



**HAL**  
open science

## Advances in specification testing

Russell Davidson, Victoria Zinde-Walsh

► **To cite this version:**

Russell Davidson, Victoria Zinde-Walsh. Advances in specification testing. Canadian Journal of Economics/Revue canadienne d'économique, 2017, 50 (5), pp.1595 - 1631. 10.1111/caje.12309 . hal-01684821

**HAL Id: hal-01684821**

**<https://amu.hal.science/hal-01684821>**

Submitted on 9 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Advances in specification testing

Russell Davidson *Department of Economics and CIREQ, McGill University; AMSE–GREQAM*

Victoria Zinde-Walsh *Department of Economics and CIREQ, McGill University*

*Abstract.* Testing the specification of econometric models has come a long way from the  $t$  tests and  $F$  tests of the classical normal linear model. In this paper, we trace the broad outlines of the development of specification testing, along the way discussing the role of structural versus purely statistical models. Inferential procedures have had to advance in tandem with techniques of estimation, and so we discuss the generalized method of moments, non parametric inference, empirical likelihood and estimating functions. Mention is made of some recent literature, in particular, of weak instruments, non parametric identification and the bootstrap.

*Résumé.* *Avancées dans les tests de spécification.* Les tests de spécification des modèles économétriques ont grandement progressé depuis les tests  $t$  et  $F$  du modèle linéaire classique. Dans ce texte, on trace à gros traits le développement du processus de test de spécification, et on discute en cours de route le rôle des modèles structurels et purement statistiques. Les procédures d'inférence ont dû progresser en tandem avec les techniques d'estimation, ce qui fait qu'on est amené à discuter de la méthode généralisée des moments, de l'inférence non-paramétrique, de la vraisemblance empirique, et des fonctions d'estimation. On mentionne certains éléments de la littérature spécialisée récente, en particulier à propos des instruments faibles, de l'identification non-paramétrique, et des techniques du bootstrap.

JEL classification: B23, C10, C12, C14

## 1. Introduction

Econometric model specification brings together economic theory, statistical methodology and data. With the growing complexity of modelling computational issues and computational power come to the forefront.

This research was supported by the Canada Research Chair program (Chair in Economics, McGill University) and by a grant from the Fonds de Recherche du Québec – Société et Culture. We are grateful to an anonymous referee for many useful suggestions that have improved the paper.

Corresponding author: Russell Davidson, russell.davidson@mcgill.ca

Initial approaches to specification focused on the broad outlines of economic theory in combination with simple statistical models, such as linear Gaussian regressions; the small data sets available would not have permitted much embellishment, and thus at the start, the issues of heteroskedasticity and autocorrelation could be treated only within classes of parsimonious parametric models. Structural models required the use of instrumental variables, and initially the main issue was the need for instruments as opposed to ordinary least squares (OLS).

Sims's (1980) critique drew attention to the fact that most structural models associated with economic theory were greatly misspecified, while statistical models, for instance, vector auto-regressions (VAR), gave a much better account of the properties of the data. Since that time, increasing computing power and greater availability of data have led to major developments in testing model specification. New conceptual approaches included testing non-nested specifications, the generalized method of moments (GMM), extensions of statistical approaches to models with general heteroskedastic and autoregressive disturbances via the use of heteroskedasticity and autocorrelation consistent covariance matrix estimators, and indirect inference. The use of artificial regressions was instrumental in constructing statistics that could be easily computed.

As it became clear that many economic models involve non-stationary variables with dynamics that could be stochastic rather than deterministic, and that neglecting this property could lead to spurious relations, new methodology that applied unit-root stochastic modelling and introduced cointegrating relations was developed, using non-stationary vector autoregression and error-correction models. Unit root and cointegration testing became a standard tool of inference in establishing the proper specification of an econometric model.

It was understood that, in many cross-sectional models, the parametric assumptions could be too restrictive, and, starting with the 1980s, econometrics saw the development of semi- and non-parametric modelling that also extended to models with stationary time-series data. One of the important developments was testing for validity of parametric specifications. Many different approaches were developed in the literature. The non-parametric methodology was recently extended to non-stationary time series.

The generalized method of moments provides a useful competitor for the likelihood approach, but both have been considerably extended over the years. The likelihood approach that relied on the knowledge of the form of the distribution can now be applied in a more general setting, thanks to empirical likelihood, which permits estimation of the probability weights along with the parameters. GMM is utilized in models with weak instruments where the number of instruments could be large, even infinite, with results available for a continuum of moments.

As the methodology of inference advances, and ever more complex specifications are studied, progress on developing meaningful results is of course dependent on the assumptions that restrict the model class. Mathematical tractability requires fairly restrictive conditions, many of which may not be supported by an

economic model that captures the relevant features of the economic theory. The most important question is that of identification: which of the features can be credibly identified by the statistical analysis of the data in a realistic model. Often one has to contend with partial identification of possibly a subset or a combination of parameters, or with describing only a set where the parameter may reside. There is also realisation that, even for identified parameters of interest, statistical inference could be impossible unless some regularising restrictions are imposed.

With the increases in complexity of statistical approaches, statistics that do not have analytically tractable finite-sample, or even asymptotic, distributions are becoming more common; however simulation and bootstrap techniques in many situations provide the ability to perform accurate inference.

This paper gives a brief discussion of these points. Other important developments in econometric model specification and testing, such as dealing with big data, multi-level modelling, network analysis and graphical models, that are starting to make inroads into econometrics are promising tools for a more realistic analysis of economic theories. A question that often plagues econometric analysis is whether a statistic that is found to be significant signals a causal effect; various approaches such as testing for Granger non-causality,<sup>1</sup> experiments and pseudo (or “natural”) experiments provide possible answers, all of which suffer from strong reliance on assumptions that are typically untestable. These issues are left outside the discussion of this paper.

Section 2 considers early econometric work on the classical linear regression with possible autocorrelation, heteroskedasticity and endogeneity. Section 3 introduces specification tests such as the  $J$  tests, Durbin–Wu–Hausman tests, or information matrix tests, and provides a brief discussion of artificial regressions. Section 4 discusses GMM and inference in the context of GMM as well as indirect inference and the HAC covariance matrix. Section 5 focuses on unit roots and cointegration. Section 6 considers non parametric approaches in specification testing, with emphasis on tests of the parametric form of the regression function; it also mentions the empirical likelihood approach. Section 7 is devoted to weak instruments and the issues of identification. The bootstrap method is discussed in Section 8. Section 9 concludes.

## 2. Classical statistics and the linear regression model

Classical statistics, and the beginnings of statistical inference, were invented in the early twentieth century. For an excellent account of these developments, and of the personalities involved, see the book by Lehmann (2011). It was at this time that the procedures we now associate with the highly restrictive classical normal linear model were developed, with the test statistics labelled  $t$  and  $F$ . These were the main tools of inference in econometrics until the second half of the century.

<sup>1</sup> The use of the term “causality” here may be misleading. Granger causality is a purely statistical notion, with little or nothing to do with causality as defined by philosophers.

### *Serial correlation*

Up through the 1970s, most econometrics textbooks began with the linear regression model, and then discussed the problems associated with it. Most important of these was probably serial correlation. Consider the linear regression model:

$$y = X\beta + u, \quad (1)$$

where the  $n$ -vector  $y$  contains the observations on the dependent variable, the  $n \times k$  matrix  $X$  those on explanatory variables, and the  $n$ -vector  $u$  contains the possibly serially correlated disturbances. The elements of the  $k$ -vector  $\beta$  and the scalar  $\rho$  are parameters.

The most commonly used test for detecting serial correlation:

$$u_t = \rho u_{t-1} + v_t; \quad v_t \text{ white noise}$$

was the Durbin–Watson test for the null hypothesis  $H_0: \rho = 0$ , proposed in Durbin and Watson (1950) and (1951). This test had many disadvantages. Under the null hypothesis, the distribution of the test statistic depends on the design matrix  $X$ , and one needs to rely on bounds for critical values which leave an ambiguous region between rejection and non-rejection. Another major disadvantage is that there can be no lagged dependent variables in the matrix  $X$ .

Durbin’s  $h$  test (1970), which allows for lagged dependent variables, was widely used, but had drawbacks as well. In particular, in some cases, the statistic cannot be computed, as it would depend on the square root of a negative number. However, Durbin’s so-called alternative procedure, which gained recognition only after being rediscovered and extended by Breusch (1978) and Godfrey (1978a) and (1978b), solved these problems. The statistic is computed by regressing the residuals from the original regression (1) on the  $X$  matrix and the lagged residuals:<sup>2</sup>

$$\hat{u}_t = X_t b + r \hat{u}_{t-1} + \text{residual}. \quad (2)$$

The statistic is the  $t$  statistic for the artificial hypothesis that  $r = 0$ . Regression (2) is an example of an **artificial regression**, a concept that will be discussed in Section 3.

### *Heteroskedasticity*

Another bugbear for econometricians at this time was heteroskedasticity. Several tests were used in order to detect it: Goldfeld–Quandt (Goldfeld and Quandt 1965), Glejser (1969), Godfrey (1978c) and Breusch–Pagan (Breusch and Pagan 1979). For Goldfeld–Quandt, the observations are ordered according to the value of some variable that is thought to be responsible for heteroskedasticity. The model is then estimated over the first and last thirds of the sample, yielding the  $F$  statistic:

2 Above, and throughout the paper, the notation “+ residual” or “+ residuals” in writing a regression is used to indicate that the regression is used in order to compute something, usually a test statistic, and is not supposed to correspond to any statistical model.

$$\frac{SSR_3/(n_3 - k)}{SSR_1/(n_1 - k)},$$

where  $SSR_1$  and  $SSR_3$  denote the sums of squared residuals from the first and last thirds of the sample and  $n_1$  and  $n_3$  denote the associated sample sizes.

A plausible model of heteroskedasticity is:

$$E(u_t^2) = h(\alpha + \mathbf{Z}_t \boldsymbol{\gamma}), \quad (3)$$

where  $h$  is a possibly nonlinear function that may only take on positive values,  $\mathbf{Z}_t$  is a  $k$ -vector of observations on exogenous or predetermined variables,  $\alpha$  is a scalar parameter and  $\boldsymbol{\gamma}$  is a  $k$ -vector of parameters. The Godfrey and Breusch–Pagan tests make use of artificial regressions based on (3), which we may write as

$$\hat{u}_t^2 = a + \mathbf{Z}_t \mathbf{c} + \text{residual}, \quad t = 1, \dots, n, \quad (4)$$

where the  $\hat{u}_t$  are OLS residuals from the original regression. An ordinary  $F$  test for the artificial hypothesis that  $\mathbf{c} = \mathbf{0}$  can be used, or alternatively  $nR^2$  from (4), where  $R^2$  is the centred coefficient of determination. Glejser replaced the regressand  $\hat{u}^2$  by the absolute value of the residual.

If heteroskedasticity was detected, it was sometimes possible to choose a function  $h$  and explanatory variables  $\mathbf{Z}_t$  and to construct a parametric skedastic function like  $h(\alpha + \mathbf{Z}_t \mathbf{c})$ . The skedastic parameters  $\mathbf{c}$  could be estimated by regressing the squared residuals on a constant and the  $\mathbf{Z}_t$ , and then the original regression re-estimated by feasible GLS.

This state of affairs, in which it was necessary to specify a skedastic function, was vastly improved by Hal White's (1980) paper, which made it possible to have asymptotically correct inference in the presence of heteroskedasticity of unknown form, by use of an HCCME (heteroskedasticity-consistent covariance matrix estimate). This opened the way for econometricians to conduct tests of specification or concerning parameters, in a manner robust to the the presence or absence of heteroskedasticity.

If it is assumed that there is no serial correlation of the disturbances in regression (1), the HCCME is a sandwich covariance matrix estimate, of the form:

$$(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \hat{\boldsymbol{\Omega}} (\mathbf{X}^\top \mathbf{X})^{-1},$$

where, in the original form, the elements of the  $n \times n$  diagonal matrix  $\hat{\boldsymbol{\Omega}}$  are the squared residuals from the original regression. It has since been seen to be preferable to adjust the squared residuals in a variety of ways—see MacKinnon and White (1985). One way is to use as diagonal elements of  $\hat{\boldsymbol{\Omega}}$  the  $\hat{u}_t^2/(n/(n-k))$ , thus incorporating a degrees-of-freedom correction. Alternatively, the diagonal elements can be the  $\hat{u}_t^2/(1-h_t)$ , or the  $\hat{u}_t^2/(1-h_t)^2$ , where  $h_t$  is the  $t^{\text{th}}$  element of the orthogonal projection matrix  $\mathbf{P}_x = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ . It was some time after the publication of White's paper that it was noticed that a good deal of White's

work had been anticipated in the statistics literature and in a different context by Eicker (1963) and (1967).

