Chapter 5
Endogeneity and Simultaneity

Karim Chalak and Alastair R. Hall

5.1 Introduction

Endogeneity and simultaneity have been at the heart of econometrics since the 1930's–40's. "Simultaneity" refers to the case in which the econometric model simultaneously determines the value of more than one, often interdependent, endogenous variable. As such, simultaneity is a fundamental property of the endogenous variables in question. "Endogeneity" reflects the characteristics of the equation being estimated, reflecting an interaction of the observability of key variables, measurement error, and functional form specification. Accordingly, the meaning of endogeneity is nuanced and context–specific.

Important foundations of econometrics were laid in the 1930's–1950's in a time when computational and data constraints limited attention to linear simultaneous equations models (LSEM). In this context, "endogeneity" is used to refer to the case in which an explanatory variable is correlated with the error, and so is a shorthand for scenarios in which ordinary least squares (OLS) estimators of the regression model parameters are inconsistent. As noted by Morgan (1990)[p.11], *"work on business cycles and on market demand analysis [...] constituted most of applied econometrics of the period up to about 1950"* and both these applications naturally lead to linear models in which endogeneity arises from simultaneity. Motivated by business cycle applications, the Cowles commission undertook pathbreaking research on estimation and inference within linear simultaneous equations models which played a key role in the emergence of econometrics as a field.[1]

Karim Chalak ✉
University of Manchester, Manchester, United Kingdom, e-mail: karim.chalak@manchester.ac.uk

Alastair R. Hall
University of Manchester, Manchester, United Kingdom, e-mail: Alastair.Hall@manchester.ac.uk

[1] Ragnar Frisch coined the term 'econometrics' in 1926, and played a key role in the founding of the Econometric Society in 1930; see Bjerkholt (2015).

Simultaneity does not necessarily imply endogeneity. For example, the outcomes of a seemingly unrelated regression model are simultaneously determined. However, the parameters of this model can be estimated consistently via OLS applied equation–by-equation. The cost to ignoring simultaneity here is a potential loss in efficiency.

Endogeneity can arise in linear models for reasons other than simultaneity. Two other leading causes are measurement error and omitted variables. In the 1930's measurement error was recognized to be a feature of economic data. Nevertheless, the implications of measurement error and simultaneity tended to be analyzed separately and, for the most part, the statistical theory of linear simultaneous equations models developed in the 1940's-1950's assumed all variables to be accurately measured. Endogeneity due to omitted variables came to the fore in the 1980's as part of the 'credibility revolution in empirical microeconomics' (Angrist and Pischke, 2010).

In the 1970's-1980's, a number of concerns were raised about the use of linear simultaneous equation models to capture the macroeconomy. In particular, the Lucas (1976) 'policy critique' led to the use of a nonlinear dynamic stochastic general equilibrium models to capture the simultaneity of macroeconomic variables. These developments ushered in the generalized method of moments (GMM), and estimation based on the information in population moment conditions. The generality of the GMM approach method led to its use in a variety of other settings in which endogeneity and/or simultaneity occur.

In this chapter we provide an overview of these and certain other related developments. An outline is as follows. Section 5.2 provides some additional background on the sources of endogeneity in linear econometric models. Section 5.3 considers estimation in simultaneous equations models, both linear and nonlinear. Section 5.4 describes the generalized method of moments inference framework. Section 5.5 discusses the developments on instrumental variables estimation of a single linear equation from 1980's onwards, inspired by the GMM and also the issues arising as part of the credibility revolution. Section 5.6 covers nonparametric models, and Section 5.7 discusses estimation and inference in models with measurement errors and other data issues. Section 5.8 reviews applications of GMM in models with endogeneity involving panel, univariate time series and spatial data. Section 5.9 offers some concluding remarks.

Given the central place that simultaneity and endogeneity occupy in econometrics, it is difficult for one chapter to comprehensively treat all the areas that are related to these notions. This includes methods, such as regression discontinuity designs and difference in differences models, that have received significant attention in recent decades. We forgo a detailed discussion of such methods here given the space constraints.

Before embarking, we note that one reflection of the importance of endogeneity and simultaneity to the development of econometrics is the number of Nobel prize winners in Economics who have contributed in some way to the narrative in this chapter. These contributors are, with year of award in parenthesis: Ragnar Frisch and Jan Tinbergen (1969), Tjalling Koopmans (1975),[2] Lawrence Klein (1980), Trygve

[2] Koopmans made a number of important contributions to the development of econometrics but received the prize for his work on the optimal allocation of resources.

Haavelmo (1989), Robert Lucas (1995), James Heckman and Daniel McFadden (2000), Robert Engle and Clive Granger (2003), Christopher Sims (2011), Lars Hansen (2013), Joshua Angrist and Guido Imbens (2021).

5.2 Endogeneity in Linear Regression Models

In this section, we present three examples of endogeneity in linear models relevant to the evolution of ideas in econometrics. To do so, it is convenient to employ the following terminology. An explanatory variable is said to be an exogenous regressor if it is uncorrelated with the error term. If all the explanatory variables are exogenous regressors then the linear model in question is referred to as an exogenous regressor model. If an explanatory variable is correlated with the error then it is said to be an endogenous regressor. If at least one explanatory variable is an endogenous regressor then the model in question is referred to as an endogenous regressor model.

Simultaneity: One of the earliest treatments to provide a method for estimating endogenous regressor models is Philip Wright's monograph *The Tariff on Animal and Vegetable Oils* published in 1928. To elaborate on his contribution, consider the situation in which up to an additive error the demand for an agricultural product is a linear combination of price and income, and the supply is a linear function of price and weather. Assuming the market clears, the quantity produced and the price are simultaneously determined as functions of income, weather and the errors–making price an endogenous regressor in the demand and supply equations.

At the time Wright was working, it was accepted that income is an exogenous regressor in the demand equation and weather is an exogenous regressor in the supply equation. It was also recognized that this information by itself is insufficient to estimate the parameters of the demand and supply equations. Wright's key insights were to recognize that the income (weather) is also a source of external information about the supply (demand) equation, and to show this information can be leveraged into estimation of the equation in question using path analysis. In effect, Wright's approach amounts to estimation of the demand (supply) equation via the method of instrumental variables (IV) with the supply (demand) factor as an instrument for price. Strangely, both at the time and for many years after, Wright's insights seem largely to have been ignored within econometrics. However, P. G. Wright (1928) is now credited with the first presentation of instrumental variables in econometrics; see Stock and Trebbi (2003).[3]

[3] The details of the method are laid out in Appendix B of Wright's monograph. While there is currently agreement that Appendix B represents the first presentation of what we now refer to as IV, there has been some controversy about who actually wrote the appendix–Philip Wright or his son Sewall Wright who had developed path analysis . Based in part on a stylometric analysis, Stock and Trebbi (2003)[p.192] conclude that *"the evidence points towards Philip being both the author of Appendix B and the person who...first derived an explicit formula for the instrumental variable estimator"*.

Measurement error: It had been understood since the late 1870's at least that measurement error in the variables potentially contaminates the ordinary least squares estimation of the coefficients of a linear relationship.[4] These early treatments were influenced by the measurement process in the physical sciences and involved what is now commonly referred to as 'classical measurement error'. In these models, the explanatory variable is subject to an additive measurement error which is uncorrelated with the underlying 'true' value of the explanatory variable and the regression error.

With classical measurement error, the explanatory variable is an endogenous regressor because the implied regression error also depends on the measurement error in the explanatory variable. Different approaches employed various assumptions to tackle the resulting endogeneity. This includes recovering the equation coefficients under the assumption that certain features of the measurement error distribution are known (see, e.g., Koopmans (1936) and Allen (1939)), bounding the coefficients under minimial uncorrelation assumptions (see Gini (1921) and Frisch (1934)), or relying on exogenous instrumental variables to recover the coefficients (see, e.g., Wald (1940) and Reiersøl (1941)). Gini (1921) and Frisch (1934) recognized that, in a simple linear regression, the specific type of endogeneity due to classical measurement error generates a bias in the OLS estimator whose direction is known to be toward zero. See Durbin (1954) for an expression of this 'attenuation bias' in the form it is typically represented in econometric textbooks today.

Omitted variables/ Selection: In the 1980's, as part of what Angrist and Pischke (2010) have termed the 'credibility revolution in empirical microeconomics', attention turned to research design and in particular the use of controlled trials or natural experiments to justify causal interpretations of inference about economic phenomena. Such inferences are typically based on linear models in which regressors are endogenous for reasons that are commonly grouped together under the umbrella title of 'omitted variables'. Earlier work distinguished these variables from measurement error or equation disturbances (see Griliches, 1974). Within the treatment effects literature, these omitted variables reflect the selection process into treatment; for example see Angrist (1990). The prototypical example of an omitted variable is a key latent variable that drives both the outcome and the selection into treatment, i.e. an unobserved confounder. Angrist and Pischke (2009, p. 62) state that the linear regression omitted variable bias formula *"is one of the most important things to know about regression."*

Understanding the nature of the omitted variables is key to constructing methods for valid inference. For example, Angrist and Kreuger (1991) consider estimation of the returns to education in which individual level data is used to regress log wage on the number of years of schooling (and controls) with the individual's ability/motivation

[4] See Adcock (1877, 1878) and Kummell (1879). These and other early treatments concentrate on regression through the origin, take the regression line to be deterministic and both variables to be subject to error. Starting from this perspective, the 'measurement error' in the dependent variable essentially translates into what today is referred to as the regression error. As a result, the term 'measurement error'–or equivalently 'errors–in–variables'–is now used to refer to models with measurement error in explanatory variables and we follow this usage below.

being treated as an omitted variable. They report estimates based on IV in which quarter of birth is used as an instrument for education, a choice justified by the assumption of no correlation between ability and date of birth and by the nature of compulsory schooling laws in the USA.

5.3 Simultaneous Equation Models

5.3.1 Linear Models

The severe socio–economic consequences of the Great Depression motivated economists to seek a more fundamental understanding of business cycles. As part of this research agenda, Tinbergen presented in 1936 a macroeconometric model for the Dutch economy, credited as the world's first applied macroeconometric model (Cornelisse and van Dijk, 2008).[5] This model consists of 22 equations in 31 variables. The League of Nations commissioned him to develop a similar, but larger, model for the USA which was published in Tinbergen (1939).

As observed by Cornelisse and van Dijk (2008), Tinbergen's analysis of the Dutch economy not only led to clear policy recommendations regarding the alleviation of unemployment but

> *"its importance for the economics profession was far more profound: for the first time the economic–policy debate had been based on empirically tested, quantitative economic analysis and not on rather informally stated economic theory, the so–called verbal approach."*(Cornelisse & van Dijk, 2008)[p.304]

While the power of Tindbergen's approach to macroeconometric policy analysis was recognized, it was also realized that the development of this approach would require a number of statistical issues to be resolved. In his seminal paper "The probability approach in econometrics", Haavelmo (1944) laid the foundations for addressing these issues. As noted by Morgan (1990),

> *"The most immediate, and probably most important, effect of Haaavelmo's paper was its influence on the work of the Cowles Commission for Research in Economics.... Haavelmo's paper had been circulated amongst econometricians in an unpublished form in 1941 and formed the basis of a new research programme initiated at Cowles by Jacob Marschak in 1943."* (Morgan, 1990)[p.251]

In his retrospective article about the commission,[6] Christ (1994) identifies the two main components of the Cowles econometric revolution as an explicit probabilistic

---

[5] See Tinbergen (1959) for an english translation of the 1936 work.

[6] As explained by Carl Christ who was part of Cowles Commission and knew Cowles personally, Alfred Cowles was an investment counselor in Colorado Springs who realized after the stock market crash of 1929 that he did not understand how the economy worked. Through a sequence of contacts with academic economists, Cowles agreed to fund an economic research institute. The Cowles Commission was founded in Colorado Springs in 1932, moving to Chicago University in 1939, and then on to Yale University in 1955, its current home; see Christ (1994).

framework and the concept of a simultaneous equations models. The pursuit of this program required the development of solutions to the so–called 'identification problem' and of methods for estimation and hypothesis testing within this framework.

Central to the Cowles Commission methodology is a division of the variables into four key categories: *endogenous*, *exogenous*, *predetermined* and *errors*. The endogenous variables consist of those variables whose outcome is simultaneously determined by the system of equations given the values of the other three types of variables. Exogenous variables consisted of variables whose outcome was determined outside the system. Predetermined variables consisted of lagged endogenous variables, and the errors capture unobservable variables.

The maroeconomic theory of the day led to a linear simultaneous equation model of the form $\mathbf{YB}' + \mathbf{X\Gamma}' = \mathbf{U}$ where $\mathbf{Y}$, $\mathbf{X}$, $\mathbf{U}$ are respectively the matrix of observations on the endogenous, exogenous/predetermined and error variables. These equations involved parameters-the elements of the matrices $\mathbf{B}$ and $\mathbf{\Gamma}$-with an economic interpretation and as a result this system of equations became known as the *structural form*. For purposes of prediction, the relevant representation is $\mathbf{Y} = \mathbf{X\Pi}' + \mathbf{V}$, known as the *reduced form*.[7] Assuming $\mathbf{B}$ is nonsingular then the parameters of two representations are linked via the equation $\mathbf{B\Pi} + \mathbf{\Gamma} = 0$.

The key question is whether it is possible to estimate uniquely $\mathbf{B}$ and $\mathbf{C}$ from a sample $\{\mathbf{Y}, \mathbf{X}\}$. Koopmans (1949) defined an equation to be *identified* if its coefficients can be uniquely determined from the observed data. For this general formulation of the model, it was recognized that the equations of the structural form are not identified because a nonsingular transformation of the structural form yields the same reduced form. It is this phenomena that the Cowles Commission referred to as the 'identification problem'.

Further restrictions are needed. There are two immediate sources of restrictions: normalizations and identities. Normalization involves fixing the coefficient on one endogenous variable in each equation to be one. In some cases, the model involves equilibrium conditions or accounting identities imply that all the coefficients in an equation are known. However typically more restrictions are needed. The Cowles Commission considered information of the following types: *exclusion restrictions* that is, that certain coefficients are zero; *linear restrictions* on the coefficients; *covariance restrictions* that is, restrictions on the covariance matrix of the errors; see Koopmans (1949), Koopmans, Rubin and Leipnik (1950), and Koopmans and Hood (1953). To illustrate the nature of these rules, we consider exclusion restrictions; a comprehensive treatment of identification in LSEM is provided by Fisher (1966).

Identification is assessed on an equation–by–equation basis. Koopmans (1949) provides two conditions for an equation to be identified based on exclusion restrictions: the *order condition* (which is necessary) and the *rank condition* (which is necessary and sufficient). The order condition states that the number of exogenous/predetermined variables excluded from the equation be at least equal to the number of the included endogenous variables minus one. If this statement holds with equality then the

[7] The terms 'structural form' and 'reduced form' to describe these sets of equations became common in the 1940's, see Morgan (1990)[footnote 16, p.181].

equation is said to be *just*–identified; if this statement holds with a strict inequality then the equation is said to be *over*–identified.

Having devised rules for assessing identification, the second component of the Cowles econometric revolution was the development of methods for both estimation of the parameters of the model and testing hypotheses regarding these parameters. For this part of the agenda, it is necessary to articulate regularity conditions on the various components of the model. The errors were assumed to be an i.i.d. sequence, and often–and always for the discussion of ML estimation–this common distribution was taken to be normal. A second key aspect of the model was the implications of a variable being exogenous or predetermined for its relationship to the error. Koopmans et al. (1950)[p.26] define exogenous variables to be *"variables that influence the endogenous variables but are not themselves influenced by the endogenous variables"*. This meant that the matrix of observations on the exogenous variables can be taken as independent of the matrix of observations on the errors. With predetermined variables, the distribution of error vector at time t is independent of past realizations of the endogenous variables (as well as the exogenous regressors). We refer the reader to Hausman (1983) for a complete accounting of the regularity conditions. In the case where $\mathbf{X}$ contains only exogenous variables, we follow Mariano (2003) and refer to the model as the *Classical* LSEM; if in addition the errors are assumed to have a normal distribution we modify the name to *Classical Normal* LSEM.

