UNIVERSIDADE DO ALGARVE

UNIDADE DE CIÊNCIAS ECONÓMICAS E EMPRESARIAIS

A GAUSS-NEWTON REGRESSION APPROACH TO TESTS OF NONNESTED HYPOTHESES IN SOME NONLINEAR ECONOMETRIC MODELS

Efigénio da Luz Rebelo

Faro, 1997

UNIVERSIDADE DO ALGARVE
UNIDADE DE CIÊNCIAS ECONÓMICAS E EMPRESARIAIS

A GAUSS-NEWTON REGRESSION APPROACH TO TESTS OF NONNESTED HYPOTHESES IN SOME NONLINEAR ECONOMETRIC MODELS

Dissertação apresentada na Unidade de Ciências Económicas e Empresariais da Universidade do Algarve com vista à obtenção do grau de doutor em Economia

por

Efigénio da Luz Rebelo

Faro, 1997

ABSTRACT

The purpose of this work is to investigate nonnested tests for competing univariate dynamic linear models with autoregressive disturbances (of order p), where the motivation for Instrumental Variable estimation is mainly due to the recognised presence of current endogenous explanatory variables, either in one or in both models.

As the (Aitken) transformation of those models yields a regression function which is nonlinear in the parameters, a Gauss-Newton Regression approach is used to obtain the results.

Estimation of the competing models either by Nonlinear Least Squares or Nonlinear Instrumental Variables under differing instrument validity assumptions is also addressed. One-Step estimation (to avoid nonlinearity) is analysed.

Emphasis is placed on those techniques which are likely to be useful to applied workers. In fact, the tests deduced are all very easy to implement in an artificial linear regression either as Lagrange Multiplier tests or based on the $C(\alpha)$ principle.

The results encompass the well known J, JA, P and PA tests for univariate spherical linear models either estimated by Ordinary Least Squares or Instrumental Variables. They can also be viewed as specialisations of the more general nonlinear corresponding versions. However as we will be able to show the J and JA tests specialisations may lead to misleading conclusions if the univariate linear model with autoregressive disturbances contain either lagged or current endogenous explanatory variables.

«Given the growth in the scope of econometric theory, the backlog of important economic issues to be resolved and the progressive availability of bigger and better data sets, together with the availability of more and more powerful computing resources, further study of econometrics is likely to prove rewarding. There is certainly no shortage of work to be done.»(Jon Stewart, 1991, pp. 305-6.)

In memorial of Aida and Aida.

BIOGRAPHIC NOTES

The author was awarded a Licenciatura degree in Economics by the Instituto Superior de Economia, Universidade Técnica de Lisboa, in July 1982.

In October 1982 he was offered a lecturership at the Universidade do Algarve, Unidade de Economia e Administração.

From October 1985 to July 1986 he attended the course work of the Master degree in Probability and Statistics offered by the Faculdade de Ciências, Universidade Clássica de Lisboa.

In October 1986 he moved to the University of Manchester, Department of Econometrics and Social Statistics, on leave of absence. In June 1987 he was awarded the Diploma for Advanced Studies in Economics and in January 1989 the Master degree in Economics specialising in Econometrics.

His Master degree was recognised by the Instituto Superior de Economia e Gestão in 1990 and he was awarded the Master degree in Métodos Matemáticos Aplicados à Economia e à Gestão by the Universidade Técnica de Lisboa.

He returned to the Universidade do Algarve in 1991. Since then he has been coordinating the whole area of Quantitative Methods and in charge of several courses both in the Licenciatura and Master degrees offered by the Unidade de Ciências Económicas e Empresariais.

ACKNOWLEDGEMENTS

The author wishes to thank his supervisors, Professors Silva Ribeiro and Len Gill for their help and guidance.

Thanks are also due to Professors Adriano Pimpão and João Albino Silva for the encouragement and advice.

His appreciation is also expressed to Professors Manuel Gomes Guerreiro and José Joaquim Laginha, two Masters who mostly have influenced the author's academic life.

The author is also indebted to Dr. Emílio Rebelo and to Mr. Acácio Lopes for their help with the typing and formatting of this work.

The research was partially carried out at the Department of Econometrics and Social Statistics of the University of Manchester. The author has particular reasons to be grateful to all the colleagues of that department.

A similar acknowledgement is also due to the Universidade do Algarve for the financial support during the period of leave of absence.

Last, but not the least, the author owes thanks to his family for the understanding and support throughout such a long period of time.

TABLE OF CONTENTS

CHAPTER 1

MOTIVATION, METHODOLOGY AND SYNOPSIS

Nonnested econometric models may arise from the existence of competing economic theories. These may postulate different definitions of variables to be used in a specification, or possibly different functional forms to explain a given phenomenon. Moreover, distinct functional forms and/or stochastic specifications may also be perfectly reasonable within a single economic theory.

Consequently, very often the econometrician is faced with several rival theoretical models and none of those models can be obtained from any of the others by the imposition of appropriate parametric restrictions.

On the other hand, in applied econometric work it is not uncommon that more than one among those (nonnested) models will be declared 'data consistent' after standard diagnostic checks (nested tests). However, in this perspective, a model is evaluated only on the basis of its own performance; such tests do not make use of the information that the model being tested is only one of the several models available to explain the phenomenon. Davidson et al (1978) and Hendry and Richardson (1983) have therefore stressed the need for an assumed model to encompass competing specifications. In other words, the need for a null model to be able to predict the performance of viable alternative specifications 'significantly well'. This is the main role of nonnested testing, where the competing models are used to provide additional checks of each other's specifications. Doing that, nonnested testing can only be a serious attempt to resolve or reconcile some of the outgoing debates in economic theory. MacKinnon (1983) (with discussion and reply), McAleer (1984) and the Journal of Econometrics special issue (1983) on nonnested specification tests provide excellent introductions to the theory and application of these tests.

Emphasising however that in reality what should be estimated is not necessarily known, one cannot expect to detect inappropriate specifications with high probability in every application. There is nothing strange, therefore, if the outcome of nonnested testing, for example between a pair of models, is a non-rejection of both models. This is certainly not the reason why nonnested testing principles have not been more widely applied.

According to Bernanke et al (1988, p.294) two main reasons for such attitude are rather the following:

«Non-nested tests have not been developed for time-series models that possess general mixtures of serial correlation, lagged dependent variables, and endogenous variables (...); (...) in more complicated practical applications, especially when one wishes to implement non-nested tests based upon maximum-likelihood techniques, these tests can be quite burdensome computationally.»

This is precisely the ultimate purpose of this work: to investigate nonnested tests for competing univariate dynamic linear models with autoregressive disturbances (of order p), where the motivation for Instrumental Variable estimation is mainly due to the recognised presence of current endogenous explanatory variables, either in one or in both models.

As the (Aitken) transformation of those models yields a regression function which is nonlinear in the parameters, a Gauss-Newton regression approach is used to obtain the results.

Estimation of the competing models either by Nonlinear Least Squares or Nonlinear Instrumental Variables under differing instrument validity assumptions will be addressed. Maximum-Likelihood estimation will not be considered: not only because the resulting tests would be computationally demanding but mainly because in many applications it may also be difficult to defend the normality assumption used in the Maximum-Likelihood test procedures. On the contrary, seeking those techniques which are easy to implement, One-Step estimation (to avoid nonlinearity) will also be analysed.

The nonnested tests that we will be able to deduce are all very easy to implement in an artificial linear regression either as Lagrange Multiplier tests or based on the $C(\alpha)$ principle.

As one can infer from Bernanke et al's (1988) statement (quoted above) research on nonnested models has concentrated mostly on models that satisfy the classical assumptions of serial independence, homoskedasticity and normality of the disturbances.

In defence of the interest of this work we should stress that departures from the classical assumptions regarding the disturbances in a linear regression model arise frequently in empirical applications. Some good examples are applications on the exchange rate (Backus (1984), on money demand (Thornton (1985), Johannes and Nasseh (1985) and Milbourne (1985)), on investment (Bernanke et al (1988)) and on employment (McAleer et al (1990)). In all these papers the authors were confronted with the necessity of applying nonnested tests to competing specifications where the disturbances exhibit autoregressive errors. Whereas in the first four papers Davidson and MacKinnon's (1981) J test (developed for spherical models) has been applied careless (see the criticism by McAleer et al (1990, pp.3623-25)), in the last two papers the authors have given an important contribution to develop valid testing procedures. However, despite the statement of Bernanke et al (1988), none of those papers propose testing procedures to handle the necessity for Instrumental Variable estimation. This will be therefore our main motivation. As a by-product, we will also derive new tests for spherical univariate linear models estimated under differing instrument validity assumptions.

The plan of the thesis is as follows:

Chapter 2 provides an introduction to univariate nonlinear regression models. In this general framework, both Least Squares and Instrumental Variable estimation will receive a fair treatment. The corresponding Gauss-Newton regressions will be deduced, their properties and most interesting applications will be discussed. The use of the Gauss-Newton regression approach for testing any type of restrictions without requiring estimation of the unrestricted model is then carefully analysed. The importance of $C(\alpha)$ tests, namely in the context of instrumental variables, justifies also the focus on One-Step estimation methods throughout this chapter.

In chapter 3 we will analyse under what conditions the use of a Gauss-Newton regression is not strictly necessary to still obtain valid asymptotic inference results for the parameters of an univariate linear model with autoregressive disturbances. We will be able to conclude that when the model is dynamic and/or it contains current endogenous explanatory variables, the use of the Gauss-Newton regression is strongly recommended because, in either case, the estimation of β and ρ will not be independent. We will also show that a simplified version of the Gauss-Newton regression (with a simplified regressand) can still be considered in both cases, but then due care and attention is needed if the model contains current endogenous explanatory variables and the disturbances exhibit an autoregressive process of order greater than one.

Obviously, all this analysis will be carried out in both the contexts of Least Squares and Instrumental Variables but also in the context of their One-Step counterparts. When Instrumental Variables should be used, the choice of the set of instruments is also under consideration.

Chapter 4, on nesting procedures, should be seen as an introduction to the set up we will be using in the last chapter (artificially nesting the nonnested models). However, testing for serial correlation and for common factor restrictions has his own merits in the context of this work: the transformed (Aitken) models to be used in chapter 5 should always be viewed as restricted versions of more general alternatives; thus, transforming the models might not be the right way to proceed. In other words, in practice, testing the restrictions imposed on the unrestricted versions of the reparametrized models should always precede the nonnested testing.

Artificially nesting the competing models is in fact a very attractive and 'didactic' way of tackling nonnested models issues. That is precisely what we will do in chapter 5.

By comparison of two Gauss-Newton regressions, one associated with the null model and the other associated with a linear combination of the two competing models, we will be able to deduce de PA and P tests. As we will learn in chapter 2, both Gauss-Newton regressions should be evaluated under restricted consistent estimates under the null and therefore we will consider both Nonlinear Restricted Least Squares and Nonlinear Restricted Instrumental Variable estimates, as appropriate. We will also relax the restrictive assumption of using a common extended set of instruments to estimate both the competing models as well as the maintained Gauss-Newton regressions. Building upon the discussion about Lagrange Multiplier tests and tests based on the $C(\alpha)$ principle (carried out on previous chapters), the consideration of differing instrument validity assumptions will not prevent us to obtain valid artificial regression test statistics.

Then, we will be able to show that the JA and J tests, not based on Gauss-Newton regressions, should be conducted with due care and attention to guarantee that the inference based upon an extended version of the null model will still yield valid asymptotic results. This will come through as a natural conclusion after fully understanding the material contained in chapter 3.

Finally, we will summarise our findings, generalising the results to higher-order autoregressive cases and specialising them to the simpler case of linear spherical models, either estimated by Ordinary Least Squares or by Instrumental Variables. This summary will clarify that the results obtained in this thesis encompass the well known results for the spherical model as well as the less known results obtained by Bernanke et al (1988) and McAleer et al (1990) for linear models with autoregressive disturbances estimated by Nonlinear Least Squares.

CHAPTER 2

UNIVARIATE NONLINEAR REGRESSION MODELS

2.1 Introduction

Consider the univariate regression model

$$y_t = x_t(\beta) + \xi_t \quad , \quad \xi_t \sim \text{IID}\left(0, w^2\right) \quad , \quad t = 1, \ldots, n \tag{2.1}$$

where y_t is the t^{th} observation on the single dependent variable, β is a k-dimensional vector of parameters, the scalar function $x_t(\beta)$, referring also to period t, is a nonlinear regression function on certain explanatory variables and/or on (some of) the parameters, and ξ_t is the disturbance term corresponding to the same time period.

In some cases, $x_t(\beta)$ may also depend on lagged values of y_t (if that is the case, model (2.1) is a dynamic model) as well as on current or lagged values of other endogenous variables.

The disturbance terms are, in turn, assumed to be independent and also identically distributed in the sense that all have mean zero and constant variance w^2.

2.2 Least Squares, Instrumental Variables and the Gauss-Newton Regression

Associated with the nonlinear regression model is an artificial regression called the Gauss-Newton Regression (G.N.R.).

To derive the G.N.R. take a first-order Taylor-series approximation to (2.1) around some parameter vector $\dot{\beta}$ to obtain

$$y_t = x_t\left(\dot{\beta}\right) + dx_t(\beta)\Big|_{\beta=\dot{\beta}}\left(\beta - \dot{\beta}\right) + higher\text{ - }order\ terms + \xi_t \tag{2.2}$$

where $x_t\left(\dot{\beta}\right)$ is the nonlinear regression function evaluated at the k-dimensional point $\dot{\beta}$ and $dx_t(\beta)$ is a k-dimensional row vector of which the i^{th} element is the derivative of $x_t(\beta)$ with respect to the i^{th} parameter.

The approximation can still be written as

$$y_t - x_t\left(\dot{\beta}\right) = dx_t(\beta)\Big|_{\beta=\dot{\beta}} . b + error\ term \tag{2.3}$$

where $x_t\left(\dot{\beta}\right)$ has been taken to the left-hand side, $\beta-\dot{\beta}$ has been replaced by b and the higher-order terms have been combined into what we name 'error term'. The left-hand side in (2.3) looks now like a residual (in period t) because $x_t\left(\dot{\beta}\right)$ represents the value predicted by the model (in period t) when $\beta=\dot{\beta}$. The number of regressors is now equal to k, the i^{th} regressor being viewed as associated with β_i.

It is usually more convenient to write (2.3) in matrix notation as

$$y - \dot{x} = \dot{X}b + \ error\ vector \tag{2.4}$$

where $y-\dot{x} \equiv y - x\left(\dot{\beta}\right)$ is a n-dimensional vector of residuals, and $\dot{X} \equiv X\left(\dot{\beta}\right)$ is a $n \times k$ matrix $X(\beta)$ evaluated at $\dot{\beta}$, with typical element

$$X_{ti}\left(\dot{\beta}\right) = \frac{\partial x_t(\beta)}{\partial \beta_i}\bigg|_{\beta=\dot{\beta}}$$

where $t=1, ..., n$ and $i=1, ..., k$.

The properties of regression (2.4) will depend on how the vector $\overset{*}{\beta}$ is obtained.

Consider first the Nonlinear Least Squares (N.L.S.) estimates $\hat{\beta}$ so that the G.N.R. becomes

$$y - \hat{x} = \hat{X}\, b + \textit{ error vector} \tag{2.5}$$

where $\hat{x} \equiv x(\hat{\beta})$ and $\hat{X} \equiv X(\hat{\beta})$.

For model (2.1), the sum-of-squares function is

$$SS(\beta) = \sum_{t=1}^{n} \left(y_t - x_t(\beta)\right)^2$$

or, in matrix notation

$$SS(\beta) = \left(y - x(\beta)\right)^T \left(y - x(\beta)\right) = \left\| y - x(\beta) \right\|^2 . \tag{2.6}$$

Therefore, minimising $SS(\beta)$ is the same as minimising the Euclidean distance between y and $x(\beta)$.

The sum-of-squares function (2.6) can be rewritten as

$$SS(\beta) = y^T y - 2y^T x(\beta) + x^T(\beta)\, x(\beta)$$

yielding first-order conditions to be satisfied by the N.L.S. estimates $\hat{\beta}$

$$-2\hat{X}^T y + 2\hat{X}^T \hat{x} = 0$$

with $\hat{x}$ and $\hat{X}$ as defined in (2.5), or, simply, after collecting terms and dropping the constant factor,

$$\hat{X}^T\left(y - \hat{x}\right) = 0\,. \tag{2.7}$$

These first-order conditions show that in the G.N.R. (2.5), the regressand, which is the residual vector $y - \hat{x}$, must be orthogonal to the matrix of derivatives $\hat{X}$. In other words, we can conclude that the Ordinary Least Squares (O.L.S.) estimate of b in regression (2.5),

$$\mathring{b} = \left(\hat{X}^T \hat{X}\right)^{-1} \hat{X}^T\left(y - \hat{x}\right) \tag{2.8}$$

must be identically zero and therefore the G.N.R. must have no explanatory power whatsoever.

In the same way every nonlinear model estimated by Least Squares (L.S.) has associated with it the version of the G.N.R. in (2.5), so does every nonlinear regression model estimated by instrumental variables (I.V.).

The Nonlinear Instrumental Variables (N.L.I.V.) estimates $\tilde{\beta}$ minimises instead the I.V. sum-of-squares function

$$IVSS(\beta) = (y - x(\beta))^T P_W (y - x(\beta)) = \left\| P_W (y - x(\beta)) \right\|^2 \tag{2.9}$$

where

$$P_W = W(W^T W)^{-1} W^T$$

is a symmetric and idempotent matrix of suitable instruments, whose number should be equal or greater than k.

The objective is therefore to minimise only the portion of the Euclidean distance between y and $x(\beta)$ that lies in the subspace of W.