### *Instrumental variables*

Simultaneous-equation models go back to the early days of econometrics. Ragnar Frisch (1934) coined the term “multicollinearity” to describe a set of more than one (hence “multi”) relation of collinearity among observable variables and a disturbance. For instance, the linear regression (1) states that the vector  $\mathbf{y}$ , the columns of the matrix  $\mathbf{X}$  and the vector  $\mathbf{u}$  are linearly dependent. It was understood that some of the variables involved in a multicollinearity<sup>3</sup> were endogenous, while others were assumed to be exogenous. Thus a multicollinearity in Frisch’s sense is, in modern terminology, a system of simultaneous equations.

Even before Frisch’s work, the technique of estimation by instrumental variables (IV) was invented by Philip Wright—or perhaps his son Sewall Wright. See the fascinating story of trying to identify which of them was responsible, or mostly responsible, for the invention in Stock and Trebbi (2003). At the end of their paper, Stock and Trebbi remark that a remaining mystery is why the invention was forgotten, to be rediscovered only many years later, first by Reiersøl (1941), whose work also had little impact at the time, and then by Durbin (1954) and Sargan (1958).

Estimation by instrumental variables was found to be necessary for consistent estimation not only in the case of endogeneity of one or more of the regressors, but also when regressors are measured with error—the famous errors-in-variables problem. Since the estimating equations for OLS can be written as:

$$\mathbf{X}^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0},$$

it follows that, if the expectation of the left-hand side, evaluated at the true value of the vector  $\boldsymbol{\beta}$ , is different from zero, then the OLS estimator is biased. If an asymptotic construction is used such that the limit of  $n^{-1}\mathbf{X}^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ , as the sample size  $n$  tends to infinity, is nonzero, then the estimator is asymptotically biased and inconsistent. If an  $n \times l$  matrix  $\mathbf{W}$ , where  $l \geq k$ , is such that:

$$\mathbb{E}(\mathbf{W}^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})) = \mathbf{0},$$

then the columns of  $\mathbf{W}$  are valid instruments and can be used in the IV estimating equations:

$$\mathbf{X}^\top \mathbf{P}_\mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}. \tag{5}$$

Here  $\mathbf{P}_\mathbf{W} \equiv \mathbf{W}(\mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top$  is the orthogonal projection matrix on to the span of the columns of  $\mathbf{W}$ . Equations (5) also express the first-order conditions for the minimisation of the IV criterion function, defined as:

$$Q(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{P}_\mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

<sup>3</sup> Regrettably, the term is often used to describe a relation of collinearity or near collinearity in just one regression equation.

If  $l = k$ , the parameters  $\beta$  are just identified. If  $l > k$ , then there are more identifying conditions than parameters to identify. The extra conditions are called over-identification restrictions and can be tested by use of the minimized value  $Q(\hat{\beta})$ , where  $\hat{\beta}$  is the IV estimator. Under the null of correct specification,  $Q(\hat{\beta})$ , divided by a consistent estimate of the variance of the disturbances, is asymptotically distributed as  $\chi^2_{l-k}$ . This test was proposed by Sargan (1958). If it leads to rejection of the hypothesis of correct specification, the reason may be that the instruments are not all valid, or that they have explanatory power for  $y$  over and above that of the variables in  $X$ . Basmann (1960) considered a somewhat modified version of the Sargan statistic, which may have better finite-sample properties.

In Durbin (1954), a null of exogeneity of the regressors in  $X$ , or at least of the absence of bias in the OLS estimator, was considered. The test statistic has an interesting form: it is based on the difference between the OLS and IV estimators. Under the null, this difference has a covariance matrix that can be estimated in various ways.

### 3. Specification tests

Until around 1980, a more or less universal feature of econometric models was that they were **structural**, in the sense that they were based on models taken from economic theory. Except for the potential problems considered in the previous section, namely serial correlation, heteroskedasticity and endogenous regressors, there was no serious attempt to test model specification, or the statistical adequacy of the models used.

In 1980, Sims published a very influential article entitled “Macroeconomics and Reality,” in which he drew attention to the widespread misspecification of many macro-models in use at that time and argued in favour of models like vector auto-regressions (VARs), which could give a much better account of the properties of the data. The rise of modern specification testing followed naturally in the wake of Sims’s call to order. There were many fairly new tests introduced, plus many extensions of existing ones.

Computing hardware slowly rose to the task of implementing econometric estimation and testing during the late 1950s and 1960s. By the mid 1970s, universities normally gave access to their mainframe computers to econometricians, who installed econometric software adapted to their purposes. It is difficult for us in the 21st century to appreciate how painful econometric computing was, even in the 1970s, when programs had to be typed on punched cards that were then submitted as batch jobs at a computing centre sometimes located on the other side of the campus.

It was, and still is, the case that the output produced by running a regression contains the value of a  $t$  statistic for each estimated parameter, as well as the value of the  $F$  statistic. They can provide evidence of the significance of the variables in the model.



The realization came that these  $t$  and  $F$  tests were meaningless if the model was misspecified, and that misspecification could take many forms, not limited to those discussed in the previous section. The classical approach to testing is to posit a model that corresponds to the alternative hypothesis and to specify the null hypothesis by imposing restrictions in the context of the alternative.

### *Non-nested hypothesis tests*

As early as 1961, the well-known statistician D. R. Cox had proposed what would now be called a non-nested hypothesis test. See also his 1962 paper. A null hypothesis is specified as a model, by which is now meant a set of data-generating processes (DGPs), a set that contains all those DGPs that satisfy the null hypothesis. Instead of embedding this set in a larger set corresponding to the alternative, a different model is taken as the alternative. The two models may overlap, but each must contain a subset the elements of which do not belong to the other model.

Cox's work was quite general and was set in the framework of likelihood theory. Pesaran and Deaton (1978) specialized Cox's approach to the case of linear regression models. Their paper may be said to have sparked the interest of econometricians in this sort of test. Then Davidson and MacKinnon (1981) devised a very simple implementation of a non-nested hypothesis test for both linear and nonlinear regressions. For the linear case, suppose that the two hypotheses for the test are as follows:

$$H_1: y = X\beta + u \quad \text{and} \quad H_2: y = Z\gamma + u,$$

with the first being the null model, the second the alternative, where in either case the disturbance vector  $u$  is white noise. From the second regression, a vector of fitted values,  $Z\hat{\gamma}$ , is computed and then introduced as an extra regressor in the null model:

$$y = X\beta + \alpha Z\hat{\gamma} + \text{residuals}. \quad (6)$$

The test statistic is the  $t$  statistic for the artificial hypothesis that  $\alpha = 0$ . This test was called the  $J$  test.

Fisher and McAleer (1981) devised a modified test, which they called the  $JA$  test, for which under the null the statistic has *exactly* Student's  $t$  distribution if the disturbances are Gaussian. Instead of the test regressor  $Z\hat{\gamma}$  in (6), the  $JA$  test uses the fitted values of the regression on  $Z$  of the fitted values  $X\hat{\beta}$  from the first regression. For the  $J$  test itself, only the asymptotic distribution is available, as standard normal. Indeed, in finite samples, the null distribution is shifted to the right relative to standard normal. The advantage of the possibility of exact inference in finite samples with the  $JA$  test is somewhat offset by the fact that the  $JA$  test is often considerably less powerful than the  $J$  test, and is never more so.

A quite general idea that arose out of the non-nested hypothesis testing literature is that of **encompassing**. If a model corresponding to a hypothesis  $H_1$ , say, is correctly specified, then, asymptotically, it is said to encompass another model

$H_2$ , which may be nested in  $H_1$  or not, and in which  $H_1$  may be nested or not, if assuming that  $H_1$  is well specified accounts for the results of estimating  $H_2$ . Specifically, in the non-nested linear regression case, this is interpreted to mean that the fitted values  $Z\hat{\gamma}$  obtained by estimating  $H_2$  directly should not be significantly different from the fitted values  $Z\tilde{\gamma}$  from the regression of  $X\hat{\beta}$  on  $Z$ , and that the probability limits as the sample size tends to infinity of  $\hat{\gamma}$  and  $\tilde{\gamma}$  should be the same. An encompassing test can therefore be based on the difference between  $\hat{\gamma}$  and  $\tilde{\gamma}$ . Such a test was proposed by Dastoor (1983) and developed further by Mizon (1984) and Mizon and Richard (1986).

### *Artificial regressions*

A suitable notation for a nonlinear regression is:

$$y = \mathbf{x}(\boldsymbol{\beta}) + u, \quad (7)$$

where the  $n$ -vector  $\mathbf{x}(\boldsymbol{\beta})$  is the vector of regression functions, which are in general nonlinear functions of the parameter  $k$ -vector  $\boldsymbol{\beta}$ , and of explanatory variables. An exceedingly useful tool in the estimation of nonlinear regression models is the **Gauss–Newton regression** (GNR). The GNR that corresponds to (7) can be written as:

$$y - \mathbf{x}(\boldsymbol{\beta}) = X(\boldsymbol{\beta})\mathbf{b} + \text{residuals}, \quad (8)$$

with the regressand and regressor matrix functions of  $\boldsymbol{\beta}$ . The regressor matrix  $X(\boldsymbol{\beta})$  is the  $n \times k$  Jacobian matrix of  $\mathbf{x}(\boldsymbol{\beta})$  with respect to  $\boldsymbol{\beta}$ —its  $(t, i)$  element is the partial derivative of  $x_t(\boldsymbol{\beta})$ , component  $t$  of  $\mathbf{x}(\boldsymbol{\beta})$ , with respect to  $\beta_i$ , component  $i$  of  $\boldsymbol{\beta}$ .

The GNR is the archetype of the class of artificial regressions. If the variables in (8) are evaluated at the nonlinear least squares (NLS) estimator  $\hat{\boldsymbol{\beta}}$ , which minimizes the sum of squared residuals from (7), then the OLS estimate of the artificial parameter vector  $\mathbf{b}$  is identically zero. Further, the OLS covariance matrix estimator from the GNR (8) is equal to the conventional NLS covariance matrix estimator. If the variables in (8) are evaluated at any  $\hat{\boldsymbol{\beta}}$  in a root- $n$  neighbourhood of  $\hat{\boldsymbol{\beta}}$ , then the GNR covariance matrix estimate is asymptotically equivalent to the NLS estimate. In this case, the vector  $\hat{\mathbf{b}}$  estimated by OLS is not zero in general, but it can be interpreted as a vector of correction terms, in the sense that the **one-step estimator** given by  $\boldsymbol{\beta} + \hat{\mathbf{b}}$  is asymptotically equivalent to the NLS estimator  $\hat{\boldsymbol{\beta}}$ . Much further information on the GNR can be found in the (1993) textbook of Davidson and MacKinnon. The methodology of artificial regressions in much more general contexts is laid out in Davidson and MacKinnon (2001).

The GNR makes it simple to implement hypothesis tests, including specification tests, with test statistics that are  $t$  statistics or  $F$  statistics based on the GNRs that correspond to the null and alternative models. For instance, if the parameter vector  $\boldsymbol{\beta}$  is partitioned as  $[\beta_1; \beta_2]$ , and one wishes to test the hypothesis that  $\beta_2 = \mathbf{0}$ , the two GNRs are formulated as:

$GNR_0$ :  $y - x(\beta_1, \mathbf{0}) = X_1(\beta_1, \mathbf{0})b_1 + \text{residuals}$ , and

$GNR_1$ :  $y - x(\beta_1, \mathbf{0}) = X_1(\beta_1, \mathbf{0})b_1 + X_2(\beta_1, \mathbf{0})b_2 + \text{residuals}$ ,

where  $X_1$  and  $X_2$  are Jacobian matrices of  $X$  with respect to  $\beta_1$  and  $\beta_2$  respectively, and where  $\beta_1$  is evaluated at either the constrained or the unconstrained estimate, or indeed anything in a root- $n$  neighbourhood of these. The test statistic is the ordinary  $F$  statistic for the artificial hypothesis that  $b_2 = \mathbf{0}$ . Different choices of  $\beta_1$  lead to numerically different statistics, but all are asymptotically equivalent to a Wald test of the restriction  $\beta_2 = \mathbf{0}$  computed using the nonlinear regression (7). In finite samples, simulation experiments suggest that the best choice of  $\beta_1$  is the constrained estimate. In that case, the test is a version of the Lagrange multiplier test. If  $\beta_2$  has only one component, it is not necessary to formulate  $GNR_0$ , as a suitable statistic is the  $t$  statistic for the single component of  $\beta_2$  in  $GNR_1$ .