The parameters can either be estimated equation–by–equation or based on estimating all equations simultaneously referred to respectively as *limited information* (LI) or *full information* (FI) estimation. We briefly review available methods within these two approaches before comparing their strengths and weaknesses. Throughout this discussion, it is assumed for ease of presentation that identification is achieved through exclusion restrictions alone unless otherwise stated.

Limited information methods: Without loss of generality, we focus on estimation of the first equation of the system. To this end, suppose that once the exclusion restrictions have been imposed this equation can be written as $\mathbf{y}_1 = \mathbf{D}_1\delta_1 + \mathbf{u}_1$ for $\mathbf{D}_1 = [\mathbf{Y}_1, \mathbf{X}_1]$ with $[\mathbf{y}_1, \mathbf{Y}_1]$ and $\mathbf{X}_1$ being the matrices of observations on the included endogenous and exogenous variables respectively, δ_1 is the parameter vector, and $\mathbf{u}_1$ the vector of observations on the error.

Indirect least squares (ILS) is a method for estimation of equations that are just–identified. The method appears in Appendix B of P. G. Wright (1928) and is also used independently by Tinbergen (1930), see Stock and Trebbi (2003). ILS exploits the mapping between the reduced form coefficients and structural parameters. The relevant reduced form coefficients are estimated via OLS and then this mapping is used to obtain the ILS estimates of δ_1.

As noted above Wrights's pathbreaking work in the 1920's on estimation of supply and demand curves seems not to have been recognized by the wider profession, and evidently not by researchers at the Cowles Commission; see Christ (1994). The method was re–discovered in the 1940's in a series of works. Reiersøl (1941) used the method of instrumental variable estimation as the solution to the problem of estimating the coefficients of an exact linear relationship involving variables that are subject to measurement error, albeit *"hidden amongst a number of extensions to Frisch's confluence analysis"* (Morgan, 1990)[p.226]. This approach was also apparently independently suggested by Geary(1942, 1943) . In his thesis, Reiersøl (1945) extended his 1941 work further and introduced the idea of 'instrumental variables', although he stated later that the term was coined by Ragnar Frisch, (see Morgan, 1990, p.228).

Reiersøl (1941) motivates his IV estimator as follows. Let $\mathbf{Z}$ be the matrix of observations on a vector z_t with the same dimension as δ_1. If $\mathbf{y}_1 = \mathbf{D}_1\delta_1 + \mathbf{u}_1$ is pre–multiplied by $\mathbf{Z}'$ then this yields a square set of equations in the unknown parameters. These equations involve the observable quantities $\mathbf{Z}'\mathbf{y}_1$ and $\mathbf{Z}'\mathbf{D}$ and the unobservable quantity $\mathbf{Z}'\mathbf{u}_1$. Reiersøl (1941) drops this unobservable term, and defines the IV estimator as the solution to $\mathbf{Z}'\mathbf{y}_1 = \mathbf{Z}'\mathbf{D}_1\hat{\delta}_1^{(IV)}$. If $\mathbf{Z}'\mathbf{D}_1$ is nonsingular then these equations yield the unique solution $\hat{\delta}_1^{(IV)} = (\mathbf{Z}'\mathbf{D}_1)^{-1}\mathbf{Z}'\mathbf{y}_1$.

For this approach to yield a consistent estimator of the parameters, the *instrument* vector, $\mathbf{z}_t$, must possess certain properties: it must be uncorrelated with the error so that $\mathbf{Z}'\mathbf{u}_1$ can be dropped without affecting large sample properties of the estimator; and it must be sufficiently related to $\mathbf{x}_t$ for the inverse to exist. These properties can be recognized as what are now commonly referred to as satisfying the *orthogonality condition* and *relevance condition* of IV estimation.

The key question is what variables can serve as an instrument. Since $\mathbf{X}_1$ are exogenous variables, they can serve as instruments–often described by saying that these variables can serve as instruments for themselves. A natural source of additional instruments–that is, instruments for the endogenous regressors–is the set of other exogenous variables in the system that are excluded from the first equation. If the first equation is identified then it follows from the order condition that there are perforce a sufficient number of exogenous variables that can serve as instruments for $\mathbf{Y}_1$. Writing $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$, if the equation is just–identified then IV is implemented with $\mathbf{Z} = [\mathbf{X}_2, \mathbf{X}_1]$. If the equation is over–identified then early practice was to choose a subset of the available instruments so that the equation is just-identified and then to implement the estimator above.

The first method to fully exploit any over–identifying restrictions is the LI maximum likelihood (LIML) estimator (Anderson and Rubin, 1949, 1950). With this approach, the parameter estimators are obtained by maximizing the likelihood function of $[\mathbf{y}_1, \mathbf{Y}_1]$ based on the reduced form representation subject to the constraints implied by the exclusion restrictions and the mapping $\mathbf{B\Pi} + \mathbf{\Gamma} = 0$.[8]

The two–stage least squares (2SLS) estimator was introduced in the 1950's seemingly independently by three authors taking three different approaches.[9] Theil (1953, 1958) obtains the estimator by pre–multiplying the (first) structural form equation by $\mathbf{X}'$ and solving the resulting equation system by Generalized Least Squares to yield the estimator $\hat{\delta}_1 = (\hat{\mathbf{D}}_1'\hat{\mathbf{D}}_1)^{-1}\hat{\mathbf{D}}_1'\mathbf{y}$ for $\hat{\mathbf{D}}_1 = [\hat{\mathbf{Y}}_1, \mathbf{X}_1]$ where $\hat{\mathbf{Y}}_1$ is the predicted

[8] See *inter alia* Mariano (2003)[pp 128–9].

[9] Anderson (2005) notes that Anderson and Rubin (1950) consider the 2SLS estimator in their derivation of the limiting distribution of LIML, but do not propose its use.

value of $\mathbf{Y}_1$ obtained by equation–by–equation estimation of the reduced form via OLS. At the time of its introduction, the simplicity of this two–stage procedure made 2SLS an attractive alternative to the more computationally demanding LIML method which involves solving a generalized eigenvalue problem.

Basmann (1957) derives the 2SLS estimator by using least squares theory to characterize the 'best' linear consistent estimators of the coefficients. Sargan (1958) derives the estimator from an IV perspective.[10] Building from the work of Geary (1948), Sargan considers estimation of the structural form equation in the over–identified case using IV with instruments for $\mathbf{Y}_1$ that are linear combinations of the available instruments, $\mathbf{X}$. Sargan (1958) presents the limiting distribution of this estimator for an arbitrary linear combination of instruments and uses the result to characterize the linear combination that is optimal in the sense that it minimizes the variance. While this optimal linear combination depends on unknown population moment matrices, the substitution of their sample analogs yields the 2SLS estimator. Sargan (1959) extends the analysis to models with autocorrelated errors.

In the years immediately following the introduction of 2SLS, Theil's formulation of the estimator seems to have received most attention: as Arellano (2002) observes, *"most of the textbooks in the 1960's routinely surveyed simultaneous equations estimators [such as 2SLS] as distinct from IV methods."* However, it is Sargan's approach that is most closely aligned with current perspectives informed by GMM: as noted by Arellano (2002)[p.450], *"Sargan was thinking in terms of moment condition, over–identifying restrictions and partially specified models"*. In these two papers in the late 1950's, Sargan has been credited with providing the *"definitive treatment of IV methods"* (see P. C. B. Phillips, 2003, p.420).

While seemingly very different approaches to estimation, it can be shown that if the equation is just–identified via exclusion restrictions then ILS, IV (with $\mathbf{Z} = [\mathbf{X}_2, \mathbf{X}_1]$), 2SLS and LIML are identical.

If the equation is over–identified via exclusion restrictions then 2SLS and LIML yield different estimators. One way to connect these two estimators is through their membership of the family of k–class estimators introduced by Theil (1961). These estimators can be defined as IV estimators with instrument matrix $\mathbf{Z}_{(k)} = k\mathbf{Z}_{2sls} + (1-k)\mathbf{Z}_{ols}$, where $Z_{2sls} = \hat{\mathbf{D}}_1$, $\mathbf{Z}_{ols} = \mathbf{D}_1$. Theil (1961) shows that LIML is a k–class estimator for a particular choice of k that is random.

The statistical properties of the 2SLS and LIML estimators have been compared in a variety of ways. The earliest comparisons are made within schemes in which the total number of exogenous variables is fixed, the sample size (T) goes to infinity and $T^{-1}\mathbf{X}'\mathbf{X}$ tends in probability to a finite, positive definite limit-often referred to now as the 'standard asymptotic analysis' (*e.g.* Mariano, 2003). Within this framework, 2SLS and LIML have the same first-order asymptotic distribution. Furthermore, LIML is asymptotically efficient among all consistent and uniformly asymptotically normal (CUAN) limited information estimators, and so 2SLS also shares this property. One way to establish this equivalence is through the framework of k–class estimators. Theil (1961) shows that all k–class estimators for which $plimT^{1/2}(k-1) = 0$ share

---

[10] Klein (1955) also shows an IV interpretation of 2SLS.

the same asymptotic distribution, a property that holds trivially for 2SLS ($k = 1$) and can be shown to hold for LIML.[11]

Given that both estimators have the same first–order asymptotic distribution, attention turned to comparing them on the basis of higher–order properties. Nagar (1959) uses asymptotic expansions to elicit the sources of potential finite sample differences between members of the family of k–class estimators within the Classical Normal LSEM. Within this approach, an asymptotic expansion is used to deduce a function of the data that approximates the estimator up to some specified order and then behaviour of the approximating function is evaluated ignoring the remainder. Originally, these methods were thought to produce approximations to the moments of the estimators, but subsequent developments reviewed below indicated this interpretation is only valid in certain cases. Therefore, we follow Mariano (2003) and refer to the moments of the approximating function as the 'asymptotic moments' of the estimator. Using this approach, Nagar (1959) shows that the asymptotic bias-the first asymptotic moment minus the true parameter value-of the 2SLS estimator increases linearly in the *degree of over-identification* of the equation minus one that is, the number of excluded exogenous variables minus the number of endogenous variables on the right-hand-side of the equation minus one.

The standard asymptotic analysis and the Nagar-type asymptotic moment calculations are based on approximations to the behaviour of the estimators. Through the 1960's–1980's, considerable attention was focused on deriving the *exact finite sample* distribution of various estimators within the Classical Normal LSEM. Early contributions concentrated on the model with a single endogenous regressor: see Richardson (1968) and Sawa (1969) for IV/2SLS and Mariano and Sawa (1972) for LIML. The extensions to the case with multiple endogenous regressors case are provided for IV/2SLS by P. C. B. Phillips (1980) and for LIML by P. C. B. Phillips (1984,1985).

These distributional results did not supplant the first–order asymptotic results as a basis for inference due to the complicated nature of these distributions and the stringency of the conditions under which the distributions are derived (such as fixed exogenous variables). However, these results do provide useful insights into the behaviour of the estimators and their comparative properties. Here we highlight a few key features that are relevant to our subsequent discussion, and refer the reader to Mariano (1983, 2003) for a more comprehensive summary and references.

For both 2SLS and LIML, the passage to the limiting distribution as $T \to \infty$ is driven by a quantity known as the *concentration parameter* (Basmann, 1963) that represents a unitless measure of the strength of the relationship between the endogenous regressors and instruments.[12] Within the Classical LSEM that satisfies the additional conditions imposed for the standard asymptotic analysis, the concentration parameter increases linearly in T. In the 1990's, concerns about the quality of identification led to analysis of IV/2SLS/LIML estimators under alternative sequences for the concentration parameter, see Section 5.5.

[11] This follows from a result due to Anderson and Rubin (1950).

[12] See *inter alia* Staiger and Stock (1997) and Stock, Wright and Yogo (2002) for further discussion.

The exact finite sample distributions of the IV/2SLS and LIML estimators are similar to Student's t and Cauchy distributions respectively. As a result, the moments of 2SLS only exist up to and including the degree of over–identification of the equation in question, but no moments of LIML exist. Notice that as a consequence if the equation of interest is just–identified then none of the moments exists for ILS, IV, 2SLS or LIML due to their equivalence. Fuller (1977) proposes a modified version of the k–class estimators (including 2SLS and LIML) whose moments exist in sufficiently large samples.

Sawa (1969) compares the finite sample distributions of the OLS and 2SLS estimator for the case with one endogenous regressor. His analysis reveals that the finite sample biases of 2SLS and OLS are in the same direction, and are of similar magnitude as the concentration parameter tends to zero *ceteris paribus*. In addition, he reports calculations that show if the number of instruments approaches the sample size then the differences in the distributions of OLS and 2SLS are 'inconspicuous'; a finding that suggests the standard first–order asymptotic analysis may be a poor guide to the finite sample behaviour of 2SLS when the degree of overidentification is large.

The non–existence of moments negates the possibility of finite sample mean–bias comparisons between 2SLS and LIML. Anderson, Kunitomo and Sawa (1982) evaluate the distribution functions of LIML and 2SLS and find that LIML tends to be median unbiased but 2SLS can exhibit substantial bias if the degree of overidentification is large. Their papers ends with the following summary

> *"the most important conclusion from the study of the LIML and [2]SLS estimators is that the [2]SLS estimator can be badly biased and in that sense its use is risky. The LIML estimator, on the other hand, has a little more variability with a slight chance of extreme values, but its distribution is centered at the parameter value."*(Anderson et al., 1982, p.1026)

These (non–) existence results also have direct implications for the interpretation of the Nagar–type expansions. As described above, the asymptotic moments are the moments of an approximation of the estimator and so may not be approximations to the exact moments of the estimator itself. For example, this is not the case for estimators for which low positive order moments do not exist, such as LIML. Sargan (1974, 1976) provides conditions under which asymptotic moments are valid approximations to the moments of the estimator. In cases where the finite moment does not exist, Sargan (1982) shows that Nagar's approach yields an approximation to the corresponding moment implied by an Edgeworth expansion approximation to the distribution function of the estimator.

Concerns about the quality of the standard first–order asymptotic distribution to the finite sample behaviour of 2SLS for large degrees of over–identification motivated the examination of the behaviour of the estimators under alternative asymptotic sequences. Bekker (1994) analyses the limiting distributions of OLS, 2SLS and LIML estimators in scenarios where the number of instruments increases to infinity with the sample size. Adapting his results to our context above of estimation of the parameters of the first equation of an LSEM, Bekker (1994) considers sequences in which the degree of over-identification (L_1, say) of the equation divided by the sample size converges to some finite constant α. Within this framework, Bekker (1994) shows that: OLS is inconsistent; 2SLS is consistent if $\alpha = 0$ but inconsistent otherwise; and

LIML is consistent. Bekker (1994) also provides a large sample distribution theory for LIML that can be leveraged to provide inference procedures for δ_1. He reports that these limiting distributions provide a better approximation to the finite sample behaviour of the estimators than the standard first–order asymptotic analysis, in which L_1 is treated as fixed, even when L_1 is small.

By definition in over–identified models more restrictions are imposed than are needed to estimate the parameter vector. Anderson and Rubin (1949) propose a likelihood ratio test of the null hypothesis that the over–identifying restrictions are consistent with the data. Basmann (1960) shows that the same test can be implemented with the 2SLS estimator replacing the LIML estimator. Sargan (1958) introduces an IV analog to this test based on the estimated sample moments associated with IV estimation that is a precursor of the GMM overidentifying restrictions test within the GMM framework, see Section 5.4.

Full–information information methods: Full–information estimation involves estimation of all the parameters of the structural form with the equation–by–equation restrictions embedded to ensure all parameters are identified. The earliest form of FI estimation to be considered by the Cowles Commission is FI maximum likelihood (FIML); see Haavelmo (1944) and Koopmans et al. (1950). However as observed by Christ (1994):

> *"Empirical work in the 1940's and early 1950's was done with mechanical desk calculators which, unlike today's cheap electronic pocket calculators, couldn't compute square roots or logarithms or exponentials with a single command. The iterative computations for FIML are so complex that the method was never applied by Cowles workers, except for an illustrative computation for a simplified model containing three equations. For most of their empirical work , the Cowles workers used LIML, which is also iterative, but much simpler. Later with the advent of electronic computers , FIML became a practical method."* (Christ, 1994, p.44)

Zellner and Theil (1962) introduce the three–stage least squares (3SLS) estimator. This estimator is derived in a similar way to Theil's derivation of 2SLS but this time applying the approach to estimation of the parameters of all equations simultaneously.