The criterion function in (2.9) can be rewritten as

$$IVSS(\beta) = y^T P_W y - 2y^T P_W x(\beta) + x^T(\beta) P_W x(\beta).$$

yielding first order conditions to be satisfied by the N.L.I.V. estimates $\widetilde{\beta}$

$$-2\widetilde{X}^T P_W y + 2\widetilde{X}^T P_W \widetilde{x} = 0$$

where $\widetilde{x} \equiv x(\widetilde{\beta})$ and $\widetilde{X} \equiv X(\widetilde{\beta})$, or more simply

$$\widetilde{X}^T P_W (y - \widetilde{x}) = 0 . \tag{2.10}$$

These first-order conditions show that the G.N.R. associated with model (2.1), when estimated by instrumental variables, should be

$$y - \widetilde{x} = P_W \widetilde{X} b + \textit{error vector} \tag{2.11}$$

so that the IV residual vector $y - \widetilde{x}$ must be orthogonal to the matrix of derivatives $\widetilde{X}$, after the latter have been projected onto the subspace of W.

Also, for this case, the O.L.S. estimate of b in regression (2.11),

$$\mathring{b} = \left(\widetilde{X}^T P_W \widetilde{X}\right)^{-1} \widetilde{X}^T P_W (y - \widetilde{x}) \tag{2.12}$$

must be identically zero and therefore G.N.R. (2.11) has no explanatory power either.

Thus, the only difference between this G.N.R. and the one associated with the N.L.S. estimates is that the regressors are now premultiplied by P_W as suggested by the first-order conditions in (2.10). Strictly speaking, there is an implicit choice here. In fact, one could argue that the alternative version, also suggested by the first-order conditions,

$$P_W\left(y - \widetilde{x}\right) = P_W \widetilde{X}\, b + \textit{ error vector} \tag{2.13}$$

would still have no explanatory power. Furthermore, one could even argue that this would be the natural version from first principles. That is to say, the one that one would obtain if we had taken a first-order Taylor-series approximation to the I.V. transformation of model (2.1) around $\widetilde{\beta}$. However, premultiplying also the independent variable by P_W is not a good idea since, then, the left-hand side of the equation would not represent I.V. residuals. For reasons that will become apparent in the next section, keeping the I.V. residuals in the left-hand side of the equation is rather convenient. That justifies our preference for the G.N.R. version in (2.11).

In turn, the choice to be made between N.L.S. and N.L.I.V. depends, of course, on the nature of the explanatory variables contained in $x_t(\beta)$.

Assuming that $x_t(\beta)$ depends only on strictly exogenous variables as defined by Engle, Hendry and Richard (1983) and that suitable regularity conditions, as defined by Davidson and MacKinnon (1993), are satisfied by model (2.1), then the N.L.S. estimator is asymptotically normal (regardless the fact that the disturbances are only assumed to be IID) and is also the best consistent and asymptotically linear estimator.[1]

On the contrary, if $x_t(\beta)$ depends also on current values of other endogenous variables, then the error terms are not independent of the regression function and its derivatives and therefore an I.V. approach is the one to be chosen. According to Davidson and MacKinnon (1993, p.224), the principal regularity conditions that are needed to guarantee that the N.L.I.V. estimator is consistent and asymptotically normal are those required for N.L.S. to be consistent and asymptotically normal, with the exception that the (violated) assumption of independence between the error term and both the regression function and its derivatives should be replaced by modified versions of the standard assumptions that validate the instrument set for the linear case.

2.3 The Gauss-Newton Regression Approach

Artificial regressions like the G.N.R.'s in (2.5) and (2.11) can be shown to be very useful as computational devices for certain purposes.

[1] For a detailed discussion on the regularity conditions and formal proofs, see Davidson and MacKinnon (1993, chapter 5, namely the theorems 5.1 and 5.2). A somewhat less technical reference is Seber and Wild (1989, chapter 12).

First of all, running the G.N.R.'s (2.5) and (2.11) by O.L.S. one can confirm if the N.L.S. and N.L.I.V. estimates, respectively, produced by proper packages, are sufficiently accurate. That will only be the case if the estimates $\overset{\circ}{b}$ produced by the G.N.R.'s are sufficiently close to zero. Strictly speaking, to make this decision, it is better to consider the corresponding t-values since these are dimensionless quantities.

Secondly, one can consider using the G.N.R.'s to find the N.L.S. and the N.L.I.V. estimates themselves. If one initiates the process by using consistent estimates $\beta^{(1)}$ to substitute for β in the G.N.R.'s and one runs the G.N.R.'s, then the estimates $\overset{\circ}{b}$ can be used to revise the initial estimates. In fact, as $\overset{\circ}{b} = \beta - \beta^{(1)}$, then $\beta^{(2)} = \overset{\circ}{b} + \beta^{(1)}$. Running again the G.N.R.'s after substituting $\beta^{(2)}$ for $\beta^{(1)}$, we can then find $\beta^{(3)} = \overset{\circ}{b} + \beta^{(2)}$ using the revised O.L.S. estimates $\overset{\circ}{b}$. Ideally, the process should continue until convergence is achieved, that is, when $\beta^{(n)} = \beta^{(n-1)}$. However, given the limitations of floating-point arithmetic on computers, one should stop the process as soon as $\overset{\circ}{b}$ is not statistically different from zero, as emphasised above.

Thirdly, one can also use the G.N.R.'s to very easily estimate consistently the covariance matrices of the N.L.S. and N.L.I.V. estimates. For a correctly specified nonlinear regression model where there is motivation for I.V., the ultimate result, as stated by Davidson and MacKinnon (1993, p.225) is given by

$$n^{\frac{1}{2}}\left(\widetilde{\beta}-\beta\right)\overset{a}{\sim} N\left(0, w^2 \operatorname*{Plim}_{n\to\infty}\left(n^{-1}X^T(\beta)P_W X(\beta)\right)^{-1}\right). \tag{2.14}$$

The covariance matrix estimate that the L.S. program will print when running G.N.R. (2.11) is, in turn, given by

$$Var\left(\overset{\circ}{b}\right) = s^2\left(\widetilde{X}^T P_W \widetilde{X}\right)^{-1} \tag{2.15}$$

where

$$s^2 = \frac{\left(y-\widetilde{x}\right)^T\left(y-\widetilde{x}\right)}{n-k} \tag{2.16}$$

because G.N.R. (2.11) has no explanatory power.

This estimate for w^2 uses exactly the same residuals as the I.V. estimate $\widetilde{w}^2$ from the original nonlinear regression in (2.1). Strictly speaking, the two estimates may only differ due to the fact that not all N.L.I.V. programs use the degrees-of-freedom adjustment. It is therefore obvious that s^2 in (2.16) consistently estimate w^2. [2]

[2] Notice that the Total Sum of Squares (*T.S.S.*) of the G.N.R. version in (2.13) would not provide a consistent estimate of w^2. This is one of the major reasons why the G.N.R. version in (2.11) is to be chosen.

On the other hand, as $\widetilde{\beta}$ is also a consistent estimate for β, $n^{-1}\widetilde{X}^T P_W \widetilde{X}$ must consistently estimate $n^{-1}X^T(\beta)P_W X(\beta)$. Thus it is clearly reasonable to use (2.15) to estimate the covariance matrix $w^2\left(X^T(\beta)P_W X(\beta)\right)^{-1}$, the finite-sample analog of the asymptotic covariance matrix which appears in (2.14).

Similarly, in a simpler N.L.S. context, running G.N.R. (2.5) to estimate consistently the covariance matrix of the N.L.S. estimates from model (2.1) is very easily seen as a perfectly valid way to proceed. Equations (2.14 - 16) should now be replaced by

$$n^{1/2}\left(\hat{\beta}-\beta\right)\overset{a}{\sim} N\left(0, w^2 \operatorname*{Plim}_{n\to\infty}\left(n^{-1}X^T(\beta)X(\beta)\right)^{-1}\right) \tag{2.17}$$

$$Var\left(\mathring{b}\right) = s^2\left(\hat{X}^T\hat{X}\right)^{-1} \tag{2.18}$$

and

$$s^2 = \frac{\left(y-\hat{x}\right)^T\left(y-\hat{x}\right)}{n-k} \tag{2.19}$$

because G.N.R. (2.5) would also have no explanatory power.

The use of (2.18) to estimate the finite-sample analog of the asymptotic covariance matrix which appears in (2.17) is therefore justifiable on the same grounds as before.

However, in contrast to the N.L.I.V. context, where the Sum of the Squared Residuals (*S.S.R.*) is not the value of the objective function, there is now a good reason to use the degrees-of-freedom adjustment. In fact, it can be shown that the bias of the Maximum Likelihood (M.L.) estimate of w^2 is, for nonlinear regression models and to order $O(n^{-1})$, the same as its exact bias for linear regression models. On the contrary, the fact that the degrees-of-freedom adjustment must be used for linear regression models estimated by O.L.S. suggests that, in large samples, s^2 in (2.19) should be approximately unbiased for nonlinear models.[3]

After this preamble to provide an introduction to the Gauss-Newton Regression Approach, we can finally embark on the study of its most interesting application for the purpose of this thesis. That is, the use of that approach for testing any type of restrictions on β without requiring estimation of the unrestricted model.

Sometimes this can be very convenient, namely when the unrestricted model is a nonlinear model whereas its restricted version is linear.

[3] The M.L. estimate is biased downward. In contrast to the same order, s^2 is unbiased (see Davidson and MacKinnon (1993, pp.166-7)).

Corresponding either to a Nonlinear Least Squares or to a Nonlinear Instrumental Variables set-up, two different cases should be considered.

In the first case, as we will see in section 2.3.1., the test statistic based on the Lagrange multiplier principle will simply be based on the corresponding G.N.R. associated with the unrestricted model, if evaluated at Restricted (Nonlinear) Least Squares or at Restricted (Nonlinear) Instrumental Variables, as appropriate.

In the second case, the test statistics, based on the $C(\alpha)$ principle, can still be very easily computed by comparison of the two corresponding G.N.R.'s associated with the restricted and unrestricted models, if both evaluated at any arbitrary restricted root-n consistent estimates under the null. This second case will be handled in section 2.3.2..

2.3.1. Tests Based on the Lagrange Multiplier Principle

Consider estimating the model

$$y = x(\beta) + \xi \qquad \xi \sim \mathrm{IID}\left(0, w^2 \mathrm{I}_n\right) \tag{2.20}$$

(which is model (2.1) in matrix form) subject to the p $(p \leq k)$ linearly independent restrictions

$$R\beta = r \tag{2.21}$$

where R is a $p \times k$ matrix of full row rank p, and r is p-dimensional vector.

Set up the Lagrangian

$$L(\beta, \lambda) \equiv (y - x(\beta))^T (y - x(\beta)) + (R\beta - r)^T \lambda \tag{2.22}$$

where λ is a p-dimensional vector of Lagrange multipliers.

Differentiating (2.22) with respect to β and λ one obtains the first-order conditions:

$$-2X^T(\hat{\beta}_R)(y - x(\hat{\beta}_R)) + R^T\hat{\lambda} = 0 \tag{2.23}$$

$$R\hat{\beta}_R - r = 0 \tag{2.24}$$

where $\hat{\beta}_R$ stands for Restricted Nonlinear Least Squares (or R.N.L.S.) and $\hat{\lambda}$ for the estimated Lagrange multipliers.

Equation (2.23) can be rewritten as

$$\tfrac{1}{2} R^T \hat{\lambda} = \hat{X}_R^T \left(y - \hat{x}_R \right) \tag{2.25}$$

where $\hat{X}_R \equiv X\left(\hat{\beta}_R\right)$ and $\hat{x}_R \equiv x\left(\hat{\beta}_R\right)$.

The term on the right-hand side of (2.25) is the so called Score vector, with asymptotic covariance given by

$$w^2 X^T(\beta) X(\beta) \tag{2.26}$$

because the vector of residuals $y - \hat{x}_R$ should converge under the null to the disturbance vector ξ.

We can therefore construct two numerically identical test statistics:

$$\tfrac{1}{4} \hat{\lambda}^T R \left(s_R^2 \hat{X}_R^T \hat{X}_R \right)^{-1} R^T \hat{\lambda} = \frac{\hat{\lambda}^T R \left(\hat{X}_R^T \hat{X}_R \right)^{-1} R^T \hat{\lambda}}{4 s_R^2} \tag{2.27}$$

and

$$\left(y - \hat{x}_R \right)^T \hat{X}_R \left(s_R^2 \hat{X}_R^T \hat{X}_R \right)^{-1} \hat{X}_R^T \left(y - \hat{x}_R \right) = \frac{\left(y - \hat{x}_R \right)^T P_{\hat{X}_R} \left(y - \hat{x}_R \right)}{s_R^2} \tag{2.28}$$

where $s_R^2 = \dfrac{(y-\hat{x}_R)^T(y-\hat{x}_R)}{n-k+p}$ (2.29)

and $P_{\hat{X}_R} = \hat{X}_R\left(\hat{X}_R^T\hat{X}_R\right)^{-1}\hat{X}_R^T$. (2.30)

The first test statistic in (2.27), is clearly a Lagrange multiplier statistic and is asymptotically distributed as $\chi^2(p)$.

Thus the second statistic too is asymptotically $\chi^2(p)$,[4] which can be very easily obtained as the Explained Sum of Squares (*E.S.S.*) from the artificial regression

$$\frac{1}{s_R}(y-\hat{x}_R) = \hat{X}_R\, b + \textit{error vector}. \qquad (2.31)$$

This is very similar to the G.N.R. in (2.4) for $\overset{*}{\beta} \equiv \hat{\beta}_R$ apart from the fact that, in regression (2.31), the restricted residuals $y-\hat{x}_R$ are each divided by the estimated standard error of the restricted regression.

On the other hand, the most relevant difference between G.N.R. (2.31) and G.N.R. (2.5) is the explanatory power of regression (2.31) in general. In other words, on the contrary of G.N.R. (2.5), the O.L.S. estimate of b in (2.31) is in general not identically zero, because the restricted residual vector is not orthogonal to the

[4] For a formal proof, see Davidson and MacKinnon (1993, pp.172-3).

matrix of derivatives of the unrestricted model, even if these are evaluated at restricted estimates. In fact, that would only be the case for a G.N.R. associated with the restricted (reparametrized) model.

Now, the *E.S.S.* of regression (2.31) is easily seen to be identical to $(n-k+p)R_u^2$ [5] of the following G.N.R.

$$y-\hat{x}_R = \hat{X}_R\, b + \textit{error vector}\,. \tag{2.32}$$

Thus, provided the restricted model still contains an intercept term, to test the restrictions one only has to compare the value of the statistic $(n-k+p)R^2$ obtained from G.N.R. (2.32) with the critical value of the $\chi^2(p)$ table, rejecting or not rejecting the null according to the usual decision rule.

However, if the restricted model does not contain an intercept term, then the restricted residuals will not have in general a zero mean. Therefore, the centred R^2 printed by the package is not necessarily equal to the uncentred R_u^2 that we need to reproduce the *E.S.S.* of regression (2.31). If that is the case, the *E.S.S.* from G.N.R. (2.31) should rather be used to make the decision, despite the fact that to run that G.N.R., it still requires a prior transformation of the restricted residuals.

[5] R_u^2 stands for uncentred R^2

To better clarify this type of tests and also to illuminate its relationship with the tests that we will discuss in the next section, consider the particular case of testing exclusion restrictions.

Let us write the null and alternative models as

$$\mathrm{H_O}:\ y = x(\beta_1, 0) + \xi \tag{2.33}$$

$$\mathrm{H_a}:\ y = x(\beta_1, \beta_2) + \xi \tag{2.34}$$

where $\xi \sim \mathrm{IID}\left(0, w^2 I_n\right)$ and β_1 and β_2 are k-p and p-dimensional vectors, respectively.

The matrix $X(\beta)$ can be partitioned conformably as $X_1(\beta)$ and $X_2(\beta)$, that is, with dimensions $n \times (k - p)$ and $n \times p$, respectively.

The G.N.R. associated with the alternative model, when evaluated at R.N.L.S. is

$$y - \hat{x}_R = \hat{X}_{1R} b_1 + \hat{X}_{2R}\ b_2 + error\ vector \tag{2.35}$$

where $\hat{x}_R \equiv x(\hat{\beta}_{1R}, 0)$, $\hat{X}_{1R} \equiv X_1(\hat{\beta}_{1R}, 0)$ and $\hat{X}_{2R} \equiv X_2(\hat{\beta}_{1R}, 0)$, whereas the G.N.R. associated with the null model is simply

$$y - \hat{x}_R = \hat{X}_{1R}\, b_1 + error\ vector \tag{2.36}$$

where both $\hat{x}_R$ and $\hat{X}_{1R}$ have the same definition as in (2.35).

Now define

$$\hat{M}_{1R} = I_n - \hat{X}_{1R}\left(\hat{X}_{1R}^T \hat{X}_{1R}\right)^{-1} \hat{X}_{1R}^T \tag{2.37}$$

the matrix that projects onto the orthogonal complement of $\hat{X}_{1R}$, so that

$$\hat{M}_{1R}\hat{X}_{1R} = 0. \tag{2.38}$$

Also,

$$\hat{M}_{1R}\left(y - \hat{x}_R\right) = \left(y - \hat{x}_R\right) \tag{2.39}$$

because the restricted estimates are such that satisfy the first-order conditions

$$\hat{X}_{1R}^T\left(y - \hat{x}_R\right) = 0. \tag{2.40}$$[6]

---

[6] That is to say, the first-order conditions imply that $y - \hat{x}_R$ already lies in the orthogonal complement of $\hat{X}_{1R}$.