By use of the GNR, it is easy to extend the  $J$  and  $JA$  tests to the case of non-nested nonlinear regressions. For a nonlinear  $J$  test of the hypotheses:

$$H_1: y = x(\beta) + u \quad \text{and} \quad H_2: y = z(\gamma) + u,$$

the testing regression is:

$$y - x(\hat{\beta}) = X(\hat{\beta})b + \alpha z(\hat{\gamma}) + \text{residuals}.$$

An example of an artificial regression outside the context of regression models is the BRMR (binary response model regression), developed independently by Davidson and MacKinnon (1984) and Engle (1984). Although the most commonly used binary response models are probit and logit, the artificial regression can be applied to any model with a binary dependent variable that can take on only the values of 0 and 1, of the form:

$$\Pr(y_t = 1) = F(X_t\beta),$$

where  $F$  is an absolutely continuous cumulative distribution function defined on the whole real axis. As usual,  $X_t$  is a vector of explanatory variables for observation  $t$  and  $\beta$  is a parameter vector, usually estimated by maximum likelihood. Observation  $t$  of the BRMR is written as:

$$\frac{y_t - F(X_t\beta)}{(F(X_t\beta)(1 - F(X_t\beta)))^{1/2}} = \frac{X_t f(X_t\beta)}{(F(X_t\beta)(1 - F(X_t\beta)))^{1/2}} b + \text{residual}, \quad (9)$$

where  $f = F'$  is the density corresponding to  $F$ . It is easy to see that the artificial regression (9) shares the properties of the GNR: if the variables are evaluated at the ML estimates  $\hat{\beta}$ , the OLS estimate of the artificial parameter vector  $b$  is zero; the OLS covariance matrix estimate from (9) is a consistent estimate of the inverse of the information matrix when the variables are evaluated at some  $\beta$  in a root- $n$  neighbourhood of  $\hat{\beta}$ , and the one-step property holds. This implies that the BRMR can be used for hypothesis testing in just the same way as the GNR.

### *From Hausman to DWH*

As mentioned above, as early as 1954, Durbin based a test for the null of exogenous regressors in a linear regression on the difference between the OLS and IV estimators. In a very influential paper, Hausman (1978) proposed a specification test of a very general nature. The test can be used whenever two estimators of a parameter or vector of parameters are available—one consistent under an alternative hypothesis, but inefficient under the null, the other efficient under the null, but inconsistent under the alternative. The test statistic is based on the difference between the two estimators.

Hausman also showed that, asymptotically, the covariance matrix of the difference is the difference between the covariance matrices of the estimator consistent under the alternative and the estimator efficient under the null. Let  $\hat{\theta}$  denote the former estimator, and  $\tilde{\theta}$  the latter. The difference vector is represented asymptotically by:

$$n^{1/2}(\hat{\theta} - \tilde{\theta}) = n^{1/2}(\hat{\theta} - \theta_0) - n^{1/2}(\tilde{\theta} - \theta_0). \quad (10)$$

For ease of notation, let  $s(\hat{\theta}) = n^{1/2}(\hat{\theta} - \theta)$ ,  $s(\tilde{\theta}) = n^{1/2}(\tilde{\theta} - \theta_0)$ , and  $t = n^{1/2}(\hat{\theta} - \tilde{\theta})$ , so that (10) can be written as:

$$s(\hat{\theta}) = s(\tilde{\theta}) + t. \quad (11)$$

The limiting distributions of all three terms in (10) or in (11) are normal with zero expectation under standard regularity conditions. Now  $t$  and  $s(\tilde{\theta})$  are uncorrelated, since otherwise  $\tilde{\theta}$  could not be efficient under the null. Indeed, if the matrix of covariances of the elements of  $s(\tilde{\theta})$  and  $t$  is nonzero, one can readily construct a linear combination of the two with smaller variance than that of  $s(\tilde{\theta})$ . If we now take the variance of each side of (11), we see that:

$$\text{var}(s(\hat{\theta})) = \text{var}(s(\tilde{\theta})) + \text{var}(t).$$

Thus the covariance matrix of  $t$  is equal to the difference between those of  $s(\hat{\theta})$  and  $s(\tilde{\theta})$ , which proves Hausman's claim. This result can also be shown directly for Durbin's test.

As was acknowledged later, in a (1973) paper Wu, too, had a version of the result of Durbin and Hausman. The specification test, whether in the context of linear regressions or more generally, is now commonly referred to as the DWH test. It can almost always be readily implemented by a suitable choice of artificial regression.

### *The information matrix test*

Somewhat in the same spirit as tests based on the difference between two estimators is the information matrix test. More than any of the other tests so far considered here, the information matrix test was conceived as a pure specification test, that is, without a formally specified alternative hypothesis. It can be applied to models estimated by maximum likelihood (ML), where, if a model is

well specified, there are at least two asymptotically equivalent estimators of the information matrix.

The log likelihood of a model can be expressed as a sum, one term for each observation in the sample. If the term from observation  $t$ ,  $t = 1, \dots, n$ , is denoted as  $\ell_t(y_t, \boldsymbol{\theta})$ , where  $\boldsymbol{\theta}$  is a  $k$ -vector of model parameters, the loglikelihood of the entire sample is the sum of the  $\ell_t(y_t, \boldsymbol{\theta})$ , where, in the time-series context,  $\ell_t(y_t, \boldsymbol{\theta})$  would be the log of the density of  $y_t$  conditional on exogenous variables and past observations.

The exponential of a term  $\ell_t(y_t, \boldsymbol{\theta})$  is a conditional density, and so it integrates to one: the relation

$$\int_{-\infty}^{\infty} \exp(\ell_t(y_t, \boldsymbol{\theta})) dy_t = 1$$

holds identically in  $\boldsymbol{\theta}$ . Differentiating this identity with respect to  $\theta_i$ , component  $i$  of  $\boldsymbol{\theta}$ , gives

$$\int_{-\infty}^{\infty} \frac{\partial \ell_t}{\partial \theta_i}(y_t, \boldsymbol{\theta}) \exp(\ell_t(y_t, \boldsymbol{\theta})) dy_t = 0,$$

which expresses the fact that the expectation of the score vector, evaluated at the true parameters, is zero.

Differentiating again, now with respect to  $\theta_j$ , gives:

$$\int_{-\infty}^{\infty} \left[ \frac{\partial^2 \ell_t}{\partial \theta_i \partial \theta_j}(y_t, \boldsymbol{\theta}) + \frac{\partial \ell_t}{\partial \theta_i}(y_t, \boldsymbol{\theta}) \frac{\partial \ell_t}{\partial \theta_j}(y_t, \boldsymbol{\theta}) \right] \exp(\ell_t(y_t, \boldsymbol{\theta})) dy_t = 0. \quad (12)$$

The identity (12), summed over  $t$ , is the information matrix equality. It is a standard result of maximum likelihood theory. It says that the expectation of the Hessian (the matrix of cross-partial derivatives) of  $\ell_t$  with respect to the parameters is equal to the negative of the expectation of the outer product of the gradient, that is, the  $k \times k$  matrix with typical element  $\partial \ell_t / \partial \theta_i(y_t, \boldsymbol{\theta}) \partial \ell_t / \partial \theta_j(y_t, \boldsymbol{\theta})$ .

The information matrix for the sample is defined as follows:

$$\mathcal{I}_{ij}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \sum_{t=1}^n \left[ \frac{\partial \ell_t}{\partial \theta_i}(y_t, \boldsymbol{\theta}) \frac{\partial \ell_t}{\partial \theta_j}(y_t, \boldsymbol{\theta}) \right]$$

The notation  $E_{\boldsymbol{\theta}}$  denotes an expectation taken using the probability distribution defined by  $\boldsymbol{\theta}$ . Given the information matrix equality, it follows that two different estimators of  $\mathcal{I}_{ij}(\boldsymbol{\theta})$ , both of them consistent, are the outer product of the gradient and the empirical Hessian:

$$\sum_{t=1}^n \left[ \frac{\partial \ell_t}{\partial \theta_i}(y_t, \boldsymbol{\theta}) \frac{\partial \ell_t}{\partial \theta_j}(y_t, \boldsymbol{\theta}) \right] \quad \text{and} \quad - \sum_{t=1}^n \frac{\partial^2 \ell_t}{\partial \theta_i \partial \theta_j}(y_t, \boldsymbol{\theta}),$$

respectively.

There are various ways, some easier than others, to estimate the covariance, under the null of correct specification, of the difference between the two esti-

mators with the non-redundant elements stacked into a vector. This allows the econometrician to construct a test with  $k(k+1)/2$  degrees of freedom.

Since the test is apparently constructed with no specific alternative in view, it is not obvious how to interpret a rejection of the null. However, Davidson and MacKinnon (1987) had shown that even tests with no explicit alternative have an implicit alternative, and so it was of interest to see what that implicit alternative might be for the information matrix test. In Davidson and MacKinnon (1992), the implicit alternatives, plural since the test has more than one degree of freedom, were elucidated for regression models. Tests of the elements of the information matrix corresponding to the regression parameters are implicitly against alternatives of heteroskedasticity of the disturbances, those corresponding to a regression parameter and the variance parameter are against alternatives of skewness, and the test corresponding to the diagonal element for the variance parameter is against an alternative of kurtosis. This paper also proposed a double-length artificial regression for the implementation of the information matrix test, and showed that it had better properties than implementations based on other artificial regressions. This was an example of something that econometricians were becoming aware of, namely that not all artificial regressions gave reliable inference in finite samples. For instance, the so-called outer-product-of-the-gradient (OPG) regression is both simple to implement and applicable quite generally to models estimated by ML. A vector of ones is regressed on the matrix with typical element  $\partial \ell_t / \partial \theta_i(y_t, \theta)$ . Unfortunately, this simple artificial regression is particularly unreliable.

#### 4. The generalized method of moments

The generalized method of moments (GMM) burst upon the world of econometrics with the seminal papers of Hansen (1982) and Hansen and Singleton (1982). According to the (ordinary) method of moments, if one has a sample of independent drawings from some distribution, one can estimate any moment of the distribution by the corresponding sample moment. When one speaks of the *generalized* method of moments, several generalizations are in fact implied. Some involve no more than relaxing regularity conditions, for instance, the assumption of IID observations. But the essential generalizations follow from three things. The first is that conditional moments may be used as well as unconditional ones, the second that moments may depend on unknown parameters, and the third that there may be more moments than parameters to estimate.

The GMM setup can be written very generally as follows. Let  $m_j(\mathbf{y}, \theta), j = 1, \dots, l$  be a set of functions of observed data  $\mathbf{y}$ , and a  $k$ -vector of parameters  $\theta, k \leq l$ , such that:

$$E_{\theta}(m_j(\mathbf{y}, \theta)) = 0. \tag{13}$$

As in the ML context, we denote by  $E_{\theta}$  an expectation computed using a DGP for which the true parameter vector is  $\theta$ . The relations (13) are called

**moment conditions.** Sometimes the model specifies that a set of *conditional* moments have zero expectation, in which case we have conditional moment conditions. GMM estimation with unconditional moment conditions replaces (13) by sample analogues. But, just as with over-identified estimation by IV, there may be more moment conditions than parameters to estimate, if  $l > k$ . In such cases, one uses a positive-definite **weighting matrix**  $A$  and seeks to minimize the GMM criterion function:

$$Q(\theta) = \mathbf{m}^\top(\mathbf{y}, \theta) \mathbf{A} \mathbf{m}(\mathbf{y}, \theta) \quad (14)$$

with respect to  $\theta$ . Here,  $\mathbf{m}(\mathbf{y}, \theta)$  is the  $l$ -vector of the  $m_j(\mathbf{y}, \theta)$ . Different choices of  $A$  lead to different estimators, which are not in general asymptotically equivalent. It can be seen that, if  $A$  is asymptotically proportional to the inverse of the covariance matrix of  $\mathbf{m}(\mathbf{y}, \theta)$ , then the asymptotic variance of the estimator is minimized.

One of the advantages of GMM as an estimation method is that it permits models which consist of a very large set of DGPs. In striking contrast to ML estimation, where the model must be completely specified, any DGP is admissible if it satisfies a relatively small number of restrictions or regularity conditions. However, these restrictions must go beyond the mere existence of the moments or conditional moments in the model. This is a consequence of the theorem of Bahadur and Savage (1956), which shows that, even for a problem as simple as estimating the expectation of a distribution from which one has an IID sample, the mere existence of the expectation is not enough for any statistical inference to be possible without further more restrictive assumptions.

Despite this limitation, it appears that the task of specification testing is considerably lightened when GMM is used rather than ML. Testable parametric restrictions may of course be of interest. But the chief specification test used with GMM estimation is a test of the over-identifying restrictions. This test can be implemented in a manner very much like the Sargan test used with IV estimation. If the weighting matrix  $A$  is not only asymptotically proportional to the inverse of the covariance matrix of  $\mathbf{m}(\mathbf{y}, \theta)$ —usually it would be a consistent estimate of that matrix—but asymptotically equal to it, then the minimized value of the GMM criterion function (14) has an asymptotic distribution of  $\chi^2_{l-k}$  if the model is well specified. The test based on this statistic is called the Hansen–Sargan test, or sometimes, unfortunately, the  $J$  test, in conflict with the non-nested hypothesis test of Davidson and MacKinnon (1981).

If the Hansen–Sargan test rejects the null of correct specification, that is, the correctness of the relations (13), even if the test has a formal implicit alternative, it can be difficult to interpret a rejection, for the same reason as with the Sargan test, namely that various misspecifications may lead to a failure of (13).

#### *GMM, instrumental variables and estimating functions*

The moments  $m_j(\mathbf{y}, \theta)$ , or generalized moments as it is useful to call them, are often sums of contributions from the individual observations of the sample. Consider the example of a linear regression. Generalized moments can be constructed

as linear combinations of the residuals. The OLS estimator of the regression  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$  is a GMM estimator, with moments the components of  $\mathbf{X}^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ . Setting them equal to zero gives the OLS estimating equations.