Rothenberg and Leenders (1964) show that if the identifying restrictions do not involve the variance-covariance matrix of the errors ($\mathbf{\Sigma}$, say) then FIML and 3SLS have the same large sample distribution, and both estimators are asymptotically efficient within the class of all CUAN estimators of δ. However, if $\mathbf{\Sigma}$ is restricted then in general FIML is at least as efficient asymptotically as 3SLS for the simple reason that FIML takes account of such restrictions and 3SLS does not.

For the case in which equations are identified by coefficient restrictions-as opposed to covariance restrictions–Madansky (1964) and Hausman (1975) show respectively that 3SLS and FIML can be interpreted as IV estimators. Hausman (1975) demonstrates that FIML uses all available *a priori* restrictions on the parameters in the construction of the instrument but 3SLS does not. It also follows from this IV interpretation that FIML and 3SLS are identical if all equations are just–identified; see Hausman (1975). Hausman, Newey and Taylor (1975) extend this IV interpretation to the case with covariance restrictions.

In terms of existence of moments, these FI estimators mimic the properties of their LI counterparts. The 3SLS estimators of the parameters of an identifed equation are finite up to and including the degree of over–identification of the equation that is, to the same order as 2SLS, see Sargan (1978). The moments of the FIML estimator do not exist, just as for LIML; a result originally presented by Sargan in a paper to the 2nd World Congress of the Econometric Society in 1970 and subsequently reproduced in Sargan (1988).

LI versus FI estimation: Aside from certain cases in which LI and FI estimators coincide,[13] the comparison between LI and FI methods depends on the object of interest. If estimation of the entire system system is the objective then FI estimators have the advantage that they are at least as efficient asymptotically as their LI counterparts; see Madansky (1964), Hausman (1975). The potential gain in efficiency comes from the expanded information set. However if interest is in the coefficients of the j^{th} equation alone then the comparison becomes more nuanced. While the potential gain in asymptotic efficiency from FI still remains, there are two considerations that speak to an advantage of LI estimation. First, FI estimation requires every equation in the system to be identified whereas LI estimation only requires this for the equation of interest. Second, the consistency and asymptotic efficiency of FI estimators is premised on the entire system being correctly specified, whereas LI is premised on the correct specification of the equation of interest. As a result, even if the equation of interest is correctly specified there is the potential with FI methods that inference about the equation of interest is compromised by mis–specification elsewhere in the system.

5.3.2 Nonlinear Models

For the most part attention in 1950's and 1960's focused on LSEMs. One exception is Sargan (1959) who develops IV methods for estimation of equations that are linear in variables but whose coefficients are nonlinear functions of an underlying parameter vector. With advances in computing in the 1970's both in hardware and algorithm design,[14] it became possible to consider models consisting of systems of nonlinear equations. Just as in linear models, identification is critical to the ability to estimate the parameters consistently and perform inference. However, unlike in linear models, two concepts of identification are required: *global* identification to ensure consistent estimation and *local first-order* identification to permit the development of the standard first–order asymptotic results upon which large sample inferences are based. While results are available for global identification for certain functional forms,[15] all the statistical analyses of estimators in nonlinear simultaneous equation

---

[13] For example see Hausman (1983).

[14] For example see Berndt, Hall, Hall and Hausman (1974).

[15] For example see Fisher (1966) and Brown (1983).

models (NSEM) are premised on high–level conditions that effectively assume the parameters to be identified.[16]

The earliest developments involve models in which each dependent variable is the subject of an equation and is explained up to an additive error by a nonlinear function of other endogenous variables, exogenous/predetermined variables and an unknown parameter vector. For this setting, Amemiya (1974) introduces a nonlinear two–stage (LI) estimator, and Jorgenson and Laffont (1974) introduce a nonlinear three–stage least squares (FI) estimator.

Subsequent treatments consider estimation of the parameters of nonlinear, simultaneous, implicit equations model (NSEM) involving endogenous and exogenous/predetermined variables. Gallant (1977) extends nonlinear 3SLS (N3SLS) estimation to this setting.[17] Amemiya (1977) develops FIML based on the assumption that the errors have a normal distribution, denoted here as NFIML. He shows that if the model is correctly specified then NFIML is CUAN and asymptotically more efficient than N3SLS. However if the error distribution is mis–specified then Amemiya (1977) argues that NFIML is inconsistent in general–in contrast to N3SLS which is consistent under less restrictive conditions on the error distribution. While this statement regarding the inconsistency of NFIML is not technically correct,[18] it remains that in systems of nonlinear, simultaneous, implicit equations the consistency of NFIML depends on the true error distribution satisfying higher–order moment restrictions which may be difficult to justify from economic theory without the assumption of normality.

While least squares and maximum likelihood are different estimation principles, both types of estimators are obtained via optimization of some function of the data. Burgete, Gallant and Souza (1982) exploit this essential structure to develop a large sample theory for a wide class of econometrics estimators of the parameters of NSEM that includes all the estimators mentioned in this section as a special case. Their approach provides one way to unify the discussion of estimation theory in nonlinear econometric models. The paper by Burgete et al. (1982) is coeval with the emergence of the GMM which provides a unified framework for discussion of estimation based on population moment conditions; see Section 5.4. It seems fair to say that the latter has become the dominant perspective on estimation in econometrics. The emergence of GMM was due in no small part to certain developments in macroeconometric modeling to which we now turn.

[16] For further discussion see Rothenberg (1971) and Bowden (1973).

[17] See Gallant and Jorgenson (1979) for a discussion of inference methods in this setting.

[18] See P. C. B. Phillips (1982).

5.3.3 Critiques of SEMs

LSEMs performed quite well in the 1950's and 1960's, in the sense of providing 'reliable' forecasts of the macroeconomy.[19] However, they did not fare so well in the 1970's. The latter decade was characterized by high unemployment and high inflation, the dual occurrence of which could not be explained by the Keynesian economics underlying the LSEMs. In fact, it was found that simple univariate time series models often provided more reliable forecasts than the LSEMs whose predictions were supposedly based on models of the macroeconomy, see Diebold (1992). This forecasting failure was accompanied by intellectual dissatisfaction at the statistical foundations of LSEMs as models of the macroeconomy. Here we highlight two critiques that are important for our subsequent discussion: Lucas's 'Policy Critique' and Sims's critique of the exogeneity assumptions.

Lucas (1976) argues that the use of LSEM for economic policy analysis is fundamentally flawed because it is premised on the assumption that the parameters of these models are invariant to the policy regime whereas in a world with forward–looking agents with rational expectations these parameters are themselves dependent on the policy regime. Sims (1980) argues that LSEMs are fundamentally flawed because[20]

> *"claims for a connection between these models and reality–the style in which "identification" is achieved for these models–is inappropriate, to the point at which claims for identification in these models cannot be taken seriously"*(Sims, 1980)[p.1]

This critique reflected in part a contemporaneous re–evaluation of the concept of exogeneity in time series models[21] to connect it to Granger non–causality (Granger, 1969) that led to the introduction of the concepts of weak–, strong– and super–exogeneity by Engle, Hendry and Richard (1983).

Together these criticisms had a significant impact on macroeconometric modelling.[22] Sims (1980) advocated the use of Vector Autoregressive (VAR) models in which all variables are treated as endogenous and each variable depends on its own lagged values and the lagged values of all other variables in the system.[23] Thus, in the terminology of LSEMs, inference is based on reduced forms. Subsequent developments led to the introduction of so–called structural VAR (SVAR) models that echo structural forms in LSEMs; see Section 5.7.1.

The Lucas critique motivated the development of macroeconometric models derived from dynamic stochastic general equilibrium (DSGE) models that involve fully–articulated preferences and technologies.[24] The first econometric implementations of these models involve maximizing a Gaussian likelihood function for the

---

[19] Lawrence Klein played a leading role in the development of macroeconometric models during this period, see Klein and Mariano (1980) for references and background.

[20] This criticism is also made by Liu (1960).

[21] See *e.g.* Sims (1977).

[22] See Diebold (1992) for a more extended discussion.

[23] See Chapter 1 of this volume.

[24] See Chapter 21 of this volume.

(potentially transformed) model variables subject to the restrictions implied by the Euler conditions. Here the estimand consists of all parameters indexing the laws of motion of the model variables including the so–called *deep structural* parameters that capture preferences and/or technology. In the case of 'linear–quadratic' DSGE models, Hansen and Sargent (1980) show this framework leads to VAR models for the economic variables in which the coefficients are subject to nonlinear restrictions. Outside this framework, except in special cases, imposition of the Euler conditions on the laws of motion for the economic variables estimation is more problematic both computationally and because consistency of the ML estimator may only be justifiable from the underlying economic model if the model variables are normally distributed.[25] Thus empirical rejections of the models may be due to the incorrect imposition of distributional assumptions purely for the purposes of implementing the ML estimation rather than mis–specification of the underlying economic model. Given these issues, attention turned to estimation of the deep structural parameters on their own based on information in the Euler condition. In an influential paper, Hansen and Singleton (1982) show the Euler condition can be leveraged into an unconditional population moment condition which can form the basis of a generalized method of moments (GMM) estimation of the deep structural parameters. They refer to this approach as generalized instrumental variable (GIV) because the moment condition is based on the orthogonality of the Euler equation residual to a vector instruments, which in this case are any variables in the agent's information set.

This type of application was one of the main motivations for the development of GMM; see Ghysels, Hall and Hansen (2002), Ghysels, Hall and Sims (2002). As is discussed below, GMM turned out to provide a statistical framework for addressing endogeneity in many other types models and applications as well.

5.3.4 The Econometrics of SEMs: What Has Aged Well?

In light of the critiques in Section 5.3.3, the fully-specified parametric SEMs in Sections 5.3.1-5.3.2 have fallen from favour as a basis for macroeconomic policy analysis. As discussed below, from the 1980's onwards attention turns toward alternative frameworks for modelling endogeneity and simultaneity among economic variables. Thus, this represents a good point in our narrative to assess the extent to which the SEM theory developed from 1940's-70's has 'aged well'.

Before doing so, we wish to reiterate a distinction made above between VARs and the LSEMs of the Cowles Commission. VAR models can be viewed as linear reduced form representations. However, they are different from the models in Section 5.3.1 in the sense that there are no exogenous variables. Crucially, the identification of the VAR parameters does not rest on endogeneity/exogeneity partition of variables that is a key ingredient of the Cowles Commission approach. Similarly, SVARs differ from structural forms for the same reason. SVAR models can be estimated by FIML, but

[25] For example, see Hansen and Singleton (1983). One remedy is to adopt a flexible functional form for the conditional distribution, see Gallant and Tauchen (1989).

identification issues re-surface for example Hamilton (1994)[p.331-333]. Thus, our assessment here does not include VAR/SVAR models.[26]

Consider first methods premised on the classical (normal) LSEM. Since a single equation in a LSEM is just an endogenous regressor model, all the LI methods described above have potential relevance in many other settings and so we postpone our assessment of their resonance in contemporary econometrics to Section 5.5. In contrast, the identification rules and FI estimation methods seem less relevant today because they more fully embrace the framework of fully-specified parametric SEM.

While the extension to NSEM allowed for more flexible functional forms, these parametric models were specified in terms of endogenous and exogenous variables and so are subject to many of the same criticisms levelled at their linear counterparts. As a result, methods designed for this framework-such as N3SLS and NFIML-have received relatively little attention since their introduction. That said the developments in statistical theory needed to accommodate the inherent nonlinearity of these models played an important role in laying the foundations of statistical analysis of the GMM estimator outlined in the next section.

5.4 Generalized Method of Moments

Generalized method of moments is introduced into econometrics by Lars Hansen in his seminal paper published in *Econometrica* in 1982. The approach taken in this paper can be viewed as marking a sea change in discussions of estimation in econometrics. The LI and FI methods described in Sections 5.3.2 and 5.3.2 were developed in the context of a specific model that specified a data generation process for endogenous variables as a function of exogenous/predetermined variables and errors. In contrast, the cornerstone to a GMM estimation is a population moment condition (pmc). Within GMM estimation the origins of the pmc are not important–what matters is that the pmc represents valid information about the model parameters of interest. The generality of this approach, allied with its computational simplicity, has meant that GMM can be–and has been–applied in a wide variety of econometric models.[27] One indication of the paper's impact can be found in the scientific background to the press release announcing the award of the 2013 Nobel prize for Economics to Eugene Fama, Hansen and Robert Shiller in which Hansen's 1982 paper is described as *"one of the most influential papers in econometrics"*.[28]

A *population moment condition* is the statement that the expectation of a vector of functions of the data variables and a subset of model parameters, θ, equals zero when evaluated at the true parameter value, θ_0, that is, $\mu(\theta_0) = 0$ say. This type of identity

[26] We refer the reader to Chapter 1 for an assessment of these models.

[27] See Hall (2005)[Table 1.1] for a list of applications in the period following the introduction of the method.

[28] See https://www.nobelprize.org/prizes/economic-sciences/2013/press-release/. The quote comes from p.24 of the Scientific background to this announcement

is referred to by Hansen (1982) as an *orthogonality condition*. Below we use these terms interchangeably as is common in the literature.

The GMM estimator, $\hat{\theta}$ say, is obtained by minimizing a quadratic form in the analogous sample moment, $\hat{\mu}(\theta)$, and a weighting matrix, $\hat{\mathbf{W}}$. Hansen (1982) shows that subject to certain regularity conditions the GMM estimator is consistent and he presents a first–order asymptotic theory that can form a basis for inference about the parameters.

As in the SEM estimators, this theory rests on the assumption that the pmc provides sufficient information to identify θ_0. In the GMM context, conditions for identification are defined in terms of the pmc. The proof of consistency of $\hat{\theta}$ is premised on global identification that is, $\mu(\theta) \neq 0$ for any $\theta \neq \theta_0$. The first–order asymptotic distribution theory is premised on the expectation of the Jacobian of the moment function being full column rank when evaluated at θ_0.[29] In this context, the parameters are said to be just–identified if the number of moment conditions, $dim\{\mu(\theta)\} = q$ say, equals the number of parameters, $dim(\theta) = p$, say; the parameters are said to be over–identified if q exceeds p.

If the parameter vector is over–identified then the first–order asymptotic distribution of the GMM estimator depends on the choice of weighting matrix. Hansen (1982) shows the optimal choice of $\hat{W}$ converges to the inverse of either the variance (i.i.d. case) or long run variance (stationary time series) of the sample moment function. This optimal choice is implemented in practice via a two–step GMM estimation.

The consistency of the GMM estimator rests critically on the validity of the pmc. If θ_0 is over–identified then Hansen (1982) shows $H_0 : \mu(\theta_0) = 0$ can be tested using minimized value of two–step GMM optimand. The resulting over–identifying restrictions test essentially generalizes to nonlinear models the homonymous test proposed by Sargan (1958).[30]

Early applications of GMM prompted extensions to the inference framework and statistical analysis provided by Hansen's original paper. One issue to receive attention was the efficiency of the estimator. Chamberlain (1987) shows that the two–step estimator GMM achieves the non–parametric efficiency bound for estimation based solely on the information in the pmc. Another consideration is what is the choice of moment pmc that yields the most efficient estimator. Although GMM technically encompasses ML, in most econometric applications GMM is employed because ML is not an option because the underlying economic model does not specify the distribution of the data.[31] Thus, in effect, the relevant question can be phrased as: what are the optimal moments to use if the most efficient choice (ML) is not available? Many early applications of GMM involved GIV, and for this scenario Hansen (1985)

[29] If the moment function is $\mathbf{m}(\theta)$–so that $\mu(\theta) = E[\mathbf{m}(\theta)]$–then the Jacobian is defined as $\mathbf{M}(\theta) = E(\partial \mathbf{m}(\theta)/\partial \theta')$.

[30] This connection was inadvertently missed in Hansen's original presentation, see Ghysels, Hall and Hansen (2002). See Arellano (2002) for a discussion of the connections between Sargan's work and GMM.

