explains why the IV route is popular for models where endogeneity is superfluous whereas measurement errors are of the key concern.

(Published as [Survey and Overview](#))

JEL B23 C13 C18 C50

Keywords Endogeneity; instrumental variables; simultaneity; omitted variable bias; multicollinearity

Authors

Duo Qin, ✉ Department of Economics, SOAS, University of London, UK, dq1@soas.ac.uk

Citation Duo Qin (2015). Resurgence of the Endogeneity-Backed Instrumental Variable Methods. *Economics: The Open-Access, Open-Assessment E-Journal*, 9 (2015-7): 1—35. <http://dx.doi.org/10.5018/economics-ejournal.ja.2015-7>

1 Introduction

Endogeneity forms almost a routine concern among empirical modellers bought up in the textbook tradition of simultaneous-equation models (SEM). A commonly taught remedy for endogeneity is the instrumental variable (IV) estimation method. Although devised over half century ago, the IV remedy has not been widely practised until recent years, accompanied by the rise to prominence models for evaluating social programmes (known also as treatment models) in microeconometrics, e.g. see Cameron (2009), and also in economic development studies, e.g. see Duflo et al. (2008). The popularity of these models has stimulated heated methodological debates over the causal validity and credibility of IV-based estimates of treatment effects, e.g. see Angrist et al. (1996), Heckman (1996; 2010), Deaton (2010) and Imbens (2010; 2014).

In the comments on Angrist et al. (1996), the following statement is particularly intriguing: ‘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). Indeed, while statisticians take as a basic prerequisite that models should specify the order of variables in ‘the essential asymmetry between cause and effect’ (Cox, 1992, p. 293), economists tend to regard asymmetric models as the ‘reduced-form’ of SEMs, where simultaneous symmetry is believed to be *the* essential representation of interdependence between economic variables. The belief is so deeply rooted that endogeneity remains a must on the checking list of applied modellers even when *a priori* economic models of interest involve only single equations.

Since many empirical studies are based on single-equation theoretical models, the IV route has been widely taken as a handy treatment of possible endogeneity in such models. 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 SEMs (Angrist and Pischke, 2009, p. 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)’ (Angrist and Pischke, 2009, pp. 114–115).

The above statement has triggered the present investigation – to track down what has led to such a shift of course. Angrist and Pischke refer to IV as ‘lingo’ (Angrist and Pischke, 2009, p.126). The primary aim of this investigation is to decipher the lingo through the changing fortunes of the IVs in tackling endogeneity and to illuminate a path through the methodological labyrinth over the valid use of IVs for applied modellers. It should be noted here that the IV method has been extended into the generalised method of moments (GMM) since Hansen’s 1982 paper. The present investigation choose to exclude, however, discussions on the GMM procedure because that technical extension neither alters nor extends the fundamental logic of the IV route in tackling endogeneity. The next section provides a historical summary of the rise and fall of the IV method in macroeconometric studies from the 1940s to the 1970s–1980s (Section 2.1) and the subsequent revitalisation of the IV route in microeconomic models from the late 1970s up to the 1990s (Section 2.2). Section 3 reviews the history by methodologically dissecting the basic models of common concern in macro and micro studies respectively, and analysing key logical ideas in applying IVs. Mathematical illustrations are kept at an as elementary as possible level so as to make the arguments easily comprehensible for applied economists whose routine econometric toolkits are built around regression models.

The dissection yields the following findings: (i) the IV method implicitly redefines the contemporaneously circular causality of simultaneous relations into an asymmetric one by means of substituting the assumed *endogenous* explanatory variable with a non-optimal conditional expectation of it, thus rejecting that variable as a valid conditional variable; (ii) the IV route falls out of favour among macro modellers because of lack of evidence against the validity of *a priori*

postulated conditional variables; (iii) endogeneity is merely a multiple-equation model assumption, often a superfluous one and definitely not the only one to represent interdependence among economic variables, as shown from the development of dynamic macroeconometric models; (iv) the IV route earns its popularity in microeconometrics by its capacity to enhance the success rate of getting statistically significant estimates for the parameter of interest through allowing, under the disguise of SEM consistent estimators, multiple ways of substituting the explanatory variable of interest by non-optimal conditional expectations of it, thanks to prevalent multicollinearity; however, (v) those IV-modified conditional variables give rise to an insurmountable difficulty – finding credible interpretation for those IV estimates, especially in the case of prognosticated OVB, the main concern in micro and development studies; and (vi) the IV route is only logically valid for models where endogeneity is superfluous and the explanatory variables of interest are latent and their observable proxies suffer from measurement errors.

The above findings raise a further question: What has led so many econometric modellers to remain with the endogeneity fallacy and to keep themselves busy generating, out of multicollinearity, endless IV-based non-optimal conditional expectations? What has prevented them from realizing what they are doing is to falsify, inadvertently, both the assumed simultaneous relationships and the conditioning power of explanatory variables of substantive interest? The concluding section seeks answers by looking back at the history again. This retrospective discussion highlights, for applied modellers, the importance of (a) raising general awareness of the conditional expectation foundation of regression models, (b) forsaking the error-term based estimator treatment route to become more proactive in empirical model design, and (c) paying greater attention to the basic purposes that econometric models are built and used for in practice.

2 History: Fifty Years of IVs in Association with SEMs

2.1 Rise and Fall of IVs in Macroeconometric Modelling

The term ‘instrumental variables’ is commonly acknowledged to come from Reiersøl’s thesis (1945) ‘Conference Analysis by Means of Instrumental Sets of

Variables', see Morgan (1990, Section 7.3). The idea of using IVs in econometrics was 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 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 errors-in-variable models. However, it was the SEM in the error-in-equation form which formed the key model of interest at the time. In fact, SEMs formed the basic model type in the formalisation of econometrics by the Cowles Commission (CC) during the 1940s, e.g. see (Bowden and Turkington, 1990, Section 1.3) and also (Qin, 1993, Chapter 3).

Arguably one of the most influential technical advances by the CC was the limited-information maximum likelihood (LIML) estimator (see Anderson and Rubin, 1949). 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 in 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 computationally convenient estimator for the 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 UK wage and price model built by Klein and Ball (1959), 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). He left aside his own IV device half way through the trial and moved on to conduct an extensive dynamic model specification search to try and improve the Klein-Ball model, aimed mainly by the OLS. It was 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 (2013, Chapter 4).

In fact, Sargan's (1964) empirical modelling success assisted simply by the OLS had been anticipated by H. Wold's 'proximity theorem' over a decade before. The theorem demonstrated that practically the bias resulting from inconsistency of the OLS estimator in an SEM should remain 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 bias would dwindle with the size of the variance of the residuals, see Wold and Juréen (1953, pp. 37–38) and also Wold and Faxér (1957). Wold remained a staunch proponent of the OLS method from the time of his first book on time-series analysis (1938). Interestingly, according to Reiersøl, Wold was 'the first opponent' of his IV method (Willassen and Reiersøl, 2000, p.118). However, Wold's disapproval of the SEM approach developed by Haavelmo and the CC group went beyond the choice of estimators. 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). He also emphasised that the choice between a causal chain model and an SEM was 'not a matter of estimation technique' (1965).

As was the case with Sargan's 1964 work, Wold's viewpoints were largely overshadowed by the Haavelmo-CC SEM approach during its consolidation period, see Qin (2014a). 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 (Fisher, 1965, p. 633). 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' (Fisher, 1965, p. 590). The latter message was clearly contrary to that taught in textbooks and put the IV method closer to a model re-specification device than an estimation method. 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 macroeconometrics led by the reformative movement of the VAR (Vector Auto-Regression) and the LSE approaches, e.g. see Qin (2013, Chapters 3 and 4). A key drive for the movement was the failures of conventionally built macroeconomic models in forecasting the turbulent economic recessions in the wake of the 1973 oil crisis. The failures were diagnosed to be mainly caused by incredible over-identification restrictions commonly used in largely static SEMs by Sims (1980), a diagnosis which effectively rejected the IV route.¹ Sims demonstrated, initially jointly with Sargent, how VAR models could do away with those restrictions and improve model forecasts substantially, e.g. see Sargent and Sims (1977) and Sims (1980). The evident improvement helped to secure the status of VARs in macroeconomics. 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 forecasting advantage of VARs over 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 and essentially static relations. What was shown in Wold's proximity theorem has finally been widely accepted even though the theorem itself has been forgotten.