Taking into consideration (2.38), (2.39) and also the well-known orthogonality condition between the regressors and the residuals of a regression run by O.L.S., it is easily seen that regression (2.35) will yield exactly the same *S.S.R.* as the regression

$$y - \hat{x}_R = \hat{M}_{1R} \hat{X}_{2R}\, b_2 + error\ vector\,. \tag{2.41}$$

On the other hand, as both regressions have the same *T.S.S.*, the *E.S.S.* of the G.N.R. in (2.35) can thus simply be given by the *E.S.S.* of regression (2.41), which is

$$\left(y - \hat{x}_R\right)^T \hat{X}_{2R}\left(\hat{X}_{2R}^T \hat{M}_{1R} \hat{X}_{2R}\right)^{-1} \hat{X}_{2R}^T\left(y - \hat{x}_R\right). \tag{2.42}$$

This quantity, if divided by any consistent estimate of w^2, provides a ratio which is asymptotically distributed as $\chi^2(p)$ under the null hypothesis. This is something that we did not stress, but was implicit, when we dealt with the general case of testing the set of restrictions in (2.21). Thus, one possibility to test the particular case of exclusion restrictions is still to use the statistic $(n - k + p)R^2$ in the G.N.R. (2.35).[7] Using this statistic, one is using s_R^2, as given in (2.29), to estimate w^2.

[7] Provided the restricted model still contains a constant term so that $R^2 = R_u^2$. If that is not the case, then a prior transformation of the regressand in (2.35) is still needed, as emphasised before.

However, a second alternative to conduct the test is highlighted by the treatment of exclusion restrictions because an explicit G.N.R., associated with the null model, is now available.

Consider then the G.N.R. in (2.36), where the *S.S.R.* equals the *T.S.S.* because this G.N.R. has no explanatory power. The difference between the *S.S.R.* from G.N.R. (2.36) and the *S.S.R.* from G.N.R. (2.35) is simply given by the *E.S.S.* from regression (2.35), because the *T.S.S.* is the same in both regressions.

In other words, the *E.S.S.* from regression (2.35) can be interpreted as the reduction in the *S.S.R.* of that regression brought about by the inclusion of $\hat{X}_{2R}$.

On the other hand, the *S.S.R.* from G.N.R. (2.35) can certainly provide a consistent estimate of w^2, because under the null the G.N.R. should have no explanatory power.

Thus, one may conclude that an ordinary pseudo-*F* test based on the statistic

$$\frac{(R.S.S.R. - U.S.S.R.)/p}{U.S.S.R./(n-k)} \tag{2.43}$$

where *R.S.S.R.* and *U.S.S.R.* stands for restricted and unrestricted sums of squared residuals from G.N.R. (2.35), is an equally valid alternative to test the exclusion restrictions[8]. And, of course, when b_2 is a scalar, the corresponding pseudo-t statistic from the G.N.R. in (2.35) is just as valid as any of the two test statistics we have been proposing.

We refer to these tests as 'pseudo' because the corresponding ratios will not actually have the F and t distributions in finite samples when $x_t(\beta)$ is nonlinear in the parameters. Moreover, we have also not excluded both the possibility of $x_t(\beta)$ containing lags of y_t and the errors being not normally distributed. However, asymptotically, $pF(p,n-k)$ and $\mathring{b}_2 / s.e.\left(\mathring{b}_2\right)$ are distributed as $\chi^2(p)$ and $N(0,1)$ under quite weak conditions and the finite-sample distributions of the 'pseudo' statistics above are frequently approximated quite well by the $F(p,n-k)$ and $t(n-k)$ distributions.

As we said, we have considered exclusion restrictions for the sake of clarity, but this in no way limits the generality of the above conclusions[9]. Even when dealing with nonlinear restrictions of the form $r(\beta)=0$ all we need to test the set of restrictions can be summarised as follows: i) Estimate the restricted (reparametrized null) model to obtain the restricted residuals and then regress those residuals on the derivatives of the unrestricted (alternative) model evaluated at restricted estimates;

[8] Even if the restricted model does not contain a constant term.

[9] The way we write the restrictions is purely a matter of parametrization.

ii) Use either the $\chi^2(p)$ statistic $(n-k+p)R^2$ given by the so obtained G.N.R. (2.32)[10], or the ordinary pseudo-F statistic with p and n-k degrees of freedom

$$\frac{U.E.S.S./p}{U.S.S.R./(n-k)} \tag{2.44}$$

where *U.E.S.S.* and *U.S.S.R.* denote the unrestricted explained sum of squares and the unrestricted sum of squared residuals from G.N.R. (2.32). That is to say, in either case, we do not need to construct the G.N.R. associated with the null model. As we will see in section 2.3.2., this is the most relevant difference between tests based on the Lagrange Multiplier principle and those based on the $C(\alpha)$ principle.

In the same way that every restricted version of a nonlinear model estimated by Least Squares can be tested by the use of an appropriate G.N.R., so can every restricted version of a nonlinear model estimated by instrumental variables. For the latter, the I.V. analog of equations (2.35-42) are as follows:

$$y-\tilde{x}_R = P_W \tilde{X}_{1R}\, b_1 + P_W \tilde{X}_{2R}\, b_2 + \textit{error vector} \tag{2.45}$$

where $\tilde{x}_R = x(\tilde{\beta}_{1R},0)$, $\tilde{X}_{1R} \equiv X_1(\tilde{\beta}_{1R},0)$ and $\tilde{X}_{2R} \equiv X_2(\tilde{\beta}_{1R},0)$;

[10] If the restricted model does not contain a constant term rather use the *E.S.S.* from the G.N.R. in (2.31).

$$y - \widetilde{x}_R = P_W \widetilde{X}_{1R}\, b_1 + \textit{error vector} \tag{2.46}$$

where both $\widetilde{x}_R$ and $\widetilde{X}_{1R}$ have the same definitions as in (2.45);

$$\widetilde{M}_{1R} = I_n - P_W \widetilde{X}_{1R}\left(\widetilde{X}_{1R}^T P_W \widetilde{X}_{1R}\right)^{-1} \widetilde{X}_{1R}^T P_W \tag{2.47}$$

The matrix that projects onto the orthogonal complement of $P_W \widetilde{X}_{1R}$, so that

$$\widetilde{M}_{1R} P_W \widetilde{X}_{1R} = 0\,; \tag{2.48}$$

Also,

$$\widetilde{M}_{1R}\left(y - \widetilde{x}_R\right) = \left(y - \widetilde{x}_R\right) \tag{2.49}$$

if the restricted estimates are such that satisfy the first-order conditions

$$\widetilde{X}_{1R}^T P_W\left(y - \widetilde{x}_R\right) = 0\,;^{11} \tag{2.50}$$

---

[11] That is to say, if the first-order conditions imply that $y - \widetilde{x}_R$ already lies in the orthogonal complement of $P_W \widetilde{X}_{1R}$. Notice, however, that this will only be the case if one has also used the same set of instruments W to obtain the R.N.L.I.V. estimates $\widetilde{\beta}_{1R}$.

Taking into consideration (2.48), (2.49) and also the orthogonality condition between the regressors and the residuals of a regression run by O.L.S., regression (2.45) can only yield exactly the same *S.S.R.* as the regression

$$y-\widetilde{x}_R=\widetilde{M}_{1R}P_W\widetilde{X}_{2R}\,b_2+\textit{error vector} \tag{2.51}$$

with exactly the same *E.S.S.* as regression (2.45), that is,

$$\left(y-\widetilde{x}_R\right)^T\widetilde{M}_{1R}P_w\widetilde{X}_{2R}\left(\widetilde{X}_{2R}^TP_W\widetilde{M}_{1R}P_W\widetilde{X}_{2R}\right)^{-1}\widetilde{X}_{2R}^TP_W\widetilde{M}_{1R}\left(y-\widetilde{x}_R\right) \tag{2.52}$$

or even

$$\left(y-\widetilde{x}_R\right)^TP_W\widetilde{X}_{2R}\left(\widetilde{X}_{2R}^TP_W\widetilde{M}_{1R}P_W\widetilde{X}_{2R}\right)^{-1}\widetilde{X}_{2R}^TP_W\left(y-\widetilde{x}_R\right) \tag{2.53}$$

taking into account (2.49).

Provided the p-dimensional vector

$$n^{-1/2}\widetilde{X}_{2R}^TP_W\left(y-\widetilde{x}_R\right)=n^{-1/2}\widetilde{X}_{2R}^TP_W\widetilde{M}_{1R}\left(y-\widetilde{x}_R\right)$$

is asymptotically normally distributed with asymptotic covariance matrix

$$w^2 \operatorname*{Plim}_{n\to\infty}\left(n^{-1}\widetilde{X}_{2R}^T P_W \widetilde{M}_{1R} P_W \widetilde{X}_{2R}\right)$$

then expression (2.53), if divided by any consistent estimate of w^2, provides a ratio which is asymptotically distributed as $\chi^2(p)$ under the null hypothesis. In fact, as Davidson and MacKinnon (1993, pp. 231-2) show, expression (2.53) turns out to be asymptotically the same as

$$\left\|P_W\left(y - x\left(\widetilde{\beta}_R\right)\right)\right\|^2 - \left\|P_W\left(y - x\left(\widetilde{\beta}\right)\right)\right\|^2 \qquad (2.54)$$

the difference between the two I.V. criterion functions associated with the restricted and unrestricted models. See also Engle (1982) for a more detailed discussion.

Bearing in mind that there is now some doubt about the appropriateness of the degrees-of-freedom adjustment, as we emphasised before (section 2.3 - cf. footnote 3), one possibility to test any set of restrictions is therefore to use the statistic nR^2 in the G.N.R. (2.45), provided the restricted model as well as the set of instruments W, used to estimate $\widetilde{\beta}_{1R}$, include a constant term.[12] Using this statistic, one is using

[12] The inclusion of a constant term in W guarantees that the fitted values and the actual values of the constant term included in the restricted model will be exactly the same. Therefore, the I.V. residuals $y - \widetilde{x}_R$ will have zero mean and $R^2 = R_u^2$. If this will not be the case because either the restricted model or the set of instruments do not contain a constant term, then the regressand of regression (2.45) should be transformed in the usual way, using now expression (2.55). Then, the *E.S.S.* of the resulting regression, rather than the statistic nR^2, should be used to perform the test.

$$\frac{1}{n}(y - \tilde{x}_R)^T (y - \tilde{x}_R) \tag{2.55}$$

to consistently estimate w^2.

It would be also valid to use the *S.S.R.* from G.N.R. (2.45) to obtain a consistent estimate of w^2, despite the fact that the regressors have been multiplied by P_W, because under the null the G.N.R. should have no explanatory power.

On the other hand, stressing the fact that both the G.N.R.'s in (2.45) and (2.46) have the same *T.S.*S., expression (2.53) can be interpreted as the reduction in the *S.S.R.* of regression (2.45) brought about by the inclusion of $P_W \tilde{X}_{2R}$. Hence, expression (2.53) is easily seen as the difference between the *S.S.R.* from G.N.R. (2.46) and the *S.S.R.* from G.N.R. (2.45). That is, the pseudo-*F*(*p*,*n*) statistic

$$\frac{(R.S.S.R.-U.S.S.R.)/p}{U.S.S.R./n} = \frac{U.E.S.S./p}{U.S.S.R./n} \tag{2.56}$$

where now *R.S.S.R.*, *U.S.S.R.* and *U.E.S.S.* denote, respectively, the restricted and unrestricted sum of squared residuals and the unrestricted explained sum of squares, all from G.N.R. (2.45), is an equally valid alternative to test the restrictions in an I.V. context.[13]

[13] Even if the restricted model and/or the set of instruments do not contain a constant term.

So far we have shown that there is essentially no difference between results for the N.L.S and N.L.I.V. versions of the G.N.R..

Before moving to the next section a last comment still in the I.V. context. We have been assuming in all the above discussion that the G.N.R. (2.45) is run by O.L.S.. One might wonder why not regress $y - \widetilde{x}_R$ on $\widetilde{X}_{1R}$ and $\widetilde{X}_{2R}$ by an I.V. procedure, using W as the matrix of instruments, avoiding the otherwise required initial step of regressing the columns of $\widetilde{X}_{1R}$ and $\widetilde{X}_{2R}$ on W.

To clarify this issue let us define

$$\breve{M}_{1R} = I_n - \widetilde{X}_{1R}\left(\widetilde{X}_{1R}^T P_W \widetilde{X}_{1R}\right)^{-1} \widetilde{X}_{1R}^T P_W \tag{2.57}$$

so that

$$\breve{M}_{1R}\widetilde{X}_{1R} = 0. \tag{2.58}$$

Also,

$$\breve{M}_{1R}\left(y - \widetilde{x}_R\right) = \left(y - \widetilde{x}_R\right) \tag{2.59}$$

if the restricted estimates are such that satisfy the first-order conditions in (2.50).[14]

Taking into consideration (2.58), (2.59) and also the I.V. analog orthogonality condition between the regressors and the residuals of a regression run by I.V., the following G.N.R.'s

$$y - \widetilde{x}_R = \widetilde{X}_{1R}\, b_1 + \widetilde{X}_{2R}\, b_2 + \ \textit{error vector} \tag{2.60}$$

and

$$y - \widetilde{x}_R = \breve{M}_{1R}\widetilde{X}_{2R}\, b_2 + \ \textit{error vector} \tag{2.61}$$

can only yield exactly the same (I.V.) *S.S.R.* and (I.V.) *E.S.S.*.

Now, the (I.V.) *E.S.S.* from G.N.R. (2.61) is easily given by

$$\breve{b}_2^T \widetilde{X}_{2R}^T \breve{M}_{1R}^T \breve{M}_{1R} \widetilde{X}_{2R} \breve{b}_2 \tag{2.62}$$

where $\breve{b}_2$ stands for I.V. estimates, that is, where

$$\breve{b}_2 = \left(\widetilde{X}_{2R}^T \breve{M}_{1R}^T P_W \breve{M}_{1R} \widetilde{X}_{2R}\right)^{-1} \widetilde{X}_{2R}^T \breve{M}_{1R}^T P_W \left(y - \widetilde{x}_R\right). \tag{2.63}$$

---

[14] As we said before , that will be the case if and only if W is the same.

Using (2.63) to substitute for $\breve{b}_2$ in (2.62), one obtains

$$\begin{aligned}\left(y-\tilde{x}_R\right)^T P_W \tilde{X}_{2R}\left(\tilde{X}_{2R}^T P_W \tilde{M}_{1R} P_W \tilde{X}_{2R}\right)^{-1} \tilde{X}_{2R}^T \breve{M}_{1R}^T \breve{M}_{1R} \tilde{X}_{2R} \\ \left(\tilde{X}_{2R}^T P_W \tilde{M}_{1R} P_W \tilde{X}_{2R}\right)^{-1} \tilde{X}_{2R}^T P_W\left(y-\tilde{x}_R\right)\end{aligned} \tag{2.64}$$

because

$$\breve{M}_{1R}^T P_W\left(y-\tilde{x}_R\right)=P_W \tilde{M}_{1R}\left(y-\tilde{x}_R\right)=P_W\left(y-\tilde{x}_R\right) \tag{2.65}$$

$$\breve{M}_{1R}^T P_W \breve{M}_{1R}=P_W \tilde{M}_{1R} P_W . \tag{2.66}$$

with $\tilde{M}_{1R}$ as defined in (2.47).

However, in (2.64)

$$\breve{M}_{1R}^T \breve{M}_{1R} \neq P_W \tilde{M}_{1R} P_W$$

even asymptotically, and therefore expression (2.64) is different from expression (2.53).

Notice finally that despite the different expressions for the E.S.S., the I.V. estimate in (2.63) is precisely the estimate one would obtain running by O.L.S. the G.N.R. in (2.51). This equality is easily proved taking into consideration (2.65) and (2.66).

To conclude, whereas the pseudo-t-statistic reported by an I.V. package is still a valid statistic to test one single restriction, one cannot use the (I.V.) E.S.S. from a G.N.R. to calculate test statistics. Therefore, for tests with more that just one degree of freedom, the use of an I.V. package is not a directly valid alternative. For the same reason, one cannot also construct pseudo-F tests obtained from I.V. estimation of a restricted and an unrestricted model.[15]

2.3.2 One-Step Estimation and Test Based on the $C(\alpha)$ Principle

In the introduction to this section we have referred to the use of the appropriate G.N.R. to find either the N.L.S. or the N.L.Ì.V. estimates themselves. Then, we have stressed that the convergence processes should stop as soon as the O.L.S. estimates of the appropriate G.N.R. (in both cases denoted by $\overset{\circ}{b}$) are not statistically different from zero.

[15] Unless the package prints the quantities in (2.54) (this is for example the case for T.S.P.). Strictly speaking, even if the package does not print those quantities there is an alternative to deduce valid pseudo-F tests. First, one would have to regress (by O.L.S.) both the restricted and unrestricted I.V. residuals on W saving the obtained *E.S.S.* from those regressions. Then, one would take the difference between those expressions to obtain expression (2.54) which is the proper expression for the numerator of the F-statistic.

Now, we want to prove that to obtain asymptotically equivalent estimates to either N.L.S. or N.L.I.V. estimates we do not actually need to carry out those processes of convergence.

With these results in mind, we will then be able to discuss a type of tests named in the literature on M.L. estimation as $C(\alpha)$ tests. The discussion of these tests will be particularly useful in the context of I.V., since its rationality will avoid the common need of considering a unique set of instruments (both to estimate the restricted model and to run the appropriate G.N.R.) when testing a set of restrictions.

2.3.2.1 One Step Estimation

Consider the model in (2.20), which is model (2.1) in matrix form.

Let $\acute{\beta}$ denote arbitrary initial estimates, which are assumed to be root-n consistent.