As with Eicker and White, it took some time before econometricians (and statisticians) realized that GMM was a different way of looking at a problem that had been studied in the statistics literature. The concept of **estimating functions** was put forward in Godambe (1960), but better developed for the purposes of econometrics in Godambe and Thompson (1978). A quasi-likelihood approach is found in Godambe and Thompson (1989).

For the OLS example above, the estimating functions are just the left-hand sides of the estimating equations. They are also called **zero functions**, that is, functions of data and parameters that have expectation zero when the true parameters are used. We can think of the residuals as *elementary* zero functions, corresponding to a single observation. They are combined into estimating functions by taking linear combinations of them with weights which come from instrumental variables that in the OLS setting are just the regressors, but, if these are endogenous or measured with error, as we saw earlier, we must replace them with valid instruments.

#### *HAC covariance matrix estimators*

Because GMM seeks as great generality as possible, it is contrary to its spirit to suppose that observations are IID. Although, as we saw earlier, use of an HCCME allows a researcher to be agnostic about whether there is heteroskedasticity; a similar approach allows for serial correlation by means of a **heteroskedasticity and autocorrelation consistent** (HAC) covariance matrix estimator. Let  $\boldsymbol{\Omega}$  be the  $n \times n$  covariance matrix of a vector of disturbances, or, more generally, of a vector  $\mathbf{u}(\mathbf{y}, \boldsymbol{\theta})$  of elementary zero functions, and let  $\mathbf{W}$  be an  $n \times l$  matrix of instrumental variables that are, asymptotically at least, uncorrelated with the zero functions. The aim is to be able to estimate the  $l \times l$  matrix  $n^{-1}\mathbf{W}^\top\boldsymbol{\Omega}\mathbf{W}$  consistently. If this can be done, then asymptotically valid inference is possible for quite general sorts of GMM models.

Analogously to the HCCME, a double sum could allow us to take account of covariances, but it is necessary to downplay or eliminate from the double sum terms for which the indices  $t$  and  $s$  are too far apart. First, define the auto-covariance matrices:

$$\boldsymbol{\Gamma}(j) \equiv \begin{cases} \frac{1}{n} \sum_{t=j+1}^n \mathbf{E}(u_t u_{t-j}) \mathbf{W}_t^\top \mathbf{W}_{t-j} & \text{for } j \geq 0, \\ \frac{1}{n} \sum_{t=-j+1}^n \mathbf{E}(u_{t+j} u_t) \mathbf{W}_{t+j}^\top \mathbf{W}_t & \text{for } j < 0, \end{cases}$$

where the expectations are taken conditional on the instruments. It is easy to check that  $\boldsymbol{\Gamma}(j) = \boldsymbol{\Gamma}^\top(-j)$ , and that:

$$n^{-1}\mathbf{W}^\top\boldsymbol{\Omega}\mathbf{W} = \sum_{j=-n+1}^{n-1} \boldsymbol{\Gamma}(j).$$



Estimates  $\hat{\Gamma}(j)$  are obtained by replacing  $E(u_t u_{t-j})$  by  $\hat{u}_t \hat{u}_{t-j}$  and similarly for  $E(u_{t+j} u_t)$ , where  $\hat{u}_t = u(y_t, \hat{\theta})$ , with  $\hat{\theta}$  a preliminary estimate of  $\theta$ , found, for instance, by minimizing the criterion function (14) with an identity weighting matrix.

The estimator known as the **Hansen–White estimator** was proposed in Hansen (1982) and White and Domowitz (1984). It can be written as

$$\hat{\Gamma}(0) + \sum_{j=1}^p (\hat{\Gamma}(j) + \hat{\Gamma}^\top(j)). \quad (15)$$

It makes use of a **lag truncation parameter**  $0 \leq p < n$ , and thus takes account only of covariances between observations separated by no more than  $p$  periods. Unfortunately the result is not guaranteed to be positive definite.

By far the most frequently used HAC estimator is the **Newey–West estimator**; Newey and West (1987). It also uses a lag truncation parameter, but the terms in (15) are multiplied by weights that decline linearly from 1 for the term  $\hat{\Gamma}(0)$  to 0 for the term with  $j = p + 1$ :

$$\hat{\Gamma}(0) + \sum_{j=1}^p \left(1 - \frac{j}{p+1}\right) (\hat{\Gamma}(j) + \hat{\Gamma}^\top(j)). \quad (16)$$

This estimator is guaranteed to be at least positive semi-definite.

Although Andrews and Monahan (1992) derived the most efficient positive semi-definite HAC estimator, it is seldom used, since it is more complicated than the Newey–West estimator, and is inclined in computation not to be positive semi-definite when  $p$  is large, on account of rounding errors in floating-point arithmetic.

In this setup, the vector  $m(y, \theta)$  in (14) is  $W^\top u(y, \theta)$  and the criterion function to be minimized with respect to  $\theta$  is:

$$u^\top(y, \theta) W (W^\top \hat{\Omega} W)^{-1} W^\top u(y, \theta), \quad (17)$$

where  $W^\top \hat{\Omega} W$  is given by (16). The covariance matrix of the estimator  $\hat{\theta}$  obtained by minimizing (17) can be estimated by:

$$\widehat{\text{var}}(\hat{\theta}) = \hat{U}^\top W (W^\top \hat{\Omega} W)^{-1} W^\top \hat{U},$$

where the  $n \times k$  matrix  $\hat{U}$  has typical element  $\partial u_t / \partial \theta_i(y_t, \hat{\theta})$ .

### *Indirect inference*

The method of **indirect inference** was developed by Smith (1993), and Gouriéroux et al. (1993). The idea is that, when a model is difficult to estimate, there may be an **auxiliary model** that is not too different from the model of interest but is much easier to estimate. For any two such models, there must exist so-called **binding functions** that relate the parameters of the model of interest to those of the auxiliary model. The idea of indirect inference is to estimate the parameters

of interest from the parameter estimates of the auxiliary model by using the relationships given by the binding functions.

A good example is using an auxiliary AR model to estimate an MA model. Durbin (1959) based his estimator for the parameters of a zero-mean MA( $q$ ) model on a system of equations linking the MA parameters to the coefficients of a truncated AR( $\infty$ ) representation; here the use of binding functions predates the formal introduction of the method of indirect inference. Several estimators can be constructed from these binding functions. Galbraith and Zinde-Walsh (1994) and (1997) evaluated MA estimators and then extended this approach to ARMA models where the binding function equations can be represented in the form:

$$\mathbf{B}\mathbf{c} - \mathbf{a} = \mathbf{0},$$

where  $\mathbf{c} = [\boldsymbol{\alpha} \ ; \ \boldsymbol{\theta}]$  is the vector of the parameters of the ARMA( $p, q$ ) model, the vector  $\mathbf{a}$  is formed by the first  $p + q$  coefficients of the auxiliary AR, and the  $(p + q) \times (p + q)$  matrix  $\mathbf{B}$  is partitioned as:

$$\mathbf{B} = \begin{bmatrix} \mathbf{I}_p & \mathbf{B}_{12} \\ \mathbf{O}_{q \times p} & \mathbf{B}_{22} \end{bmatrix},$$

with the  $\mathbf{B}_{..}$  submatrices having the coefficients of the AR as the non-zero entries. The  $q \times q$  matrix  $\mathbf{B}_{22}$  plays a special role, as it is identical to the resultant of the AR( $p$ ) and MA( $q$ ) polynomials in the ARMA( $p, q$ ) model. This makes it possible to test the specification of ARMA, since the hypothesis that there are common roots in the AR and MA parts can be formally stated as testing the restriction  $g(\mathbf{a}) = \det \mathbf{B}_{22} = 0$ . The test uses the Wald statistic:

$$\frac{g(\hat{\mathbf{a}})^2}{\mathbf{G}^\top(\hat{\mathbf{a}})\mathbf{V}(\hat{\mathbf{a}})\mathbf{G}(\hat{\mathbf{a}})},$$

where  $\hat{\mathbf{a}}$  is the vector of estimated truncated AR parameters,  $\mathbf{G}(\hat{\mathbf{a}})$  is the vector of derivatives  $(\partial g / \partial \mathbf{a})(\hat{\mathbf{a}})$ , and  $\mathbf{V}(\hat{\mathbf{a}})$  is the covariance matrix of the AR estimated coefficients. Another application of indirect inference to the estimation of MA models is found in Ghysels et al. (2003).

This illustrates the typical set-up for indirect inference: restrictions on the parameters of interest are expressed through the auxiliary model parameters using the binding function, then the estimators and the estimated variance from the auxiliary model is used in forming the statistics, whether they be likelihood ratio type statistics, or Wald.

Indirect inference is applied in complex situations where direct inference may simply be difficult (as in the ARMA example), but could also be impossible. A current literature extends this methodology to dynamic stochastic general equilibrium models where it may well be argued that complete specification may not be achievable, see for instance the discussion of the difficulties and pitfalls in Dridi et al. (2007) where nevertheless a partial indirect inference approach is successfully explored.

In summary, the GMM methodology has proved to be very useful in parametric models. Extensions of GMM to wider model classes, for instance for functions in Hilbert space that require a continuum of moment conditions, were developed in Carrasco and Florens (2000) and subsequent literature.

## 5. Unit roots and cointegration

The phenomenon of **spurious regressions** was brought to the attention of econometricians by Granger and Newbold (1974), who used simulation methods that were very crude by today's standards. If  $x_t$  and  $y_t$  are time series that are entirely independent of each other, we might hope that running the simple linear regression:

$$y_t = \beta_1 + \beta_2 x_t + v_t \tag{18}$$

would usually produce an insignificant estimate of  $\beta_2$  and an  $R^2$  near 0. However, this is so only under quite restrictive conditions on the nature of the  $x_t$  and  $y_t$ . In particular, if  $x_t$  and  $y_t$  are independent random walks, the  $t$  statistic for  $\beta_2 = 0$  does not follow Student's  $t$  or the standard normal distribution, even asymptotically. Instead, its absolute value tends to become larger and larger as the sample size  $n$  increases. Ultimately, as  $n \rightarrow \infty$ , it rejects the null hypothesis that  $\beta_2 = 0$  with probability 1. Moreover, the  $R^2$  does not converge to 0 but to a random, positive number that varies from sample to sample. When a regression model like (18) appears to find relationships that do not really exist, it is called a spurious regression.

Spurious regressions are obviously traps for valid, or asymptotically valid, inference. They can arise more generally than in the case of independent random walks, in particular with series which, although asymptotically stationary, are strongly serially correlated. Despite this, many econometricians thought that the most probable danger was when time series had a **unit root**. Consider an ARMA process  $y_t$  generated from a white-noise process  $u_t$ :

$$(1 - A(L))y_t = (1 + B(L))u_t. \tag{19}$$

Here  $A(L)$  and  $B(L)$  are lag polynomials with no constant term. If all the roots of the polynomial  $1 - A(z)$  are outside the unit circle in the complex plane,  $y_t$  is stationary,<sup>4</sup> or, as we prefer to say,  $I(0)$ —integrated to order 0. If it has one or more roots on the unit circle, these are unit roots, and if there is one or more root inside the unit circle, the process is explosive.

Although many tests of the null hypothesis that a given series has a unit root have been proposed, the simplest and most widely-used tests for unit roots are variants of ones developed by Dickey and Fuller (1979). Consider the AR(1) model:

<sup>4</sup> Strictly speaking,  $y_t$  is not stationary unless it is initialized from the stationary distribution. It is merely asymptotically stationary. But here we follow the usual convention of ignoring this distinction, which in any case has no bearing on what follows.

$$y_t = \rho y_{t-1} + u_t, \quad t = 1, \dots, n, \quad (20)$$

where  $u_t$  is white noise. When  $\rho = 1$ , this model has a unit root and is in fact a random walk process. A testing regression for the null of a unit root can be based on (20), transforming the regressand to  $\Delta y_t$ :

$$\Delta y_t = \alpha y_{t-1} + \text{residual}, \quad (21)$$

where  $\alpha = \rho - 1$ . A statistic to test the null of a unit root is  $n$  times the OLS estimate of coefficient of  $y_{t-1}$ . Another is the  $t$  statistic. Distributions of these statistics are valid only under very restrictive conditions. Other testing regressions include more regressors than just the lagged dependent variable:

$$\begin{aligned} \Delta y_t &= \gamma_0 + \alpha y_{t-1} + \text{residual}, \\ \Delta y_t &= \gamma_0 + \gamma_1 t + \alpha y_{t-1} + \text{residual}, \\ \Delta y_t &= \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \alpha y_{t-1} + \text{residual}. \end{aligned} \quad (22)$$

The presence of the additional regressors leads to greater robustness against the presence of deterministic time trends, but also to lower power. Again, possible test statistics are  $n$  times the OLS estimate of  $\alpha$ , known as the  $z$  statistic, and the  $t$  statistic for  $\alpha$ , known as the  $\tau$  statistic.