Contemporaneous simultaneity was absent in the initial VAR specification. Instead, it was implicit in the covariance matrix of the VAR residual terms. Such simultaneity was restored in structural VARs subsequently, which were developed mainly to appease the strong SEM conviction in the profession. Nevertheless, the FIML, instead of IV estimators, was usually applied for structural VARs, indicating a common faith in both the symmetric model formulation and the choice

¹ I owe this interpretation to Ron Smith.

of conditional variables in individual equations. In fact, most empirical VAR modellers paid far less attention to individual parameter estimates than to the overall model performance in the form of shock-based impulse analyses. However, such analyses were predicated on imposing certain restrictions, e.g. orthogonal restrictions, on the covariance matrix of the error terms such that these terms would form a recursive scheme. The imposition followed *de facto* the principle of Wold's causal ordering, e.g. see Sims (1980), and consolidated VARs into asymmetrically specified cause-effect models.

The LSE approach shared with the VAR approach on setting statistically innovation error terms as a primary modelling criterion. But unlike the latter, the LSE approach put great emphasis on single-equation based conditional models, which were seen as reduced from the joint distribution of all the variables concerned, a distribution underlying the SEM, e.g. see Hendry and Richard (1982). The approach effectively extended Wold's ideas of causal chain models with white-noise residuals. As mentioned earlier, the LSE approach largely grew from Sargan's 1964 paper. One important turning points of its initial growth was a shift of 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 naming exhibited a close association to Sargan's IV research and also Hendry's early interest in comparing various estimators for SEMs, see e.g. see Ericsson and Hendry (2004). The programme was, however, 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 (Ericsson and Hendry, 2004, p.752). To a certain extent, the history of GIVE through its subsequent versions known as PcGIVE repeated Sargan's 1964 trial – the central role of IV estimators being soon replaced by that of dynamic specification searches. In fact, a switch from the estimator-centric strategy to active dynamic model design marked the rise of the LSE approach, e.g. see Qin (2013, Chapter 4).

The LSE approach and the VAR approach have won over the majority of applied macroeconomists. In so doing, concerns over endogeneity of static SEMs have drastically dwindled. Unsurprisingly, empirical studies using IV-based estimators to tackle *a priori* fixed parameters pertinent to assumed endogenous explanatory variables have faded gradually.

2.2 Revitalization of IV Methods in Microeconometrics

Just as endogeneity and static SEMs were overshadowed by the development of dynamic macroeconometrics during the 1970s, a wave of concern arose over possible endogeneity in regression models explaining limited dependent variables (LDV), i.e. variables of the truncated or dichotomous type. Such models grew from microeconomic studies using household surveys of the cross-section data type.²

One of the leading research fields at the time was labour economics. A well-known research topic was female labour participation decisions. From the stance of textbook microeconomics, such decisions should be based on the interdependence between wage rate and the choice to work (a binary or dummy variable) and/or the number of work hours (a truncated variable), whereas empirical studies of this interdependence was hindered by one data difficulty – lack of information on the offered wage rates for those who did not work. The difficulty gave rise to the concern over selection bias, i.e. the estimated wage effects using the observed wage information might be biased with respect to the entire working population.

The Tobit method was applied to early labour supply models. An alternative two-step estimation procedure was devised by Heckman to circumvent the selection-bias issue, which he viewed as a special type of OVB, during the 1970s (e.g. see Heckman, 1974; 1976). In a subsequent paper, Heckman generalised the early labour supply models into an SEM involving an endogenous dummy variable (1978).³ Here, the dummy variable equation in the SEM was given the 2SLS interpretation of the IV method. In other words, the purpose of this equation was to filter out the assumed endogeneity by a set of IVs rather than seriously explaining the dummy variable *per se* (see the next section for a more detailed description).

² Although this paper does not cover the earlier history of microeconometrics, it is interesting to cite R. Stone, who was engaged in extensive research in household expenditure data during the 1950s, on his recollection about IVs: ‘I did experiment with instrumental variables but I never had any luck with them and they never appeared in my published works’ (Stone and Pesaran, 1991, p. 103) (I thank Ron Smith for providing me with this historical fact).

³ 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.

Heckman's exploration was strengthened, around the same time, by similar studies on 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 microeconometrics.

Empirical findings of the subsequent decade or so, however, presented weak and inconclusive support for the IV (or 2SLS-based) treatment of endogeneity 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 less than five cases. In comparison, there are far more cases applying Heckman's two-step procedure for the treatment of selection bias but not simultaneity 'bias'.⁴ Of the few which use *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 generally statistically insignificant, e.g. see Addison and Portugal (1989). These findings corroborate what has already been found by macro modellers.

It was not until the early 1990s that the IV route finally gained significantly growing popularity in empirical studies. The boom was associated with models designed for evaluating the effects of social programmes from individual based cross-section data. The objective of those models oriented the econometricians' attention towards measurement issues pertinent to a single parameter, the parameter of a dummy (binary) explanatory variable representing the specific programme of prior interest. Under the influence of Robin's treatment model (1989) taken from biometric research, randomisation of treatments was seen as a key prerequisite. But it was a prerequisite which social programmes could seldom satisfy. Hence, concerns over selection bias came to the fore. Clearly, observed social outcomes could be due either to the programme of interest or to other related factors when the programme treatment had not been randomly assigned nor tightly controlled. Inadequate consideration of those factors could result in OVB in

⁴ 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 the SEM specification rather than possible selection bias, which was effectively an OVB, see Heckman (1979).

the estimated treatment effect. This case was referred to as ‘selection on observables’ following Barnow et al. (1981). It was also likely that factors causing selection bias were unobservable, i.e. the case of selection on unobservables. Remedies to the latter case led to the revival of Heckman’s SEM involving an endogenous treatment dummy variable, e.g. see Heckman and Hotz (1989). Consequently, the IV route became seen as an ideal and general approach in tackling the problem of an ‘endogenous’ treatment variable, 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 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 regarded 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 for the IV method as a way to randomise the sample of veterans 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. 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 extend, in collaboration with Imbens, a general framework for measuring ‘local average treatment effects’ (LATE). In particular, Angrist and Imbens (1991) argued for the IV route 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 the fundamental threat, and thus the IV route offered a simple and general solution. In particular, if IVs could be selected in such a way that they were uncorrelated with those potentially relevant omitted variables, it would free modellers from worries over any latent and unobservable OVB when relatively simple programme

evaluation models were specified. Subsequently, they justified the 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 worked wonders for popularising the endogeneity-OVB-driven IV route among applied modellers. IV estimators have now become almost routinely applied 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 programme evaluation models has stimulated more technical research, such as extension of the IV route to panel data models or to a system of equations, e.g. see Wooldridge (1995; 1996).

Accruing empirical results, however, have attracted concerns over the substantive credibility of the IV route. Essentially, the concerns are focused on the ‘identification’ capacity of IV estimators with respect to particular treatment effects of interest, e.g. see Angrist et al. (1996) and the comments following that paper, Manski (1993; 2000), Deaton (2010), Heckman (2010) and Imbens (2010). The scepticism expressed in these works is reminiscent of Sims’s critique of the ‘incredible identification’ of traditional SEMs in macroeconometrics (1980, p.1). Meanwhile, the practical drawback of the frequent lack of substantively convincing and statistically powerful IV estimators has been increasingly recognised, e.g. see Blundell and Costa Dias (2009), Khandker et al. (2010, Chapter 6). Although econometricians have been responsive in devising various tests for weak IVs, e.g. see Stock and Yogo (2005), Andrews et al. (2007), the drawback runs much deeper than can be rescued by these tests. It touches fundamentally on the limited capacity of causal inference using the estimation-based remedy for assumed SEMs in microeconometrics, e.g. see Heckman (2005; 2008) and Chen and Pearl (2012). However, much of this literature is too recent to fit properly under a historical lens.