[31] For example see Hall (2005)[Chapter 3.8] or Hall (2015) for further discussion and references.

provides a method for characterizing the greatest lower efficiency bound of the GIV estimator over all possible sets of instruments.[32]

In the late 1980's - early 1990's concerns started to be raised about the quality of the first–order asymptotic theory as an approximation to the finite sample behaviour of the GMM estimator in various econometric models given the sample sizes available. Starting with Tauchen (1986), a number of simulation studies were published during this period; the July 1996 issue of *Journal of Business and Economic Statistics* contained seven papers devoted to the topic. A review of the findings is provided by Hall (2005)[Chapter 6.3] who concludes

> *"It seems fair to say that the evidence is mixed. In some models of interest the approximation can be good at sample sizes of 100, and in others it is bad even at samples a hundred times larger....Perhaps only one thing can be said for certain, that is finite sample behaviour is far more complex than would be predicted by [first–order] asymptotic theory."*(Hall, 2005, p.230)

The accumulated simulation evidence stimulated four types of development: alternative estimators; data-based methods for moment selection; alternative approximations to the limiting distribution theory; and estimation based on an increasing number of moment conditions as the sample size increases. Often these extensions were developed first in the context of the linear IV model and then subsequently extended to nonlinear models. Here we focus on the more general case. In the next section, as part of our discussion of a single linear equation estimated via IV, we return to some of these extensions because firstly linearity sometimes facilitates the derivation of stronger results and secondly to highlight some interesting connections to the Classical LSEM analysis.

Various alternative methods have been proposed for translating information in pmcs into estimates of the parameters. Chief among these are the continuous–updating GMM (CUGMM) (Hansen, Heaton and Yaron, 1996), Empirical Likelihood (Qin and Lawless, 1994) and Exponential Tilting (Kitamura and Stutzer, 1997) estimators. All three are special cases of two more general classes of estimators known as minimum discrepancy (Kitamura, 2007) and Generalized Empirical Likelihood (Smith, 1997). While these estimators are first–order equivalent to the (two–step) GMM estimator they can be shown to have some second–order advantages over GMM, see Newey and Smith (2004). Nevertheless, all are numerically more complicated to implement than GMM. In addition, concerns have been raised about whether GEL estimators have finite sample moments, see next section.[33] These considerations may explain in part the relatively limited impact of these alternative estimators on empirical practice in econometrics.

In practice, the underlying economic model may or may not be correctly specified and if true implies a candidate set of possible moment conditions upon which to base estimation. Data–based moment selection strategies are designed to provide a way of choosing the 'best' subset to use. Proposed methods vary in how they define 'best'.

[32] Also see Hansen, Heaton and Ogaki (1988).

[33] Hausman, Lewis, Menzel and Newey (2011) propose a modified version of the CUGMM estimator that possesses moments.

Andrews (1999) analyses the properties of data–based moment selection strategies based on the over–identifying restrictions test statistic. He introduces an information criterion (MSC) and compares its properties to sequential testing strategies. By construction, these methods focus on selecting the largest subset of pmc's from the candidate set that are satisfied at the same parameter value based on the orthogonality condition. Such a strategy can lead to the inclusion of redundant moment conditions.[34] While there is no cost to the inclusion of redundant moment conditions in terms of first–order asymptotic behaviour, their inclusion can affect the quality of this theory as an approximation to finite sample behaviour; see Hall and Peixe (2003). For GIV estimation, Hall and Peixe (2003) propose a sequential strategy based on first choosing the subset of pmc's that satisfy the orthogonality condition via MSC and then selecting from this subset using their information criterion designed to exclude redundant moments and so identify the *relevant* subset of pmcs. In cases where there is a subset of pmc's that identify the parameters and whose validity is not in question, Cheng and Liao (2015) provide a method for moment selection based on orthogonality and relevance in a single pass based on an adaptive Lasso–type criterion. Donald, Imbens and Newey (2009) also consider methods for instrument selection within the GIV framework. In their set–up, the candidate set of moment conditions are taken to be valid and the 'best' instrument is defined as the the one that minimizes the asymptotic mean–squared error of the GMM estimator. Their approach builds from the one presented by Donald and Newey (2001) for the linear IV model; see next section. Following the application of such data–based model selection methods, it would be desirable to perform inference based on standard methods as if the chosen moments had been selected *a priori*. While simulation evidence reported in many of the the studies cited above is generally supportive of this approach, Pötscher (1991) demonstrates that there are circumstances where this approach is not valid; concerns that may have limited applications of these data–based model selection in practice.

In the 1990's concerns began to be raised about whether the pmcs in many models of interest only *weakly identify* the parameters, *e.g.* see Nelson and Startz (1990) and Bound, Jaeger and Baker (1995). In broad terms, weak identification captures the situation in which the value of the moment function is 'relatively' insensitive to the parameter value and so is not informative about θ_0. To understand the consequences of weak identification, it is necessary to devise an appropriate mathematical model to capture this behaviour in a useful way. Following the analyses in Staiger and Stock (1997) (IV in linear models) and Stock and Wright (2000) (GMM in nonlinear models), weak identification is captured using a Pitman–drift term. In the case where the entire parameter vector is weakly identified then the population moment function $\mu(\theta)$ is assumed to equal a Pitman drift term that is proportional to one over the square root of the sample size N. Following the introduction of weak identification, it became customary to refer to the identification conditions underpinning the standard first–order asymptotic analysis as *strong identification*.

Stock and Wright (2000) extend this basic set–up to allow only a sub–vector of θ_0 to be weakly identified that is, $\theta_0 = (\alpha_0', \beta_0')'$ where α_0 is weakly identified and β_0

[34] A subset of pmcs are redundant if they do not contribute to the asymptotic efficieny of the GMM estimator given the other pmcs included. See Breusch, Qian, Schmidt and Wyhowski (1999).

would be strongly identified *if* α_0 were known. Stock and Wright (2000) show that GMM estimators of α_0 are random in the limit, the GMM estimator of β_0 are square root of N consistent but have a limiting distribution that is different from that derived under the assumption θ_0 is strongly identified. As a result, the standard confidence intervals based on inverting the corresponding GMM–based t–statistic are invalid, see Dufour (1997).

These findings motivated the development of alternative inference procedures. The resulting procedures deliver confidence sets for θ_0 by inverting a statistic whose limiting distribution is invariant to the quality of the identification. For nonlinear models, Stock and Wright (2000) suggest basing inference on the so–called S–statistic which is defined as the sample size times the CUGMM minimand, using the fact that if the moment pmc holds then the S-statistic converges in distribution to χ^2_q random variable. While valid, the degrees of freedom of the S–statistic exceed the number of parameters and so suggests scope for sharper inferences. To this end, Kleibergen (2005) proposes basing inference on his K–statistic which has a limiting χ^2_p distribution. The K–statistic is a modified version of GMM Lagrange Multiplier statistic for testing a point null hypothesis about the parameter vector in which the usual sample Jacobian is replaced by an alternative derived from CUGMM estimation. Kleibergen (2005) shows that this CUGMM–based Jacobian estimator is asymptotically uncorrelated with the sample moment function, and as a result the limiting distribution of the K–statistic is robust to the nature of the identification. As noted by Kleibergen (2005), a comparison of the power properties of the S– and K–statistics is nuanced depending not just on the comparative degrees of freedom but also on the validity of the pmc. Kleibergen (2005) provides extensions of his method that take into account information in the over–identifying restrictions, and also extensions that allow for only part of the parameter vector to be weakly identified. Unlike the traditional Wald confidence intervals–estimator $\pm\, 2\times$ the standard error– the confidence sets described above can be empty or disjoint sets or unbounded. Implementation may also face computational constraints that limit the size of p for which these methods are feasible.

One response to concerns about the quality of identification is to allow the number of moment conditions to increase with the sample size. Han and Phillips (2006) develop a first–order asymptotic theory for GMM estimators under very general conditions. Their analysis highlights how the properties of the estimator depend crucially on the way in which information about the parameters accrues as the sample size increases. Newey and Windmeijer (2009) derive the limiting distribution of GEL estimators under many weak moment asymptotics.[35] Compared to the standard first–order asymptotic analysis, a key difference is in the formula for the variance. As Newey and Windmeijer (2009) observe, their results can be viewed as an extension of Bekker (1994)'s results for LIML (see Section 5.3).

Strong and weak identification are just two such scenarios. Other scenarios have been suggested, such as "nearly–weak identification" , which as the name suggests

[35] Newey and Windmeijer (2009) also consider a jack–knife version of GMM but show it is less efficient than the GEL estimator under many weak moment asymptotics. See Section 5.5 for discussion of jack–knife IV in linear models.

lies in between strong and weak identification. For ease of presentation, we discuss the relationship between these three identification scenarios in the context of the linear IV model in the next section.

Aged well?: Over the forty plus years since its introduction, GMM has become an important part of the econometrician's toolkit. The strength of the approach is its basis on the information about the parameters in a pmc. As a result it provides a framework for estimation based on information deduced from economic theory without the need to fully specify the underlying economic model. In particular, the GMM framework provides a convenient way for estimating models involving endogeneity and simulateneity.

While there are some limitations to the original GMM inference framework, such as the assumption of strong identification, econometricians have been creative and successful in devising inference methods that are valid under more general scenarios. GMM has 'aged well'.

5.5 IV Estimation of the Endogenous Regressor Model

Endogenous regressor models arise in econometrics outside of the macroeconometric SEM context, as illustrated in Section 5.2. These contexts, such as the credibility revolution, raised new questions about inference procedures in endogenous regressor models. To a large part the answers were developed using a GMM perspective on IV estimation, and so in this section we also revisit some of the issues from the previous section on GMM.

For ease of presentation, we concentrate on the case in which the model involves a single endogenous regressor. We refer to this model as the 'structural equation' (of interest) below and denote its components by: y_1 (the dependent variable), y_2 (the endogenous regressor), $\mathbf{x}$ (the exogenous regressors/controls), and u (the error); in addition, $\mathbf{z} = (\mathbf{w}', \mathbf{x}')$ denotes the instruments.

Testing endogeneity: Outside of the structure of the Cowles Commission's LSEM with its imposed partition of the data vector into endogenous and exogenous variables, it may be desired to test whether y_2 is actually an endogenous regressor. Wu (1973) first introduced suitable tests based on a comparison of OLS and 2SLS estimators of the coeffcient on y_2. Hausman (1978) provides an alternative way of calculating the test based on the OLS estimators of the regression of y_1 on y_2, $\mathbf{x}$ and $\hat{v}_2$, where $\hat{v}_2$ is the prediction error from the regression of y_2 on the instruments. Under the null hypothesis that y_2 is an exogenous regressor the coefficient on $\hat{v}_2$ is zero. The OLS estimators of the remaining coefficients are numerically equivalent to the 2SLS estimators of the structural equation, see Hausman (1978). This alternative method for implementing 2SLS is an example of what has become known as the *control function* approach for estimation in the presence of endogeneity, see

Wooldridge (2015).

IV and GMM: Taking a GMM perspective, $\mathbf{z}$ is a valid instrument if it satisfies both the orthogonality condition, $E[\mathbf{z}u] = 0$, and the condition for global and first–order local identification, which in this case are the same: $rank\{E[\mathbf{z}(y_2, \mathbf{x})']\}$ equals the number of regression coefficients. The latter being commonly referred to as the relevance condition in the linear IV context.

As noted in Section 5.3, Sargan's presentation of 2SLS takes what would now be considered a GMM perspective. From that perspective, if u is conditionally homoscedastic given $\mathbf{z}$ then the (two–step) GMM is identical to the 2SLS estimator. Absent that condition, the 2SLS estimator is (potentially) less efficient asymptotically than the two–step GMM estimator. Pursuing the connection between the earlier LSEM methods and GMM further, Hansen et al. (1996) show that under the conditional homoscedasticity restriction, then their CUGMM estimator is identical to LIML. This equivalence implies that CUGMM does not have finite moments in this setting. On the basis of an extensive simulation study, Guggenberger (2008) conjectures a similar conclusion holds for the wider class of GEL estimators.

Instrument selection: Donald and Newey (2001) propose a method for selecting among valid instruments based on minimizing the second–order asymptotic mean–square error of either 2SLS or LIML. Their simulation evidence suggests that instrument selection generally leads to an improvement in estimator performance. Within their framework, the parameters are strongly identified and the number of instruments is allowed to increase with the sample size: for 2SLS this rate must be less than the square root of the sample size; and for LIML less than the sample size. Carrasco (2012) considers 2SLS and LIML estimation in the case where the number of instruments may exceed the sample size. She shows how regularization methods can be used to implement 2SLS but the same approach does not work with LIML. She provides conditions under which the regularized 2SLS estimator is square root of the sample size consistent with an asymptotic distribution that is normal and achieves the semi–parametric efficiency bound. In cases where the regression of y_2 on the instruments is sparse, Belloni, Chen, Chernozhukov and Hansen (2012) advocate use of instrument selection based on Lasso estimation and provide conditions under which this instrument selection does not affect the first–order asymptotic properties of the post–Lasso IV estimators of the parameters in the structural equation.

Weak identification: Staiger and Stock (1997) analyze the limiting behaviour of k–class estimators of the structural equation under weak identification. Within their framework weak identification is captured by assuming that in the population 'first-stage regression' of y_2 on the instruments, the coefficient on $\mathbf{w}$ is given by $\pi = \pi_N = N^{-1/2}\eta$ for some non–null vector of bounded constants η. In this context, weak identification derives from the properties of the instrument and so instruments are a commonly referred to as being weak

Staiger and Stock (1997) show that the k–class estimator of the coefficient on y_2 in the structural equation is not consistent and instead converges to a non–standard distribution. They further show that the asymptotic distributions of the 2SLS and LIML estimators of this parameter are different. This is in marked contrast to the behaviour of the estimators described in Section 5.3 within the Classical Normal LSEM within which the parameters are strongly identified. A key difference between the two settings is in the implied behaviour of the concentration parameter, κ_N say: under strong identification κ_N increases linearly with the sample size (N) and this drives the convergence to the first–order asymptotic distributions described in Section 5.3 but under weak identification κ_N converges to a bounded constant.

There are other interesting parallels to the finite sample results derived in the 1970's. Staiger and Stock (1997) show that the weak identification asymptotics deliver a limiting distribution for the 2SLS estimator that is the same as the finite sample distribution of the estimator derived within the Classical Normal LSEM. For LIML, the equivalence is to the finite sample distribution of the so–called LIMLK estimator, which is LIML implemented with a known error variance matrix.[36]

A number of identification–robust inference procedures have been proposed in the context of the linear IV setting. Dufour (1997) and Staiger and Stock (1997) show that a valid confidence interval for β_0, the coefficient on the endogenous regressor, can be obtained by inverting the AR statistic (Anderson and Rubin, 1949) for testing the point null hypothesis $\beta_0 = \beta_*$. Kleibergen (2002) proposes that inferences are based on his K–statistic, see Section 5.4. He shows that the resulting confidence interval for β_0 is centred on the LIML estimator of this parameter, and so can be viewed as an identification–robust method for performing inference based on LIML. This interval is therefore also non–empty: in contrast, confidence intervals constructed from the AR statistic can be empty; *e.g.* see Dufour (1997) or Staiger and Stock (1997). Moreira (2003) proposes a conditional likelihood approach to inference. Within his framework, the conditional likelihood ratio (CLR) test statistic has a non–standard distribution with simulated critical values provided by Moreira (2003). Simulation evidence reported in his paper shows that inferences based on the CLR statistic tend to be sharper than their counterparts based on either the K– or AR–statistics.

Within the linear IV model, the nature of the identification reflects the relationship between the endogenous regressors and instruments. It is therefore possible to assess nature of this relationship directly. Several tests have been proposed,[37] but within the single endogenous regressor model, the simplest approach is to base this inference on the first–stage regression. Within this framework, the coefficients of the structural equation are only identified if $\pi \neq 0$. Thus the null of no identification ($\pi = 0$) can be tested using the associated F–statistic. Early applications of this test used conventional critical values, but Stock et al. (2002) demonstrate that rejection does not guarantee conventional first–order asymptotic inference procedures are reliable. Instead, based on an extensive simulation study, they provide new guidelines based on whether the focus of inference is point estimation or hypothesis testing. These point to using

[36] A similar result applies for modified LIML estimators such as the one proposed by Fuller (1977), see Staiger and Stock (1997).