There is no obvious way to choose between the  $z$  and  $\tau$  statistics, and they have non-standard asymptotic distributions, different both for the two versions, and different also for each of the testing regressions in (21) and (22). They are usually expressed as distributions of functionals of Brownian motion—see for instance Phillips (1987)—and are numerically simulated.

If the process  $y_t$  is a more general ARMA process than AR(1), as in (19), the residuals in the testing regressions are serially correlated. Account can be taken of this by including one or more lags of the regressand in the testing regression. If this succeeds in making the residuals look like white noise, then fortunately the asymptotic distributions are unchanged. Tests involving the lagged regressand are called augmented Dickey–Fuller (ADF) tests. There is no uniformly most powerful statistic against the alternative that  $\rho < 1$  or  $\alpha < 0$ . Instead, **point-optimal** tests have been devised—see Elliott et al. (1996).

The literature on unit-root testing is vast, and here only the basic notions have been treated. Other forms of non-stationarity, such as structural breaks as well as fractional unit roots or near-unit roots, can also be damaging for inference.

### *Cointegration*

Economic theory often suggests that certain pairs of economic variables should be linked by a long-run equilibrium relationship. Although the variables may drift away from equilibrium for a while, economic forces may be expected to act so as to restore equilibrium. But, if these variables are I(1), they either diverge from one another, like the variables in a spurious regression, or else they co-move; that is, they diverge together. This can happen if there is a linear combination of the I(1) variables that is I(0), in which case they are said to be cointegrated.

The earliest reference to cointegration appears to be Granger (1981). The best-known paper is Engle and Granger (1987), and two relatively accessible articles are Hendry (1986) and Stock and Watson (1988).

When testing for cointegration, it is important to note that the null hypothesis is that of *no* cointegration. If two variables, each  $I(1)$ , are not cointegrated, there are two unit root processes involved, whereas when they are cointegrated, only one unit root process drives both variables. The null of no cointegration postulates that there are two.

Cointegration establishes a long-run relation between variables, but short-run deviations from the equilibrium path are of interest as well. The **error correction model** (ECM) provides a way to take account of  $I(0)$  terms, such as differenced and lagged differenced variables, as well as terms that reflect the cointegrating relations. These error correction terms may have an important impact, and testing for an ECM specification is useful for econometric analysis and forecasting.

A considerably more general approach to cointegration testing was developed by Johansen (1988) and (1991). It is based on the estimation of a **vector autoregression** (VAR), by Gaussian maximum likelihood. The method allows for several possibly cointegrated variables, not just two. Denote this number by  $m$ . The **cointegrating rank**,  $r$ , is then the number of linearly independent linear combinations of the  $m$  variables that are  $I(0)$ ; clearly  $0 \leq r \leq m$ . The case with  $r=0$  implies no cointegration at all, while  $r=m$  means that all the variables are in fact  $I(0)$ . Between these limiting cases, a set of null hypotheses about the value of  $r$  can be formulated. As  $r$  decreases towards zero, each hypothesis is successively less restricted than its predecessor, and so may serve as the alternative hypothesis for its predecessor. Alternatively, a hypothesis about  $r$ ,  $r > 0$ , can be tested against the alternative that  $r=0$ .

Consider the following VAR in differences of a set of time series  $Y = [y_1 \dots y_m]$ :

$$\Delta Y = \Delta Y_1 \Gamma_1 + \dots + \Delta Y_{p-1} \Gamma_{p-1} - Y_p \Pi + U. \quad (23)$$

Here the subscripts  $1, \dots, p$  denote lags. The lag truncation parameter  $p$  is chosen, as for an ADF test, so as to make the residuals look like white noise. If  $r=0$ , the matrix  $\Pi$  of coefficients of the lagged variables in levels may have rank  $m$ ; if there is no cointegration,  $\Pi$  must be zero. If the rank of  $\Pi$  is  $r$ , then  $r$  is the cointegrating rank.

It turns out that maximizing the log likelihood under the different constraints imposed by fixing  $r$  can be performed by solving an eigenvalue-eigenvector problem for an  $m \times m$  positive definite matrix  $A$ , of which we denote the eigenvalues, in increasing order, by  $\lambda_i$ ,  $i = 1, \dots, m$ . Given a true value of  $r$ , the  $r$  smallest eigenvalues should not be significantly different from 0, and the corresponding eigenvectors span the  $r$ -dimensional space of cointegrating vectors, that is, linear combinations that are  $I(0)$ .

Test statistics are conveniently computed as likelihood ratio statistics. To test the null that  $r = r_1$ ,  $0 \leq r_1 < m$ , against the alternative that  $r = r_2$ ,  $r_1 < r_2 \leq m$ , the LR test statistic is:

$$LR = -n \sum_{i=r_1+1}^{r_2} \log \lambda_i.$$

Johansen's methodology is commonly used in practice, but, as with unit root testing, the literature on cointegration is vast.

## 6. Non parametric approaches in specification testing

It appears that Bierens (1982a) was the first econometrician to make a strong argument for non parametric model specification in econometrics, and for developing tests that, by not restricting the model to be parametric, would thus be consistent. Another paper by Bierens (1982b) is entitled "Tests of model specification in the absence of alternative hypotheses," presumably to indicate a departure from the approach of Davidson and MacKinnon (1981), the title of whose paper mentions "...model specification in the presence of alternative hypotheses."

Of course, earlier it was known that some classical non parametric distribution-free tests, such as the sign and rank tests are highly effective when applicable; unfortunately, fairly stringent restrictions such as exchangeability make this type of test unavailable in various settings. A modern exploration of these classical tests in time series models can be found in Dufour and Hallin (1987) and Campbell and Dufour (1995).

There is a vast statistical literature that predates the adoption of the non-parametric approaches in econometrics; it starts with the use of the empirical distribution function in the 1930s for Kolmogorov and Smirnov tests—Kolmogorov (1933) and Smirnov (1948). The introduction of kernel and similar estimators for density functions, by Parzen (1962) and Rosenblatt (1956), and for the conditional mean functions, by Nadaraya (1964) and Watson (1964), provided the basis for testing based on these types of estimators.

Ullah (1985) introduced consistent kernel-based tests of conditional mean specification; the statistic compared the sums of squared residuals from the parametric model and from the model with the conditional mean estimated by a nonparametric kernel estimator. The asymptotic distribution of the test statistic was established in the general case by Fan and Li (2002).

The first kernel-based statistic for specification testing in econometric time-series models was probably by Robinson (1989), who established many of the fundamental asymptotic results for kernel estimators. He points out that non parametric methods can achieve efficiency gains relative to an incorrect parametric specification, but also notes that, if the parametric model is correct, it is always possible to do better with it relative to the non parametric approach. In view of the fact that non parametric approaches rely on the selection of bandwidth and

lag length the question is raised as to whether a comparable representation could be achieved by enriching the specification of a parametric model. The so-called semi-non parametric approach pursues this path—see for instance Gallant and Nychka (1987).

Of the non parametric and semi-non parametric approaches, the former is based on evaluating functionals without explicitly considering the nonparametric function; the latter uses point-wise evaluation of the function. Both were explored in the context of testing parametric forms of conditional distributions. Andrews (1997) developed a conditional Kolmogorov test that did not require any local non parametric modelling. In local approaches to testing the parametric conditional distribution in the binary index model, Stute and Zhu (2005), construct kernel-based statistics, while Tuvaandorj and Zinde-Walsh (2014) proposed statistics that use generalized functions. Rothe and Wied (2013) provide an adaptive test and an extensive comparison of statistics. Recent advances on testing the parametric form of the conditional quantile function are in Escanciano and Goh (2014).

A non parametric approach that is only starting to make inroads into in econometric practice is based on wavelets, in particular, on wavelet shrinkage. Greenblatt (1996) outlines this approach. There have been a few applications: Ramsay and Lampart (1998), who study the relationship between the money supply and output; Davidson, Labys, and Lesourd (1998), who examine the evolution over time of commodity prices, and Stengos and Sun (2001), who develop a wavelet-based specification test for regression functions.

After the initial theoretical advances of the 1980s and the rapidly following developments in the 1990s, non parametric specifications and the corresponding tests made their way into applied work, although it took some time to convince applied researchers that parametric specifications should not be taken for granted. Nowadays it is common practice to plot the distributions or densities of data, or residuals, often with multidimensional graphics, in order to judge the suitability of a particular parametric approach. However, numerous examples make it clear that such tools may still fail unless the model is correctly specified.

A very striking example is provided in Horowitz and Härdle (1996), where the model for technology adoption in German manufacturing was estimated by probit and produced results fairly plausible at first glance. However, the non parametric specification revealed that the true probability distribution was bimodal. Consequently, little value could be attached to parametric characteristics of the technology choice.

#### *Tests for the conditional mean*

The problem most often tackled in the specification testing literature is to test a parametric specification of the regression function in a general model:

$$Y = g(X) + U, \tag{24}$$

where the random variables  $Y$  and  $X$  are jointly distributed, and the null hypothesis is  $g(x) = \gamma(x, \theta)$  for some known function  $\gamma$  and parameter vector  $\theta$  for almost all  $x$ . The alternative is that there exists no  $\theta$  such that  $g(x) = \gamma(x, \theta)$  for almost all  $x$  in the support of  $X$ .

The test statistic first introduced by Bierens (1982a) in the IID case, and later extended to stationary time series, uses least-squares residuals  $\hat{u}_j$  and weights  $w(X, \tau)$ , for instance Fourier transform weights  $w(X, \tau) = \exp(i\tau X)$  ( $i^2 = -1$ ), or some transformed version. The **integrated conditional moment (ICM)** statistic is:

$$\hat{T}_n = \int \left| \frac{1}{\sqrt{n}} \sum_{j=1}^n \hat{u}_j w(X, \tau) \right|^2 d\tau.$$

Distributional results in Bierens and Ploberger (1997) establish power against sequences of  $n^{-1/2}$  drifting alternatives. The limit distribution is a mixture of independent  $\chi_1^2$ , and so simulations are required to control the test size. These types of statistics, in the terminology of Fan and Li (2000), are of Bierens type. The book by Bierens (2017) provides an overview; it includes applications to specification tests in mixed proportional hazard, competing risk models, and models of auctions.

Kernel-based tests were proposed by Härdle and Mammen (1993) and Hong and White (1995); a particularly popular test statistic was introduced in Li and Wang (1998) and Zheng (1996); it is further developed for a novel application to propensity score testing in Shaikh et al. (2009). The test statistic is:

$$I_n^a = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i=1}^n \hat{u}_i \hat{u}_j K\left(\frac{X_i - X_j}{h}\right), \quad (25)$$

with the  $\hat{u}_j$  residuals from the parametric model,  $h$  a bandwidth, and  $K$  a kernel function. The local tests have power against alternatives only at a nonparametric rate, like  $(1/(nh^d))^{1/2}$  for instance. The asymptotic distribution is Gaussian; this follows as a result of the slower convergence rate, since then the estimation error in the parameter is negligible asymptotically, and the estimated parametric function is treated as the true parametric function. The derivation of the limit distribution relies on the asymptotics for degenerate U statistics—see Hall (1984).

Fan and Li (2000) show that if the bandwidth  $h$  is fixed, the kernel-based statistic can be interpreted as being of the Bierens type, with the corresponding asymptotic power properties. In a finite sample, of necessity the bandwidth is finite. The properties of the test can then be interpreted according to the behaviour of the bandwidth as the sample size goes to infinity. Fan and Li demonstrate that, despite the asymptotic superiority of the  $\sqrt{n}$  Bierens-type tests, the kernel-based tests may enjoy better power when the alternative exhibits spikes; the power comparisons will depend on whether we are more concerned with low or high frequency functional alternatives.

There are several issues that affect adoption of test statistics based on kernel and other local estimators. Some of the problems can be solved by improving the



choices made by the econometrician: the particular form of the statistic, choices of bandwidth (number of terms in the series, wavelet scale) and the kernel or basis functions (see Ellison and Ellison 2000, Stengos and Sun 2001, Escanciano and Goh 2014, and discussions of differing bandwidths and kernels for discrete covariates in Li and Racine 2007). Adaptive choices are explored to optimize the rate for the testing problem starting with Horowitz and Spokoiny (2001). Typically the bootstrap is employed in the testing; Lee and Ullah (2001) and Li and Wang (1998) focus on the performance of the bootstrap.

The kernel-based statistics recently were extended to non-stationary regressors. Several papers developed the distribution of statistics similar to (25) to test the null hypothesis of a parametric conditional regression function, when the regressor could be non-stationary. In Wang and Phillips (2012), the statistic is given by:

$$S_n = \sum_{t=1}^n \sum_{s \neq t=1}^n \hat{u}_{t+1} \hat{u}_{s+1} K \left( \frac{X_t - X_s}{h} \right),$$

where the  $\hat{u}_t$  are the residuals of the parametric model and  $K$  is a kernel function with bandwidth  $h$ .