The G.N.R. in (2.4) becomes

$$y - \acute{x} = \acute{X}\ b + error\ vector \tag{2.67}$$

where $\acute{x} \equiv x\left(\acute{\beta}\right)$, $\acute{X} \equiv x\left(\acute{\beta}\right)$ and $\acute{b} = \beta - \acute{\beta}$.

The O.L.S. estimates from this regression are

$$\mathring{b} = \left(\acute{X}^T \acute{X}\right)^{-1} \acute{X}^T\left(y - \acute{x}\right) \tag{2.68}$$

and thus the one-step (O.S.) estimator is simply given by

$$\overset{\#}{\beta} = \acute{\beta} + \mathring{b}. \tag{2.69}$$

Taking a first-order Taylor-series approximation to $x\left(\acute{\beta}\right)$ around β, yields

$$x\left(\acute{\beta}\right) \approx x(\beta) + X(\beta)\left(\acute{\beta} - \beta\right) \tag{2.70}$$

where $x\left(\acute{\beta}\right) \equiv \acute{x}$.

Substituting (2.20) and (2.70) into (2.68), the O.L.S. estimates $\mathring{b}$ can be approximated by

$$\mathring{b} \approx \left(\dot{X}^T, \dot{X}\right)^{-1} \dot{X}^T\left(x(\beta) + \xi - x(\beta) - X(\beta)\left(\dot{\beta} - \beta\right)\right)$$
$$\approx \left(\dot{X}\dot{X}\right)^{-1} \dot{X}^T \xi - \left(\dot{X}^T \dot{X}\right)^{-1} \dot{X}^T X(\beta)\left(\dot{\beta} - \beta\right) \tag{2.71}$$

Now, inserting appropriate powers on n so that all quantities become $O(1)$, one obtains

$$n^{\frac{1}{2}}\mathring{b} \approx \left(n^{-1}\dot{X}^T \dot{X}\right)^{-1} n^{-\frac{1}{2}} \dot{X}^T \xi - \left(n^{-1} \dot{X}^T \dot{X}\right)^{-1} n^{-1} \dot{X}^T X(\beta)\, n^{\frac{1}{2}}\left(\dot{\beta} - \beta\right)$$
$$\overset{a}{=} \left(n^{-1}X^T(\beta)X(\beta)\right)^{-1} n^{-\frac{1}{2}} X^T(\beta)\xi - n^{\frac{1}{2}}\left(\dot{\beta} - \beta\right) \tag{2.72}$$

where $\overset{a}{=}$ means asymptotically equivalent, since

$$n^{-1}\dot{X}^T \dot{X} \overset{a}{=} n^{-1}\dot{X}^T X(\beta) \overset{a}{=} n^{-1}X^T(\beta)X(\beta)$$

and also

$$n^{-\frac{1}{2}}\dot{X}^T \xi \overset{a}{=} n^{-\frac{1}{2}} X^T(\beta)\xi$$

as a consequence of the consistency of $\dot{\beta}$.

Taking into account (2.69) and (2.72) one can then easily conclude that

$$
\begin{aligned}
n^{\frac{1}{2}}\overset{\#}{\beta} &= n^{\frac{1}{2}}\acute{\beta} + n^{\frac{1}{2}}\mathring{b} \\
&\overset{a}{=} n^{\frac{1}{2}}\acute{\beta} + \left(n^{-1}X^{T}(\beta)X(\beta)\right)^{-1} n^{-\frac{1}{2}}X^{T}(\beta)\xi - n^{\frac{1}{2}}\left(\acute{\beta}-\beta\right) \\
&\overset{a}{=} \left(n^{-1}X^{T}(\beta)X(\beta)\right)^{-1} n^{-\frac{1}{2}}X^{T}(\beta)\xi + n^{\frac{1}{2}}\beta
\end{aligned}
$$

That is,

$$
n^{\frac{1}{2}}\left(\overset{\#}{\beta}-\beta\right) \overset{a}{=} \left(n^{-1}X^{T}(\beta)X(\beta)\right)^{-1} n^{-\frac{1}{2}}X^{T}(\beta)\xi \tag{2.73}
$$

or after taking the probability limit of $n^{-1}X^{T}(\beta)X(\beta)$,

$$
n^{\frac{1}{2}}\left(\overset{\#}{\beta}-\beta\right) \overset{a}{\sim} N\left(0, w^{2}\operatorname*{Plim}_{n\to\infty}\left(n^{-1}X^{T}(\beta)X(\beta)\right)^{-1}\right) \tag{2.74}
$$

because, as (2.73) shows, the O.S. estimator $\overset{\#}{\beta}$, as defined in (2.69) and obtained after running G.N.R. (2.67) by O.L.S., is asymptotically equivalent to the N.L.S. estimator $\hat{\beta}$ and it must therefore have the same asymptotic distribution (cf. expression (2.17)).

The interest of one-step estimation in the context of L.S. relies on the fact that it is sometimes easier to obtain root-n consistent but inefficient estimates than to obtain N.L.S. estimates. This will be, for example, the case if the nonlinear model is simply a restricted version of a linear model subject to nonlinear restrictions. In these circumstances, the initial unrestricted estimates can very easily be obtained and then used as initial root-n consistent estimates to construct the G.N.R. in (2.67).

One-step estimation can also be useful in the I.V. context as we will see in this subsection. A detailed discussion in this context seems unnecessary since the main results are not essentially different from those obtained for the L.S. case. We will therefore restrict ourselves to mention the analog I.V. expressions.

Consider that one wants to estimate model (2.1) by N.L.I.V. using W as the set of instruments. Consider also that, for any particular reason, one has already available the I.V. estimates $\acute{\beta}$, which have been obtained using W_D, a different set of instruments.

It is easily seen that whereas expressions (2.69-70) remain the same, the expressions (2.67-8) and (2.71-73) should now be replaced by the following ones:

$$y - \acute{x} = P_W \acute{X} b + error \tag{2.75}$$

$$\mathring{b} = \left(\acute{X}^T P_W \acute{X}\right)^{-1} \acute{X}^T P_W \left(y - \acute{x}\right) \tag{2.76}$$

$$\mathring{b} \approx \left(\acute{X}^T P_W \acute{X}\right)^{-1} \acute{X}^T P_W \xi - \left(\acute{X}^T P_W \acute{X}\right)^{-1} \acute{X}^T P_W X(\beta)\left(\acute{\beta} - \beta\right) \tag{2.77}$$

$$n^{1/2}\overset{\circ}{b} \overset{a}{=} \left(n^{-1}X^T(\beta)P_W X(\beta)\right)^{-1} n^{-1/2} X^T(\beta)P_W\xi - n^{1/2}\left(\acute{\beta}-\beta\right) \tag{2.78}$$

and

$$n^{1/2}\left(\overset{\#}{\beta}-\beta\right) \overset{a}{=} \left(n^{-1}X^T(\beta)P_W X(\beta)\right)^{-1} n^{-1/2} X^T(\beta)\xi\,. \tag{2.79}$$

It should therefore come as no surprise to learn that

$$n^{1/2}\left(\overset{\#}{\beta}-\beta\right) \overset{a}{\sim} N\left(0, w^2 \operatorname*{Plim}_{n\to\infty}\left(n^{-1}X^T(\beta)P_W X(\beta)\right)^{-1}\right) \tag{2.80}$$

because, as (2.79) shows, the O.S. estimator $\overset{\#}{\beta}$ obtained after running G.N.R. (2.75) by O.L.S., is asymptotically equivalent to the N.L.I.V. estimator $\widetilde{\beta}$ and it must therefore have the same asymptotic distribution (cf. expression (2.14)).

It is worth noting that the equivalence between the O.S. estimates based on the appropriate G.N.R. and either the N.L.S. or the N.L.I.V. estimates is only valid asymptotically. In finite samples the O.S. estimates may actually differ greatly. Thus, O.S. estimation makes more sense for large samples so that the initial consistent estimates are not far from the true values, and also where a nonlinear procedure may be too expensive.

2.3.2.2 Tests Based on the $C(\alpha)$ Principle

Once again consider the models in (2.33) and (2.34)

H_0: $y = x(\beta_1, 0) + \xi$

H_a: $y = x(\beta_1, \beta_2) + \xi$.

The G.N.R. associated with the alternative model, when evaluated at restricted arbitrary root-n consistent estimates (under the null) is either

$$y - \acute{x}_R = \acute{X}_{1R}\, b_1 + \acute{X}_{2R}\, b_2 + error\ vector \qquad (2.81)$$

or

$$y - \acute{x}_R = P_W\, \acute{X}_{1R}\, b_1 + P_W\, \acute{X}_{2R}\, b_2 + error\ vector \qquad (2.82)$$

according to the context (either L.S. or I.V., respectively).

In turn, the G.N.R.'s associated with the null model are, respectively

$$y - \acute{x}_R = \acute{X}_{1R}\, b_1 + error\ vector \qquad (2.83)$$

and

$$y - \acute{x}_R = P_W \acute{X}_{1R} b_1 + error\ vector\,. \tag{2.84}$$

In the above G.N.R.'s, $\acute{x}_R$, $\acute{X}_{1R}$ and $\acute{X}_{2R}$ have similar definitions as before. That is,

$$\acute{x}_R \equiv \acute{x}\left(\acute{\beta}_{1R},0\right),\ \acute{X}_{1R} \equiv X_1\left(\acute{\beta}_{1R},0\right) \text{ and } \acute{X}_{2R} \equiv X_2\left(\acute{\beta}_{1R},0\right).$$

Let us first consider the L.S. context.

Define,

$$\acute{M}_{1R} = I_n - \acute{X}_{1R}\left(\acute{X}_{1R}^T \acute{X}_{1R}\right)^{-1} \acute{X}_{1R}^T \tag{2.85}$$

so that

$$\acute{M}_{1R} \acute{X}_{1R} = 0\,. \tag{2.86}$$

However,

$$\acute{M}_{1R}\left(y-\acute{x}_R\right) \neq \left(y-\acute{x}_R\right) \tag{2.87}$$

because, in general,

$$\acute{X}_{1R}^T\left(y-\acute{x}_R\right) \neq 0. \tag{2.88}$$

That is, $\acute{\beta}_{1R}$ will not in general satisfy the first-order conditions for N.L.S. estimates of the restricted model.

Taking into consideration (2.86) and also the orthogonality condition between the regressors and the residuals of a regression run by O.L.S., it is still true that regression (2.81) will yield exactly the same *S.S.R.* as the regression

$$\acute{M}_{1R}\left(y-\acute{x}_R\right) = \acute{M}_{1R}\,\acute{X}_{2R}\,b_2 + error\ vector. \tag{2.89}$$

However, now the *T.S.S.*'s of the two regressions will in general be different and thus, so it will be their *E.S.S.*'s.

To obtain the *E.S.S.* of regression (2.81) let us first obtain its *S.S.R.*.

In both regressions (2.81) and (2.89)

$$\mathring{b}_2 = \left(\acute{X}_{2R}^T \, \acute{M}_{1R} \, \acute{X}_{2R} \right)^{-1} \acute{X}_{2R}^T \, \acute{M}_{1R} \left(y - \acute{x}_R \right) \tag{2.90}$$

and the *S.S.R.* from regression (2.81) can be obtained as the difference between the *T.S.S.* and the *E.S.S.* from regression (2.89),

$$\left(y - \acute{x}_R \right)^T \acute{M}_{1R} \left(y - \acute{x}_R \right) - \mathring{b}_2^T \, \acute{X}_{2R}^T \, \acute{M}_{1R} \, \acute{X}_{2R} \, \mathring{b}_2 \tag{2.91}$$

or substituting for (2.90),

$$\begin{aligned} &\left(y - \acute{x}_R \right)^T \acute{M}_{1R} \left(y - \acute{x}_R \right) - \left(y - \acute{x}_R \right)^T \acute{M}_{1R} \, \acute{X}_{2R} \\ &\qquad \left(\acute{X}_{2R}^T \, \acute{M}_{1R} \, \acute{X}_{2R} \right)^{-1} \acute{X}_{2R}^T \, \acute{M}_{1R} \left(y - \acute{x}_R \right) \end{aligned} \tag{2.92}$$

Finally, to obtain the *E.S.S.* from regression (2.81), take the difference between its *T.S.S.* and *S.S.R.*,

$$\begin{aligned} &\left(y - \acute{x}_R \right)^T \left(y - \acute{x}_R \right) - \left(y - \acute{x}_R \right)^T \acute{M}_{1R} \left(y - \acute{x}_R \right) + \\ &\qquad + \left(y - \acute{x}_R \right)^T \acute{M}_{1R} \, \acute{X}_{2R} \left(\acute{X}_{2R}^T \, \acute{M}_{1R} \, \acute{X}_{2R} \right)^{-1} \acute{X}_{2R}^T \, \acute{M}_{1R} \left(y - \acute{x}_R \right) \\ &= \left(y - \acute{x}_R \right)^T \acute{P}_{1R} \left(y - \acute{x}_R \right) + \\ &\qquad + \left(y - \acute{x}_R \right)^T \acute{M}_{1R} \, \acute{X}_{2R} \left(\acute{X}_{2R}^T \, \acute{M}_{1R} \, \acute{X}_{2R} \right)^{-1} \acute{X}_{2R}^T \, \acute{M}_{1R} \left(y - \acute{x}_R \right) \end{aligned} \tag{2.93}$$

where

$$\acute{P}_{1R} = \acute{X}_{1R}\left(\acute{X}_{1R}^{T}\,\acute{X}_{1R}\right)^{-1}\acute{X}_{1R}^{T}\,. \tag{2.94}$$

The first term of expression (2.93) is easily seen to be the *E.S.S.* from regression (2.83) which is in general not zero because $\acute{X}_{1R}$ has still some explanatory power for $y-\acute{x}_R$ (cf. (2.88)).

The difference between the *E.S.S.*'s from regressions (2.81) and (2.83) is in turn the second term of expression (2.93). That is,

$$\left(y-\acute{x}_R\right)^{T}\acute{M}_{1R}\,\acute{X}_{2R}\left(\acute{X}_{2R}^{T}\,\acute{M}_{1R}\,\acute{X}_{2R}\right)^{-1}\acute{X}_{2R}^{T}\,\acute{M}_{1R}\left(y-\acute{x}_R\right) \tag{2.95}$$

can still be interpreted as the increase in the *E.S.S.* (or the reduction in the *S.S.R.*) of regression (2.81) brought about by the inclusion of $\acute{X}_{2R}$.

Apart from the fact that this last expression is evaluated at restricted root-n consistent estimates ($\acute{\beta}_{1R}$,0) rather than at the restricted N.L.S. estimates ($\tilde{\beta}_{1R}$,0), the expression looks like expression (2.42).[16]

---

[16] Notice that expression (2.42) could be rewritten as $\left(y-\hat{x}_R\right)^{T}\hat{M}_{1R}\hat{X}_{2R}\left(\hat{X}_{2R}^{T}\hat{M}_{1R}\hat{X}_{2R}\right)^{-1}\hat{X}_{2R}^{T}\hat{M}_{1R}\left(y-\hat{x}_R\right)$, since $\left(y-\hat{x}_R\right)=\hat{M}_{1R}\left(y-\hat{x}_R\right)$ (cf.(2.39)).

Now, from the last subsection we know that

$$n^{1/2}\left(\overset{\#}{\beta}-\beta\right)\overset{a}{=}n^{1/2}\left(\hat{\beta}-\beta\right) \tag{2.96}$$

where $\overset{\#}{\beta}=\overset{\prime}{\beta}+\mathring{b}$ is the O.S. estimator, $\mathring{b}$ is the O.L.S. estimate of b from (2.81) and $\hat{\beta}$ is the unrestricted N.L.S. estimate. Since $\overset{\prime}{\beta}_{2R}=0$, we have that $\overset{\#}{\beta}_{2R}=\mathring{b}_2$ and thus

$$n^{1/2}\left(\mathring{b}_2-\beta_2\right)\overset{a}{=}n^{1/2}\left(\hat{\beta}_2-\beta_2\right). \tag{2.97}$$

It should therefore come as obvious that a pseudo-*F* test for $b_2=0$ from G.N.R. (2.81) is equivalent to a test for $\beta_2=0$ and asymptotically the same as the pseudo-*F* test previously described and based on the L.M. principle (cf. (2.43)). Also, when b_2 is a scalar, the corresponding pseudo-*t* statistic from G.N.R. (2.81) can only be as equally valid as it was when based in the L.M. principle. Notice, however, that there is a crucial argument here. That is, both statistics use now the *S.S.R.* from regression (2.81) to estimate w^2. This must therefore be a consistent estimate for w^2. In fact, despite (2.88), under the null hypothesis $\beta_2=0$,

$$\underset{n\to\infty}{\text{Plim}}\left[n^{-1}\overset{\prime}{X}{}_{1R}^{T}\left(y-\overset{\prime}{x}_R\right)\right]=0 \tag{2.98}$$

which is simply a consequence of the consistency of $\acute{\beta}_{1R}$ under the null. In other words, under the null, the G.N.R. (2.81) should have no explanatory power asymptotically.

On the contrary, the $(n-k+p)R_u^2$ from G.N.R. (2.81) is not a valid statistic since the equality between expressions (2.43) and (2.44), obtained in subsection 2.3.1, does not hold now in general. As we stressed before, $\acute{X}_{1R}$ still has some ability to explain $y-\acute{x}_R$ in general, and thus expression (2.93) is in general different from (2.95).

As an alternative, one could however construct a valid test statistic as $(n-k)R_u^2$ from (2.81) minus $(n-k)R_u^2$ from (2.83), which would still have a $\chi^2(p)$ distribution.[17]

Whichever the choice, one can conclude that, unlike the tests based on the L.M. principle, these type of tests (based on the $C(\alpha)$ principle) always require estimation of two G.N.R.'s - one associated with the unrestricted model and other associated with the null model. Both G.N.R.'s must be evaluated at restricted consistent root-n estimates under the null. Therefore, like for the tests based on the L.M. principle, estimation of the unrestricted model is never required.