[37] See *inter alia* Cragg and Donald (1993), Shea (1997) and Hall, Rudebusch and Wilcox (2006).

much larger critical values, often summarized by the statement that 'the first–stage F needs to be at least 10'–a rule–of–thumb that conveys the spirit of the adjustment but grossly over simplifies Stock, Wright and Yogo's recommendations.[38]

For the case where there are multiple endogenous regressors, Stock et al. (2002) advocate testing for weak identification based on the Cragg–Donald (1993) statistic for testing the rank of a matrix. Similar issues arise regarding the choice of critical value as described above for the F–statistic; see Stock et al. (2002). Kleibergen and Paap (2006) provide a rank–based statistic that can be used in this context when the data are not i.i.d.

Estimation with many instruments: Chao and Swanson (2005) derive the probability limit of the 2SLS and LIML estimators in cases where the number of instruments is allowed to increase at a rate equal to or less than the sample size. Their analysis reveals that whether or not the estimator is consistent depends crucially on the rate which the concentration parameter increases with the sample size. For 2SLS, consistency requires the concentration parameter to increase at a faster rate than the number of instruments, whereas the consistency of LIML requires only that the concentration parameter increases faster than the square root of the number of instruments.[39] However, the LIML estimator is calculated under the assumption that the errors are conditionally homoscedastic. If there is neglected heteroscedasticity, LIML becomes inconsistent as the number of instruments becomes large, see Bekker and van der Ploeg (2005). If the heteroscedasticity is accommodated through the use of CUGMM estimation then then asymptotically valid inference can be performed using the methods in Newey and Windmeijer (2009).

Split–sample and jackknife IV: As noted above, 2SLS exhibits a finite–sample bias. Angrist and Krueger (1995) observe that this finite sample bias of the 2SLS can be traced to the correlation between the estimated coefficients used to construct the first–stage predicted values and the errors of the structural equation. To break this correlation, Angrist and Krueger (1995) propose *Split–sample IV (SSIV)* estimation. This involves randomly splitting the sample into two halves with one half used to calculate the first–stage coefficients and the other half used to estimate the parameters of the structural equation. The latter estimation only differs from the second–stage of 2SLS in the construction of the predicted value of $y_{1,i}$: for SSIV this predicted value is calculated using the estimated first–stage coefficient from the other half of the sample. If the instruments are strong then the asymptotic variance of SSIV is double that of 2SLS due to the sample splitting. However if the instruments are weak then Angrist and Krueger (1995) show SSIV has more attractive finite sample bias properties than 2SLS and argue these can outweigh the efficiency loss in practice.

---

[38] Also see Olea and Pflueger (2013).

[39] Chao and Swanson (2005) show that a bias–corrected version of 2SLS exhibits similar behaviour to LIML.

An unattractive feature of SSIV is its dependence on the sample split. Angrist, Imbens and Kreuger (1999) propose a jackknife IV estimator (JIVE) that achieves the same goal of breaking the correlation between the first–stage coefficient estimators and the error in the equation of interest but without the arbitrary data split. Parenthetically, we note that this estimator had previously been introduced by G. D. A. Phillips and Hale (1977) but at a time prior to the concerns about weak identification which motivated SSIV. Like SSIV, JIVE differs from 2SLS in how the predicted value of the endogenous regressors is calculated. With JIVE, the predicted value of $y_{2,i}$ is calculated using the estimated first–stage coefficients based on a sample consisting of all observations *except for the* i^{th}. Angrist et al. (1999) show that JIVE has the same first–order asymptotic distribution as 2SLS under strong identification and so is more efficient than SSIV. They report simulation evidence that in models with many weak instruments JIVE dominates 2SLS in terms of median–bias and confidence interval coverage rate, and is similar to the LIML estimator in terms of these two metrics.

Other identification scenarios: Under strong identification the concentration parameter goes to infinity at the same rate as the sample size, and under weak identification the concentration parameter converges to a bounded constant. Clearly, the behaviour allowed by Chao and Swanson (2005) allows for different rates of increase between these two extremes. Hahn and Kuersteiner (2002) explore the behaviour of the 2SLS estimator with a fixed number of instruments in the case where $\pi_N = N^{-c}\eta$ for $0 \leq c < \infty$. Their framework encompasses strong identification ($c = 0$) and weak identification ($c = 1/2$); they refer to the range in between as *nearly–weak identification*. They show that $c = 1/2$ represents a point of discontinuity in the sense that the limiting behaviour of the 2SLS estimator is qualitatively the same under strong and nearly–weak identification but different for weak identification. This discontinuity can be explained in terms of the concentration parameter: if $0 \leq c < 1/2$ then the concentration parameter tends to infinity with N albeit at a rate that depends on c.

Aged well? As demonstrated above, there is a natural connection between IV/2SLS/LIML estimation in a LSEM and GMM/CUGMM estimation of the endogenous regressor model. As a result, aspects of the earlier literature resonate in contemporary applications. At the same time, some details are different. In the Cowles Commission approach, the choice of instrument is based on exclusion restrictions defined via reference to a fully-specified LSEM. In more modern times, the choice of instruments is justified via a conditional moment restrictions deduced from economic theory. In the earlier LSEM context, the standard asymptotic analysis is premised on strong identification and a fixed number of instruments, a framework that was subsequently found to provide a poor approximation to the finite sample behaviour in certain settings of interest. The GMM perspective has allowed alternative approximations to be developed based on many and/or weak instruments.

The exact finite sample theory of LI estimators in the LSEM is derived under conditions that were satisfied by neither the macroeconometric variables of the 1940's nor by variables in subsequent applications (such as in the credibility revolution). However, while this theory did not deliver new inference procedures, it does provide useful insights into the behaviour of 2SLS/LIML estimators, and it is striking how many of these insights carry onto the GMM/CUGMM estimators in more general contexts.

In summary, IV estimation of the endogenous regressor model has been an important part of econometrics since the earliest days of the field. Over the years implementation of IV-based inferences may have changed but the essence of the method has remained the same. This endurance testifies to the method having 'aged well'.

5.6 Nonparametric Models

Although the developments of LSEM brought in key insights and methods for tackling endogeneity, the simplicity of the linear model sometimes masks both the scope and limitations of the available methods. Along with the advances in the estimation of parametric nonlinear methods, economists turned their attention to adopting, and building on, the statistical literature on nonparametric regression estimation and flexible curve fitting. We refer the reader to Chapter 11 for a discussion of several aspects of estimation and inference in nonparametric models. The last three decades also witnessed significant progress in understanding identification in nonparametric systems with endogeneity and/or simultaneity. Early applications of nonparametric methods in economics included the estimation of consumer demand and Engel curves; see e.g. the discussion in Banks, Blundell and Lewbel (1997). This subsection discusses some of these developments.

5.6.1 Conditional Exogeneity in Nonseparable Systems

Consider extending the single linear equation discussed in Section 5.5 to its nonparametric analogue, $y_1 = r(y_2, \mathbf{x}, u)$ where we maintain that u is a scalar for simplicity. We note two distinguishing features of this nonparametric structural function. First, similar to nonlinear models, the 'effect' of y_2 on y_1 is no longer a constant. Instead, this effect can depend on the dosage, i.e. on the value of y_2, and on the value of the covariates $\mathbf{x}$. Second, this effect exhibits unobserved heterogeneity in that it can vary with u. Specifically, assuming differentiability, the marginal effect of y_2 on y_1 (the derivative of r with respect to y_2) given x can vary with y_2 and u. Attention often focuses on the identification of features of the distribution of this marginal effect. Among these, the conditional average marginal effect given $\mathbf{x}$ and the average marginal effect are particularly salient. These quantities obtain by averaging the marginal effect over the distribution of $u|\mathbf{x}$ and $(u, \mathbf{x})$ respectively. When the structural function is linear, these average effects coincide with the constant slope coefficient on y_2.

Whereas conditional means such as $E(y_1|y_2, \mathbf{x})$ are readily estimable, they generally fail to identify the conditional average structural function. Nevertheless, contrasting these quantities reveal a natural extension of the notion of conditional exogeneity, namely that u is independent of y_2 conditional on $\mathbf{x}$. Under this condition, and given regularity conditions, the derivative of the conditional mean $E(y_1|y_2, \mathbf{x})$ identifies the conditional average marginal effect given $\mathbf{x}$. See, e.g. Blundell and Powell (2003), Altonji and Matzkin (2005), and White and Chalak (2013). As in the linear case, the presence of an omitted variable can lead to a failure of conditional exogeneity. In this case, a nonparametric regression will fail to identify the average marginal effect of y_2 on y_1. Chalak (2019) derives a nonparametric analogue of the omitted variable bias formula. Here too, the bias depends on the effect of the omitted variable on the outcome y_1 and on the dependence of the omitted variable on the endogenous variable y_2.

5.6.2 Separable Equations

5.6.2.1 Separability in the Outcome Equation

The literature explores the identifying power of requiring the structural function to be additively separable, with u entering linearly. This removes the unobserved heterogeneity in the effect of y_2 on y_1. Nevertheless, under this assumption, one can rely on instrumental variables $\mathbf{z} = (\mathbf{w}', \mathbf{x}')'$ that satisfy the familiar notion of mean exogeneity from parametric models, $E(u|\mathbf{z}, \mathbf{x}) = E(u|\mathbf{x})$. Newey and Powell (2003) show that in this case the identification of the structural function reduces to the existence of a unique solution of an integral equation or, equivalently, to the conditional distribution of y_2 given $\mathbf{z}$ being complete in $\mathbf{w}$. See also Darolles, Fan, Florens and Renault (2011) who cast this problem as the solution of an ill–posed inverse problem and propose a regularized estimation procedure.

5.6.2.2 Triangular Systems and the Control Function Approach

Alternatively, the control function approach restores the general nonseparable equation for y_1 and tackles endogeneity by relying on the assumption that the equation for $y_2 = q(\mathbf{w}, \mathbf{x}, v)$ is either separable (e.g., Newey, Powell and Vella (1999)) or nonseparable yet monotonic in the unobservable v (e.g., Imbens and Newey (2009)). In particular, if the instrument is exogenous given the covariates then conditioning on v and the covariates ensures that y_2 is conditionally exogenous. The literature makes use of the monotonicity or separability of q to estimate v using the y_2 equation and to then use this estimate to control for the endogeneity of y_2, analogously to the linear equation analysis.

In the fully nonseparable case, Schennach, White and Chalak (2012) show that local indirect least square methods generally fail to recover average marginal effects. Nevertheless, they can be used to test the hypothesis of a zero effect.

5.6.3 Binary Treatment

The case of a binary treatment y_2 occupies a central space in the literature. Here, $r(1,\mathbf{x},u)$ and $r(0,\mathbf{x},u)$ play the role of 'potential outcomes' and interest often centers on the identification of the conditional average treatment effect (CATE) and the average treatment effect (ATE). These parameters average the difference in the potential outcomes over the distributions of $u|\mathbf{x}$ and $(u,\mathbf{x})$ respectively. When y_2 is conditionally exogenous, the potential outcomes are strongly 'ignorable,' , i.e. they are conditonally independent of y_2 given the covariates (see, Rosenbaum and Rubin (1983)). The identification of CATE and ATE proceeds analogously to the discussion above on measuring marginal treatment effects. Rosenbaum and Rubin (1983) demonstrate that the propensity score, defined as the probability of receiving the treatment given the covariates, plays a central role here. In particular, provided the propensity score is bounded away from 0 and 1, conditioning on this scalar, rather than on the vector of covariates, suffices to ensure that y_2 is conditionally exogenous.

Imbens and Angrist (1994) consider the case in which y_2 is endogenous but an exogenous instrument is available. Provided the treatment obeys a 'monotonicity' assumption, Imbens and Angrist (1994) show that the Wald estimand recovers a local average treatment effect (LATE). To illustrate, consider the case of a binary instrument as in Angrist, Imbens and Rubin (1996). The population can then be partitioned into always takers, never takers, compliers, and defiers. Monotonicity rules out the defier subpopulation. In this case, the Wald estimand recovers LATE, the average treatment effect for the subpopulation of compliers.

Vytlacil (2002) shows that monotonicity is equivalent to imposing a threshold crossing restriction on the treatment selection equation, akin to the earlier literature on discrete choice (see McFadden (1974)). Using this representation, Heckman and Vytlacil (2005) show that local indirect least squares estimand recovers the marginal treatment effect (MTE) parameter, which measures the average treatment effect for those who are indifferent as to whether to select into the treatment at a given value of the instrument. In turn, MTE serves as a building block unifying several treatment effect parameters.

5.6.4 Partial Identification

As the above discussion reveals, standard estimands, such as the Wald estimand, (point) identify different local average treatment effects, depending on which instrument is used and even on which value of the instrument is considered. This distinction does not arise in, e.g., linear models where parameters of interest are constants and defined a priori. Accordingly, a strand of the literature fixes the parameters of interest, e.g. ATE, at the outset and characterizes the sharp identification region for these parameters, i.e., the tightest set of parameter values that are consistent with the data under the imposed assumptions. Manski (1990) provides such as an analysis for ATE under various assumptions on the outcome and treatment.

In the case of binary y_1, y_2, and an exogenous instrument that is excluded from the y_1 equation, Balke and Pearl (1997) characterize the identification region for ATE without requiring monotonicity. Manski and Pepper (2000) weaken the standard instrumental variable assumption that the conditional mean potential outcome does not depend on the instrument given the covariates to require that it is instead monotonic in the instrument. They derive the identification region for ATE under this assumption and other restrictions on the outcome and treatment. See also Chesher and Rosen (2017) who study the partial identification of structures in generalized instrumental variable models, extending notions of observational equivalence in complete models (see, e.g., Hurwicz (1950)).

The developments in partial identification ushered in research on estimation and inference for identified sets or partially identified parameters. This includes work that directly extends the assumption of global identification in GMM that is, $\mu(\theta) \neq 0$ for any $\theta \neq \theta_0$, underlying consistency of $\hat{\theta}$. In particular, Chernozhukov, Hong and Tamer (2007) dispense with the assumption of a unique, i.e. point–identified, θ_0 and consider a set Θ_0 of minimizers of an econometric criterion function. The paper develops estimators and confidence regions as suitable contour sets of the sample analogue criterion function. When interest attaches to a partially identified parameter, rather than the identification set for it, the confidence region for the set may be conservative. Imbens and Manski (2004) put forward confidence intervals that cover the partially identified parameter directly with a prespecified probability. Other key work include Beresteanu and Molinari (2008) who make use of random set theory to tackle identification and inference in partially identified models and Kline and Tamer (2016) who study a Bayesian approach to inference in models characterized by a known mapping from estimable reduced-form parameter to the identified set for a partially identified parameter.

5.6.5 Nonparametric Systems of Equations

The above results often focuses on a single outcome equation and, possibly, on a selection equation for the treatment. More generally, LSEM generalize to nonparametric systems involving J structural equations. When the direction of the dependence generates a sequential ordering among the system variables, with predecessor variables affecting successor variables, the system admits no feedback loops and is said to be recursive. Reviving S. Wright (1921)'s work on path analysis, Pearl (2000) studies nonparametric recursive systems and demonstrates the usefulness of graphical methods in characterizing the identification of causal effects. In particular, graphical criterion, such as d–separation, can help determine whether conditioning on a certain set of variables suffices to ensure the conditional exogeneity of a variable or, more generally, to secure identification of a causal effect without recourse to strong restrictions on the structural functions.

As discussed in White and Chalak (2009), recursive systems rule out notions of simultaneity and equilibrium which are central in economics. Matzkin (2008) studies identification in simultaneous nonparametric systems and shows that under certain monotonicity and independence conditions, identification can be characterized by a restriction on the rank of a matrix, echoing earlier results on identification in linear systems discussed in Section 5.3. See also Angrist, Graddy and Imbens (2000) for work tackling instrumental variable estimation of a nonparametric demand and supply model.