3 Anatomy of the IV Route for Endogeneity

3.1 3.1. Fundamental Role of IVs for SEMs in Macroeconometrics

Let us start the anatomy from the basic textbook exposition of the IV methods. Write the simplest identifiable two-equation SEM as:

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

where z_{it} are assumed exogenous. It is well-known, due originally to Haavelmo (1943), that single-equation OLS estimates, $\hat{\alpha}_1$, are inconsistent with model (1) because of $cov(x_t, u_{1t}) \neq 0$ and $cov(y_t, u_{2t}) \neq 0$. These correlations are the result of the assumed simultaneous relationship or circular causality of $y_t \leftrightarrow x_t$. Here, x_t and y_t are referred to as endogenous variables.

The IV treatment of the OLS inconsistency is best demonstrated via the 2SLS. It suffices here to look only at the 2SLS representation of the first equation in (1):

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

where Ψ denotes a set of IVs which are significantly correlated with x_t and not directly part of the explanatory variables to y_t in the second equation in (2).⁵

Empirical validation of the IV estimate, $\tilde{\alpha}_1$, over the OLS estimate, $\hat{\alpha}_1$, is commonly based on the Durbin-Wu-Hausman endogeneity test, e.g. see Hausman (1978). Endogeneity confirmed by the test depends essentially on the IV equation of the first stage of the 2SLS *not* giving a very good fit of x_t so as to sustain $\tilde{\alpha}_1$ being significantly different from $\hat{\alpha}_1$ in (1). Obviously, the closer x_t^* resembles to x_t the smaller the difference between $\tilde{\alpha}_1$ and $\hat{\alpha}_1$. This simple fact rules out the IV equation being the optimal predictor of x_t , thus denying the existence of a

⁵ This is used here for simplicity; the normally required exclusion condition is weaker than the present assumption.

uniquely ‘best’ x_t^* . In the second stage, the replacement of x_t by a non-optimally generated regressor, x_t^* , effectively redefines the postulated causal relationship $x_t \rightarrow y_t$ by $x_t^* \rightarrow y_t$. Obviously, the redefinition modifies and breaks $y_t \leftrightarrow x_t$ of SEM (1).

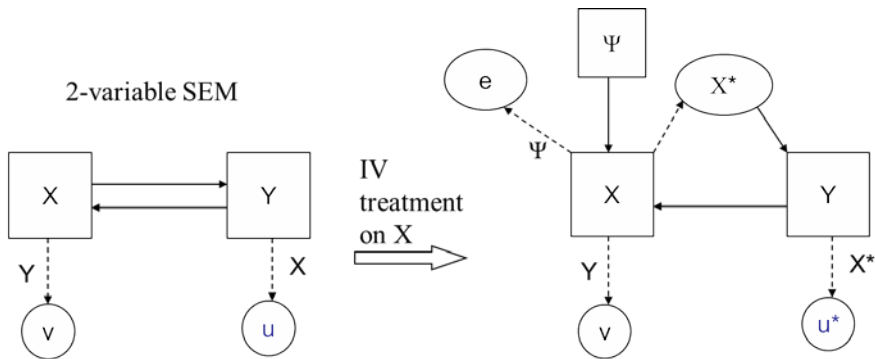
The non-optimal and non-unique nature of $x_t^* = E(x_t|\Psi)$ explains why there remains widespread distrust of IV-based estimates among many applied economists. Surprisingly, few have delved into the substantive consequence of the redefinition, although the issue of how to conduct appropriately statistical inference for models involving predictor-generated regressors has long been studied, e.g. see Pagan (1984). Interestingly, the redefinition is effectively exposed by Brito and Pearl (2002) outside the econometric circle, see also Brito (2010). Following their approach, let us ponder over the redefinition with the help of flow graphs.⁶

The left panel of Figure 1 depicts the interdependence in a simple SEM, where squares indicate observed variables and circles denote error terms assumed to satisfy the statistical innovation property.⁷ The dotted arrows indicate products which are model-derived rather than 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 into 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: a non-optimal predictor-generated x_t^* and a none innovative and substantial residual term, e_t , indicated by an oval shape in contrast to those circle shaped residuals. The absence of statistical selection criteria for filtering multiple non-

⁶ The term ‘flow graphs’ is used here to differentiate these graphs from the causal path diagram popularised by Pearl’s 2009 book. A distinctive feature of the present graphs is the exposition of the role of the error terms, a feature which is absent from path diagrams. This feature is designed to highlight the nature of the IV route, i.e. generating non-optimal predictors to modify initially postulated conditional variables.

⁷ This assumption is stronger than the classical assumptions normally used in regression models in that it requires the model to be optimal predictor.

Figure 1: Flow Graphs of the IV Redefinition of SEM



optimal x_t^* leaves the task to substantive justifications for the redefinition of $x_t \rightarrow y_t$ by $x_t^* \rightarrow y_t$.

One obvious justification is that x_t contains sizeable measurement errors, the original justification for the IV invention. But that is a too remote from the case of the OLS inconsistency with SEMs. The right panel of Figure 1 suggests that x_t serves as the instrument or medium for Ψ in transmitting its causal effect to y_t . This justification, however, undermines the practical relevance of SEMs as it endorses Wold's causal chain specification and also approves of his argument that the problem of an SEM lies in its inadequately formulated causal relations. Notice that this justification runs into a basic logic problem: Ψ , by the textbook definition, ought NOT to be selected out of any substantive reasoning and hence should not be capable of shouldering the responsibility of being the ultimate cause. We are forced back to the first justification of measurement errors.

The difficulty in substantively justifying the IV treatment of endogeneity in SEMs helps explain why it has failed to prevail in macroeconomic modelling. Accruing empirical evidence suggests overwhelmingly that static SEMs are frequently dynamically inadequately specified rather than postulated on erroneous conditional variables. The rising dominance of VARs serves as the best example in two respects. First, OLS is shown to be a generally adequate estimator here

because of the primary criterion of dynamic specification – securing statistically innovation error terms; this is a powerful corroboration of Wold’s proximity theorem. Second, the spirit of mutual causality of the variables of substantive interest is retained although endogeneity is assumed absent in simple VARs, e.g. a bivariate VAR:

$$(3) \quad \begin{pmatrix} y \\ x \end{pmatrix}_t = A_0 + \sum_{i=1}^s A_i \begin{pmatrix} y \\ x \end{pmatrix}_{t-i} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}_t \quad \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}_t = \mathbf{v}_t \sim IN(0, \Sigma_v)$$

Notice that simultaneity is tucked in the covariance of the error terms, Σ_v . The resulting treatment is focused on orthogonalising Σ_v . Such orthogonalising methods essentially break the circular causality in (3) by adding a dynamic order to x_t and y_t so as to remove their symmetric status. Unsurprisingly, these methods have aroused contentions. What is uncontentious is that the IV route does not help.