---

[17] Consideration of the same small sample correction guarantees a unique consistent estimate of w^2, which is strictly the same as for the F-test proposed above.

As we have emphasised before, these conclusions are perfectly valid even if one wants to test any set of (nonlinear) restrictions. The way one considers the restrictions is just a question of reparametrization.

We also said that a particular reasonable motivation for O.S. estimation would exist if the nonlinear model is simply a restricted version of a linear model subject to nonlinear restrictions. In this case, restricted estimates may be obtained substituting for the unrestricted O.L.S. estimates in the nonlinear set of restrictions. These restricted estimates, which would also be root-n consistent under the null could then be used as initial estimates to construct the appropriate G.N.R.'s..

Motivation for O.S. estimation, and therefore for $C(\alpha)$ tests, is however most present in the I.V. context. In this context it can easily happens that the regression function for the alternative model depends on more strictly exogenous variables than the regression function for the null model. In this circumstances the matrix of instruments W, that one would have to use to estimate the G.N.R. associated with the alternative model, would have more columns than the one actually used to estimate the null, say W_R. It could even be the case that consistent estimation of the null model does not require the use of I.V.. Moreover, and most interesting for the sake of this thesis, if the two models are nonnested models, then the recommended sets of instruments to be used will certainly be different sets, say W_{R1} and W_{R2}. Thus, to compute a test based on the L.M. principle, one would have to reestimate the null model again, using the most comprehensive set of instruments W . Using a test based on the $C(\alpha)$ principle, as we will see, is still a strictly valid alternative to avoid reestimation of the null model if one uses the proper G.N.R.'s to conduct the tests.

The treatment of this case is similar to the L.S. case and the final result is not fundamentally different. However, to make sure that there is no misleading generalisation of the equations presented for the L.S. case, we will repeat the reasoning.

Consider then G.N.R.'s (2.82) and (2.84) previously defined. Define now

$$\acute{M}_{1R} = I_n - P_W \acute{X}_{1R}\left(\acute{X}_{1R}^T P_W X_{1R}\right)^{-1} \acute{X}_{1R}^T P_W \tag{2.99}$$

so that

$$\acute{M}_{1R} P_W \widetilde{X}_{1R} = 0 \tag{2.100}$$

but still

$$\acute{M}_{1R}\left(y - \acute{x}_R\right) \neq \left(y - \acute{x}_R\right) \tag{2.101}$$

because, in general, $\acute{\beta}_{1R}$ will not satisfy the first-order conditions for N.L.I.V. estimation of the restricted model. That is,

$$\acute{X}_{1R}^{T} P_W \left(y - \acute{x}_R \right) \neq 0 . \tag{2.102}$$

Certainly this will be the case if to estimate the null model one has used either W_R, W_{R1} (or no set of instruments at all), rather than W, the comprehensive set of instruments.

Taking into consideration (2.100) and also the orthogonality condition between the regressors and the residuals of a regression run by O.L.S., it is still true that regression (2.82) will yield exactly the same *S.S.R.* as the regression

$$\acute{M}_{1R} \left(y - \acute{x}_R \right) = \acute{M}_{1R} P_W \acute{X}_{2R} b_2 + \textit{error vector} . \tag{2.103}$$

The two regressions will also in general have different *T.S.S.*'s and thus also different *E.S.S.*'s.

In both regressions,

$$\mathring{b}_2 = \left(\acute{X}_{2R}^{T} P_W \acute{M}_{1R} P_W \acute{X}_{2R} \right)^{-1} \acute{X}_{2R}^{T} P_W \acute{M}_{1R} \left(y - \acute{x}_R \right) \tag{2.104}$$

and thus, the difference between the *T.S.S.* and the *E.S.S.* from regression (2.103) is given by

$$
\begin{aligned}
&\left(y-x_R^{'}\right)^T M_{1R}^{'}\left(y-x_R^{'}\right)-\left(y-x_R^{'}\right)^T M_{1R}^{'}\, P_W\, X_{2R}^{'} \\
&\qquad \left(X_{2R}^{'T}\, P_W\, M_{1R}^{'}\, P_W\, X_{2R}^{'}\right)^{-1} X_{2R}^{'T}\, P_W\, M_{1R}^{'}\left(y-x_R^{'}\right)
\end{aligned}
\tag{2.105}
$$

which is still the *S.S.R.* from regression (2.82).

Finally, taking the difference between the *T.S.S.* and the *S.S.R.* from regression (2.82) one similarly obtains

$$
\begin{aligned}
&\left(y-x_R^{'}\right)^T P_{1R}^{'}\left(y-x_R^{'}\right)+\left(y-x_R^{'}\right)^T M_{1R}^{'}\, P_W\, X_{2R}^{'} \\
&\qquad \left(X_{2R}^{'T}\, P_W\, M_{1R}^{'}\, P_W\, X_{2R}^{'}\right)^{-1} X_{2R}^{'T}\, P_W\, M_{1R}^{'}\left(y-x_R^{'}\right)
\end{aligned}
\tag{2.106}
$$

where

$$
P_{1R}^{'} = P_W\, X_{1R}^{'}\left(X_{1R}^{'T}\, P_W\, X_{1R}^{'}\right)^{-1} X_{1R}^{'T}\, P_W . \tag{2.107}
$$

Given the new definition of $P_{1R}^{'}$, this is the I.V. analog expression of (2.93). A similar interpretation can therefore be offered in the I.V. context for the above expression. The first term of expression (2.106) is easily still seen as the *E.S.S.* from regression (2.84) which is in general not zero because $P_W\, X_{1R}^{'}$ has still some explanatory power for $y - x'_R$ (cf. (2.102)). Also, the difference between the *E.S.S.*'s from regressions (2.82) and (2.84) is given by the second term of expression (2.106),

$$\left(y - x'_R\right)^T M'_{1R} P_W X'_{2R}\left(X'^T_{2R} P_W M'_{1R} P_W X'_{2R}\right)^{-1} X'^T_{2R} P_W M'_{1R}\left(y - x'_R\right) \qquad (2.108)$$

which can again be interpreted as the increase in the *E.S.S.* (or the reduction in the *S.S.R.*) of regression (2.82) brought about by the inclusion of $P_W X'_{2R}$.

From the last subsection we also know that

$$n^{1/2}\left(\overset{\#}{\beta} - \beta\right) \overset{a}{=} n^{1/2}\left(\tilde{\beta} - \beta\right) \qquad (2.109)$$

where $\overset{\#}{\beta} = \beta' + \overset{\circ}{b}$ is the O.S. estimator, $\overset{\circ}{b}$ is the O.L.S. estimate of b from (2.82) and $\tilde{\beta}$ is the unrestricted N.L.I.V. estimate.

Therefore

$$n^{1/2}\left(\overset{\circ}{b}_2 - \beta_2\right) \overset{a}{=} n^{1/2}\left(\tilde{\beta}_2 - \beta_2\right) \qquad (2.110)$$

since for $\overset{\#}{\beta}_{2R} = \overset{\circ}{b}_2$ for $\beta'_{2R} = 0$.

Taking now into consideration the consistency of $\acute{\beta}_{1R}$ under the null, then

$$\underset{n\to\infty}{\text{Plim}}\left[n^{-1}\acute{X}_{1R}^{T}P_{W}\left(y-\acute{x}_{R}\right)\right]=0 \tag{2.111}$$

despite the fact that in general the first-order conditions will not be satisfied.

Therefore, either a pseudo-F statistic for $b_2 = 0$ from G.N.R. (2.82) (a pseudo-t statistic, if b_2 is a scalar) or the difference between nR_u^2 from G.N.R. (2.82) and nR_u^2 from G.N.R. (2.84) would be valid testing procedures also in the I.V. context.

On the contrary, the nR_u^2 from G.N.R (2.82) would not be a valid statistic since the equality in (2.56), obtained in subsection 2.3.1, does not hold in general in this context.

2.3.3 Final Comments

Two last comments before we put an end to this chapter.

1. The choice among tests often depends on which test has a finite-sample distribution better approximated by its large-sample distribution.

For example, among the tests based on the L.M. principle, there is some evidence that the F-test should be preferred to the corresponding version based on nR_u^2 (see Kiviet (1986)).

On the other hand, as we said before, O.S. estimation and therefore testing procedures based on the $C(\alpha)$ principle makes more sense for large samples. The fact that the restricted initial estimates are root-n consistent under the null does not prevent them from being extremely inefficient in finite-samples. Therefore, the O.S. estimates may still differ greatly from the true values of the parameters to be estimated.

The use of testing procedures based on the $C(\alpha)$ principle is however most interesting when testing restrictions in the I.V. context because it allows us to relax the use of a unique set of instruments. Finite Sample Theory and Monte Carlo evidence suggest in fact that an 'excessive' number of instruments to estimate a model by I.V. increases the bias of the I.V. estimator (see Davidson and MacKinnon (1993, chapter 7) and the references within). This is an inevitable consequence of the I.V. estimator approaching the O.L.S. estimator as $P_W X$ approaches X. Moreover, when the 'extra' instruments have little ability to explain the offending regressors, the I.V. estimators may be extremely inefficient. This harmful effect may be important when forcing the use of comprehensive set of instruments to reestimate the null model. This argument is against the use of L.M. procedures in the context of nonnested models estimated by I.V., specially if the sample size is large so that the initial root-n consistent estimates are likely to be close to the true values of the parameters to be estimated.

2. The methodology so far discussed is also strictly valid for the construction of diagnostic tests. In this case, if the model to be tested is model (2.1), the G.N.R.'s in (2.35), (2.45) and (2.81-4) should be replaced, respectively, by

$$y - \hat{x} = \hat{X}b + \hat{Z}c + \textit{ error vector} \tag{2.112}$$

$$y - \tilde{x} = P_W \tilde{X}b + P_W \tilde{Z}c + \textit{ error vector} \tag{2.113}$$

$$y - \acute{x} = \acute{X}b + \acute{Z}c + \textit{ error vector} \tag{2.114}$$

$$y - \acute{x} = P_W \acute{X}b + P_W \acute{Z}c + \textit{ error vector} \tag{2.115}$$

$$y - \acute{x} = \acute{X}b + \textit{ error vector} \tag{2.116}$$

$$y - \acute{x} = P_W \acute{X}b + \textit{ error vector} \tag{2.117}$$

with the $n \times p$ Z matrix to be evaluated at N.L.S. $\hat{\beta}$, N.L.I.V. $\tilde{\beta}$ or at any root-n consistent estimates under the null β. This matrix Z must also satisfy the same regularity conditions as $X(\beta)$ and might or might not actually depend on β. Also, in the G.N.R.'s (2.112) and (2.114) it is assumed that $n^{-1}Z^{T}\xi$ would tend to a zero vector, whereas in the G.N.R.'s (2.113) and (2.115) W must be a valid instrument set for the null model.

Implicitly, of course, Z must correspond to the matrix X_2 for some unrestricted model that includes model (2.1) as a special case. This will be, for example, the case for nonnested models, where the comprehensive model is simply an extended version of the null model under testing.

CHAPTER 3

LINEAR REGRESSION MODELS WITH AUTOREGRESSIVE ERRORS

3.1 Introduction

Most of the issues involved in estimating models with, and testing for, serial correlation will be clarified by the discussion of the AR(1) case. This is also by far the most popular error process in applied econometric work. We will restrict our attention to first-order autoregressive errors. There will be no loss of generality since all the results carry over to higher order processes in an obvious fashion.

Consider the model

$$y_t = x_t\,\beta + u_t \quad ; \quad u_t = \rho\, u_{t-1} + \xi_t \quad ; \quad \xi_t \sim \text{IID}\,(0, w^2) \qquad (3.1)$$

where x_t is row t of matrix X, $t = 1, 2, ..., n$, and β is a k-dimensional vector.

Because

$$u_{t-1} = y_{t-1} - x_{t-1}\,\beta \tag{3.2}$$

model (3.1) can be rewritten as

$$y_t = x_t\,\beta + \rho\,(y_{t-1} - x_{t-1}\,\beta\,) + \xi_t \quad ;\ \xi_t \sim \text{IID}\,(0, w^2) \tag{3.3}$$

which is spherical, that is with well-behaved disturbance terms, but nonlinear in the parameters since the regression function is

$$x_t'(\beta,\rho) = x_t\,\beta + \rho\,(y_{t-1} - x_{t-1}\,\beta\,) \tag{3.4}$$

depending on β as well as on ρ in a nonlinear fashion.

It is therefore justifiable to appeal for a nonlinear estimation method and for the use of the corresponding G.N.R. to make valid inference about the model. Happily, the results we have already obtained in the previous chapter will allow us to handle most of the issues under analysis in this chapter (as well as in the following chapters).

3.2 Least Squares Estimation

Provided X does not contain any current endogenous variable, assuming the stationarity condition $|\rho| < 1$ and dropping the first observation for both y and all the variables in X, it seems natural to estimate model (3.3) by N.L.S. and to make inference about it by using the corresponding G.N.R., as firstly advocated by Davidson and MacKinnon (1980 b).

The N.L.S. may be obtained by minimising, with respect to (β, ρ) the criterion function

$$[y - x'(\beta,\rho)]^T [y - x'(\beta,\rho)]$$

$$= \{y - [X\beta + \rho\,(y_{-1} - X_{-1}\,\beta\,)]\}^T \{y - [X\beta + \rho\,(y_{-1} - X_{-1}\,\beta\,)]\} \qquad (3.5)$$

where y_{-1} has typical element y_{t-1} and X_{-1} has typical row x_{t-1}.

The first-order conditions are then given by

$$\left[X - \hat{\rho}\,X_{-1} \qquad y_{-1} - X_{-1}\,\hat{\beta}\right]^T \left\{y - \left[X\hat{\beta} + \hat{\rho}\left(y_{-1} - X_{-1}\hat{\beta}\right)\right]\right\} = 0 \qquad (3.6)$$

and the corresponding G.N.R. is

$$y-\left[X\hat{\beta}+\hat{\rho}\left(y_{-1}-X_{-1}\hat{\beta}\right)\right]=\left[X-\hat{\rho}X_{-1} \quad y_{-1}-X_{-1}\hat{\beta}\right]\begin{bmatrix} b \\ r \end{bmatrix}+error\ \ vector \quad (3.7)$$

or

$$y_t-\hat{\rho}\,y_{t-1}-\left(x_t-\hat{\rho}\,x_{t-1}\right)\hat{\beta}=\left(x_t-\hat{\rho}\,x_{t-1}\right)b+r\left(y_{t-1}-x_{t-1}\hat{\beta}\right)+error\ term\ ;$$

$$t=2,\ \ldots,\ n \qquad (3.8)$$

where

$$y_t-\hat{\rho}\,y_{t-1}-\left(x_t-\hat{\rho}\,x_{t-1}\right)\hat{\beta}=\hat{\xi}_t$$

$$x_t-\hat{\rho}\,x_{t-1}=\left.\frac{\partial x_t'(\beta,\rho)}{\partial\beta}\right|_{(\beta,\rho)=\left(\hat{\beta},\hat{\rho}\right)}$$

$$y_{t-1}-x_{t-1}\hat{\beta}=\left.\frac{\partial x_t'(\beta,\rho)}{\partial\rho}\right|_{(\beta,\rho)=\left(\hat{\beta},\hat{\rho}\right)}=\hat{u}_{t-1}$$

and ^ denotes, as before, N.L.S. estimates.

This regression can be used as the basis for an algorithm to minimise the criterion function in (3.5): the convergence process to be stopped as soon as the O.L.S.

estimates $\overset{\circ}{b}$ and $\overset{\circ}{r}$ are not statistically different from zero, that is to say, when the G.N.R. has no explanatory power.

Under suitable regularity conditions plus the assumptions described above, the N.L.S. estimators for β and ρ will be consistent, asymptotically efficient and asymptotically normal with a covariance matrix consistently estimated by the estimated covariance matrix of the O.L.S. estimators for b and r in the G.N.R..

In fact, a particularisation of the general result given in equation (2.17) yields

$$(n-1)^{\frac{1}{2}}\begin{bmatrix}\hat{\beta}-\beta\\ \hat{\rho}-\rho\end{bmatrix}\overset{a}{\sim}N\left(\begin{bmatrix}0\\0\end{bmatrix}, w^2 \underset{n\to\infty}{\text{Plim}}\left\{(n-1)^{-1}\begin{bmatrix}(X-\rho X_{-1})^T(X-\rho X_{-1}) & (X-\rho X_{-1})^T u_{-1}\\ u_{-1}{}^T(X-\rho X_{-1}) & u_{-1}{}^T u_{-1}\end{bmatrix}\right\}^{-1}\right) \quad (3.9)$$

whereas the covariance estimate produced by the G.N.R. will be given by

$$\frac{\hat{\xi}^{\,T}\hat{\xi}}{n-k-2}\begin{bmatrix}\left(X-\hat{\rho}X_{-1}\right)^T\left(X-\hat{\rho}X_{-1}\right) & \left(X-\hat{\rho}X_{-1}\right)^T\hat{u}_{-1}\\ \hat{u}_{-1}{}^T\left(X-\hat{\rho}X_{-1}\right) & \hat{u}_{-1}{}^T\hat{u}_{-1}\end{bmatrix}^{-1} \quad (3.10)$$

where u_{-1} has typical element u_{t-1} and $\hat{u}_{-1}$ has typical element $\hat{u}_{t-1}$.