5.6.6 Beyond parametric methods: what has aged well?

A main feature of nonparametric models is their ability to accomodate unoberseved heterogeneity. Because this is a common feature of many economic models, nonparametric models have 'aged well' and have become increasingly popular especially with the assimilation of machine learning techniques in econometrics; see Chapter 12. The advances discussed in this section demonstrate that methods that make use of covariates to ensure conditional exogeneity exhibit remarkable robustness to functional form and generalize well to nonparametric structures that accommodate unobserved heterogeneity. At the same time, work extending instrumental variable methods, a cornerstone toolkit in econometrics and applied work, to account for unobserved heterogeneity, revealed a wider scope of instrumental variables methods and a more nuanced interpretation of the objects that these methods recover. Taken together, this work revealed the important role that notions of additive separability and monotonicity, which are implicitly assumed in linear models, play in hindering or facilitating the identification of useful objects in nonparametric models. With unobserved heterogenous present, the literature began to place more emphasis on what can be learned about certain features of various effects under given assumptions, and on characterizing and estimating the resulting identification sets. While important progress has been made generalizing identification results in simultaneous systems, there remains room to further study identification and estimation in the most general form of these systems. As methods get refined and extended, the 'nonparametric approach' will likely continue to be play a key role in flexibly accommodating unobserved heterogeneity.

5.7 Measurement Error and Other Data Issues

As discussed in the Introduction, endogeneity can arise due to measurement error. We discuss linear, nonlinear, and nonparametric measurement error models and briefly describe other data issues, such as data combination, that can also lead to endogeneity.

5.7.1 Measurement Error in Linear Systems

Consider the single linear equation for y_1, with $y_2, ..., y_J$ serving as explanatory variables and u as equation disturbance. Suppose that the econometrician observes error–laden measurements of the expanatory variables. In its most basic additive form, this takes the form $y_j = y_j^* + \varepsilon_j$ for $j = 2, ..., J$ where ε_j denotes the measurement error and the star notation indicates that the econometrician does not directly observe the realizations of the explanatory variables y_j^*. The classical measurement error model assumes that ε_j is uncorrelated with y_j^* and with the disturbance u. To fix ideas, set $J = 2$. Then, after substituting for $y_2^* = y_2 - \varepsilon_2$ in the y_1 equation, endogeneity arises due to the correlation between y_2 and the implied regression error $u - \beta_0 \varepsilon_2$ which arises due to their joint dependence on ε_2. As is evident from this description, the resulting bias of the OLS estimator depends on the measurement error variance.

Koopmans (1936) considers inference within the multiple regression model in which all explanatory variables are subject to measurement error. He proposes a framework in which the true values of the variables are treated as unknown parameters to be estimated along with the usual regression parameters, and the variance–covariance matrix of the measurement errors is known up to scale. He proposes ML-based confidence intervals for the regression coefficients and also examines how measurement errors impact inferences based on procedures that ignore their presence. Allen (1939) proposes a method for adjusting least squares estimates of the coefficient in regression through the origin based on the observed series, with the adjustment based on assumptions about aspects of the measurement error distribution.

Reiersøl (1941) and Wald (1940) developed instrumental variable methods to tackle the resulting endogeneity. For the simple regression model, Wald (1940) shows that the slope parameter can be estimated consistently by splitting the observations into two groups and taking the ratio of the difference in the group means of the dependent variable to the difference in the group means of the regressor. Durbin (1954) shows this is equivalent to estimating via IV estimation using a dummy variable for group membership as instrument. A natural candidate instrumental variable is a second measure whose measurement error is uncorrelated with the first measure's error as well as with the latent explanatory variable and the disturbance u.

Interestingly, Gini (1921) and Frisch (1934) proposed an earlier different approaches to estimating β_0. In particular, they derive bounds on β_0, providing one of the earliest partial identification results in econometrics. While β_0 cannot be consistently estimated under the classical measurement error model alone, the Gini–Frisch bounds establish that β_0 must lie on the line segment with one endpoint corresponding to the slope coefficient in the 'forward' regression of y_1 on y_2 and the other endpoint corresponding to the inverse of the slope coefficient in the 'reverse' regression of y_2 on y_1. Klepper and Leamer (1984) extend this analysis to the case in which $J > 1$, and demonstrate that the multivariable case poses significant difficulties for identification. Leamer (1987) and Chalak and Kim (2020) consider the case of a system of equations, demonstrating the gain that arises when analyzing multiple equations jointly. Chalak and Kim (2021) show that relaxing the exclusion restriction, to allow both y_2^* and y_2 to enter the y_1 equation, invalidates the Gini–Frisch bounds . This demonstrates the crucial role that the exclusion restriction plays for identification and the paper puts forward a number of auxiliary assumptions that can be used to recover useful bounds.

The classical model assumes that the unobserved variables and errors are pairwise uncorrelated. Another strand of the literature explores the identification gain that arises when strengthening this assumption, for example, to assume that these variables are jointly independent. Reiersøl (1941, 1950) and Geary (1942) show that higher order moments, such as the skewness of the distribution, can be used to identify and consistently estimate β_0. On the one hand, this approach requires that higher order moments carry useful information, which rules out that y_2^* is normally distributed for example. On the other hand, it does not require the availability of instruments or second measurements.

In the 2000's, measurement error has also emerged as an important factor in the estimation of impulse response functions in macroeconometrics. As noted in Section 5.3.3, Sims (1980) advocated the use of VAR models as a basis for macroeconomic analysis in place of LSEMs. Impulse response functions (IRFs) capture how the value of a VAR innovation at time t affects the contemporaneous and future values of the system variables. While IRFs reveal the dynamic interrelationships between system variables, they do not as a rule reveal how the system responds to structural (exogenous) shocks such as tax changes or monetary policy shocks.

Initial responses to the failure of VAR models to identify structural shocks were two fold: (i) assume the system is generated by a structural VAR (SVAR) model; (ii) the narrative approach. SVARs are an extension of VAR models in which a system variable at time t can depend on the contemporaneous value of the other system variables. Within this approach, economic theory is used to deduce conditions on the coefficients under which the structural shock of interest can be identified from the SVAR innovations. The narrative approach involves the estimation of the structural shock from some external data source and then its incorporation in the model. As pointed out by Mertens and Ravn (2013), these estimated shocks are most likely subject to measurement error that may, if ignored, undermine inferences. However, they demonstrated that the two methods can be combined to provide a third way commonly referred to as SVAR–IV.[40] The essence of SVAR–IV is that these narrative variables can serve as instruments that can be used to identify the structural shock within the SVAR model.

[40] This acronym is used by Stock and Watson (2018).

Local projections (Jordà, 2005) provide an alternative method for estimation of impulse responses. In line with this approach, Stock and Watson (2018) provide a framework that connects the impulse responses to the dynamic causal impact of a structural shock on a system variable. Under certain assumptions, they show that the h-period ahead impulse response equals the coefficient in the population regression through the origin of the system variable in period $t+h$ on the structural shock in period t. If the structural shock is observed then this coefficient can be consistently estimated via OLS, but if the shock series is only measured with error then this is no longer the case for the reasons discussed above. However, the estimated shock series can be used as an instrument in a local projection (LP–IV) estimation (Jordà, Schularick and Taylor, 2015). Stock and Watson (2018) provide conditions under which LP–IV estimator is consistent for the the impulse response.

5.7.2 Measurement Error in Nonlinear and Nonparametric Systems

Several of the methods discussed in Section 5.7.1 extend to nonlinear and nonparametric models. Chen, Hong and Tamer (2005) demonstrate how the availability of auxiliary validation data can be used to point identify the parameters in nonlinear models with measurement error. The key assumption they employ is that the conditional distribution of the latent variable given its observed measure coincide in the population and the auxiliary data. Schennach (2014) considers nonlinear models defined by moment conditions involving both observed and latent variables, which encompass measurement error models. The paper demonstrates how using a least favorable entropy–maximizing distribution can yield equivalent moment conditions that involve only observable variables.

Extensions to nonparametric separable equations $y_1 = \tilde{r}(y_2^*) + u$ with $y_2 = y_2^* + \varepsilon_2$, includes the work of Fan and Truong (1993) who assume that the measurement error ε_2 has a known distribution and make use of independence assumptions and deconvolution methods to identify and estimate the model. Schennach and Hu (2013) do not require that the distribution of ε_2 is known but maintain that y_2^*, u, and ε_2 are mutually independent, mirroring the assumptions imposed in work invoking higher order methods in linear models discussed in Section 5.7.1. Remarkably, they show that this model is generally point identified except in special yet central cases, including that of a linear $\tilde{r}$ with normally distributed variables.

Several papers derive bounds on quantities of interest in nonparametric models. Horowitz and Manski (1995) consider settings in which realizations of the data can be either erroneous or error free. They consider the contaminated and corrupted sampling cases, defined by whether the error occurs independently from the sample realization. The paper derives bounds on means and other distributional features by imposing a maximal error rate. Chesher (1991) studies the consequences of measurement error by providing small error variance approximations to regressions functions and nonparametric distributions involving mismeasured variables. Returning to the original insights underlying the Gini–Frisch bounds, Chalak (2024) generalizes these bounds to nonparametric nonseparable models $y_1 = r(y_2^*, u)$ with $y_2 = y_2^* + \varepsilon_2$. The nonparametric Gini–Frisch bounds obtain under stochastic monotonicity conditions and make use of nonparametric forward and reverse regressions to bound the average marginal effect of y_2^* on y_1.

5.7.3 Two–sample methods

A further complication arises when data on y_1 and y_2 appear in separate data sets. Angrist and Krueger (1992) consider such a setting when estimating the effect of age at school entry on educational attainment. They point out that, while the data on y_1 and y_2 are not linked, instrumental variables z, such as quarter of birth dummies, may be available in both data sets. In this case, one can estimate the slope coefficient on y_2 via instrumental variable methods by separately estimating the covariance matrix of y_1 and z and that of y_2 and z. Inoue and Solon (2010) show that in the two–sample context, the instrumental variable and two stage least squares estimators are numerically distinct and that the latter estimator is asymptotically efficient relative to the former. For the partially linear model, D'Haultfœuille, Gaillac and Maurel (2025) use tools from optimal transport theory to bound the slope coefficient on y_2 in the y_1 equation when the instrument may fail the exclusion restriction. See also Manski and Tamer (2002) who show how monotonicity and mean independence conditions can be used to bound the nonparametric regression $E(y_1|y_2, x)$ when precise data on either y_1 or y_2 (but not both) is available, with only interval data available for the other variable.

5.7.4 What aspects of measurement error models aged well?

That error in variables affects inference in linear regression analysis, and can lead to attenuation bias, has now been understood for more than a century. Nevertheless, this remains an active area of research, underscoring that measurement error models have 'aged well.' Recent methods demonstrate the consequences of measurement error in more general nonparametric models, often focusing on the case of a scalar mismeasured variable. Several of the insights on using higher order moments proved helpful in this respect. Here too, the separability and/or monotonicity of the functions in the unobservables plays an important role in facilitating identification. Despite these advances, more work lies ahead, including to better understand the consequences of measurement error in general models with multiple mismeasured variables.

5.8 Endogeneity in Other Settings

Endogeneity naturally arises in econometric models in a number of settings. Here we briefly consider three scenarios of empirical relevance: dynamic linear panel data models, mixed regressive spatial autoregressive models, and linear time series models. Since all three are covered elsewhere in this volume, we seek here only to provide further illustrations of how endogeneity and simultaneity occur in econometric models and to relate the proposed estimation methods in these contexts to the methods discussed above.

5.8.1 Dynamic Linear Panel Data Models

Suppose that observations are indexed by (i,t) where i and t indicate respectively the cross–sectional and time dimensions, and let $y_{i,t}$ denote the dependent variable. Further assume that cross–sectional observations, N, becomes asymptotically large while the number of observations over time, T, remains fixed. Within this setting, the error is assumed to consist of two components: the idiosyncratic error, indexed by (i,t), assumed uncorrelated with the explanatory variables; and the 'fixed effect', indexed by i alone, that may be correlated with the explanatory variables. By definition, the presence of a fixed effect means that some or all of the explanatory variables are endogenous regressors.

One potential way forward is to estimate the time–differenced version of the equation via OLS. However, if $y_{i,t-1}$ is included as an explanatory variable in the original model then its time- difference is an endogenous regressor in the transformed model, and so this strategy does not yield consistent estimators. To circumvent this problem, Anderson and Hsiao (1981) propose using an IV approach to estimation of the time–differenced version of the model. Their approach involves using $y_{i,t-2}$ (or its time-difference) as an instrument.

In a very influential paper, Arellano and Bond (1991) point out that in cases where $T > 4$, the panel data structure implies that longer lags of $y_{i,t}$ are also available as instruments, and propose a GMM approach to estimation that fully exploits this information. However, this generic approach has two drawbacks. First, while the inclusion of these additional pmcs has the potential to increase the asymptotic efficiency of the estimator, there is likely a diminishment in the incremental information gains as longer lags are used which may manifest itself adversely in the finite sample behaviour of the estimator, see Roodman (2009) for further discussion and solutions. Second, Blundell and Bond (1998) demonstrate that if the model is estimated via Arellano & Bond's method then parameters are weakly identified in certain parts of the parameter space. These concerns lead Blundell and Bond (1998) to propose a 'systems GMM' estimator that is based on augmenting the Arellano & Bond type pmcs with an additional set that can be shown to hold under a specific assumption regarding the initial condition, $y_{i,1}$. Inference can then be based on standard first–order asymptotic GMM theory premised on strong identification . Their approach provides an example of circumventing the problems caused by weak identification through the inclusion of additional information. We concur with the sentiment expressed in Chapter 2 of this volume that these methods 'have been a success and have aged really well'.

5.8.2 Univariate Linear Time Series Models

Following the pioneering work of Box and Jenkins (1970), autoregressive moving average (ARMA) models have become popular forecasting economic time series. In this forecasting context, the parameters are typically of little interest in their own right. However, there are some cases in which there is a natural interest in some or all of the AR parameters. These parameters can, of course, be obtained by estimating an ARMA model. As an alternative, IV methods can be used. Originally proposed in the control theory literature, [41] this approach had some currency in the context of testing for unit roots in macroeconometric series, a topic that received a lot of attention following the publication of the study by Nelson and Plosser (1982) study.[42] Fuller (1976) and Dickey and Fuller (1979) consider the case in which a AR(p+1) process is estimated and it is desired to test whether the process contains a unit root that is, it is generated by an ARIMA(p,1,0) process. In its simplest form, their test is based on the regression of the first difference of the series on its lagged level and the lagged first differences up to lag $p-1$. The validity of their approach rests critically on the lagged first differences being exogenous regressors in his model. However, these variables become endogenous regressors if the true data generation process is ARIMA(p,1,q) with $q > 0$. Valid inference based on statistics with the Dickey–Fuller distributions can be restored in a number of ways: (i) by inverting the MA polynomial and basing inference on an AR($p_T + 1$) model where $p_T \to \infty$ with the sample size, T, (Said and Dickey, 1984); (ii) estimating an ARMA model (Said and Dickey, 1985); (iii) estimating an AR(1) model by OLS but applying a non–parametric correction to the test statistics that removes the "bias" caused by omitting the other components of the model in the estimation (P. C. B. Phillips, 1987); (iv) estimating the AR coefficients of the ARMA model via IV with instruments constructed from appropriately chosen lagged values of the series (Pantula and Hall, 1987). Both (ii) and (iv) require (some) knowledge of p, q which makes these approaches less attractive. Methods (i) and (iii) have become the standard procedures for testing the null hypothesis of a unit root in economic time series.

[41] See Stoica, Söderstroöm and Friedlander (1985).

[42] IV methods have also been applied in the estimation of linear models with MA errors that arise due to multi–period conditional moment restrictions but in this context the regressors are exogenous and the issue is efficiency, for example see Hansen and Singleton (1996). Kuersteiner (2001) considers optimal IV estimators for the parameters of ARMA models.

5.8.3 Mixed Regressive Spatial Autoregressive Models

In cross–sectional spatial data, the variables are indexed by location i, say. Mixed regressive spatial autoregressive (MRSAR) models explain the dependent variable y_i as a linear function of the exogenous variables specific to location i, $\mathbf{x}_i$, an error term specific to location i, and also the dependent variables and (potentially) the exogenous variables specific to other locations. The spatial dependence is typically controlled via a fixed spatial weights matrix that is chosen by the practitioner *a priori*. The spatial spillover between locations means that the outcomes of the dependent variables at all locations are simultaneously determined as functions of the exogenous variables and errors at all locations.