Gao et al. (2009) looked at testing in a model with martingale difference disturbances and random walk regressors, while Wang and Phillips (2012) consider a general setting. In particular, they do not require independence of  $X$  and the model disturbances  $u_t$ . Under the null, the convergence of the standardized statistic to standard normal is established; the non-stationarity requires developing novel approaches to the limit theory for the U-statistic, in particular reflecting the self-intersecting properties of the non-stationary  $X$ , in order to be able to capture the distribution of  $K((X_t - X_s)/h)$  in the sample. The rate for the local alternatives in the non-stationary case is not independent of the form of the alternative and Chen et al (2015) make a study of the importance of the local deviations from the null.

### *Empirical likelihood*

Maximum likelihood provides a very powerful framework for estimation and testing, but it is resolutely parametric. Owen (1988) proposed a non parametric likelihood approach, which he called empirical likelihood. The method is used both for estimation and inference; the construction of confidence sets as well as of hypothesis tests.

Empirical likelihood shares a number of optimality properties with parametric likelihood. In particular, likelihood ratio tests can be carried out in the usual way, by use of a statistic that is equal to twice the difference between the empirical log likelihood maximized unconstrained (under the alternative) and maximized under the constraints of the null hypothesis. Under the null, the asymptotic distribution of the empirical likelihood ratio (ELR) test is chi-squared with degrees of freedom equal to the number of constraints.

For an IID sample of size  $n$ , each observation additively contributes to the empirical log likelihood (EL) the log of a parameter associated with the observation, and interpreted as a probability. This is analogous to what happens with a parametric model with a discrete dependent variable. The problem of maximizing the EL is thus as follows:

$$\max_{p_i, i=1, \dots, n} \sum_{i=1}^n \log p_i \text{ subject to } \sum_{i=1}^n p_i = 1.$$

It is easy to see that the solution to this problem is  $p_i = 1/n$  for all  $i$ . The maximized value is then  $-n \log n$ . The maximizing probabilities are the weights given to the observations in the empirical distribution function (EDF) of the sample.

To test the hypothesis that the expectation of the underlying distribution is equal to a given value  $\mu$ , the restricted EL requires solving the problem:

$$\max_{p_i} \sum_{i=1}^n \log p_i \text{ subject to } \sum_{i=1}^n p_i = 1 \text{ and } \sum_{i=1}^n p_i(y_i - \mu) = 0, \quad (26)$$

where the  $y_i$  are the observations in the sample. The discrete distribution characterised by assigning the maximizing  $p_i$  for problem (26) to  $y_i$  does indeed have expectation  $\mu$  by virtue of the two constraints. The Lagrangian for (26) is:

$$\sum_{i=1}^n \log p_i - \lambda \left( \sum_{i=1}^n p_i - 1 \right) - \gamma \left( \sum_{i=1}^n p_i(y_i - \mu) \right),$$

where  $\lambda$  and  $\gamma$  are Lagrange multipliers. The first-order conditions for the maximization are:

$$p_i = \frac{1}{\lambda + \gamma(y_i - \mu)}, \quad \sum_{i=1}^n p_i = 1, \quad \sum_{i=1}^n p_i(y_i - \mu) = 0.$$

It is not always possible to find a set of probabilities  $p_i$  that satisfy these conditions and also the requirement that  $0 \leq p_i \leq 1$ —for that, it is necessary that the origin lies in the convex hull of the  $y_i - \mu$ . But, if we assume that a feasible solution exists, then the probabilities are  $p_i = 1/(n(1 + \hat{\gamma}(y_i - \mu)))$  where  $\hat{\gamma} = \gamma/n$ . There is no analytic solution for  $\gamma$  of the first-order conditions, but the solution is easy to find numerically by a one-dimensional root-finding algorithm, after which the constrained maximum of the EL can be found.

Owen was primarily interested in constructing confidence intervals for parameters, see Owen (1990). This can be done by (numerically) solving for the values of  $\mu$  for which the ELR statistic is equal to a critical value given, for an asymptotic confidence interval, by the appropriate quantile of the  $\chi_1^2$  distribution.

Empirical likelihood emerged as a serious competitor for GMM, both for estimation and inference, with the work of Qin and Lawless (1994). Suppose that we have an IID sample,  $\{y_i\}$ ,  $i = 1, \dots, n$ , and that we wish to estimate a parameter  $k$ -vector  $\theta$ , having at our disposal  $l$  unbiased estimating functions,  $l \geq k$ , as elements of the  $l$ -vector  $\mathbf{m}(y, \theta)$ , such that, for a DGP with true parameter vector  $\theta$ ,  $E(\mathbf{m}(Y, \theta)) = \mathbf{0}$ , where  $Y$  denotes a random variable of which the  $y_i$  are

realizations. The Lagrangian for the empirical likelihood problem with this set-up is:

$$\sum_{i=1}^n \log p_i - \lambda \left( \sum_{i=1}^n p_i - 1 \right) - n \boldsymbol{\gamma}^\top \sum_{i=1}^n p_i \mathbf{m}(y_i, \boldsymbol{\theta}).$$

For a given  $\boldsymbol{\theta}$ , the probabilities that satisfy the corresponding first-order conditions are:

$$p_i(\boldsymbol{\theta}) = \frac{1}{n(1 + \boldsymbol{\gamma}^\top(\boldsymbol{\theta})\mathbf{m}(y_i, \boldsymbol{\theta}))},$$

where  $\boldsymbol{\gamma}(\boldsymbol{\theta})$  is the vector of Lagrange multipliers that satisfy the first-order conditions. The ELR statistic against an unconstrained alternative is then

$$ELR(\boldsymbol{\theta}) = \sum_{i=1}^n \log(1 + \boldsymbol{\gamma}^\top(\boldsymbol{\theta})\mathbf{m}(y_i, \boldsymbol{\theta})). \quad (27)$$

In the just identified case, with  $l = k$ , it is readily seen that the estimator  $\hat{\boldsymbol{\theta}}$  that minimises (27) is the solution of the estimating equations  $\sum_{i=1}^n \mathbf{m}(y_i, \boldsymbol{\theta}) = 0$ .

The main result of Qin and Lawless is that, for  $l > k$ ,  $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$  is asymptotically distributed as  $N(\mathbf{0}, \mathbf{V})$ , where  $\mathbf{V}$  is an easily estimated sandwich covariance matrix,  $n^{1/2}\boldsymbol{\gamma}(\hat{\boldsymbol{\theta}})$  is asymptotically  $N(\mathbf{0}, \mathbf{U})$ , with  $\mathbf{U}$  also readily estimated, and the EDF defined by the estimated probabilities  $p_i(\hat{\boldsymbol{\theta}})$  is consistent under the null. In order to test a hypothesis like  $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ , the ELR statistic is shown by Qin and Lawless to be  $2(ELR(\hat{\boldsymbol{\theta}}) - ELR(\boldsymbol{\theta}_0))$ , with the function  $ELR$  of (27).

The approach of Qin and Lawless was extended to cover conditional moment conditions, like those used for GMM estimation, by Kitamura et al. (2003) and Donald et al. (2003). The behaviour of EL, some variants of generalized empirical likelihood (GEL), along with various versions of GMM, was studied by Newey and Smith (2004) by means of stochastic expansions. There are at least two major advantages to the EL approach: bias due to a high degree of over-identification is considerably less than with any version of GMM, and there is no need to estimate a covariance matrix, as that estimation is done implicitly in the computation of the ELR statistic.

## 7. Weak instruments

A call to order not too dissimilar to the one by Sims (1980) was made by Staiger and Stock (1997). They found a number of papers that in 1997 were recent and that used IV estimation (two-stage least squares or 2SLS) for estimation and testing, but where the instruments used were “weak,” in the sense that they were poorly correlated with the endogenous regressors in the equation being estimated. They re-analyzed the data used in a paper by Angrist and Krueger (1991) on the returns to education. Angrist and Krueger had used quarter of birth and its interaction with covariates as instruments for education in an earnings equation. Staiger

and Stock found that the  $F$  statistic for the first-stage regression in the two-stage procedure was less than 5 and concluded that this was enough to invalidate the usual asymptotic inference. They develop a different asymptotic construction, which holds the first-stage  $F$  constant as the sample size tends to infinity, and find that this alternative asymptotic approach agrees much better with finite-sample simulation results.

In Stock et al. (2002), the ideas of Staiger and Stock were extended to GMM estimation, and the adjective “weak” applied to the identification afforded by the instruments, rather than to the instruments themselves. A considerable literature followed, in which a rather simple model was treated. It can be written as:

$$y_1 = \beta y_2 + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u}_1 \text{ and} \quad (28)$$

$$y_2 = \mathbf{W}\boldsymbol{\pi} + \mathbf{u}_2. \quad (29)$$

Here  $y_1$  and  $y_2$  are  $n$ -vectors of observations on endogenous variables,  $\mathbf{Z}$  is an  $n \times k$  matrix of observations on exogenous variables and  $\mathbf{W}$  is an  $n \times l$  matrix of instruments such that  $\mathcal{S}(\mathbf{Z}) \subset \mathcal{S}(\mathbf{W})$ , where the notation  $\mathcal{S}(\mathbf{A})$  means the linear span of the columns of the matrix  $\mathbf{A}$ . The disturbances are assumed to be serially uncorrelated and, for many of the analytical results, normally distributed. We assume that  $l > k$ , so that the model is either exactly identified or, more commonly, overidentified. It is desired to perform reliable inference about  $\beta$ , the coefficient of the endogenous regressor in (28).

An LM test was independently proposed by Kleibergen (2002) and Moreira in a 2001 working paper, and Moreira (2003) also proposed a conditional likelihood ratio (CLR) test. It was immediately apparent that these tests gave vastly more reliable inference than the conventional  $t$  test, which of course is a Wald test. More theoretical work followed, with optimality results in Andrews et al., (2006) and Hillier (2009a), who gave a complete elucidation of the finite-sample properties of the Kleibergen–Moreira LM test and the CLR test, including their power functions, under the assumption of Gaussian disturbances. The CDF of the CLR test when there are more than one endogenous regressor was obtained in Hillier (2009b).

Exceedingly useful for the analysis of the system (28) and (29) is a fact that the 2SLS and LIML estimators of  $\beta$  are deterministic functions of just six quadratic forms—see Mariano and Sawa (1972). They are:

$$y_1^\top \mathbf{P}_1 y_1, \quad y_1^\top \mathbf{P}_1 y_2, \quad y_2^\top \mathbf{P}_1 y_2, \quad y_1^\top \mathbf{M}_W y_1, \quad y_1^\top \mathbf{M}_W y_2 \text{ and } y_2^\top \mathbf{M}_W y_2.$$

Here  $\mathbf{P}_1 = \mathbf{P}_W - \mathbf{P}_Z$ , where  $\mathbf{P}_W$  and  $\mathbf{P}_Z$  are the orthogonal projections on to the spans of the columns of  $\mathbf{W}$  and  $\mathbf{Z}$ , respectively and  $\mathbf{M}_W = \mathbf{I} - \mathbf{P}_W$ . Under the assumption of Gaussian disturbances, Davidson and MacKinnon (2008) showed that these six quantities can be expressed as deterministic functions of eight mutually independent random variables, four of them standard normal and four  $\chi^2$ , and the model parameters. As will be seen later, this considerably alleviates the computational burden of bootstrapping the tests.

### *Identification*

It was with the turn of the century that econometricians returned their attention to structural models, after two decades of focus on specification testing. The link with economic theory was again emphasized, but combined with a need for statistical adequacy, as Sims's (1980) admonitions were not forgotten. This in turn led to a concern for identification, already evoked by the weak identification problem with IV estimation. Some models were found to be unidentified, others partially identified, others set-identified, and so on.

A fundamental concept turns out to be **non parametric identification**, for which the best reference is the Handbook chapter by Rosa Matzkin (2007), one of the pioneers in this field. Non parametric identification within a class of non parametric functions or distributions implies identification of any parametric subset of this class, but the reverse is not necessarily true.

The sort of issue to be addressed in discussing non parametric identification is illustrated by the model of the previous subsection, (28) and (29). If the matrix  $\mathbf{W}$  is not observed, then a sample of just  $y_1$  and  $y_2$  is enough to identify the joint distribution of these two variables. But this is not enough to identify  $\beta$ , unless  $y_2$  is not correlated with  $u_1$ .

Thus even in parametric models, full parameter identification may rely on assumptions that are overly strong and may be especially inappropriate when dealing with missing data or interval data, as well as structural models defined by some equilibrium conditions (Manski 2003). A more general specification may permit identifying only subsets of parameters, or some combinations of parameters, or domains in the space where the parameters may reside (for instance, Manski and Tamer 2002). Inference for such specifications can be based on testing moment conditions (including inequalities) for the identified set of parameters, possibly against a non parametric alternative—see Andrews and Barwick (2012), Bugni et al. (2015), among others.

## **8. The bootstrap**

The **bootstrap** was introduced into the statistical literature by the seminal paper of Efron (1979). What Efron proposed would in modern terminology be called the resampling bootstrap. A good number of different, simulation-based procedures subsequently arose that were either called bootstrap procedures, or recognized as similar to the bootstrap. Of these, the parametric bootstrap is the most important that does not involve resampling.