It is obvious that $\dfrac{\hat{\xi}^{\,T}\hat{\xi}}{n-k-2}$ consistently estimate w^2. Note that since the G.N.R. (3.8) has no explanatory power, the *R.S.S.* will be exactly the same as the *T.S.S.* in that G.N.R.. Also, $\hat{\rho}$ and $\hat{\beta}$ consistently estimate ρ and β. Therefore, n-1 times the covariance matrix estimate in (3.10) estimates consistently the covariance matrix of $(n-1)^{\frac{1}{2}}\begin{bmatrix}\hat{\beta}-\beta \\ \hat{\rho}-\rho\end{bmatrix}$. In other words, the covariance matrix estimate produced by the G.N.R. estimates consistently the covariance matrix of $\begin{bmatrix}\hat{\beta}-\beta \\ \hat{\rho}-\rho\end{bmatrix}$, which is the matrix whose estimation we are interested in.

One could also make valid inference about the N.L.S. estimates using the following simplified version of the G.N.R. (say S.G.N.R.)

$$y_t - \hat{\rho}\, y_{t-1} = \left(x_t - \hat{\rho}\, x_{t-1}\right) \mathrm{b} + r\left(y_{t-1} - x_{t-1}\hat{\beta}\right) + error\ term\ ;\ t = 2, \ldots, n \qquad (3.11)$$

which would report the same estimates for the variances of the O.L.S. estimates of b and r since both regressions have the same regressors and also the same residuals.

The second part of this statement is not so obvious, but it can easily be proved. First notice that the O.L.S. estimate $\overset{\circ}{r}$ in the G.N.R. (3.8) is the same as the one we obtain running the S.G.N.R. (3.11) (say $\overset{\circ}{r}_S$) since

$$M_{\left(X-\hat{\rho}X_{-1}\right)}\left(X-\hat{\rho}X_{-1}\right)=0.\text{[1]} \qquad (3.12)$$

In other words, $\overset{\circ}{r}_S$ also equals zero in the S.G.N.R.(3.11). Secondly, the O.L.S. estimate of b in the S.G.N.R.(3.11) (say $\overset{\circ}{b}_S$) is precisely $\hat{\beta}$, the N.L.S. estimate. In fact, in (3.8), the O.L.S. estimate of b, which is identically zero, is given by

$$\begin{aligned}\overset{\circ}{b} &= \left(\left(X-\hat{\rho}X_{-1}\right)^T M_{\hat{u}_{-1}}\left(X-\hat{\rho}X_{-1}\right)\right)^{-1}\left(X-\hat{\rho}X_{-1}\right)^T M_{\hat{u}_{-1}}\left(y-\hat{\rho}y_{-1}-\left(X-\hat{\rho}X_{-1}\right)\hat{\beta}\right)\\ &= \left(\left(X-\hat{\rho}X_{-1}\right)^T M_{\hat{u}_{-1}}\left(X-\hat{\rho}X_{-1}\right)\right)^{-1}\left(X-\hat{\rho}X_{-1}\right)^T M_{\hat{u}_{-1}}\left(y-\hat{\rho}y_{-1}\right)-\hat{\beta}\end{aligned}$$

where $M_{\hat{u}_{-1}}$ is the matrix that projects onto the orthogonal complement of $\hat{u}_{-1}$.

Therefore

$$\hat{\beta}=\left(\left(X-\hat{\rho}X_{-1}\right)^T M_{\hat{u}_{-1}}\left(X-\hat{\rho}X_{-1}\right)\right)^{-1}\left(X-\hat{\rho}X_{-1}\right)^T M_{\hat{u}_{-1}}\left(y-\hat{\rho}y_{-1}\right)$$

which is the O.L.S. estimate $\overset{\circ}{b}_S$ in the S.G.N.R.(3.11). Hence, the residuals in (3.11) are just the regressand in (3.8), or the residuals in (3.11) are just the residuals in (3.8) since the G.N.R.(3.8) has no explanatory power.

---

[1] $M_{\left(X-\hat{\rho}X_{-1}\right)}$ is the matrix that projects onto the orthogonal complement of $\left(X-\hat{\rho}X_{-1}\right)$.

Note that apart from the necessary regularity conditions mentioned in section 2.2., the stationarity condition must hold. Otherwise standard results about N.L.S., in particular the asymptotic normality theorem, would no longer apply. In practice the N.L.S. estimate $\hat{\rho}$ may be greater than 1, or although smaller, still very close to 1 (in absolute value); one should treat this as evidence of model inadequacy.

On the other hand, losing one observation makes no difference asymptotically.[2] We are therefore assuming the sample size is reasonably large.

Under these circumstances, however, it may not be worth the time to obtain N.L.S. estimates, because O.S. estimates are asymptotically equivalent and thus quite an adequate alternative.

Consider then $\left(\overset{'}{\beta},\overset{'}{\rho}\right)$ any vector of root-n consistent estimates for model (3.3). It is known from subsection 2.3.2.1. that

$$\left(\overset{\#}{\beta},\overset{\#}{\rho}\right)=\left(\overset{'}{\beta},\overset{'}{\rho}\right)+\left(\overset{\circ}{b},\overset{\circ}{r}\right)\overset{a}{=}\left(\hat{\beta},\hat{\rho}\right) \tag{3.13}$$

[2] On the contrary setting the first observation equal to zero would imply that u_1 rather than ξ_1, is the error term for observation 1. It would therefore no longer be appropriate simply to use N.L.S. to estimate model (3.3) as heteroskedasticity would have been created (observation 1 would have variance $w^2/\left(1-\rho^2\right)$ rather than w^2).

where $\left(\overset{\#}{\beta},\overset{\#}{\rho}\right)$ are the O.S. estimates, $\left(\hat{\beta},\hat{\rho}\right)$ are the N.L.S. estimates, $\left(\mathring{b},\mathring{r}\right)$ are the O.L.S. estimates of (b,r) in the following G.N.R.

$$y_t - \dot{\rho}\, y_{t-1} - \left(x_t - \dot{\rho}\, x_{t-1}\right)\dot{\beta} = \left(x_t - \dot{\rho}\, x_{t-1}\right)b + r\left(y_{t-1} - x_{t-1}\,\dot{\beta}\right) + error\ term\ ;$$

$$t = 2,\ \ldots,\ n \qquad (3.14)$$

and $\overset{a}{=}$ means asymptotically equivalent.

Once again, using the corresponding S.G.N.R. version

$$y_t - \dot{\rho}\, y_{t-1} = \left(x_t - \dot{\rho}\, x_{t-1}\right)b + r\left(y_{t-1} - x_{t-1}\,\dot{\beta}\right) + error\ term\ ;\ t = 2,\ \ldots,\ n \qquad (3.15)$$

would produce exactly the same O.L.S. estimate of r and therefore the same O.S. estimate of ρ. However, using the S.G.N.R. version would have the advantage of producing (directly) the O.S. estimate $\overset{\#}{\beta}$ as being $\mathring{b}_S$, the O.L.S. estimate of b in this simplified version of the G.N.R..

In fact, in (3.15),

$$\mathring{b}_S = \left[\left(X - \dot{\rho}\, X_{-1}\right)^T M_{\dot{u}_{-1}} \left(X - \dot{\rho}\, X_{-1}\right)\right]^{-1} \left(X - \dot{\rho}\, X_{-1}\right)^T M_{\dot{u}_{-1}} \left(y - \dot{\rho}\, y_{-1}\right) \qquad (3.16)$$

where $\dot{u}_{-1}$ has typical element $\dot{u}_{t-1}$ and $M_{\dot{u}_{-1}}$ is the matrix that projects onto the orthogonal complement of $\dot{u}_{-1}$.

On the other hand, in the G.N.R. (3.14),

$$\mathring{b} = \left[\left(X-\dot{\rho} X_{-1}\right)^T M_{\dot{u}_{-1}} \left(X-\dot{\rho} X_{-1}\right)\right]^{-1}\left(X-\dot{\rho} X_{-1}\right)^T M_{\dot{u}_{-1}}\left[\left(y-\dot{\rho} y_{-1}\right)-\left(X-\dot{\rho} X_{-1}\right)\dot{\beta}\right]$$

$$= \left[\left(X-\dot{\rho} X_{-1}\right)^T M_{\dot{u}_{-1}} \left(X-\dot{\rho} X_{-1}\right)\right]^{-1}\left(X-\dot{\rho} X_{-1}\right)^T M_{\dot{u}_{-1}}\left(y-\dot{\rho} y_{-1}\right)-\dot{\beta}. \quad (3.17)$$

As from (3.13), $\overset{\#}{\beta} = \dot{\beta} + \mathring{b}$, that is $\mathring{b} = \overset{\#}{\beta} - \dot{\beta}$, one may conclude that

$$\overset{\#}{\beta} = \left[\left(X-\dot{\rho} X_{-1}\right)^T M_{\dot{u}_{-1}} \left(X-\dot{\rho} X_{-1}\right)\right]^{-1}\left(X-\dot{\rho} X_{-1}\right)^T M_{\dot{u}_{-1}}\left(y-\dot{\rho} y_{-1}\right) \quad (3.18)$$

which is expression (3.16), the O.L.S. estimate $\mathring{b}_S$ in (3.15).

Also, as $\mathring{b} = \overset{\#}{\beta} - \dot{\beta}$ in (3.14), the O.L.S. residuals from that regression will be given by

$$y_t - \dot{\rho} y_{t-1} - \left(x_t - \dot{\rho} x_{t-1}\right)\dot{\beta} - \left(x_t - \dot{\rho} x_{t-1}\right)\mathring{b} - \mathring{r}\left(y_{t-1} - x_{t-1}\dot{\beta}\right)$$

$$= y_t - \overset{'}{\rho} y_{t-1} - \left(x_t - \overset{'}{\rho} x_{t-1}\right)\overset{'}{\beta} - \left(x_t - \rho x_{t-1}\right)\overset{\#}{\beta} + \left(x_t - \overset{'}{\rho} x_{t-1}\right)\overset{'}{\beta} - \overset{\circ}{r}\left(y_{t-1} - x_{t-1}\overset{'}{\beta}\right)$$

$$= y_t - \overset{'}{\rho} y_{t-1} - \left(x_t - \overset{'}{\rho} x_{t-1}\right)\overset{\#}{\beta} - \overset{\circ}{r}\left(y_{t-1} - x_{t-1}\overset{'}{\beta}\right) \qquad (3.19)$$

which are also the O.L.S. residuals from regression (3.15) since in that regression , as we have shown, $\overset{\circ}{b}_S = \overset{\#}{\beta}$ and the O.L.S. estimate of r is the same as the one produced by (3.14).

Therefore, the O.S. estimators for β and ρ which are consistent, asymptotically efficient and asymptotically normal (cf.(2.74)), will have its covariance matrix consistently estimated by the estimated covariance matrix of the O.L.S. estimates of b and r , no matter the version of the G.N.R. one may use (either (3.14) or (3.15)).[3]

Of course, the fact that O.S. estimators based on either version of the G.N.R. are asymptotically equivalent to N.L.S. estimators does not imply that the former will have finite-sample properties similar to those of the latter. Even under the crucial assumption that the model is correctly specified, a great deal depends on the quality of the initial consistent estimates (as emphasised in subsection 2.3.2.1.).

The question is therefore how to obtain those initial consistent estimates. Previously we have assumed that X does not contain any current endogenous variable. If we

[3] The covariance estimate produced by both G.N.R.'s is similar to the one given in (3.10) (with $\wedge$ replaced by $'$ everywhere).

assume that X does not contain any lagged dependent variable either, the answer to the question is quite obvious.

First, one should run by O.L.S. the regression

$$y_t = x_t\,\beta + error\ term\ ; \quad t = 1, 2, ..., n \tag{3.20}$$

to obtain inefficient but consistent (initial) estimates of β. Then, using the O.L.S. residuals

$$\hat{u}_t = y_t - x_t\,\hat{\beta}\ ;\ t = 1, 2, ..., n \tag{3.21}$$

which are consistent, one should run the O.L.S. regression

$$\hat{u}_t = \rho\,\hat{u}_{t-1} + error\ term\ ;\ t = 2, ..., n \tag{3.22}$$

to obtain the (initial) consistent estimate of ρ. .

Estimating ρ in this way implies that $\hat{u}_{-1}$ will be orthogonal to the regressand in G.N.R.(3.14). In fact, $\hat{\rho}$ has been obtained by minimising the criterion function

$$SS = \left(\hat{u} - \rho\,\hat{u}_{-1}\right)^T \left(\hat{u} - \rho\,\hat{u}_{-1}\right)$$

$$= \left[y - X\dot{\beta} - \rho\left(y_{-1} - X_{-1}\dot{\beta}\right)\right]^T \left[y - X\dot{\beta} - \rho\left(y_{-1} - X_{-1}\dot{\beta}\right)\right] \tag{3.23}$$

with first-order conditions

$$\left(y_{-1} - X_{-1}\dot{\beta}\right)^T \left[y - X\dot{\beta} - \dot{\rho}\left(y_{-1} - X_{-1}\dot{\beta}\right)\right] = 0. \tag{3.24}$$

Despite that, the O.L.S. estimate $\mathring{r}$ in the G.N.R.(3.14) will still in general be different from zero. It would be identically zero, if and only if either $\dot{u}_{-1}$ is also orthogonal to $X - \dot{\rho}X_{-1}$ or $X - \dot{\rho}X_{-1}$ is also orthogonal to the regressand.

However, as X only comprises exogenous variables, u_{-1} will be independent of X, which implies that

$$\operatorname*{Plim}_{n\to\infty}\left[(n-1)^{-1}\left(X - \dot{\rho}X_{-1}\right)^T \dot{u}_{-1}\right] = 0. \tag{3.25}$$

In these circumstances, as we will now show , the estimate of the variance of the O.L.S. estimate $\mathring{b}_R$ given by the following restricted version of the S.G.N.R. (say R.S.G.N.R.)

$$y_t - \dot{\rho}\, y_{t-1} = \left(x_t - \dot{\rho}\, x_{t-1}\right) \mathrm{b} + \textit{error term} \;;\; t = 2, \ldots, n \tag{3.26}$$

will be asymptotically the same as the estimate given by both $\mathring{b}$ and $\mathring{b}_S$ in G.N.R.(3.14) and in its simplified version, the S.G.N.R. in (3.15), respectively.

In the last two cases, the estimated covariance matrix of $\mathring{b}$ (also of $\mathring{b}_S$) is given by the upper left-hand block of the matrix

$$\frac{SSR\left(\mathring{b}, \mathring{r}\right)}{n-k-2}\begin{bmatrix} \left(X - \dot{\rho} X_{-1}\right)^T\left(X - \dot{\rho} X_{-1}\right) & \left(X - \dot{\rho} X_{-1}\right)^T \dot{u}_{-1} \\ \dot{u}_{-1}^{\,T}\left(X - \dot{\rho} X_{-1}\right) & \dot{u}_{-1}^{\,T} \dot{u}_{-1} \end{bmatrix}^{-1} \tag{3.27}$$

where, as we have already shown, $SSR\left(\mathring{b}, \mathring{r}\right) = SSR\left(\mathring{b}_S, \mathring{r}_S\right)$ (the sum-of-squares of the residuals in (3.19)), whereas the estimated covariance matrix of $\mathring{b}_R$ from regression (3.26) is simply given by

$$\frac{SSR\left(\mathring{b}_R\right)}{n-k-1}\left[\left(X - \dot{\rho} X_{-1}\right)^T\left(X - \dot{\rho} X_{-1}\right)\right]^{-1}. \tag{3.28}$$

Because of (3.25), $(n-1)^{-1}$ times the second factor in (3.27) will be asymptotically blockdiagonal and therefore the estimated covariance matrix of both $\mathring{b}$ and $\mathring{b}_S$ in the two versions of the G.N.R. reduces to

$$\frac{SSR\left(\mathring{b},\mathring{r}\right)}{n-k-2}\left[\left(X-\acute{\rho}X_{-1}\right)^{T}\left(X-\acute{\rho}X_{-1}\right)\right]^{-1}. \tag{3.29}$$

Moreover, is now reasonably clear that asymptotically, the O.L.S. estimates of b in (3.16) and (3.17) collapse, respectively, to

$$\begin{aligned}\mathring{b}_S &= \left[\left(X-\acute{\rho}X_{-1}\right)^{T}\left(X-\acute{\rho}X_{-1}\right)\right]^{-1}\left(X-\acute{\rho}X_{-1}\right)^{T}\left(y-\acute{\rho}y_{-1}\right) \\ &= \mathring{b}_R\end{aligned} \tag{3.30}$$

and

$$\begin{aligned}\mathring{b} &= \left[\left(X-\acute{\rho}X_{-1}\right)^{T}\left(X-\acute{\rho}X_{-1}\right)\right]^{-1}\left(X-\acute{\rho}X_{-1}\right)^{T}\left(y-\acute{\rho}y_{-1}\right)-\acute{\beta} \\ &= \overset{\#}{\beta}-\acute{\beta} = \mathring{b}_R-\acute{\beta}\end{aligned} \tag{3.31}$$

That is, the O.S. estimate $\overset{\#}{\beta}$ given by G.N.R.(3.14) which is in general the same as the O.L.S. estimate $\mathring{b}_S$ in the S.G.N.R.(3.15), is also asymptotically the same as the O.L.S. estimate $\mathring{b}_R$ that one would obtain running the R.S.G.N.R.(3.26).

On the other hand , the residuals are exactly the same in (3.14) and (3.15) and equal to

$$y_t - \overset{\circ}{\rho}\, y_{t-1} - \left(x_t - \overset{\circ}{\rho}\, x_{t-1}\right) \overset{\#}{\beta} \tag{3.32}$$

because, taking into account (3.25), $\overset{\circ}{r}$ will be asymptotically zero in (3.19).

These are precisely the residuals we would obtain running by O.L.S. the regression in (3.26) since $\overset{\circ}{b}_R$ in that regression equals $\overset{\#}{\beta}$, as equation (3.31) shows.