Various approaches have been proposed for consistently estimating the parameters of the model. Kelejian and Prucha (1998) propose an IV/2SLS that exploits the orthogonality of exogenous variables and their spatial lags to the error. Lee (2007) shows that if the pmcs implicit in IV/2SLS estimation are augmented by pmcs based on the second moments of the errors then the resulting GMM estimator can be asymptotically as efficient as the ML estimator derived under the assumption that the errors have a joint normal distribution. Lee and Yu (2014) consider an extension of this GMM method to spatial dynamic panel data models. We refer the reader to Chapter 7 of this volume for an appraisal of how these methods fit into the evolution of spatial econometrics.

5.9 Concluding Remarks

In this chapter, we review how endogeneity and simultaneity have played an important role in the evolution of econometrics as a field. In line with a theme of this volume, we use this section to pull together some of our earlier comments about what methods/ideas have “aged well” and what ideas less so, recognizing that perspectives on this question could well change as the subject continues to evolve.

This chapter began with a review of the statistical theory of estimation and inference in LSEM developed during the 1940’s–70’s and motivated to a large extent by business cycle applications. So it seems reasonable to start our appraisal by considering resonance of this body of work in the econometrics of 2026.

Due to computational constraints, LI methods were often the only feasible option in the 1940’s–1950’s and the development of IV/LIML/2SLS methods was a crucial first step in the empirical implementation of LSEMs in practice. To this day, IV estimation of the linear endogenous regressor model remains an important framework for applied econometric analysis in a variety of settings. As a result, many of the LI estimation methods developed in the 1940’s–1970’s are still relevant today even if the equation of interest is not viewed as part of an explicitly fully–specified SEM. As discussed, 2SLS was developed from a number of different perspectives and of these, Sargan’s approach is most in line with current thinking, being a direct forerunner of the GMM approach to estimation. It is also striking how aspects of the finite sample theory for IV/2SLS/LIML estimators reappear in the weak instrument asymptotics literature and also in the discussion around existence of finite sample moments of GMM and GEL estimators.

Subsequently, it became feasible to implement FI (3SLS/FIML) estimation of LSEMs which at the time was a welcome development given the potential efficiency gains of FI methods over their LI counterparts. However, with the fall from grace of LSEMs in macroeconomics, the interest in 3SLS/FIML methods is arguably less today than it was in the 1960s. Furthermore, given the forceful critiques by Lucas (1976) and Sims (1980), it is hard to envisage a resurrection of LSEMs as a basis for macroeconomic forecasting and policy analysis.

Computational advances in the 1970's made it possible to move beyond linear models of the relationships between economic variables. However, these nonlinear SEMs, while allowing more general functional forms, shared many of the features for which LSEMs were to be criticized for at the end of that decade, such as the designation of systems variables as endogenous or exogenous. As a result, interest in the use of NSEMs in macroeconomic modeling was largely swept away by the emergence of VAR and DSGE models. That said, the statistical theory derived for NSEM estimators laid foundations for the corresponding analysis of GMM–based statistics in nonlinear models.

While GMM may have its origins in macroeconomic applications, its basis on the information in a pmc has made GMM a popular way to estimate models with endogeneity and simultaneity in other areas of application. The strength of GMM is its basis on the information of pmc. However, this can also be a weakness because not all pmcs are equally informative. This has driven econometricians to consider carefully how parameters are identified. This drive is evident in the credibility revolution during which attention turned to research design and in particular the use of controlled trials or natural experiments to justify causal interpretations of inference about economic phenomena.

The credibility revolution made significant progress robustifying earlier econometric approaches. Extensions of instrumental variable methods and the control function approach to nonparametric models with heterogeneity revealed a significantly wider scope for these methods and introduced more nuanced interpretations of the parameters they recover. This work enabled generating useful empirical estimates under significantly more general specifications than before. Moreover, work on partial identification clarified the credibility of the estimates by characterizing the mapping from the observed data and imposed assumptions to the identification regions of parameters. Extending earlier work on path analysis and LSEM, recent work dedicated more attention to understanding the identifying power of exclusion, sign, and shape restrictions in general models. Similarly, significant progress has been made in generalizing the classical measurement error model to accommodate nonparametric and heterogeneous effects.

Recent work also integrated machine learning techniques into standard forecasting and econometric methods. The continued advances in computation and artificial intelligence will likely leave their mark on various econometric methods, including those tackling simultaneity and measurement error.

References

Adcock, R. J. (1877). Note of the method of least squares. *The Analyst*, *4*(6), 183–184.

Adcock, R. J. (1878). A problem in least squares. *The Analyst*, *5*(2), 53–54.

Allen, R. G. D. (1939). The assumptions of linear regression. *Economica*, *6*(22), 191–201.

Altonji, J. G. & Matzkin, R. L. (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica*, *73*(4), 1053–1102.

Amemiya, T. (1974). The nonlinear two-stage least sqaures estimator. *Journal of Econometrics*, *2*(1), 105–110.

Amemiya, T. (1977). The maximim likelihood and nonlinear three-stage least squares estimator in the nonlinear simultaneous equations models. *Econometrica*, *45*(1), 955–68.

Anderson, T. W. (2005). Origins of the limited information maximum likelihood and two-stage least squares estimators. *Journal of Econometrics*, *127*(1), 1–16.

Anderson, T. W. & Hsiao, C. (1981). The asymptotic properties of estimates of the parameters of single equation in a complete system of stochastic equations. *Journal of the American Statistical Association*, *76*(375), 598–606.

Anderson, T. W., Kunitomo, N. & Sawa, T. (1982). Evaluation of the distribution function of the limited information maximum likelihood estimator. *Econometrica*, *50*(4), 1009–27.

Anderson, T. W. & Rubin, H. (1949). Estimation of the parameters of single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, *20*(1), 46–63.

Anderson, T. W. & Rubin, H. (1950). The asymptotic properties of estimates of the parameters of single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, *21*(4), 570–82.

Andrews, D. W. K. (1999). Consistent moment selection procedures for generalized method of moments estimation. *Econometrica*, *67*(3), 543–564.

Angrist, J. D. (1990). Lifetime earnings and the vietnam era draft lottery: evidence from social secirity administrative records. *American Economic Review*, *80*(3), 313–36.

Angrist, J. D., Graddy, K. & Imbens, G. W. (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *The Review of Economic Studies*, *67*(3), 499–527.

Angrist, J. D., Imbens, G. W. & Kreuger, A. B. (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics*, *14*(1), 57–67.

Angrist, J. D., Imbens, G. W. & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455.

Angrist, J. D. & Kreuger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, *106*(4), 979–1014.

Angrist, J. D. & Krueger, A. B. (1992). The effect of age at school entry on educational attainment: An application of instrumental variables with moments from two samples. *Journal of the American Statistical Association*, *87*(418), 328–336.

Angrist, J. D. & Krueger, A. B. (1995). Split-sample instrumental variables estimates of the return to schooling. *Journal of Business and Economic Statistics*, *13*(2), 225–35.

Angrist, J. D. & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Angrist, J. D. & Pischke, J.-S. (2010). The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, *24*(2), 65–98.

Arellano, M. (2002). Sargan's instrumental variables estimation and gneeralized method of moments. *Journal of Business and Economic Statistics*, *20*(4), 450-459.

Arellano, M. & Bond, S. R. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies*, *58*, 277–297.

Balke, A. & Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, *92*(439), 1171–1176. doi: 10.1080/01621459.1997.10474074

Banks, J., Blundell, R. & Lewbel, A. (1997). Quadratic engel curves and consumer demand. *The Review of Economics and Statistics*, *79*(4), 527–539. doi: 10.1162/003465397557015

Basmann, R. L. (1957). A generalized classical method of linear estimation of coefficients in a structural equation. *Econometrica*, *25*(1), 77–83.

Basmann, R. L. (1960). On finite sample distributions of generalized classical linear identifiability test statistics. *Journal of the American Statistical Association*, *55*(292), 650–9.

Basmann, R. L. (1963). Remarks concerning the exact finite sample distributions of GCL estimators in econometric statistical inference. *Journal of the American Statistical Association*, *58*(304), 943–76.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, *62*(3), 657–81.

Bekker, P. A. & van der Ploeg, J. (2005). Instrumental Variable estimation based on grouped data. *Statistica Neerlandica*, *59*(3), 239–67.

Belloni, D., Chen, D., Chernozhukov, V. & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, *80*, 2369–2430.

Beresteanu, A. & Molinari, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica*, *76*(4), 763–814.

Berndt, E. K., Hall, B. H., Hall, R. E. & Hausman, J. A. (1974). Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement*, *3*(4), 653-65.

Bjerkholt, O. (2015). How it all began: the first meeting of the Econometric Society, lausanne,1931. *The European Journal of the History of Economic Thought*, *22*(6), 1149–78.

Blundell, R. & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, *87*, 115–143.

Blundell, R. & Powell, J. (2003). Endogeneity in nonparametric and semiparametric regression models. In M. Dewatripont, L. P. Hansen & S. J. Turnovsky (Eds.), *Advances in economics and econometrics: Theory and applications, eighth world congress* (pp. 312–357). Cambridge University Press.

Bound, J., Jaeger, D. A. & Baker, R. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endoenous explanatory variable is weak. *Journal of the American Statistical Association*, *90*(430), 443–50.

Bowden, R. (1973). The theory of parametric identification. *Econometrica*, *41*(6), 1069–74.

Box, G. E. P. & Jenkins, G. M. (1970). *Time series analysis: forecasting and control*. Holden Day.

Breusch, T. S., Qian, H., Schmidt, P. & Wyhowski, D. (1999). Redundancy of moment conditions. *Journal of Econometrics*, *91*, 89–111.

Brown, B. W. (1983). The identification problem in systems nonlinear in variables. *Econometrica*, *51*(1), 175–96.

Burgete, J. F., Gallant, A. R. & Souza, G. (1982). On unification of the asymptotic theory of nonlinear econometric models. *Econometric Reviews*, *1*(2), 151–190.

Carrasco, M. (2012). A regularization approach to the many instruments problem. *Journal of Econometrics*, *170*, 383–398.

Chalak, K. (2019). Identification of average effects under magnitude and sign restrictions on confounding. *Quantitative Economics*, *10*(4), 1619–1657.

Chalak, K. (2024). Nonparametric gini-frisch bounds. *Journal of Econometrics*, *238*(1), 105560.

Chalak, K. & Kim, D. (2020). Measurement error in multiple equations: Tobin's q and corporate investment, saving, and debt. *Journal of Econometrics*, *214*(2), 413–432. doi: 10.1016/j.jeconom.2019.08.001

Chalak, K. & Kim, D. (2021). Measurement error without the proxy exclusion restriction. *Journal of Business & Economic Statistics*, *39*(1), 200–216.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, *34*, 305–334.

Chao, J. & Swanson, N. (2005). Consistent estimation with a large number of weak instruments. *Econometrica*, *73*, 1673–1692.

Chen, X., Hong, H. & Tamer, E. (2005). Measurement error models with auxiliary data. *The Review of Economic Studies*, *72*(2), 343–366.

Cheng, X. & Liao, Z. (2015). Select the valid and relevant moments: an information-based lasso for gmm with many moments. *Journal of Econometrics*, *186*, 443–464.

Chernozhukov, V., Hong, H. & Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, *75*(5), 1243–1284.

Chesher, A. (1991). The effect of measurement error. *Biometrika*, *78*(3), 451–462.

Chesher, A. & Rosen, A. M. (2017). Generalized instrumental variable models. *Econometrica*, *85*(3), 959–989.

Christ, C. F. (1994). The Cowles Commission's contributions to econometrics at chicago, 1939-1955. *Journal of Economic Literature*, *32*(1), 30–59.

Cornelisse, P. A. & van Dijk, H. (2008). Jan Tinbergen (1903-1994). In S. N. Durlauf & L. E. Blume (Eds.), *The new palgrave dictionary of economics* (pp. 302–308). Palgrave MacMillan.

Cragg, J. G. & Donald, S. G. (1993). Testing identifiability and specification in Instrumental Variables models. *Econometric Theory*, *9*(2), 222–40.

Darolles, S., Fan, Y., Florens, J.-P. & Renault, E. (2011). Nonparametric instrumental regression. *Econometrica*, *79*(5), 1541–1565.

D'Haultfœuille, X., Gaillac, C. & Maurel, A. (2025, January). Partially linear models under data combination. *The Review of Economic Studies*, *92*(1), 238–267.

Dickey, D. A. & Fuller, W. A. (1979). Distribution of the estimators for an autoregressive process with a unit root. *Journal of the American Statistical Association*, *74*(366), 427–31.

Diebold, F. X. (1992). The past, present and future of macroeconomic forecasting. *Journal of Economic Perspectives*, *12*(2), 175–92.

Donald, S. G., Imbens, G. W. & Newey, W. K. (2009). Choosing instrumental variables in conditional moment restrictions models. *Journal of Econometrics*, *152*, 28-36.

Donald, S. G. & Newey, W. K. (2001). Choosing the number of instruments. *Econometrica*, *69*(5), 1161-1192.

Dufour, J.-M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica*(6), 1365–1387.

Durbin, J. (1954). Errors in variables. *Review of the International Statistical Institute*, *22*(1), 23–32.

Engle, R. F., Hendry, D. F. & Richard, J.-F. (1983). Exogeneity. *Econometrica*, *51*(2), 277-304.

Fan, J. & Truong, Y. K. (1993). Nonparametric regression with errors in variables. *The Annals of Statistics*, *21*(4), 1900–1925.

Fisher, F. M. (1966). *The identification problem in econometrics*. McGraw-Hill.

Frisch, R. A. K. (1934). *Statistical confluenece analysis by means of complete regression systems*. Univ. Okonomiske Institutt, Oslo.

Fuller, W. A. (1976). *Introduction to statistical time series*. Wiley.

Fuller, W. A. (1977). Some properties of a modification of the limited information estimator. *Econometrica*, *45*(4), 939–53.

Gallant, A. R. (1977). Three-stage least squares estimation of a system of simultaneous, nonlinear, implicit equations. *Journal of Econometrics*, *5*(1), 77–88.

Gallant, A. R. & Jorgenson, D. W. (1979). Statistical inference for a system of simultaneous, non-linear, implicit equations in the context of instrumental variables estimation. *Journal of Econometrics*, *11*(2-3), 275–302.

Gallant, A. R. & Tauchen, G. (1989). Semi-nonparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications. *Econometrica*, *57*(5), 1091-120.

Geary, R. C. (1942). Inherent relations between random variables. *Proceedings of the Royal Irish Academy (Section A)*, *47*(6), 63–76.

Geary, R. C. (1943). Relations between statistics: the general and the sampling problem when the samples are large. *Proceedings of the Royal Irish Academy (Section A)*, *49*(10), 177–96.

Geary, R. C. (1948). Studies in the relations between economic time series. *Journal of the Royal Statistical Society, Series B*, *10*(1), 140–58.

Ghysels, E., Hall, A. R. & Hansen, L. P. (2002). Interview with Lars Hansen. *Journal of Business and Economic Statistics*, *20*(4), 442-447.

Ghysels, E., Hall, A. R. & Sims, C. A. (2002). Interview with Christopher Sims. *Journal of Business and Economic Statistics*, *20*(4), 448-449.

Gini, C. (1921). Sull'interpolazione di una retta quando i valori della variabile indipendente sono affetti da errori accidentali. *Metron*, *1*(3), 63–82.

Granger, C. W. J. (1969). Investigating causal relations by ecoometric models and cross-spectral methods. *Econometrica*, *37*(3), 424-38.

Griliches, Z. (1974). Errors in variables and other unobservables. *Econometrica*, *42*(6), 971–998.

Guggenberger, P. (2008). Finite sample evidence suggesting a heavy tail problem of the generalized empirical likelihood estimator. *Economeric Reviews*, *26*(4-6), 526-541.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, *12*(Suppl.), 1–115.

Hahn, J. & Kuersteiner, G. (2002). Discontinuities of weak instrument limiting distributions. *Economics Letters*, *55*, 325-31.

Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press.

Hall, A. R. (2015). Econometricians have their moments: GMM at 32. *Economic Record*, *91*(Special issue, June), 1-24.