It took some years for econometricians to become aware of the bootstrap. At the World Congress of the Econometric Society in Tokyo in 1995, Joel Horowitz gave an invited talk on the bootstrap in econometrics. The main part of the content of his talk can be found in Horowitz (2000). A point he made strongly was that, although the bootstrap could be an incredibly powerful tool for reliable inference, it was all too easy to use it wrongly and get completely misleading “results.”

Later, econometricians followed the lead of statisticians in making extensive use of the bootstrap, and misuse became less frequent. The traditional problems of regression models, heteroskedasticity and serial correlation, were made accessible to bootstrapping. The **pairs bootstrap**—Freedman (1981)—was designed to be robust to heteroskedasticity, but the strangely named **wild bootstrap**—see Wu (1986), Liu (1988), Mammen (1993) and Davidson and Flachaire (2008)—is also robust to heteroskedasticity and is now preferred. Serial correlation was and is a harder nut to crack. The first attempts were various forms of the **block bootstrap**—see Künsch (1989), Politis and Romano (1994) and Gonçalves and Kilian (2004). Another approach is the **sieve bootstrap**, which is a semi-non parametric method—see Bühlmann (1997, 2002), Choi and Hall (2000) and Park (2002). An important step was taken by Park (2003) in showing how to use the bootstrap for more reliable unit root tests.

Weak instruments also posed a challenge to the bootstrap. Davidson and MacKinnon (2008, 2010) studied a number of plausible procedures and found one that not only works well but also is easy to compute, as explained earlier.

The bootstrap is a very general statistical technique. The properties of an unknown DGP that one wants to study are estimated as the corresponding properties of the **bootstrap DGP**, which is an estimate of the unknown DGP. Thus the bootstrap can be the basis for estimating the bias, the variance, the quantiles, and so on, of an estimator, test statistic or any other random quantity of interest. Although the bootstrap is most often implemented by simulation, conceptually simulation is not an essential element of the bootstrap.

The various sorts of bootstrap in use differ among themselves on account of the way in which the unknown DGP is estimated, which in turn depends on the model that represents the null hypothesis. For instance, the DGPs of a model to which classical maximum likelihood is applied are completely characterized by their parameter values. It is therefore appropriate to use as a bootstrap DGP an element of the model, whose parameters are the ML estimates—this is an instance of a **parametric bootstrap**. Clearly the optimality properties of ML imply the optimality of this bootstrap DGP. On the other hand, the much less restrictive assumptions that econometricians prefer to use nowadays mean that a purely parametric bootstrap DGP is not suitable, and is certainly not optimal.

### *The golden rules of bootstrapping*

The most straightforward form of bootstrap inference proceeds by the computation of a bootstrap  $P$  value for the null hypothesis under test. Bootstrap confidence sets are not difficult to construct but may impose a heavier computational burden than a  $P$  value. Once the statistic used to test the null has been computed and the appropriate sort of bootstrap DGP has been determined, and if, as is almost universally the case, simulation is used to implement the bootstrap principle, the procedure for computing the bootstrap  $P$  value follows the following pattern.

- (i) Compute the test statistic from the original sample; call its realized value  $\hat{\tau}$ .
- (ii) Determine the realizations of all other data-dependent things needed to set up the bootstrap DGP.
- (iii) Generate  $B$  bootstrap samples, and for each one, compute a realization of the bootstrap statistic,  $\tau_j^*, j=1, \dots, B$ .
- (iv) Compute the simulated bootstrap  $P$  value as the proportion of bootstrap statistics  $\tau_j^*$  that are more extreme than  $\hat{\tau}$ .

The bootstrap test rejects the null hypothesis at significance level  $\alpha$  if the  $P$  value is less than  $\alpha$ .

In Davidson (2007) and (2016), it is proposed that the choice of bootstrap DGP should be governed by the **golden rules** of bootstrapping, as follows:

**Golden rule 1:**

The bootstrap DGP must belong to the model that represents the null hypothesis.

**Golden rule 2:**

Unless the test statistic is pivotal for the null model, the bootstrap DGP should be as good an estimate of the true DGP as possible, under the assumption that the true DGP belongs to the null model.

The point of the first rule is that the property of the unknown DGP that we wish to estimate is the distribution of the test statistic under the null. The second rule seems quite innocuous, but many bootstrap DGPs that satisfy the first rule violate the second, for example by using parameters estimated under the alternative. In many cases, as with weak instruments, the reliability of the bootstrap is very sensitive to whether or not the second rule is obeyed.

A statistic is said to be **pivotal** with respect to a model if its distribution is the same for all the DGPs in the model. Beran (1988) showed that, in the general case in which the chosen statistic is not pivotal, the bootstrap performs an operation of **pre-pivoting**, whereby the bootstrap  $P$  value is closer to being pivotal than the original non-pivotal statistic. Hall's 1992 book discusses the asymptotic bootstrap **refinements** that arise if a statistic is asymptotically pivotal.

It make intuitive sense, therefore, to consider bootstrap **iteration**. Beran introduced the **double bootstrap**, in which the bootstrap  $P$  value is itself bootstrapped. The double bootstrap is much more computationally intensive than the single bootstrap, and further iterations increase the burden exponentially. However, bootstrap iteration is feasible in some cases, and it is an open question in what circumstances the iterative scheme converges and, if so, to what sort of limiting distribution.

**9. Concluding remarks**

Appropriate model specification, and the corresponding interpretation of the results in empirical work, is of paramount importance. The trade-offs between

parsimony and complexity of econometric specifications, in terms of the ability to obtain meaningful results, now comes into sharper focus with the availability of larger data sets and computing capacity. Model specification that makes full use of these developments and attempts to reduce the number of arbitrary assumptions now often gives more tenuous results than the more parsimonious specifications of simpler models. Specification tests provide the necessary guidance about the reliability of the models used; this paper outlines some widely used methods of specification testing.

## References

- Andrews, D. W. K. (1997) "A conditional Kolmogorov test," *Econometrica* 65, 1097–128
- Andrews, D. W. K., and P. J. Barwick (2012) "Inference for parameters defined by moment inequalities: A recommended moment selection procedure," *Econometrica* 80, 2805–26
- Andrews, D. W. K., and J. C. Monahan (1992) "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator," *Econometrica* 60, 953–66
- Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006) "Optimal two-sided invariant similar tests for instrumental variables regression," *Econometrica* 74, 715–52
- Angrist, J. D., and A. B. Krueger (1991) "Does compulsory school attendance affect schooling and earnings?," *The Quarterly Journal of Economics* 106, 979–1014
- Bahadur, R. R., and L. J. Savage (1956) "The nonexistence of certain statistical procedures in nonparametric problems," *The Annals of Statistics* 27, 1115–22
- Basmann, R. L. (1960) "On finite sample distributions of generalized classical linear identifiability tests statistics," *Journal of the American Statistical Association* 55, 650–59
- Beran, R. (1988) "Pivoting test statistics: A bootstrap view of asymptotic refinements," *Journal of the American Statistical Association* 83, 687–97
- Bierens, H. (1982a) "Consistent model specification tests," *Journal of Econometrics* 20, 105–34
- (1982b) "Tests of model specification in the absence of alternative hypotheses," *Journal of Econometrics* 20, 105–34
- (2017) *Econometric Model Specification: Consistent Model Specification Tests and Semi-Nonparametric Modeling and Inference*. World Scientific
- Bierens, H., and W. Ploberger (1997) "Asymptotic theory of integrated conditional moment tests," *Econometrica* 65, 1129–51
- Breusch, T. S. (1978) "Testing for autocorrelation in dynamic linear models," *Australian Economic Papers* 17, 334–55
- Breusch, T. S., and A. R. Pagan (1979) "A simple test for heteroskedasticity and random coefficient variation," *Econometrica* 47, 1287–94
- Bugni, F., I. Canay, and X. Shi (2015) "Specification tests for partially identified models defined by moment inequalities," *Journal of Econometrics* 185(1), 259–82
- Bühlmann, P. (1997) "Sieve bootstrap for time series," *Bernoulli* 3, 123–48
- (2002) "Bootstraps for time series," *Statistical Science* 17, 52–72
- Campbell, B. and J.-M. Dufour (1995) "Exact non-parametric orthogonality and random walk tests," *The Review of Economics and Statistics* 77, 1–16
- Carrasco, M., and J.-P. Florens (2000) "Generalisation of GMM to a continuum of moment conditions," *Econometric Theory* 16, 797–834



- Chen, J., J. Gao, D. Li, and Z. Liu (2015) "Specification testing in nonstationary time series models," *Econometrics Journal* 18, 117–36
- Choi, E., and P. Hall (2000) "Bootstrap confidence regions computed from auto-regressions of arbitrary order," *Journal of the Royal Statistical Society, Series B* 62, 461–77
- Cox, D. R. (1961) "Tests of separate families of hypotheses," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1, 105–23
- (1962) "Further results on tests of separate families of hypotheses," *Journal of the Royal Statistical Society, Series B* 24, 406–24
- Dastoor, N. K. (1983) "Some aspects of testing non-nested hypotheses," *Journal of Econometrics* 21, 213–28
- Davidson, R. (2007) "Bootstrapping econometric models," *Quantile* 3, 13–36 (in Russian)
- (2016) "Computing, the bootstrap, and economics." Presidential address to the Canadian Economics Association, *Canadian Journal of Economics* 48, 1195–214
- Davidson, R., and E. Flachaire (2008) "The wild bootstrap, tamed at last," *Journal of Econometrics* 146, 162–69
- Davidson, R., W. Labys, and J.-B. Lesourd (1998) "Wavelet analysis of commodity price behaviour," *Computational Economics* 11, 103–28
- Davidson, R., and J. G. MacKinnon (1981) "Several tests for model specification in the presence of alternative hypotheses," *Econometrica* 49, 781–93
- (1984) "Convenient specification tests for logit and probit models," *Journal of Econometrics* 25, 241–62
- (1987) "Implicit alternatives and the local power of test statistics," *Econometrica* 55, 1305–29
- (1992) "A new form of the information matrix test," *Econometrica* 60, 145–57
- (1993) *Estimation and Inference in Econometrics*. New York: Oxford University Press
- (2001) "Artificial regressions." In B. Baltagi, ed., *A Companion to Econometric Theory*, chapter 1, pp. 16–37. Oxford: Blackwell Publishers
- (2008) "Bootstrap inference in a linear equation estimated by instrumental variables," *Econometrics Journal* 11, 443–77
- Davidson, R., and J. G. MacKinnon (2010) "Wild bootstrap tests for IV regression," *Journal of Business and Economic Statistics* 28, 128–44
- Delgado, M. A., and T. Stengos (1994) "Semiparametric specification testing of non-nested econometric models," *The Review of Economic Studies* 61, 291–303
- Dickey, D. A., and W. A. Fuller (1979) "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association* 74, 427–31
- Donald, S. G., G. W. Imbens, and W. K. Newey (2003) "Empirical likelihood estimation and consistent tests with conditional moment restrictions," *Journal of Econometrics* 117, 55–93
- Dridi, R., A. Guay, and E. Renault (2007) "Indirect inference and calibration of dynamic stochastic general equilibrium models," *Journal of Econometrics* 136, 397–430
- Dufour, J.-M., and M. Hallin (1987) "Tests non paramétriques optimaux pour le modèle autoregressif d'ordre un," *Annales d'économie et statistique* 5, 411–34
- Durbin, J. (1954) "Errors in variables," *Review of the International Statistical Institute* 22, 23–32
- (1959) "Efficient estimation of parameters in moving-average models," *Biometrika* 46, 306–16
- (1970) "Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables," *Econometrica* 38, 410–21