Therefore, as the different corrections for the degrees of freedom are also negligible asymptotically, $SSR\left(\overset{\circ}{b}, \overset{\circ}{r}\right) = SSR\left(\overset{\circ}{b}_R\right)$, and we must conclude that inference based upon model (3.26) is perfectly valid, provided X only comprises exogenous variables.

Under this crucial condition, it is now easily seen that in the context of N.L.S. it would be perfectly valid, as well, to make asymptotic inference based upon the restricted version of (3.11)

$$y_t - \hat{\rho}\, y_{t-1} = \left(x_t - \hat{\rho}\, x_{t-1}\right) \mathrm{b} + \mathit{error\ term}\ ;\ t = 2, \ldots, n. \tag{3.33}$$

Hence, provided X only comprises exogenous variables there is no real advantage in using a G.N.R. to make inference based on N.L.S. or O.S. estimates since (3.33) and (3.26) , respectively, are perfectly valid to do so. The difference between the final O.L.S. regression to be run after a Cochrane-Orcutt procedure as in (3.33) and the simple O.S. estimation as in (3.26) will be confined to the small sample performance[4] as asymptotically they are equivalent.

3.3 Instrumental Variable Estimation

Let us now assume that X contains either current endogenous variables or lagged dependent variables or even variables of both types. Whichever the case, the O.L.S. regression (3.20) does not produce consistent initial estimates of β.

There are two different sources of inconsistency, one from the fact that X may contain lagged dependent variables in the presence of serially correlated disturbances, and the other from the fact that X may contain current endogenous variables. These deserve separate treatment.

Consider the former case first.

One possible way to overcome the problem is, as before, to use N.L.S. estimates of the transformed model in (3.3). Note that the presence of y_{-1} as well as of lagged dependent variables in X is now irrelevant given the sphericality of the model.

---

[4] Even the N.L.S. estimates from (3.33) entail some loss of finite sample performance relative to a full M.L. estimation, because of the difference in the treatment of starting values.

In other words, the transformation of the model gets rid of the serial correlation problem which was, in the presence of lagged dependent variables, the source of the correlation between those variables and the disturbances u_t in model (3.1).

Therefore, as long as we are prepared to apply N.L.S., does not really matter whether the original model contains lagged dependent variables or not.

Provided X does not contain any current endogenous variable, assuming stationarity conditions not only for ρ but also for the coefficients of the lagged dependent variables [5] and dropping as many observations for y and all the variables in X as the order of the highest lag, the N.L.S. estimates will be consistent, asymptotically normal and asymptotically efficient. However, in the presence of lagged dependent variables in X, inference about the N.L.S. estimates based upon regression (3.33) is no longer valid and consequently we must use the G.N.R. or its simplified version, namely to obtain the correct estimates of the variances.

On the other hand, if we prefer to apply a O.S. estimator, dynamics will imply a different solution, eventually the necessity for I.V. estimation to find initial consistent root-n estimates of β and ρ.

A proper discussion of all these aspects is the main objective of the next section.

---

[5] The roots of the polynomial which defines the difference equation in y must lie outside the unit circle.

3.3.1 The Dynamic Adjustment Model

Consider the general formulation in (3.1), in which now X implicitly contains one or more lagged dependent variables. Either using the G.N.R. (3.8) or its simplified version, the S.G.N.R.(3.11), the reported estimated covariance matrix for the N.L.S. estimates of β and ρ will be given by

$$\frac{SSR\left(\overset{\circ}{b},\overset{\circ}{r}\right)}{n-k-2}\begin{bmatrix}\left(X-\hat{\rho}X_{-1}\right)^{T}\left(X-\hat{\rho}X_{-1}\right) & \left(X-\hat{\rho}X_{-1}\right)^{T}\hat{u}_{-1}\\ \hat{u}_{-1}^{T}\left(X-\hat{\rho}X_{-1}\right) & \hat{u}_{-1}^{T}\hat{u}_{-1}\end{bmatrix}^{-1} \tag{3.34}$$

where, as we have shown before, $SSR\left(\overset{\circ}{b},\overset{\circ}{r}\right)=SSR\left(\overset{\circ}{b}_{S},\overset{\circ}{r}_{S}\right)$ also in the context of L.S..

On the other hand, the estimated covariance matrix for $\overset{\circ}{b}_{R}$ from regression (3.33) is simply given by

$$\frac{SSR\left(\overset{\circ}{b}_{R}\right)}{n-k-1}\left[\left(X-\hat{\rho}X_{-1}\right)^{T}\left(X-\hat{\rho}X_{-1}\right)\right]^{-1}. \tag{3.35}$$

The presence in X of any lagged dependent variable or any other variable that may be correlated with u_{-1}, invalidates the orthogonality condition

$$\operatorname*{Plim}_{n\to\infty}\left[(n-1)^{-1}\left(X-\hat{\rho}X_{-1}\right)^T\hat{u}_{-1}\right]=0 \tag{3.36}$$

and therefore the block-diagonality of (3.34).

Hence, the estimated covariance matrix of $\overset{\circ}{b}$ (also of $\overset{\circ}{b}_S$) in either the G.N.R. (3.8) or in its simplified version (3.11) should be written as

$$\frac{SSR\left(\overset{\circ}{b},\overset{\circ}{r}\right)}{n-k-2}\left[\left(X-\hat{\rho}X_{-1}\right)^T M_{\hat{u}_{-1}}\left(X-\hat{\rho}X_{-1}\right)\right]^{-1}. \tag{3.37}$$

The comparison between this expression and (3.35) justifies our earlier statement that inference based on regression (3.33) might be misleading. The reason is that

$$\left[\left(X-\hat{\rho}X_{-1}\right)^T M_{\hat{u}_{-1}}\left(X-\hat{\rho}X_{-1}\right)\right]^{-1}-\left[\left(X-\hat{\rho}X_{-1}\right)^T\left(X-\hat{\rho}X_{-1}\right)\right]^{-1} \tag{3.38}$$

is Positive-Semidefinite and in turn $SSR\left(\overset{\circ}{b},\overset{\circ}{r}\right)=SSR\left(\overset{\circ}{b}_R\right)$ since the G.N.R.(3.8)will have no explanatory power and, in the R.S.G.N.R.(3.33), $\overset{\circ}{b}_R=\hat{\beta}$, the N.L.S. estimate.

Therefore, apart the negligible asymptotic effect of using different degrees of freedom in the denominator of the statistic that estimates w^2, one may conclude that the standard errors reported by regression (3.33) will be too small even asymptotically. [6]

Let us now turn our attention to the O.S. estimation of the dynamic adjustment model.

First note that model (3.3) can be rewritten as

$$y_t - \rho\, y_{t-1} = \left(x_t - \rho\, x_{t-1}\right)\beta + \xi_t \quad ; \quad \xi_t \sim \mathrm{IID}(0, w^2)\,. \tag{3.39}$$

Hence, if we are able to find a consistent estimate $\dot{\rho}$ to substitute for ρ, we can then find a consistent estimate of β, from O.L.S. applied to the R.S.G.N.R. in (3.26).

But, as we have been emphasising, if X comprises lagged dependent variables or any other variable correlated with u_{-1}, the estimation of β is not independent of the estimation of ρ. Therefore, the initial root-n consistent estimates $\dot{\beta}$ and $\dot{\rho}$ should be used in either the G.N.R. (3.14) or in its simplified version, the S.G.N.R. in (3.15) so that valid (asymptotic) inference can be conducted based upon them [7]. On the contrary, (asymptotic) inference simply based upon the R.S.G.N.R. version in

---

[6] As a matter of fact, the difference in the different corrections for the degrees of freedom will reinforce this effect for finite samples.

[7] The proof for the O.S. case is on the same lines as for N.L.S. presented earlier.

(3.26) will be misleading since that regression will report a wrong estimate of the covariance matrix of $\overset{\#}{\beta}$.

It might seem that we could obtain consistent estimates of both β and ρ estimating by O.L.S.

$$y_t = \rho\, y_{t-1} + x_t\beta + x_{t-1}\gamma + \xi_t \quad ; \quad \xi_t \sim \text{IID}(0, w^2) \tag{3.40}$$

where $\gamma = -\rho\beta$.

However, as some of the coefficients in this model may not be identifiable, that will rarely be the case. For example, a constant term might be included in X and therefore in X_{-1} ; Then, one of the parameters in β and one of the parameters in γ cannot be separately identified.

To avoid this specific situation, consider the alternative specification of model (3.3),

$$y_t = \alpha + x_t\beta + \rho\,(y_{t-1} - \alpha - x_{t-1}\,\beta\,) + \xi_t$$

or

$$y_t = \alpha\left(1-\rho\right) + \rho\, y_{t-1} + x_t\beta + x_{t-1}\gamma + \xi_t \quad ; \quad \xi_t \sim \text{IID}\left(0, w^2\right) \tag{3.41}$$

If all the lags of the dependent variable included in X are greater than one, the estimated coefficient of y_{-1} provides a consistent estimate of ρ. Unfortunately, the absence of the first lag in X is not so common.

For example, consider that X comprises the two regressors y_{-1} and z with corresponding coefficients β_1 and β_2. Model (3.41) becomes

$$y_t = \alpha\left(1-\rho\right) + \rho\, y_{t-1} + \beta_1 y_{t-1} + \beta_2 z_t - \rho\beta_1 y_{t-2} - \rho\beta_2 z_{t-1} + \xi_t$$

or

$$y_t = \alpha\left(1-\rho\right) + \left(\rho+\beta_1\right) y_{t-1} - \rho\beta_1 y_{t-2} + \beta_2 z_t - \rho\beta_2 z_{t-1} + \xi_t \qquad (3.42)$$

whose corresponding unrestricted version is given by

$$y_t = \delta_1 + \delta_2 y_{t-1} + \delta_3 y_{t-2} + \delta_4 z_t + \delta_5 z_{t-1} + \xi_t . \qquad (3.43)$$

As there are five regression coefficients in this unrestricted version, one more than in the restricted model (3.42), there are several ways to obtain a consistent estimate of ρ using the consistent unrestricted O.L.S parameter estimates. The easiest is to obtain

$$\dot{\rho} = -\frac{\mathring{\delta}_5}{\mathring{\delta}_4} \tag{3.44}$$

from the last two relationships $\delta_4 = \beta_2$ and $\delta_5 = -\rho\beta_2$.

Since in many cases the original model will have more than one regressor like z, the increasing degree of arbitrariness used to estimate ρ is a drawback of this procedure.

A completely different approach would be to run regression (3.20) by I.V. to first obtain consistent estimates of β and then to use the I.V. residuals in the usual way to estimate ρ consistently (see Hatanaka (1980)).

Formally, to obtain $\dot{\beta}$, the I.V. estimate of β, we can run instead by O.L.S. the regression

$$y_t = \mathring{x}_t \beta + error\ term \tag{3.45}$$

where $\mathring{x}_t$ is row t of matrix $\mathring{X} = P_W X$ and P_W projects orthogonally onto W, a suitable matrix of instruments. Then, use the I.V. residuals

$$\dot{u}_t = y_t - x_t \dot{\beta} \tag{3.46}$$

which are consistent, to run a second O.L.S. regression

$$\dot{u}_t = \rho \dot{u}_{t-1} + error\ term \tag{3.47}$$

to obtain a consistent estimate of ρ.[8]

Estimating ρ in this way once again implies that $\dot{u}_{-1}$ will be orthogonal to the regressand in the G.N.R. (3.14), since $\dot{\rho}$ is still obtained by minimising the criterion function in (3.23).

The only difference is that given the initial I.V. estimation of β, that criterion function now incorporates I.V. rather than O.L.S. residuals.

Despite that, as we stressed before, the O.L.S. estimate $\mathring{r}$ in the G.N.R.(3.14) (and the O.L.S. estimate $\mathring{b}_S$ in the G.N.R.(3.15)) will be neither in general nor asymptotically equal to zero if X comprises lagged dependent variables.

All this has been carefully proved in the context of L.S. estimation and now easily generalised to a context in which an initial consistent estimate of β is obtained by I.V..

---

[8] The number of observations that we lose in all the regressions involved (regression (3.20) included) is given by the highest order of the lags considered as instruments. A valid alternative would be to set equal to zero the unobserved values as necessary: that will not affect the consistency of the initial estimates.

Dynamics is mostly relevant in the sense that invalidates the inference commonly based upon the restricted version either in (3.26) or in (3.33), that is, regardless we may use O.S. or N.L.S. estimation.

To conclude this section a final note. In equation (3.45) the instrumental variables (the columns of $\overset{\circ}{X}$) are the columns of predicted values from the O.L.S. regressions of each column of X on W. Including in W at least all the exogenous variables in X, so that for those variables the predicted values are actually their actual values, allows us to interpret the O.L.S. regression (3.45) as being a regression where only the lagged dependent variables have been substituted by instrumental variables. The fact that other variables than the exogenous variables in X (and some of their lags) might be included in the set of instruments brings back the question of arbitrariness. An element of it is indeed also involved in this procedure, because the initial consistent estimates will depend on the instruments used.

Before commenting further on this issue, let us bring the other source of inconsistency referred to, that is, the case in which the inconsistency of the O.L.S. estimation of regression (3.20) is rather due to the presence of current endogenous variables in the matrix X.

3.3.2 Current Endogenous Variables as Explanatory Variables

This is a more fundamentally different case since (even) the application of N.L.S. to model (3.3) would not guarantee consistent estimates. The reason is that the existing correlation between the endogenous elements in x_t and u_t may still remain between those elements and ξ_t.

Therefore, model (3.3) has to be estimated by N.L.I.V..

The N.L.I.V. estimates may be obtained by minimising with respect to (β, ρ) the criterion function

$$\left(y - x'(\beta,\rho)\right)^T P_W \left(y - x'(\beta,\rho)\right)$$

$$= \left\{y - \left[X\beta + \rho\ (y_{-1} - X_{-1}\ \beta\ )\right]\right\}^T P_W \left\{y - \left[X\beta + \rho\ (y_{-1} - X_{-1}\ \beta\ )\right]\right\} \qquad (3.48)$$

Which is the I.V. analog version of the criterion function given in (3.5).

The first-order conditions for this minimisation are given by

$$\left[X - \tilde{\rho} X_{-1} \quad y_{-1} - X_{-1}\tilde{\beta}\right] P_W \left\{y - \left[X\tilde{\beta} + \tilde{\rho}\ \left(y_{-1} - X_{-1}\tilde{\beta}\right)\right]\right\} = 0 \qquad (3.49)$$

and the corresponding G.N.R. is

$$y-\left[X\tilde{\beta}+\tilde{\rho}\left(y_{-1}-X_{-1}\tilde{\beta}\right)\right]=P_{W}\left[X-\tilde{\rho}X_{-1} \quad y_{-1}-X_{-1}\tilde{\beta}\right]\begin{bmatrix}b\\r\end{bmatrix}+error\ \ vector \quad (3.50)$$

or

$$y_{t}-\tilde{\rho}y_{t-1}-\left(x_{t}-\tilde{\rho}\ x_{t-1}\right)\ \tilde{\beta}=\left(\overset{\circ}{x}_{t}-\tilde{\rho}\overset{\circ}{x}_{t-1}\right)b+r\left(\overset{\circ}{y}_{t-1}-\overset{\circ}{x}_{t-1}\tilde{\beta}\right)+error\ term \qquad (3.51)$$

where

$$y_{t}-\tilde{\rho}\ y_{t-1}-\left(x_{t}-\tilde{\rho}\ x_{t-1}\right)\ \tilde{\beta}=\tilde{\xi}_{t}$$

$$\overset{\circ}{x}_{t-1}-\tilde{\rho}\ \overset{\circ}{x}_{t-1}=\left.\frac{\partial \overset{\circ}{x}'_{t}(\beta,\rho)}{\partial\beta}\right|_{(\beta,\rho)=\left(\tilde{\beta},\tilde{\rho}\right)}$$

$$\overset{\circ}{y}_{t-1}-\overset{\circ}{x}_{t-1}\tilde{\beta}=\left.\frac{\partial\ \overset{\circ}{x}'_{t}(\beta,\rho)}{\partial\rho}\right|_{(\beta,\rho)=\left(\tilde{\beta},\tilde{\rho}\right)}$$

$\overset{\circ}{x}_{t}$, $\overset{\circ}{x}_{t-1}$ and $\overset{\circ}{y}_{t-1}$ denotes the corresponding row of $\overset{\circ}{X}=P_{W}X$, $\overset{\circ}{X}_{-1}=P_{W}X_{-1}$ and $\overset{\circ}{y}_{-1}=P_{W}y_{-1}$ respectively, $\overset{\circ}{x}'_{t}$ the corresponding row of $\overset{\circ}{X}'=P_{W}X'$, and ~ denotes, as before, N.L.I.V. estimates.

The G.N.R. in (3.51) is the one to be used for an algorithm to minimise the criterion function in (3.48) where the convergence process is to be stopped as soon as the O.L.S. estimates $\mathring{b}$ and $\mathring{r}$ are not statistically different from zero.

Under suitable regularity conditions mentioned in section 2.2., given a suitable set of instruments W, assuming $|\rho| < 1$ and losing some observations for both y and all the variables in X [9] the N.L.I.V. estimators will be consistent and asymptotically normal with a covariance matrix consistently estimated by the estimated covariance matrix of the O.L.S. estimators for b and r in the G.N.R. (3.51).