Hall, A. R. & Peixe, F. (2003). A consistent method for the selection of relevant instruments. *Economeric Reviews*, *91*(3), 269-287.

Hall, A. R., Rudebusch, G. D. & Wilcox, D. W. (2006). Judging instrument relevance in Instrumental Variables estimation. *International Economic Review*, *37*(2), 283-298.

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ, U. S. A.: Princeton University Press.

Han, C. & Phillips, P. C. B. (2006). Gmm with many moment conditions. *Econometrica*, *74*, 147–192.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*(4), 1029-54.

Hansen, L. P. (1985). A method for calculating bounds on the asymtptoic covariance matrices of Generalized Method of Moments estimators. *Journal of Econometrics*, *30*, 203-238.

Hansen, L. P., Heaton, J. & Ogaki, M. (1988). Efficiency bounds implied by multi–period conditional moment restrictions. *Journal of the American Statistical Association*, *83*(403), 863–871.

Hansen, L. P., Heaton, J. & Yaron, A. (1996). Finite sample properties of some alternative GMM estimators obtained from financial market data. *Journal of Business and Economic Statistics*, *14*(3), 262–280.

Hansen, L. P. & Sargent, T. J. (1980). Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and Control*, *2*, 7-46.

Hansen, L. P. & Singleton, K. J. (1982). Generalized instrumental variables estimators of nonlinera rational expectations models. *Econometrica*, *50*(5), 1269-86.

Hansen, L. P. & Singleton, K. J. (1983). Stochastic consumption, risk aversion and the temporal behaviour of asset returns. *Journal of Political Economy*, *91*(2), 249-65.

Hansen, L. P. & Singleton, K. J. (1996). Efficient estimation of linear-asset pricing models with moving average errors. *Journal of Business and Economic Statistics*, *14*(1), 53-68.

Hausman, J. A. (1975). An instrumental variable approach to full information estimators for linear and certain nonlinear econometric models. *Econometrica*, *43*(4), 727-38.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*(6), 1251-71.

Hausman, J. A. (1983). Specification and estimation of simultaneous equation models. In Z. Griliches & M. D. Intriligator (Eds.), *Handbook of econometrics, volume 1* (pp. 392–448). North-Holland.

Hausman, J. A., Lewis, R., Menzel, K. & Newey, W. K. (2011). Properties of the cue estimator and a modification with moments. *Journal of Econometrics*, *165*(1), 45-57.

Hausman, J. A., Newey, W. K. & Taylor, W. E. (1975). Efficient estimation and identification of simultaneous equation models with covariance restrictions. *Econometrica*, *55*(4), 849-74.

Heckman, J. J. & Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, *73*(3), 669–738.

Horowitz, J. L. & Manski, C. F. (1995). Identification and robustness with contaminated and corrupted data. *Econometrica*, *63*(2), 281–302.

Hurwicz, L. (1950). Generalization of the concept of identification. In T. C. Koopmans (Ed.), *Statistical inference in dynamic economic models* (Vol. 10, pp. 245–257). New York: John Wiley & Sons.

Imbens, G. W. & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*(2), 467–475.

Imbens, G. W. & Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, *72*(6), 1845–1857.

Imbens, G. W. & Newey, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, *77*(5), 1481-1512.

Inoue, A. & Solon, G. (2010, August). Two-sample instrumental variables estimators. *The Review of Economics and Statistics*, *92*(3), 557–561.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, *95*(1), 161–182.

Jordà, O., Schularick, M. & Taylor, A. M. (2015). Betting the house. *Journal of International Economics*, *96*(Supplement 1), S2–S18.

Jorgenson, D. W. & Laffont, J. J. (1974). Efficient estimation of nonlinear simultaneous equations with additive disturbancess. *Annals of Economic and Social Measurement*, *3*(4), 615-40.

Kelejian, H. H. & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, *17*(1), 99–121.

Kitamura, Y. (2007). Empirical likelihood methods in econometrics: theory and practice. In R. Blundell, W. K. Newey & T. Personn (Eds.), (p. 174-237). Cambridge University Press.

Kitamura, Y. & Stutzer, M. (1997). An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, *65*(4), 861–874.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, *70*, 1781–1803.

Kleibergen, F. (2005). Testing parameters in gmm without assuming that they are identified. *Econometrica*, *73*, 1103–1124.

Kleibergen, F. & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, *133*, 97–126.

Klein, L. R. (1955). On the interpretation of theil's method of estimating economic relationships. *Metroeconomica*, *7*(3), 147–53.

Klein, L. R. & Mariano, R. S. (1980). The et interview: Professor l. r. klein. *Econometric Theory*, *3*(3), 409–60.

Klepper, S. & Leamer, E. E. (1984). Consistent sets of estimates for regressions with errors in all variables. *Econometrica*, *52*(1), 163–183.

Kline, B. & Tamer, E. (2016). Bayesian inference in a class of partially identified models. *Quantitative Economics*, *7*(2), 329–366.

Koopmans, T. C. (1936). *Linear regression analysis of economic time series*. Netherlands Economic Institute, Haarlem.

Koopmans, T. C. (1949). Identification problems in economic moel construction. *Econometrica*, *17*(2), 125–144.

Koopmans, T. C. & Hood, W. C. (1953). The estimation of simultaneous linear economic relationships. In T. C. Koopmans & W. C. Hood (Eds.), *Studies in econometric methods* (p. 112-99). Wiley, New York.

Koopmans, T. C., Rubin, H. & Leipnik, R. B. (1950). Measuring the equation systems of dynamic economics. In T. C. Koopmans (Ed.), *Statistical inference in dynamic economic models* (pp. 302–308). Wiley, New York.

Kuersteiner, G. (2001). Optimal instrumental variables estimation for arma models. *Journal of Econometrics*, *104*, 359–405.

Kummell, C. H. (1879). Reduction of oobservation equations that contain more than one observed quantity. *The Analyst*, *6*(4), 97–105.

Leamer, E. E. (1987). Errors in variables in linear systems. *Econometrica*, *55*(4), 893–909.

Lee, L.-F. (2007). Gmm and 2sls estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, *137*, 489–514.

Lee, L.-F. & Yu, J. (2014). Efficient gmm estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics*, *180*, 174–197.

Liu, T.-C. (1960). Underidentification, structural estimation and forecasting. *Econometrica*, *28*(4), 855–65.

Lucas, R. E. (1976). Econometric policy evaluation: a critique. In K. Brunner & A. H. Meltzer (Eds.), *Carnegie-rochester conference series on public economics* (Vol. 1, pp. 19–46). North Holland.

Madansky, A. (1964). On the efficiency of three-stage least squares estimation. *Econometrica*, *32*(1/2), 51–6.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review*, *80*(2), 319–323.

Manski, C. F. & Pepper, J. V. (2000). Monotone instrumental variables: With an application to the returns to schooling. *Econometrica*, *68*(4), 997–1010.

Manski, C. F. & Tamer, E. (2002, March). Inference on regressions with interval data on a regressor or outcome. *Econometrica*, *70*(2), 519–546.

Mariano, R. S. (1983). Analytical small-sample distribution theory in econometrics: the simultaneous-equations case. *International Economic Review*, *23*(3), 503–33.

Mariano, R. S. (2003). Simultaneous equation model estimators. In B. H. Baltagi (Ed.), *A companion to theoretical econometrics* (pp. 122–143). Blackwell Publishng.

Mariano, R. S. & Sawa, T. (1972). The exact finite-sample distribution of the limited information maximum likelihood estimator in the case of two included endogenous variables. *Journal of the American Statistical Association*, *67*(337), 159–163.

Matzkin, R. L. (2008). Identification in nonparametric simultaneous equations models. *Econometrica*, *76*(5), 945–978.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.

Mertens, K. & Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the united states. *American Economic Review*, *103*(4), 1212–47.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica*, *71*(4), 1027–1048.

Morgan, M. S. (1990). *The history of econometric ideas*. Cambridge University Press.

Nagar, A. L. (1959). The bias and moment matrix of the general k-class estimators of the parameters in structural equations. *Econometrica*, *27*(4), 575–95.

Nelson, C. R. & Plosser, C. R. (1982). Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics*, *10*(2), 139–62.

Nelson, C. R. & Startz, R. (1990). The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. *Journal of Business*, *63*(1), S125–40.

Newey, W. K. & Powell, J. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, *71*(5), 1565–1578.

Newey, W. K., Powell, J. & Vella, F. (1999). Identification and estimation of triangular simultaneous equation models. *Econometrica*, *67*(3), 565–603.

Newey, W. K. & Smith, R. J. (2004). Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica*, *72*(2), 219–255.

Newey, W. K. & Windmeijer, F. (2009). Generalized method of moments with many wek moment conditions. *Econometrica*, *77*(3), 687–719.

Olea, J. L. M. & Pflueger, C. (2013). A robust test for weak inistruments. *Journal of Business and Economic Statistics*, *31*(3), 358-69.

Pantula, S. G. & Hall, A. R. (1991). Testing for unit roots in autoregressive moving average models. *Journal of Econometrics*, *48*(3), 325–54.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.

Phillips, G. D. A. & Hale, C. (1977). The bias of instrumental variable estimators in simultaneous equation systems. *International Economic Review*, *18*(1), 219–28.

Phillips, P. C. B. (1980). The exact finite sample density of instrumental variable estimators in an equation with n+1 endogenous variables. *Econometrica*, *48*(4), 861–878.

Phillips, P. C. B. (1982). On the consistency of nonlinear fiml. *Econometrica*, *50*(5), 1307–24.

Phillips, P. C. B. (1984). The exact distribution of liml:i. *International Economic Review*, *25*(1), 249–61.

Phillips, P. C. B. (1985). The exact distribution of liml:ii. *International Economic Review*, *26*(1), 21–36.

Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica*, *55*(2), 277–301.

Phillips, P. C. B. (2003). In memory of John Denis Sargan. *Econometric Theory*, *19*(3), 416–22.

Pötscher, B. M. (1991). Effects of model selection on inference. *Econometric Theory*, *7*(2), 163–185.

Qin, J. & Lawless, J. (1994). Empirical likelihood and generalized estimating equations. *Annals of Statistics*, *22*(1), 300–325.

Reiersøl, O. (1941). Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica*, *9*(1), 1–24.

Reiersøl, O. (1945). Confluence analysis by means of instrumental sets of variables. *Arkiv for Mathematik, Astronomi och Fysik*, *32*(1), 4–119.

Reiersøl, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica*, *18*(4), 375–389.

Richardson, D. H. (1968). The exact distribution of a structural coefficient estimator. *Journal of the American Statistical Association*, *63*(324), 1214–26.

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, *71*(1), 135–158.

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Rothenberg, T. J. (1971). Identification of parameteric models. *Econometrica*, *39*(3), 577–91.

Rothenberg, T. J. & Leenders, C. T. (1964). Efficient estimation of simultaneous equation systems. *Econometrica*, *32*(1/2), 57–76.

Said, S. E. & Dickey, D. A. (1984). Testing for unit roots in ARMA models of unknown order. *Biometrika*, *71*(3), 599–607.

Said, S. E. & Dickey, D. A. (1985). Hypothesis testing in ARIMA(p,1,q) models. *Journal of the American Statistical Association*, *80*(390), 369-74.

Sargan, J. D. (1958). Estimation of economic relationships using instrumental variables. *Econometrica*, *26*(3), 393–415.

Sargan, J. D. (1959). Estimation of relationships with autocorrelated residuals by the use of instrumental variables. *Journal of the Royal Statistical Society, Series B*, *21*(1), 91–105.

Sargan, J. D. (1974). The valdity of Nagar's expansion for the moments of econometric estimators. *Econometrica*, *42*(1), 169–76.

Sargan, J. D. (1976). Econometric estimators and the Edgeworth approximation. *Econometrica*, *44*(3), 421–48.

Sargan, J. D. (1978). On the existence of the moments of the 3sls estimator. *Econometrica*, *46*(6), 1329–50.

Sargan, J. D. (1982). On Monte Carlo estimates of moments that are infinite. In R. L. Basmann & G. F. Rhodes (Eds.), *Advances in econometrics: A research annual* (pp. 267–299). JAI Press.

Sargan, J. D. (1988). The finite sample distribution of fiml estimators. In E. Maasoumi (Ed.), *Contributions to econometrics: John Denis Sargan, volume 2* (pp. 57–75). Cambridge University Press.

Sawa, T. (1969). The exact finite sampling distribution of ordinary least squares and two stage least squares estimator. *Journal of the American Statistical Association*, *64*(327), 923–36.

Schennach, S. M. (2014). Entropic latent variable integration via simulation. *Econometrica*, *82*(1), 345–385.

Schennach, S. M. & Hu, Y. (2013). Nonparametric identification and semiparametric estimation of classical measurement error models without side information. *Journal of the American Statistical Association*, *108*(501), 177–186.

Schennach, S. M., White, H. & Chalak, K. (2012). Local indirect least squares and average marginal effects in nonseparable structural systems. *Journal of Econometrics*, *166*(2), 282–302.

Shea, J. (1997). Instrument relevance in multivariate linear models: a simple measure. *The Review of Economics and Statistics*, *79*(2), 348–52.

Sims, C. A. (1977). Exogeneity and causal orderng in macroeconomic models. In C. A. Sims (Ed.), *New methods in business cycle research: proceedings from a conference* (pp. 23–44). Federal Reserve Bank of Minneapolis.

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

Smith, R. J. (1997). Alternative semi-parametric likelihood approaches to generalized method of moments estimation. *Economics Journal*, *107*(March), 503-519.

Staiger, D. & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*(3), 557–586.

Stock, J. H. & Trebbi, F. (2003). Retrospectives: who invented instrumental variable regression. *The Journal of Economic Perspectives*, *17*(3), 177–194.

Stock, J. H. & Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economics Journal*, *128*(610), 917–948.

Stock, J. H. & Wright, J. (2000). Gmm with weak identification. *Econometrica*, *68*(5), 1055–1096.

Stock, J. H., Wright, J. H. & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics*, *20*(4), 518-529.

Stoica, P., Söderstroöm, T. & Friedlander, B. (1985). Optimal instrumental variable estimates of ar parameters of an arma process. *IEEE Transactions on Automatic Control*, *AC-30*(11), 1066-74.

Tauchen, G. (1986). Statistical properties of Generalized Method of Moments estimators of structural parameters obtained from financial market data. *Journal of Business and Economic Statistics*, *4*(4), 397-416.

Theil, H. (1953). *Estimation and simultaneous correlation in complete equations*. Central PLanning Bureau, The Hague.

Theil, H. (1958). *Economic forecasts and policy* (first ed.). North Holland Publishing Company, Amsterdam.

Theil, H. (1961). *Economic forecasts and policy* (second ed.). North Holland Publishing Company, Amsterdam.

Tinbergen, J. (1930). Bestimmung und deutung von angebotskurven ein beispiel. *Zeitschrift für Nationalökonomie*, *1*, 669–79.

Tinbergen, J. (1939). *Statistical testing of business cycle theories. i: A method and its application to investment activity. ii: Business cycles in the united sates of america*. Geneva: League of Nations.

Tinbergen, J. (1959). An economic policy for 1936. In L. Klaassen, L. Koyck & H. Witteveen (Eds.), *Jan Tinbergen: Selected papers.* Amsterdam: North Holland.

Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*, *70*(1), 331–341.

Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics*, *11*(3), 284–300.

White, H. L. & Chalak, K. (2009). Settable systems: An extension of pearl's causal model with optimization, equilibrium, and learning. *Journal of Machine Learning Research*, *10*(8), 1759–1799.

White, H. L. & Chalak, K. (2013). Identification and identification failure for treatment effects using structural systems. *Econometric Reviews*, *32*(3), 273–317.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *The Journal of Human Resources*, *50*(2), 420–445.

Wright, P. G. (1928). *The tariff on animal and vegetable oils*. MacMillan.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, *20*(7), 557–585.

Wu, D.-M. (1973). Alternative tests of independence between stochastic regressors and disturbances. *Econometrica*, *41*(4), 733-750.

Zellner, A. & Theil, H. (1962). Three-stage least squares: simultaneous estimation of simultaneous equations. *Econometrica*, *30*(1), 54–78.

Index