Model (3) is regarded as characterising the joint distribution of x_t and y_t under the LSE approach. Since most of empirical studies are based on partial equilibrium theories, the LSE approach seeks to reduce (3) into a conditional model of the substantive interest. Let us assume one lag dynamically adequate for simplicity. A one-lag VAR can be transformed into:

$$(4a) \quad y_t = \alpha_{10} + \alpha_{11,1}y_{t-1} + \alpha_{12,0}x_t + \alpha_{12,1}x_{t-1} + \varepsilon_{1t}$$

$$(4b) \quad x_t = \alpha_{20} + \alpha_{21,1}y_{t-1} + \alpha_{22,1}x_{t-1} + \varepsilon_{2t}$$

according to the *a priori* interest in $x_t \rightarrow y_t$. The LSE approach further investigates, by *a posteriori* exogeneity tests, the possibility of reducing (4b) to a marginal model:

$$(4b') \quad x_t = \alpha_{20} + \alpha_{22,1}x_{t-1} + \varepsilon_{2t}$$

Notice that the error covariance in (3) is now transformed into a contemporaneous relationship between x_t and y_t in conditional model (4a), the focal attention of the LSE approach, and that x_t is assumed weakly exogenous with respect to $\alpha_{12,0}$ (see Engle et al. 1983). Notice also how the static part of model (4) resembles (2),

except for the non-optimal predictor-generated x_t^* . This difference can, however, be eliminated without resorting to simultaneous interdependence. Let us illustrate this by a counter-example on exogeneity given in Hendry's 1995 textbook (Section 5.2). Suppose that y_t represents coffee price and x_t frost, and that our substantive interest is to forecast coffee price. Clearly, coffee price and frost do not form any interdependent relationship and thus a single-equation dynamic model such as (4a) should suffice. Nevertheless, frost can be 'endogenized' if the current frost situation has to be predicted. In other words, we can extend a single-equation model, say (4a), into a two-equation model similar to (2):

$$(5) \quad \begin{aligned} y_t &= \alpha_{10} + \alpha_{11,1}y_{t-1} + \alpha_{12,0}x_t^e + \alpha_{12,1}x_{t-1} + \varepsilon_{1t} \\ x_t^e &= E(x_t | Z_t) = \beta'Z_t \end{aligned}$$

where the variable set, Z_t , may include lagged exogenous variables. However, no sensible modeller would choose to treat Z_t as a set of IVs in (5), since any non-optimal predictors of x_t^e would defeat the modelling attempt. This example shows that the IV route is not a panacea for endogeneity; neither is endogeneity an econometric synonym for simultaneous interdependence of economic variables.

3.2 Selection Bias, Endogenized Dummy Variable and IVs in Microeconometrics

Let us start from a simple LDV model of y_i , a truncated variable (assuming the truncation occurs at 0 for simplicity), explained by w_i , in a cross-section data setting:

$$(6) \quad \begin{aligned} y_i &= \alpha w_i + u_i && \text{if } y_i > 0 \\ y_i &= 0 && \text{otherwise} \end{aligned}$$

Under Heckman's two-step procedure, the second equation in (6) is extended through a binary variable, d_i , to represent the selection process which generates the truncated outcome:

$$(7) \quad \begin{aligned} y_i &= \alpha w_i + u_i & y_i > 0 \\ d_i &= \gamma Z_i + e_i & d_i \in \{0,1\} \end{aligned}$$

When the two error terms are *assumed* correlated: $cov(u_i e_i) \neq 0$, the selection process described by the second equation of (7) will affect outcome, y_i , of the first equation, because:

$$(8) \quad \begin{aligned} E(y_i | d_i > 0) &= \alpha w_i + E(u_i | -e_i < \gamma Z_i) = \alpha w_i + \beta [f(\gamma Z_i)] \\ \Rightarrow y_i &= \alpha w_i + \beta [f(\gamma Z_i)] + \varepsilon_i \end{aligned}$$

where function $f(\gamma Z_i)$ relates to the conditional expectation of the second equation of (7), e.g. the inverse Mill's ratio derived from a probit regression in the Heckman procedure. Model (8) shows how a simple LDV model in (6) may be missing an additional regressor which is correlated with w_i due to 'selection bias' defined by the second equation in (7). Since the bias implies $cov(w_i Z_i) \neq 0$, the OLS estimate of α using the first equation in (7) would obviously suffer from OVB when β turns out to be statistically significant in (8). In practice, such significance has been amply shown to live on multicollinearity between w_i and Z_i . The finding is related to the dominantly linear distribution of the inverse Mill's ratio, e.g. see Puhani (2002).

Notice that there is a significant difference between the Heckman procedure and Tobit. While Tobit aims at correcting the truncation bias due to the modelled variable, the Heckman procedure is focused on the possible truncation effect of the explanatory variable of interest. This difference can be easily seen if we extend the root model of (8), i.e. (6), into a multiple regression model by adding a set of control variables, X_i :

$$(6') \quad \begin{aligned} y_i &= \alpha w_i + \tau X_i + u_i & \text{if } y_i > 0 \\ y_i &= 0 & \text{otherwise.} \end{aligned}$$

The parameter vector, τ , is out of the radar of the Heckman correction. In other words, the possibility of τ also suffering from selection bias is disregarded in (8')

$$(8') \quad y_i = \alpha w_i + \tau X_i + \beta [f(\gamma Z_i)] + \varepsilon_i.$$

Indeed, numerous extant empirical results demonstrating the Heckman selection effect are centered on the OVB effect of β on α only. A key reason underlying this difference is Heckman's concern over possible endogeneity bias in models involving truncated explanatory variables. In fact, w_i is assumed as an endogenous truncated variable in (Heckman, 1978). Let us follow Heckman's approach and modify (8') into an SEM:

$$(9) \quad \begin{aligned} y_i &= \alpha w_i + \tau X_i + \varepsilon_i \\ w_i &= \phi' X_i^w + \beta [f(d_i^*)] + e_i \end{aligned}$$

where $X_i \neq X_i^w$ and $f(d_i^*)$ is the inverse Mill's ratio derived from a probit regression like the second equation in (7). Noticeably, both the probit regression and the second equation in (9) are now *instrumental* for the consistent estimation of α , the key parameter of interest. Comparison of (9) with (1) tells us that (9) is a pseudo SEM in that y_i does not simultaneously explain w_i . The 'endogeneity' interpretation of w_i is thus superfluous. In fact, (9) is closer to (2) than (1). The foregoing discussion has already shown us that a viable justification for IVs is the measurement error case. Reflecting on model (9) makes us realise that the key problem to be treated here is measurement errors presumably induced by selection bias. In other words, concerns over selection bias make modellers distrust the observed w_i and seek the latent and bias-free \hat{w}_i by taking the effect of d_i into explicit consideration. Naturally, \hat{w}_i should not be the optimal predictor of w_i . Moreover, d_i^* need not be the optimal predictor of d_i as its instrumental effect on \hat{w}_i is usually negligible because, as shown from (8') above, the OVB correction via β is limited to one or two parameters in ϕ , a parameter vector which is deemed substantively uninteresting. The essence of model (2) remains dominant in (9).⁸

The methodology of model (9) has been relayed into programme evaluation modelling methods. Programme evaluation models take a similar form as (9), only a binary dummy variable, d_i , becomes the key explanatory variable of interest:

⁸ For an empirical illustration of this point, see Qin et al. (2014).

$$(10) \quad \begin{aligned} y_i &= \beta d_i + \tau X_i + \varepsilon_i \\ d_i &= \gamma Z_i + e_i. \end{aligned}$$

In (10), ε_i is commonly referred to as the ‘idiosyncratic treatment effect’, and the modellers’ central objective is to measure and interpret its parameter, β , as the average treatment effect (ATE). Since the outcome of a social programme can seldom be attributed solely to d_i , X_i is used to represent ‘selection on the observables’, i.e. to cover other relevant factors which may be correlated with d_i . Meanwhile, modelling ‘selection on the unobservables’ leads to the specification of d_i as an ‘endogenous regressor’, e.g. see Angrist et al. (1996). The predicament here is that Z_i is unobservable since all the observables are already covered by X_i . An alternative single-equation specification is to decompose ε_i into two parts, one for $cov(d_i \varepsilon_i) \neq 0$ following the endogeneity assumption of d_i and the other the leftover residual:

$$(11) \quad \begin{aligned} y_i &= \beta d_i + \tau X_i + \varepsilon_i \\ &= \beta d_i + \tau X_i + f(d_i | Z_i) + \varepsilon_i^* = \beta d_i + \tau X_i + \lambda' Z_i + \varepsilon_i^* \end{aligned}$$