- Durbin, J., and G. S. Watson (1950) "Testing for serial correlation in least squares regression I," *Biometrika* 37, 409–28
- (1951) "Testing for serial correlation in least squares regression II," *Biometrika* 38, 159–78
- (1971) "Testing for serial correlation in least squares regression III," *Biometrika* 58, 1–19
- Efron, B. (1979) "Bootstrap methods: Another look at the jackknife," *The Annals of Statistics* 7, 1–26
- Eicker, F. (1963) "Asymptotic normality and consistency of the least squares estimators for families of linear regressions," *The Annals of Mathematical Statistics* 34, 447–56
- (1967) "Limit theorems for regressions with unequal and dependent errors." In L. M. Le Cam and J. Neyman, eds., *Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1, 59–82
- Elliott, G., T. J. Rothenberg, and J. H. Stock (1996) "Efficient tests for an autoregressive unit root," *Econometrica* 64, 813–36
- Ellison, G., and S. F. Ellison (2000) "A simple framework for nonparametric specification testing," *Journal of Econometrics* 96, 1–23
- Engle, R. F. (1984) "Wald, likelihood ratio and Lagrange multiplier tests in econometrics." In Z. Griliches and M. D. Intriligator, eds., *Handbook of Econometrics*, Vol. 2, ch. 13, pp 775–826. Amsterdam: North-Holland
- Engle, R. F., and C. W. J. Granger (1987) "Co-integration and error correction: Representation, estimation and testing," *Econometrica* 55, 251–76
- Escanciano J. C., and S. C. Goh (2014) "Specification analysis of linear quantile models," *Journal of Econometrics* 178, 495–507
- Fan, Y., and Q. Li (2000) "Consistent model specification tests: Nonparametric versus Bierens' tests," *Econometric Theory* 16, 1016–41
- Fisher, G. R., and M. McAleer (1981) "Alternative procedures and associated tests of significance for non-nested hypotheses," *Journal of Econometrics* 16, 103–19
- Freedman, D. A. (1981) "Bootstrapping regression models," *The Annals of Statistics* 9, 1218–28
- Frisch, R. (1934) "Statistical confluence analysis by means of complete regression systems," publication no. 5, Institute of Economics, Oslo University
- Galbraith, J. W., and V. Zinde-Walsh (1994) "A simple noniterative estimator for moving average models," *Biometrika* 81, 143–55
- (1997) "On some simple, autoregression-based estimation and identification techniques for ARMA models," *Biometrika* 84, 685–96
- Gallant, A. R., and D. W. Nychka (1987) "Semi-nonparametric maximum likelihood estimation," *Econometrica* 55, 363–90
- Gao, J., M. King, Z. Lu, and D. Tjøstheim (2009) "Specification testing in nonlinear and nonstationary time series autoregression," *The Annals of Statistics* 37, 3893–928
- Ghysels, E., L. Khalaf, and C. Vodounou (2003) "Simulation based inference in moving average models," *Annales d'Économie et de Statistique* 69, 85–99
- Glejser, H. (1969) "A new test for heteroskedasticity," *Journal of the American Statistical Association* 64, 316–23
- Godambe, V. P. (1960) "An optimum property of regular maximum likelihood estimation," *The Annals of Mathematical Statistics* 31, 1208–11
- Godambe, V. P., and M. E. Thompson (1978) "Some aspects of the theory of estimating equations," *Journal of Statistical Planning and Inference* 2, 95–104
- (1989) "An extension of quasi-likelihood estimation (with discussion)," *Journal of Statistical Planning and Inference* 22, 137–72

- Godfrey, L. G. (1978a) "Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables," *Econometrica* 46, 1293–301
- (1978b) "Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables," *Econometrica* 46, 1303–10
- (1978c) "Testing for multiplicative heteroskedasticity," *Journal of Econometrics* 8, 227–36
- Goldfeld, S. M., and R. E. Quandt (1965) "Some tests for homoskedasticity," *Journal of the American Statistical Association* 60, 539–47
- Gonçalves, S., and L. Kilian (2004) "Bootstrapping autoregressions with conditional heteroskedasticity of unknown form," *Journal of Econometrics* 123, 89–120
- Gouriéroux, C., A. Monfort, and E. Renault (1993) "Indirect inference," *Journal of Applied Econometrics* 8, S85–S118
- Granger, C. W. J. (1981) "Some properties of time series data and their use in econometric model specification," *Journal of Econometrics* 16, 121–30
- Granger, C. W. J., and P. Newbold (1974) "Spurious regressions in econometrics," *Journal of Econometrics* 2, 111–20
- Greenblatt, S. A. (1996) "Wavelets in econometrics." In M. Gilli, ed., *Computational Economic Systems*, Vol. 5 of *Advances in Computational Economics*. Kluwer Academic Publishers
- Hall, P. (1984) "Central limit theorem for integrated squared error of multivariate nonparametric density estimators," *Journal of Multivariate Analysis* 14, 1–16
- (1992) *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag
- Hansen, L. P. (1982a) "Large sample properties of generalized method of moments estimators," *Econometrica* 50, 1029–54
- Hansen, L. P., and K. J. Singleton (1982) "Generalized instrumental variables estimators of nonlinear rational expectations models," *Econometrica* 50, 1269–86
- Härdle, W., and E. Mammen (1993) "Comparing nonparametric versus parametric regression fits," *The Annals of Statistics* 21, 1926–47
- Hausman, J. A. (1978) "Specification tests in econometrics," *Econometrica* 46, 1251–72
- Hendry, D. F. (1986) "Econometric modelling with cointegrated variables: An overview," *Oxford Bulletin of Economics and Statistics* 48, 201–12
- Hillier, G. (2009a) "Exact properties of the conditional likelihood ratio test in an IV regression model," *Econometric Theory* 25, 915–57
- (2009b) "On the conditional likelihood ratio test for several parameters in IV regression," *Econometric Theory* 25, 305–35
- Hong, Y., and H. White (1995) "Consistent specification testing via nonparametric series regression," *Econometrica* 63, 1133–59
- Horowitz, J. L. (1995) "bootstrap methods in econometrics: Theory and numerical performance," paper presented at the 7th World Congress of the Econometric Society, Tokyo
- (2000) "The bootstrap." In J. J. Heckman and E. E. Leamer, eds., *Handbook of Econometrics*, Vol. 5. Elsevier
- Horowitz, J. L., and W. Härdle (1996) "Direct semiparametric estimation of single-index models with discrete covariates," *Journal of the American Statistical Association* 91, 1632–40
- Horowitz, J. L., and V. G. Spokoiny (2001) "An adaptive, rate-optimal test of a parametric mean regression model against a nonparametric alternative," *Econometrica* 69, 599–631
- Johansen, S. (1988) "Statistical analysis of cointegrating vectors," *Journal of Economic Dynamics and Control*, 12, 231–54

- (1991) “Estimation and hypothesis testing of cointegration in Gaussian vector autoregressive models,” *Econometrica* 59, 1551–80
- Kitamura, Y., G. Tripathi, and H. Ahn (2003) “Empirical likelihood-based inference in conditional moment restriction models,” *Econometrica* 72, 1667–714
- Kleibergen, F. (2002) “Pivotal statistics for testing structural parameters in instrumental variables regression,” *Econometrica* 70, 1781–803
- Kolmogorov, A. (1933) “Sulla determinazione empirica di una legge di distribuzione,” *Giornale dell’Istituto Italiano degli Attuari* 4, 83–91
- Künsch, H. R. (1989) “The jackknife and the bootstrap for general stationary observations,” *The Annals of Statistics* 17, 1217–41
- Lee, T.-H., and A. Ullah (2001) “Nonparametric bootstrap tests for neglected nonlinearity in time series regression models,” *Journal of Nonparametric Statistics* 13, 425–51
- Lehmann, E. L. (2011) *Fisher, Neyman, and the Creation of Classical Statistics*. Springer
- Li, Q., and J. Racine (2007) *Nonparametric Econometrics: Theory and Practice*. Princeton, NJ: Princeton University Press
- Li, Q., and S. Wang (1998) “A simple consistent bootstrap test for a parametric regression function,” *Journal of Econometrics* 87, 145–65
- Liu, R. Y. (1988) “Bootstrap procedures under some non-i.i.d. models,” *The Annals of Statistics* 16, 1696–708
- MacKinnon, J. G., and H. White (1985) “Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties,” *Journal of Econometrics* 29, 305–25
- Mammen, E. (1993) “Bootstrap and wild bootstrap for high dimensional linear models,” *The Annals of Statistics* 21, 255–85
- Manski, C. F. (2003), *Partial Identification of Probability Distributions*, Springer-Verlag
- Manski, C. F., and E. Tamer (2002) “Inference on regressions with interval data on a regressor or outcome,” *Econometrica* 70, 519–46
- Mariano, R. S., and T. Sawa (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, 159–63
- Matzkin, R. (2007) “Nonparametric Identification.” In J. J. Heckman and E. E. Leamer, eds., *Handbook of Econometrics*, Vol. 6B, ch. 73, pp. 5307–68. Elsevier
- Mizon, G. E. (1984) “The encompassing approach in econometrics.” In K. F. Wallis and D. F. Hendry, eds., *Quantitative Economics and Econometric Analysis*. Oxford: Basil Blackwell
- Mizon, G. E., and J.-F. Richard (1986) “The encompassing principle and its application to testing non-nested hypotheses,” *Econometrica* 54, 657–78
- Moreira, M. J. (2003) “A conditional likelihood ratio test for structural models,” *Econometrica* 71, 1027–48
- Nadaraya, E. A. (1964) “On estimating regression,” *Theory of Probability and its Applications* 9, 141–42
- Newey, W. K., and R. J. Smith (2004) “Higher order properties of GMM and generalised empirical likelihood estimators,” *Econometrica*, 72, 219–55
- Newey, W. K., and K. D. West (1987) “A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix,” *Econometrica* 55, 703–708
- Owen, A. B. (1988), “Empirical likelihood ratio confidence intervals for a single functional,” *Biometrika* 75(2), 237–49
- (1990) “Empirical likelihood ratio confidence regions,” *The Annals of Statistics* 18, 90–120

- Park, J. Y. (2002) "An invariance principle for sieve bootstrap in time series," *Econometric Theory* 18, 469–90
- (2003) "Bootstrap unit root tests," *Econometrica* 71, 1845–95
- Parzen, E. (1962) "On the estimation of a probability density and the mode," *The Annals of Mathematical Statistics* 33, 1065–76
- Pesaran, M. H., and A. S. Deaton (1978) "Testing non-nested nonlinear regression models," *Econometrica* 46, 677–94
- Phillips, P. C. B. (1987) "Time series regression with a unit root," *Econometrica* 55, 277–301
- Politis, D. N., and J. P. Romano (1994) "The stationary bootstrap," *Journal of the American Statistical Association* 89, 1303–13
- Qin, J., and J. Lawless (1994) "Empirical likelihood and general estimating equations," *The Annals of Statistics* 22, 300–25
- Ramsey, J. B., and C. Lampart (1998) "The decomposition of economic relationships by time scale using wavelets: Expenditure and income," *Studies in Nonlinear Dynamics and Econometrics*, 3(1), 23–42
- Reiersøl, O. (1941) "Confluence analysis by means of lag moments and other methods of confluence analysis," *Econometrica* 9, 1–24
- Robinson, P. (1989) "Hypothesis testing in semiparametric and nonparametric models for econometric time series," *The Review of Economic Studies* 58, 437–53
- Rosenblatt, M. (1956) "Remarks on some nonparametric estimates of a density function," *The Annals of Mathematical Statistics* 27, 832–37
- Rothe, C., and D. Wied (2013) "Misspecification testing in a class of conditional distributional models," *Journal of the American Statistical Association* 108(501), 314–24
- Sargan, J. D. (1958) "The estimation of economic relationships using instrumental variables," *Econometrica* 26, 393–415
- Shaikh, A., M. Simonsen, E. Vytlacil, and N. Yildiz (2009) "A specification test for propensity score using its distribution conditional on participation," *Journal of Econometrics* 151, 33–46
- Sims, C. A. (1980) "Macroeconomics and reality," *Econometrica* 48, 1–48
- Smirnov, N. (1948) "Table for estimating the goodness of fit of empirical distributions," *The Annals of Mathematical Statistics* 19, 279–81
- Smith, A. A., Jr. (1993) "Estimating nonlinear time-series models using simulated vector autoregressions," *Journal of Applied Econometrics* 8, S63–S84
- Stengos, T., and Y. Sun (2001) "A consistent model specification test for a regression function based on nonparametric wavelet estimation," *Econometric Reviews* 20, 41–60
- Staiger, D., and J. H. Stock (1997) "Instrumental variables regressions with weak instruments," *Econometrica* 65, 557–86
- Stock, J. H., and F. Trebbi (2003) "Who invented instrumental variable regression?," *The Journal of Economic Perspectives*, 17-3, 177–94
- Stock, J. H., and M. W. Watson (1988) "Variable trends in economic time series," *The Journal of Economic Perspectives* 2, 147–74
- Stock, J. H., J. H. Wright and M. Yogo (2002) "A survey of weak instruments and weak identification in generalized method of moments," *The Journal of Business and Economic Statistics* 20, 518–29
- Stute, W. and L.-X. Zhu (2005) "Nonparametric checks for single-index models," *The Annals of Statistics* 33(3), 1048–83
- Tuvaandorj, P., and V. Zinde-Walsh (2014) "Limit theory and inference about conditional distributions." In T. B. Fomby, J. Y. Park, and Y. Chang, eds., *Advances in Econometrics*, Vol. 33, ch. 12, pp. 397–424. Essays in Honor of Peter C. B. Phillips

- Ullah, A. (1985) "Specification analysis of econometric models," *Journal of Quantitative Economics* 2, 187–209
- Wang, Q., and P. C. B. Phillips (2012) "A specification test for nonlinear nonstationary models," *The Annals of Statistics* 40, 727–58
- Watson, G. S. (1964) "Smooth regression analysis," *Sankhya: The Indian Journal of Statistics, Series A*, 26, 359–72
- White, H. (1980) "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica* 48, 817–38
- White, H., and I. Domowitz (1984) "Nonlinear regression with dependent observations," *Econometrica* 52, 143–61
- Wu, C. F. J., (1986) "Jackknife, bootstrap and other resampling methods in regression analysis," *The Annals of Statistics* 14, 1261–95
- Wu, D.-M. (1973) "Alternative tests of independence between stochastic regressors and disturbances," *Econometrica* 41, 733–50
- Zheng, J. X. (1996) "A consistent test of functional form via nonparametric estimation techniques," *Journal of Econometrics* 75, 263–89