Obviously, a programme evaluation model omitting Z_i in (11) would suffer from OVB should λ be significant. In other words, the OLS estimator of β in (10), denoted by $\hat{\beta}$, is biased from $\hat{\beta}^*$, the OLS estimator of β in the second equation of (11). An IV remedy for this is to find a set of IVs, Ψ_i , which is correlated with d_i but uncorrelated with Z_i . As shown by McFadden (1999), an IV estimator, $\tilde{\beta}$, can be easily derived from multiplying Ψ_i to the lower right equation in (11):

$$(12) \quad \Psi_i' y_i = \beta \Psi_i' d_i + \tau \Psi_i' X_i + \lambda' \Psi_i' Z_i + \Psi_i' \varepsilon_i^*.$$

The derivation exploits $cov(\Psi_i, Z_i) = 0$ by choice of Ψ_i . This estimator is recommended as the consistent ATE estimate of β . However, $\tilde{\beta}$ does not correct the OVB, $\hat{\beta}^* - \hat{\beta} \neq 0$, due to the unobservables, Z_i , *per se*, since $\Psi_i \perp Z_i$. All it

does is to replace the supposedly biased d_i by a non-optimal predictor of it, $d_i^* = E(d_i | \Psi_i)$, so as to get rid of the bias by redefining the conditional variable of interest. In fact, the 2SLS representation of the IV route brings us back to model (9):

$$(13) \quad y_i = \beta d_i^* + \tau X_i + \nu_i, \quad d_i^* = E(d_i | \Psi_i).$$

Since d_i^* is neither similar to d_i nor unique, it is no wonder debatable whether $\tilde{\beta}$ can be justified as measuring the desired *ATE*. The best example here is the explicit labelling of $\tilde{\beta}$ as ‘*LATE*’ (local *ATE*) to differentiate it from *ATE*, when Ψ_i is chosen following the grouping estimator by Wald (1940), which was proposed for tackling measurement errors. The differentiation corroborates forcefully our earlier conclusion that the IV treatment is only effective for measurement errors. Indeed, substantial corrections to ‘randomise’ d_i via Ψ_i could result in $\tilde{\beta}$ being much too ‘local’ to be credibly informative about the real impact of d_i .

3.3 Synopsis and Reflection

The nature of the IV route is to generate sub-optimal predictors of the assumed endogenous explanatory variable of interest in a theoretical model, and redefine the model as conditioning on one such generated predictor rather than the original explanatory variable. The non-optimal nature of the IV-modified regressors tells us that the IV route cannot be regarded as an innocuous estimation method because its selection cannot be based solely on statistical criteria. Rather, it is an implicit and indirect model re-specification device. This helps explain why statisticians hold aloof from the IV route.⁹

Recognition the IV’s role in model re-specification enables us to see what theoretical models the IV treatment is really applicable. We can classify these

⁹ In a popular book, *Causality: Models, Reasoning, and Inference*, by Pearl, ‘instrumental variables’ are categorised as a ‘causal concept’ (2009, p. 40).

models into three types according to the explanatory variable of interest: (I) SEMs where the explanatory variable of interest and the explained variable are assumed to be interdependent; (II) single-equation models where the explanatory variable of interest is susceptible to the threat of OVB; (III) errors-in-variable models where the explanatory variable of interest is latent and the available proxy is believed to contain measurement errors. The foregoing autonomy tells us the following.

- (a) The IV route is not applicable to model type I unless there is convincing evidence against the ‘interdependence’ hypothesis in favour of asymmetrically conditioning the explained variable on an IV-generated non-optimal predictor of the originally postulated explanatory variable of interest.
- (b) The IV route is not applicable to type II unless modellers are ready to revise their causal hypothesis by redefining the explanatory variable of interest to certain non-optimal predictors of it. Such redefinition will affect the interpretation of the parameter of interest.
- (c) The IV route is applicable only to type III, provided there is adequate evidence showing that the chosen IVs are capable of filtering out the measurement errors under concern. Here, the key practical difficulty is where to find such IVs. The search can be significantly reduced if the source of measurement errors is known, for instance, selection bias. Such reduction overrules the textbook criteria of choosing IVs by statistical correlation only. It turns the IV equation at the first stage of the 2SLS into a substantively meaningful one, since its ultimate aim is to generate a better proxy of the latent explanatory variable of interest than the unfiltered proxy.

The above analysis should help redirect applied modellers’ attention, when the IV route is tried, from getting consistent estimates for *a priori* given parameters of interest onto selecting conditional variables which are not only substantively interpretable but also data-permissible. The selection entails more routine use of statistical model evaluation criteria than what is taught in textbook econometrics.

In fact, there exists, outside the econometrics circle, a rich literature on how to make statistical analysis useful for causal analysis. For example, Cox (1992) presents a succinct, balanced and jargon-free summary of the key points including the use of ‘derived explanatory variables’, latent variables, the effects of quasi-experimental designs and the role of regression analysis, in subsequent to his

equally excellent summary on the role of statistical models (1990). Unfortunately, much of the statisticians' views have been screened out of econometrics by a thick barrier – endogeneity or simultaneous causality, e.g. see Heckman (2005; 2008).

However, our analysis reveals clearly how the IV route makes econometric modellers abandon simultaneous interdependence and commit themselves unequivocally to asymmetric models conditioning on IV-modified regressors. These non-optimal regressors deviate significantly from the theoretically postulated explanatory variables of interest and are produced effectively by exploiting multicollinearity. This gives rise to the following question: What has led so many econometric modellers to hold such an incredulously self-contradictory position and remaining indulged in myriads of 'endogenous regressors'?

4 Retrospection on IV's Resurgence

Seeking answers to the above question brings us back to the history described in Section 2. Clearly, macroeconometric movements into dynamic modelling contribute decisively to the diminishing role of endogeneity and its IV treatment. Whereas in microeconometrics, static models still prevail owing to heavy dependence on cross-section data evidence. However, most of the empirical micro models are based on partial-equilibrium theories. There is seldom substantive need to model the interdependence between the modelled variable and the explanatory variable of interest. Why has endogeneity retained its central status when it is actually a superfluous add-on?

There are at least two discernible reasons. First, SEMs and endogeneity have been taught as the hallmark of econometrics. This breeds an unwritten common understanding that an element of endogeneity would raise the persuasive power of empirical model results. Second, the endogeneity-backed IV route can help enhance the chance of finding a statistically significant parameter of interest, thanks to abundant multicollinearity among economic variables.¹⁰ This practical bonus is particularly enticing when interdependence is not part of the substantive

¹⁰ For a more detailed discussion on the history of multicollinearity and structural models, see Qin (2014b), and for a recent empirical illustration, see Qin et al. (2014).

interest nor is the purpose of empirical studies to find best possible predictors for the modelled variable.

But neither reason can defend the unsatisfactory state of having multiple and non-optimal ways of determining the estimates of key parameters of interest. The growing dissatisfaction is evident not only from the unceasing debates over the credibility of IV-based empirical results but also from the rising trend that increasingly many modellers are going for panel data and better designed surveys instead of playing around with estimator choice. But why has the dissatisfaction not yet challenged endogeneity – the superfluous foundation of the IV route?

Retrospective search for answers points to widespread weak understanding or neglect of the nature of regression models – conditional expectation, in spite of the fact that regression has remained the love of econometrics. That is most evident from the self-inconsistent jargon, ‘endogenous regressor’. While regression defines a regressor its conditional variable status, the ‘endogenous’ attribute rejects its validity. Historically, the idea of interpreting economic interdependence as joint probability distribution was easily embraced by the profession soon after Haavelmo’s 1944 monograph, in contrast to Wold’s vehement advocacy of conditional expectations. Issues pertinent to how to systematically reduce joint distributions to conditional-expectation based regression models were left unexplored till much later, e.g. see Hendry and Richard (1982). In microeconometrics, the adoption of the notion of conditional mean independence in the recent decades is actually pertinent to the issue of regression model closure with respect to OVB, rather than issues deriving directly from the joint distribution underlying a classical SEM. On the other hand, few empirically operational economic models are postulated explicitly in conditional expectation terms, no matter how partial their equilibrium basis is. That leaves empirical modellers the freedom to decide whether those causal premises could be directly translated into conditional expectations. When empirical evidence suggests OVB, or absence of ‘conditional mean independence’, endogeneity comes in handy to keep those premises their maintained status. Here, the econometric duality of endogeneity sustains its enduring appeal. While superfluous and hence apparently harmless to partial-equilibrium based causal premises, endogeneity is fundamental for justifying multiple IV-based estimators as equally consistent estimates for those parameters of interest which are defined in theoretical models highly incomplete for empirical purposes.

Indeed, the fundamental aspect of endogeneity is so essential for the estimation-centred textbook econometrics that concerns over correlation of ‘endogenous regressors’ with error terms have grown almost into a professional paranoia. Few realise that the correlation is as illusive and superfluous as the assumed endogeneity in initially single-equation based theoretical models. The frost example and model (5) in Section 3.1 has demonstrated clearly how the correlation is generated by the assumed endogeneity. Essentially, the paranoia grows out of the fallacy that independent error terms exist prior to model specification and carry certain ‘structural’ interpretation similar to other economic variables. Take the programme evaluation models for example. As mentioned in the previous section, the error term, ε_i , in (10) is commonly regarded as the idiosyncratic *treatment* effect, i.e. the effect by d_i . No justification is given as why ε_i does not possibly contain the ‘idiosyncratic effect’ of individual variables in X_i or other unspecified effects due either to their negligible significance or to modellers’ incomplete knowledge. In fact, it is practically impossible to validate the argument of endogeneity bias on the ground of correlation between a regressor and the error term in a multiple regression setting, especially when the model fit remains relatively low. Notice how much the basis of the IV treatment for ‘selection on the unobservables’ is weakened once ε_i is viewed as a model-derived compound of unspecified miscellaneous effects. In general, error terms of statistical models are derived from model specification. As such, they are unsuitable for any ‘structural’ interpretation, e.g. see Qin and Gilbert (2001).

It may be debatable whether interdependence postulated in theoretical models can be translated generally into conditional expectation models in practice. However, once the IV route is adopted, such models are unquestionably translated into conditional expectation models in which the *a priori* assumed endogenous explanatory variables are deemed as invalid conditional variables and replaced by their non-optimal predictor generated variables. Screening these IV-modified variables entails explicit regression model selection criteria. The single rule of getting statistically significant estimates within the expected value ranges for the parameters of interest is clearly inadequate. A recent empirical study of labour supply models using multiple cross-section survey data sets by Qin et al. (2014) reveals just how serious this inadequacy is and how obstructive the endogeneity-backed IV route can be to data discovery when the *a priori* invalidation of the conditional status of a key explanatory variable turns out to be empirically

unsustainable. Subsample parameter constancy serves as a crucial model selection criterion in their study. In general, since it is vital to verify empirically the conditional status of the explanatory variables of interest, formal variable selection tests should be called for. Here, the Cox procedure (1961; 1962) for testing non-nested conditional models fits in handily. The procedure makes diagnostic statistics on the error terms an essential part of the selection criteria.¹¹

Understanding the IV route from the above angle further suggests that a practically more general approach is to directly translate economic causal premises, especially those based on the partial equilibrium notion, into conditional expectation models and experiment with various non-nested model designs.¹² This approach not only dispenses with superfluous endogeneity, along with its unnecessary entanglement with OVB as well as the barrier between statisticians and econometricians, but also turns the prerequisite of conditional mean independence into an operational procedure of seeking the minimum model closure, e.g. see Olsen and Morgan (2005). In doing so, the production of fragile empirical evidence for the purpose of maintaining *a priori* incompletely defined causal premises should be significantly reduced.

However, such a direct approach is unlikely to become widely accepted, as long as the current trend keeps growing, i.e. doing cross-section survey data based econometric analyses for the purpose of finding demonstrable empirical support for a single or very few possible causes, e.g. the case of programme evaluation studies. When policy dispute is the top agenda while forecasting and thus model precision is apparently irrelevant, the appeal of IV-based consistent estimators will remain strong. Although it goes beyond the scope of the present investigation to evaluate different purposes under which econometrics is practised, highlighting how the IV route is related to a particular purpose should help deepen readers' understanding of the basic role that IVs actually play beyond the estimation box. The ups and downs of IVs have been riding on two broadly contrasting purposes – realism versus impressionism. History shows how the endogeneity based IV route

¹¹ The econometric concept of 'encompassing' has evolved essentially from the Cox procedure, e.g. see Bontemps and Mizon (2008).

¹² It should be noted that the concept of causation is shown to have two varieties – 'dependence' and 'production', e.g. see Hall (2004), and that the statistical notion of conditional expectation is closely related to dependence if not so to production.

suits well to the latter but not the former. In particular, the history of macroeconometrics shows that IVs are unlikely to provide applied modellers with better (as judged by model forecasting performance) contemporaneously conditional variables than those postulated on the basis of substantive knowledge; whereas the history of microeconometrics shows that IVs are effective tools for furnishing, out of large data sets of a multitude of loosely related factors, causally simple premises with consistent empirical measures such that the premises can be turned into eloquently illustrated stories.

Acknowledgement: This paper is built on ‘A history of polyvalent structural parameters: the case of instrument variable estimators’, a SOAS working paper (No. 183) jointly written with Yanqun Zhang. I am grateful for her to allow me to adapt freely from that paper and to help me with various queries I raised to her. My thanks also go to Olav Bjerkholt, Chris Gilbert, David Hendry, Haibing Gu, Sophie Van Hullen, Risa Morimoto, Ron Smith and anonymous referees from *Economics: The Open-Access, Open-Assessment E-Journal* for all their invaluable support, comments and suggestions.

References

- Addison, J.T., and Portugal, P. (1989). Job displacement, relative wage changes, and duration of unemployment. *Journal of Labor Economics* 7: 281–302. <https://ideas.repec.org/a/ucp/jlabec/v7y1989i3p281-302.html>
- Aldrich, J. (1993). Reiersøl, Geary and the idea of instrumental variables. *The Economic and Social Review* 24: 247–273. <http://core.kmi.open.ac.uk/download/pdf/30658.pdf>
- Amemiya, T. (1974). Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica* 42: 999–1012. http://econpapers.repec.org/article/ecmemetrp/v_3a42_3ay_3a1974_3ai_3a6_3ap_3a999-1012.htm
- 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: 347–351. http://www.jstor.org/stable/2285311?seq=1#page_scan_tab_contents
- Anderson, T. W., and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics* 20: 46–63. <http://projecteuclid.org/euclid.aoms/1177730090>

- Andrews, D.W.K., Moreira, M. J., and Stock, J.H. (2007). Performance of conditional Wald tests in IV regression with weak instruments. *Journal of Econometrics* 139: 116–132. <https://ideas.repec.org/a/eee/econom/v139y2007i1p116-132.html>
- Angrist, J. (1990). Lifetime earnings and the Vietnam era draft lottery: Evidence from social security administrative records. *American Economic Review* 80: 313–336. <https://ideas.repec.org/a/aea/aecrev/v80y1990i3p313-36.html>
- Angrist, J., and Imbens, G. (1991). Sources of identifying information in evaluation models. NBER Technical Working Papers 0117, National Bureau of Economic Research.
- Angrist, J., Imbens, G., and Rubin, D. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91: 444–455. http://www.jstor.org/stable/2291629?seq=1#page_scan_tab_contents
- Angrist, J., and Krueger, A. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics* 106: 979–1014. <https://ideas.repec.org/a/tpr/qjecon/v106y1991i4p979-1014.html>
- Angrist J.D., and Pischke, J. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.
- Barnow, B.S., Cain, G.G., and Goldberger, A.S. (1981). Selection on observables. *Evaluation Studies Review Annual* 5: 43–59.
- Blundell, R., and Costa Dias, M. (2009). Alternative approaches to evaluation in empirical microeconomics. *The Journal of Human Resources* 44: 565–640. <https://ideas.repec.org/a/uwp/jhriss/v44y2009i3p565-640.html>
- Bontemps, C., and Mizon, G.E. (2008). Encompassing: Concepts and implementation. *Oxford Bulletin of Economics and Statistics* 70: 721–750. <https://ideas.repec.org/a/bla/obuest/v70y2008is1p721-750.html>
- Bowden, R, and Turkington, D.A. (1990). *Instrumental variables*. Cambridge University Press.
- Brito, C. (2010). Instrumental sets. In R. Dechter, H. Geffner, and J.Y. Halpern (Eds.), *Heuristics, probability and causality: A tribute to Judea Pearl*. London: College Publications.
- Brito, C., and Pearl, J. (2002). Generalized instrumental variables. In *Uncertainty in artificial intelligence, proceedings of the eighteenth conference*. San Francisco: Morgan Kaufmann. <https://dslpitt.org/uai/papers/02/p85-brito.pdf>
- Cameron, A.C. (2009). Microeconometrics: Current methods and some recent developments. In K. Patterson., and T.C. Mills (Eds.) *Palgrave handbook of econometrics*, vol. 2. Houndmills: Palgrave MacMillan.

- Chen, B., and Pearl, J. (2012). Regression and causation: A critical examination of econometrics textbooks. *Real-World Economics Review* 65: 2–20.
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2338705
- Christ, C.F. (1966). *Econometric models and methods*. New York: John Wiley & Sons.
- Cox, D.R. (1961). Tests of separate families of hypotheses. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, vol. 1*. Berkeley: University of California Press.
- Cox, D.R. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society, Series B* 24: 406–424.
http://www.jstor.org/stable/2984232?seq=1#page_scan_tab_contents
- Cox, D.R. (1992). Causality: Some statistical aspects. *Journal of Royal Statistical Society Series A* 155: 291–301.
http://www.jstor.org/stable/2982962?seq=1#page_scan_tab_contents
- Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature* 48: 424–455.
<https://www.aeaweb.org/articles.php?doi=10.1257/jel.48.2.424>
- Duflo, E., Glennerster, R., and Kremer, M. (2008). Using randomization in development economics research: A toolkit. In T.P. Schultz, and J.A. Strauss (Eds.), *Handbook of development economics*. Elsevier.
- Duncan, G.M., and Leigh, D.E. (1985). The endogeneity of union status: An empirical test. *Journal of Labor Economics* 3: 385–402.
http://www.jstor.org/stable/2534847?seq=1#page_scan_tab_contents
- Durbin, J. (1954). Errors in variables. *Review of the International Statistical Institute* 22: 23–32.
<http://www.jstor.org/discover/10.2307/1401917?sid=21105234495131&uid=62&uid=2129&uid=2&uid=26292&uid=5910216&uid=26293&uid=3&uid=3737864&uid=70&uid=67>
- Engle, R.F., Hendry, D.F., and Richard, J.-F. (1983). Exogeneity. *Econometrica* 51, 277–304. http://www.jstor.org/stable/1911990?seq=1#page_scan_tab_contents
- Ericsson, N.R., and Hendry, D.F. (2004). The ET interview: Professor David F. Hendry. *Econometric Theory* 20: 745–806.
<http://dido.econ.yale.edu/korora/et/interview/hendry.pdf>
- Fisher, F.M. (1965). Dynamic structure and estimation in economy-wide econometric models. In J.S. Duesenberry, G. Fromm, L.R. Klein, and E. Kuh (Eds.), *The Brookings quarterly econometric model of the United States*. Chicago: Rand McNally.

- Gilbert, C.L. (1989). LSE and the British approach to time series econometrics. *Oxford Economic Papers* 41: 108–128.
http://www.jstor.org/stable/2663185?seq=1#page_scan_tab_contents
- Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica* 11: 1–12.
http://www.jstor.org/stable/1905714?seq=1#page_scan_tab_contents
- Hall, N. (2004). Two concepts of causation. In N. Hall, J. Collins, and L.A. Paul (Eds.), *Causation and counterfactuals*. The MIT Press.
- Hansen, L.P. (1982) Large sample properties of generalized methods of moments estimators. *Econometrica*, 50: 1029–1054.
http://faculty.chicagobooth.edu/john.cochrane/teaching/35904_Asset_Pricing/Hansen%20econometrica%20GMM.pdf
- Hausman, J. (1978). Specification tests in econometrics. *Econometrica* 46: 1251–1271.
<http://econweb.tamu.edu/keli/Hausman%201978.pdf>
- Heckman, J. (1974). Effects of child-care programs on women's work effort. *Journal of Political Economy* 82: S136–S163.
<https://ideas.repec.org/a/ucp/jpolec/v82y1974i2ps136-s163.html>
- Heckman, J. (1976). A life-cycle model of earnings, learning, and consumption. *Journal of Political Economy* 84: S11–S44.
http://www.jstor.org/stable/1831101?seq=1#page_scan_tab_contents
- Heckman, J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46: 931–959.
<http://www.columbia.edu/~pjd1/mypapers/mycurrentpapers/mycurrentpapers/dummyheckman.pdf>
- Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica* 47: 153–161. <https://ideas.repec.org/a/ecm/emetrp/v47y1979i1p153-61.html>
- Heckman, J. (1996). Comment (to Angrist et al., 1996). *Journal of the American Statistical Association* 91: 459–462.
- 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: 356–398.
- Heckman, J., and Hotz, V.J. (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: 862–874.
<http://athens.src.uchicago.edu/jenni/dvmaster/FILES/Heckman-Hotz89f.pdf>

- 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: 457–480.
https://ideas.repec.org/a/cup/etheor/v19y2003i03p457-480_19.html
- Hendry, D.F., and Richard, J-F. (1982). On the formulation of empirical models in dynamic econometrics. *Journal of Econometrics* 20: 3–33.
<https://ideas.repec.org/a/eee/econom/v20y1982i1p3-33.html>
- Imbens, G. (2010). Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature* 48: 399–423.
<https://ideas.repec.org/a/aea/jecclit/v48y2010i2p399-423.html>
- Imbens, G. (2014) Instrumental variables: An econometrician’s perspective. *NBER Working Paper Series*, no 19983, National Bureau of Economic Research, Cambridge, MA. <http://www.nber.org/papers/w19983.pdf>
- Imbens, G., and Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica* 62: 467–475.
<https://ideas.repec.org/a/ecm/emetrp/v62y1994i2p467-75.html>
- Khandker, S.R., Koolwal, G.B., and Samad, H.A. (2010). *Handbook on impact evaluation: Quantitative methods and practices*. The World Bank.
- Klein, L.R. (1955). On the interpretation of Theil’s method of estimation of economic relations. *Metroeconomica* 7: 147–153.
<http://onlinelibrary.wiley.com/doi/10.1111/j.1467-999X.1955.tb00750.x/full>
- Klein, L.R., and Ball, R.J. (1959). Some econometric of the determination of absolute wages and prices. *Economic Journal* 69: 465–482.
- Lee, L-F. (1978). Unionism and wage rates: A simultaneous equations model with qualitative and limited dependent variables. *International Economic Review* 19: 415–433.
http://econpapers.repec.org/article/ieriecrev/v_3a19_3ay_3a1978_3ai_3a2_3ap_3a415-33.htm
- Lee, L-F., Maddala, G., and Trost, R. (1980). Asymptotic covariance matrices of two-stage probit and two-stage tobit methods for simultaneous equations models with selectivity. *Econometrica* 48: 491–503.
http://www.jstor.org/stable/1911112?seq=1#page_scan_tab_contents
- Maddala, G.S. (1983). *Limited-dependent and qualitative variables in economics*. New York: Cambridge University Press.
- Manski, C.F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* 60: 531–542.
<https://ideas.repec.org/a/bla/restud/v60y1993i3p531-42.html>

- Manski, C.F. (2000). Identification problems and decisions under ambiguity: Empirical analysis of treatment response and normative analysis of treatment choice. *Journal of Econometrics* 95: 415–542.
<https://ideas.repec.org/a/eee/econom/v95y2000i2p415-442.html>
- McFadden, D. (1999). Chapter 4. Instrument variables. Lecture notes. Econ. 240B.
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: 462–465.
- Morgan, M.S. (1990). *The history of econometric ideas*. Cambridge University Press.
- Nelson, F., and Olson, L. (1978). Specification and estimation of a simultaneous equation model with limited dependent variables. *International Economic Review* 19: 695–709.
http://econpapers.repec.org/article/ieriecrev/v_3a19_3ay_3a1978_3ai_3a3_3ap_3a695-709.htm
- Olsen, W., and Morgan, J. (2005). A critical epistemology of analytical statistics: Addressing the sceptical realist. *Journal for the Theory of Social Behaviour* 35: 255–284. <http://onlinelibrary.wiley.com/doi/10.1111/j.1468-5914.2005.00279.x/abstract>
- Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review* 25: 221–247.
http://econpapers.repec.org/article/ieriecrev/v_3a25_3ay_3a1984_3ai_3a1_3ap_3a221-47.htm
- 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.
http://econpapers.repec.org/article/blajecsur/v_3a14_3ay_3a2000_3ai_3a1_3ap_3a53-68.htm
- Qin, D. (1993). *Formation of econometrics: A historical perspective*. Oxford University Press.
- Qin, D. (2013). *A History of econometrics: The reformation from the 1970s*. Oxford University Press.
- Qin, D. (2014a). Consolidation of the Haavelmo-Cowles commission research program. *Econometric Theory*.
<http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=9289978&fileId=S0266466614000309>
- Qin, D. (2014b). Inextricability of confluence and autonomy in econometrics. *Oeconomia* 4: special issue on the ‘History of econometrics’. <http://oeconomia.revues.org/883>

- Qin, D., and Gilbert, C.L. (2001). The error term in the history of time series econometrics. *Econometric Theory* 17: 424–450.
<http://journals.cambridge.org/action/displayFulltext?type=1&fid=65914&jid=ECT&volumeId=17&issueId=02&aid=65913>
- Qin, D., van Huellen, S., and Wang, Q.-C. (2014). What happens to wage elasticities when we strip playometrics? Revisiting married women labour supply model. Department of Economics Working Paper Series, No. 190, The School of Oriental and African Studies, London.
<http://www.soas.ac.uk/economics/research/workingpapers/file97784.pdf>
- Reiersøl, O. (1945). Confluence analysis by means of instrumental sets of variables. *Arkiv for Matematik, Astronomi och Fysik* 32A: 1–119.
- Robins, J.M. (1989). The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In L. Sechrest, H. Freeman, and A. Mulley (Eds.), *Health service research methodology: A focus on AIDS*. Washington, D.C., NCHSR, U.S. Public Health Service.
- Sargan, J.D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica* 26: 393–415.
http://www.jstor.org/stable/1907619?seq=1#page_scan_tab_contents
- 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. http://www.jstor.org/stable/2983930?seq=1#page_scan_tab_contents
- 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*. London: Butterworth.
- Sargent, T.J., and Sims, C.A. (1977). Business cycle modelling without pretending to have too much a priori economic theory. In *New methods in business cycle research: Proceedings from a conference*. Federal Reserve Bank of Minneapolis.
- Sims, C.A. (1980). Macroeconomics and reality. *Econometrica* 48: 1–48.
<https://ideas.repec.org/a/ecm/emetrp/v48y1980i1p1-48.html>
- Stern, S. (1989). Measuring the effect of disability on labor force participation. *The Journal of Human Resources* 24: 361–395.
http://econpapers.repec.org/article/uwpjhriss/v_3a24_3ay_3a1989_3ai_3a3_3ap_3a361-395.htm
- Stock, J., and Yogo, M. (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.

- Stock, J.H., and Trebbi, F. (2003). Retrospectives: Who invented instrumental variable regression? *Journal of Economic Perspectives* 17: 177–194.
<https://ideas.repec.org/a/aea/jecper/v17y2003i3p177-194.html>
- Stone, R., and Pesaran, M.H. (1991). The ET interview: Professor Sir Richard Stone. *Econometric Theory* 7: 85–123.
http://econpapers.repec.org/article/cupetheor/v_3a7_3ay_3a1991_3ai_3a01_3ap_3a85-123_5f00.htm
- Theil, H. (1953). Estimation and simultaneous correlation in complete equation systems. The Hague: Central Planning Bureau (mimeographed memorandum).
http://link.springer.com/chapter/10.1007/978-94-011-2546-8_6#page-1
- Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics* 11: 284–300.
http://www.jstor.org/stable/2235677?seq=1#page_scan_tab_contents
- Willassen, Y., and Reiersøl, O. (2000). The ET interview: Professor Olav Reiersøl. *Econometric Theory* 16: 113–125. <http://korora.econ.yale.edu/et/interview/reiersol.pdf>
- Wold, H.O.A. (1938). *A study in the analysis of stationary time series*. Stockholm: Almqvist & Wiksell.
- Wold, H.O.A. (1954). Causality and econometrics. *Econometrica* 22: 162–177.
http://www.jstor.org/stable/1907540?seq=1#page_scan_tab_contents
- 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.
http://www.jstor.org/stable/2342961?seq=1#page_scan_tab_contents
- Wold, H.O.A. (1960). A generalization of causal chain models (Part III of a Triptych on Causal Chain Systems). *Econometrica* 28: 443–463.
http://www.jstor.org/stable/1907733?seq=1#page_scan_tab_contents
- Wold, H.O.A. (1961). Construction principles of simultaneous equations models in econometrics. *Bulletin of the Institute of International Statistics* 38: 111–136.
- Wold, H.O.A. (1965). A letter report to Professor P. C. Mahalanobis. In C.R. Rao (Ed.), *Essays on econometrics and planning*. Oxford: Pergamon Press.
- Wold, H., and Faxér, P. (1957). On the specification error in regression analysis. *Annals of Mathematical Statistics* 28: 265–267.
http://www.jstor.org/stable/2237040?seq=1#page_scan_tab_contents
- Wold, H.O.A., and Juréen, L. (1953). *Demand analysis: A study in econometrics*. New York: Wiley and Sons.

- Wooldridge, J.M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics* 68: 115–132.
<https://ideas.repec.org/a/eee/econom/v68y1995i1p115-132.html>
- Wooldridge, J.M. (1996). Estimating systems of equations with different instruments for different equations. *Journal of Econometrics* 74: 387–405.
<https://ideas.repec.org/a/eee/econom/v74y1996i2p387-405.html>
- Wooldridge, J.M. (2010). *Econometric analysis of cross section and panel data, second edition*. The MIT Press.
- Wright, P.G. (1928). *The tariff on animal and vegetable oils*. New York: MacMillan.

Please note:

You are most sincerely encouraged to participate in the open assessment of this article. You can do so by either recommending the article or by posting your comments.

Please go to:

<http://dx.doi.org/10.5018/economics-ejournal.ja.2015-7>

The Editor