The O.L.S. estimate of b in the G.N.R., which is identically zero, will be given by

$$\mathring{b} = \left[\left(\mathring{X} - \tilde{\rho}\mathring{X}_{-1}\right)^{T} M_{\left(\mathring{y}_{-1} - \mathring{X}_{-1}\tilde{\beta}\right)}\left(\mathring{X} - \tilde{\rho}\mathring{X}_{-1}\right)\right]^{-1}\left(\mathring{X} - \tilde{\rho}\mathring{X}_{-1}\right)^{T} M_{\left(\mathring{y}_{-1} - \mathring{X}_{-1}\tilde{\beta}\right)}\left[\left(y - \tilde{\rho}y_{-1}\right) - \left(X - \tilde{\rho}X_{-1}\right)\tilde{\beta}\right]$$

$$= \left[\left(\mathring{X} - \tilde{\rho}\mathring{X}_{-1}\right)^{T} M_{\left(\mathring{y}_{-1} - \mathring{X}_{-1}\tilde{\beta}\right)}\left(\mathring{X} - \tilde{\rho}\mathring{X}_{-1}\right)\right]^{-1}\left(\mathring{X} - \tilde{\rho}\mathring{X}_{-1}\right)^{T} M_{\left(\mathring{y}_{-1} - \mathring{X}_{-1}\tilde{\beta}\right)}\left(y - \tilde{\rho}y_{-1}\right) - \tilde{\beta} \quad (3.52)$$

since

[9] The number of observations that we lose is given by the highest order of the lags considered as instruments.

$$\left(\mathring{X}-\tilde{\rho}\mathring{X}_{-1}\right)^T M_{\left(\mathring{y}_{-1}-\mathring{X}_{-1}\tilde{\beta}\right)}\left(X-\tilde{\rho}X_{-1}\right)=\left(\mathring{X}-\tilde{\rho}\mathring{X}_{-1}\right)^T M_{\left(\mathring{y}_{-1}-\mathring{X}_{-1}\tilde{\beta}\right)}\left(\mathring{X}-\tilde{\rho}\mathring{X}_{-1}\right) \tag{3.53}$$

as P_W is idempotent.[10]

Therefore

$$\tilde{\beta}=\left[\left(\mathring{X}-\tilde{\rho}\mathring{X}_{-1}\right)^T M_{\left(\mathring{y}_{-1}-\mathring{X}_{-1}\tilde{\beta}\right)}\left(\mathring{X}-\tilde{\rho}\mathring{X}_{-1}\right)\right]^{-1}\left(\mathring{X}-\tilde{\rho}\mathring{X}_{-1}\right)^T M_{\left(\mathring{y}_{-1}-\mathring{X}_{-1}\tilde{\beta}\right)}\left(y-\tilde{\rho}y_{-1}\right) \tag{3.54}$$

which is just the O.L.S. estimate $\mathring{b}_S$ in the S.G.N.R. version

$$y_t-\tilde{\rho}\,y_{t-1}=\left(\mathring{x}_t-\tilde{\rho}\;\mathring{x}_{t-1}\right)b+r\left(\mathring{y}_{t-1}-\mathring{x}_{t-1}\tilde{\beta}\right)+error\ term. \tag{3.55}$$

On the other hand, the O.L.S estimate of r in the G.N.R., which is also identically zero, will be given by

$$\mathring{r}=\left[\left(\mathring{y}_{-1}-\mathring{X}_{-1}\tilde{\beta}\right)^T M_{\left(\mathring{X}-\tilde{\rho}\mathring{X}_{-1}\right)}\left(\mathring{y}_{-1}-\mathring{X}_{-1}\tilde{\beta}\right)\right]^{-1}\left(\mathring{y}_{-1}-\mathring{X}_{-1}\tilde{\beta}\right)^T M_{\left(\mathring{X}-\tilde{\rho}\mathring{X}_{-1}\right)}\left[\left(y-\tilde{\rho}y_{-1}\right)-\left(X-\tilde{\rho}X_{-1}\right)\tilde{\beta}\right]$$

---

10 $M_{\left(\mathring{y}_{-1}-\mathring{X}_{-1}\tilde{\beta}\right)}$ is the matrix that projects onto the orthogonal complement of $\left(y_{-1}-X_{-1}\tilde{\beta}\right)$ that lies in the subspace of W.

$$=\left[\left(\overset{\circ}{y}_{-1}-\overset{\circ}{X}_{-1}\tilde{\beta}\right)^{T} M_{\left(\overset{\circ}{X}-\tilde{\rho}\overset{\circ}{X}_{-1}\right)}\left(\overset{\circ}{y}_{-1}-\overset{\circ}{X}_{-1}\tilde{\beta}\right)\right]^{-1}\left(\overset{\circ}{y}_{-1}-\overset{\circ}{X}_{-1}\tilde{\beta}\right)^{T} M_{\left(\overset{\circ}{X}-\tilde{\rho}\overset{\circ}{X}_{-1}\right)}\left(y-\tilde{\rho}y_{-1}\right) \quad (3.56)$$

since

$$\left(\overset{\circ}{y}_{-1}-\overset{\circ}{X}_{-1}\tilde{\beta}\right)^{T} M_{\left(\overset{\circ}{X}-\tilde{\rho}\overset{\circ}{X}_{-1}\right)}\left(X-\tilde{\rho}X_{-1}\right)\tilde{\beta}=\left(\overset{\circ}{y}_{-1}-\overset{\circ}{X}_{-1}\tilde{\beta}\right)^{T} M_{\left(\overset{\circ}{X}-\tilde{\rho}\overset{\circ}{X}_{-1}\right)}\left(\overset{\circ}{X}-\tilde{\rho}\overset{\circ}{X}_{-1}\right)\tilde{\beta} \quad ,$$

as P_W is idempotent,

$$= 0, \text{ as } M_{\left(\overset{\circ}{X}-\tilde{\rho}\overset{\circ}{X}_{-1}\right)}\left(\overset{\circ}{X}-\tilde{\rho}\overset{\circ}{X}_{-1}\right)=0.^{11} \quad (3.57)$$

Therefore, $\overset{\circ}{r}$ in the G.N.R. is precisely the same as $\overset{\circ}{r}_S$ in the S.G.N.R.(3.55).

All this allows us to conclude that the residuals in the G.N.R., which are just the regressand, are given by

$$y_t-\tilde{\rho}y_{t-1}-\left(x_t-\tilde{\rho}x_{t-1}\right)\tilde{\beta} \quad (3.58)$$

whereas the residuals in the simplified version are given by

[11] $M_{\left(\overset{\circ}{X}-\tilde{\rho}\overset{\circ}{X}_{-1}\right)}$ is the matrix that projects onto the orthogonal complement of $\left(X-\tilde{\rho}X_{-1}\right)$ that lies in the subspace of W.

$$y_t - \tilde{\rho} y_{t-1} - \left(\overset{\circ}{x}_t - \tilde{\rho} \overset{\circ}{x}_{t-1}\right)\tilde{\beta}. \tag{3.59}$$

In other words, the residuals are not the same in the two regressions and therefore if we are using a two-stage N.L.S. procedure we must not consider using (3.55) to base our inference about the N.L.I.V. estimates. If we do, we do not estimate w^2 consistently.

Note that premultiplying P_W by the regressand in both regressands would make the residuals equal. In that case, however, none of the regressands would estimate w^2 consistently.

Also, it might seem easier to regress the regressand on the regressors by an I.V. procedure, using W as matrix of instruments. This would avoid the initial stage of regressing the regressors on W and the residuals reported by the package would be the correct ones (given in (3.58)). However, this is not as good idea as it seems since one could then not use the explained sum of squares reported by the package to calculate test statistics. [12]

Previously we discarded the restricted simplified version of the G.N.R., because it is not valid to base inference on it when X contains either lagged dependent or

---

[12] This statement has been fully justified in the previous chapter (subsection 2.3.1). However the issue will be recovered in chapter 4, when dealing with testing procedures in the context of this particular model.

current endogenous variables.[13] However, now we do not discard the simplified version of the I.V.-based G.N.R., despite the fact that it is not valid to base inference on it when X contains current endogenous variables and we use a two-stage L.S. procedure.

Nevertheless, from now on we will mostly consider the (complete) G.N.R. version. Obviously, this version 'encompasses' the others in the sense that it is generally valid.

As before, the nonlinearity of the estimation process can still be avoided considering O.S. estimators.

The G.N.R. to be used will be the I.V. analog of (3.14),

$$y_t - \overset{\prime}{\rho} y_{t-1} - \left(x_t - \overset{\prime}{\rho} x_{t-1} \right) \overset{\prime}{\beta} = \left(\mathring{x}_t - \overset{\prime}{\rho} \mathring{x}_{t-1} \right) b + r \left(\mathring{y}_{t-1} - \mathring{x}_{t-1} \overset{\prime}{\beta} \right) + \textit{error term} \tag{3.60}$$

where

$$\left(\overset{\#}{\beta}, \overset{\#}{\rho} \right) = \left(\overset{\prime}{\beta}, \overset{\prime}{\rho} \right) + \left(\mathring{b}, \mathring{r} \right) \overset{a}{=} \left(\tilde{\beta}, \tilde{\rho} \right). \tag{3.61}$$

---

[13] Actually, the statement concerning the presence of current endogenous variables will be fully justified in the next section, when discussing the set of instruments one should include in W.

The O.S. estimators so obtained for β and ρ can be proved to be consistent and asymptotically normal and will have its covariance matrix consistently estimated by the estimated covariance matrix of the O.L.S. estimators for b and r (see section (2.3.1)).

The initial consistent root-n estimates can be obtained as in the previous (dynamic) case, running regression (3.20) by I.V. to first obtain consistent estimates of β and then using the I.V. residuals to estimate ρ consistently.

Including in W at least all the exogenous variables in X allows us this time to interpret the O.L.S. regression (3.45) as being a regression where only the current endogenous variables have been substituted by instrumental variables.

It remains to discuss which set of instruments should we use whenever an I.V. approach is in order, that is, either in the dynamic model with serially correlated disturbances or in the presence of current endogenous regressors in the equation.

3.3.3 The Choice of the Set of Instruments

For asymptotic efficiency, one always wants the instrument set to be used in the G.N.R. to include all the exogenous variables but also all the lagged variables that appear in the regression function $x_t'(\beta,\rho)$.

To clarify ideas, consider for example the simpler model

$$y_{1,t} = \beta_0 + \beta_1 z_{1,t} + \beta_2 z_{2,t} + \beta_3 y_{1,t-1} + \beta_4 y_{2,t} + u_t \quad ; \quad u_t = \rho\, u_{t-1} + \xi_t \quad ;$$

$$\xi_t \sim \text{IID}\left(0, w^2\right) \qquad (3.62)$$

where X contains three exogenous variables (a constant and two z's) one lagged dependent variable (the first lag of y_1) and one current endogenous variable (y_2).

In this case, the regression function $x_t'(\beta,\rho)$ will be given by

$$\begin{aligned} x_t'(\beta,\rho) = {} & \beta_0 + \beta_1 z_{1,t} + \beta_2 z_{2,t} + \beta_3 y_{1,t-1} + \beta_4 y_{2,t} + \\ & + \rho\left(y_{1,t-1} - \beta_0 - \beta_1 z_{1,t-1} - \beta_2 z_{2,t-1} - \beta_3 y_{1,t-2} - \beta_4 y_{2,t-1}\right) \end{aligned}$$

$$\begin{aligned} = {} & \beta_0(1-\rho) + \beta_1 z_{1,t} + \beta_2 z_{2,t} - \rho\beta_1 z_{1,t-1} - \rho\beta_2 z_{2,t-1} + (\beta_3 + \rho) y_{1,t-1} - \\ & - \rho\beta_3 y_{1,t-2} + \beta_4 y_{2,t} - \rho\beta_4 y_{2,t-1} \end{aligned} \qquad (3.63)$$

One should therefore use

$$W = \left\{1, z_1, z_2, z_{1,-1}, z_{2,-1}, y_{1,-1}, y_{1,-2}, y_{2,-1}\right\} \qquad (3.64)$$

as the set of instruments to form the instrumental variables in both the G.N.R.'s (3.51) and (3.60).

Note that including in W at least all these variables it allows us to interpret those G.N.R.'s in the following way

$$y_t - \overset{*}{\rho} y_{t-1} - \left(x_t - \overset{*}{\rho} x_{t-1}\right)\overset{*}{\beta} = \left(\overset{\circ}{x}_t - \overset{*}{\rho} x_{t-1}\right) b + r\left(y_{t-1} - x_{t-1}\overset{*}{\beta}\right) + error\ term \qquad (3.65)$$

where $y_t = y_{1,t}$; $y_{t-1} = y_{1,t-1}$; $\overset{\circ}{x}_t = \left[1 \quad z_{1,t} \quad z_{2,t} \quad y_{1,t-1} \quad \overset{\circ}{y}_{2,t}\right]$;

$x_{t-1} = \left[1 \quad z_{1,t-1} \quad z_{2,t-1} \quad y_{1,t-2} \quad y_{2,t-1}\right]$

and $*$ denotes either N.L.I.V. or any other root-n consistent estimate, respectively.

In other words this choice of W generates G.N.R.'s where only the current endogenous variable y_2 has been substituted by an instrumental variable $\overset{\circ}{y}_2$. In the O.S. context, however, one first has to find initial consistent root-n estimates of β to obtain consistent residuals to then find an initial consistent root-n estimate of ρ.

As regression (3.62) includes a current endogenous variable but also a lagged dependent variable in the presence of serial correlated disturbances, one therefore should form instrumental variables for both y_2 and $y_{1,-1}$. Obviously, the set of instruments to be used at this stage must not include $y_{1,-1}$. It must not include either the lagged dependent variable $y_{1,-2}$ or the lagged endogenous variable $y_{2,-1}$ since either of them is correlated with the disturbances.[14]

Hence, we are left with

$$W_r = \left\{1, z_1, z_2, z_{1,-1}, z_{2,-1}\right\} \tag{3.66}$$

that is, with only five valid instruments to estimate five parameters. Consequently, the parameters will be exactly identified and the validity of the instruments cannot be tested. Also, in the set (3.64), the presence of the first lag of y_2 in some way would incorporate some level of explanation of y_2 itself. In other words, there was no need for the inclusion of extra exogenous variables to form the instrumental variable $\overset{\circ}{y}_2$ to be used in the G.N.R.'s. Now, not considering $y_{2,-1}$ in the reduced set (3.66) one has lost that level of explanation when forming the instrumental variable for y_2 at the first stage.

The remedy is therefore obvious: consider some extra exogenous variables that one suspects explain y_2 and include those variables and also their first lag in both

[14] If $y_{2,t}$ is correlated with u_t, $y_{2,t-1}$ is correlated with u_{t-1} .Thus, both $y_{2,t-1}$ and u_t depend on u_{t-1}; therefore, $y_{2,t-1}$ and u_t are also correlated. In what respects $y_{1,t-2}$ (or any other lagged dependent variable) its correlation with u_t is obvious.

sets.[15] The overidentifying restrictions implied by the extended reduced set can and should then be tested at the first stage.

Obviously, if $y_{1,-1}$ was not in equation (3.62), the set of instruments to form the instrumental variable for y_2 in the G.N.R.'s would not comprise $y_{1,-2}$ but would still comprise $y_{1,-1}$ as well as $y_{2,-1}$; the set of instruments to be used at the first stage in a O.S. context would still be the reduced set (3.66).

As in the previous case, this set might also be extended with some extra exogenous variables and their first lag. Even though now the overidentifying restrictions implied by the reduced set (3.66) could be tested,[16] looking for some level of explanation of y_2 is still a valid argument on the grounds of efficiency.[17]

Once again if we would consider using these extra variables, they should also be included in the more comprehensive set to be used in the G.N.R.'s.

On the other hand, if y_2 was the variable missing in equation (3.64), there was no need for I.V. estimation in the G.N.R.'s. The set of instruments to be used at the

[15] If we are implicitly considering a reduced form for y_2, the transformed model should also reflect that information.

[16] Note that now we have a set comprising five instruments to estimate only four parameters in (3.62).

[17] One should however be aware that gains in efficiency are eventually obtained at the cost of an increased bias. One should therefore not exaggerate in the number of instruments. That is the ultimate reason why we do not recommend including lags of higher order in the set (3.64), despite the fact that they are valid instruments to be used in the G.N.R.'s.

first stage would simply be the reduced set in (3.66) whose validity could still be tested.[18]

Finally, one can make more explicit the lack of independence, in either case, between the estimation of β and ρ, which therefore justifies the necessity of including $\dot{u}_{-1}$ in the G.N.R. version to be estimated.

As we have explained earlier (in the context of the dynamic adjustment model), when lagged dependent variables are the only offending regressors, the orthogonality conditions ((3.25) and (3.36)) are not satisfied and therefore inference must not be based upon the R.S.G.N.R.'s. When I.V. estimation of the G.N.R.'s is also required, what is also at stake is the analog I.V. orthogonality condition

$$\operatorname*{Plim}_{n\to\infty}\left(n^{-1}X^T P_W \dot{u}_{-1}\right)=0, \tag{3.67}$$

or simply

$$\operatorname*{Plim}_{n\to\infty}\left(n^{-1}\, W^T \dot{u}_{-1}\right)=0. \tag{3.68}$$

It is easily seen that condition (3.68) will not be satisfied if y_2 is also included in *X* because *W* (in 3.64) would then comprise the lags $y_{1,-1}$, $y_{1,-2}$ and $y_{2,-1}$. Neither

[18] In this case, as we have shown, also the regressand of the G.N.R.'s could simply be $y_t-\dot{\rho}\, y_{t-1}$.

would it be satisfied if y_2 was the only offending regressor in *X* as $y_{1,-1}$ and $y_{2,-1}$ would still remain in *W*.

3.4 Conclusions

The discussion conducted in this chapter definitely highlights the advantage of using the G.N.R. approach, introduced in chapter 2, to make valid inference about a model like (3.1), wherever either lagged dependent variables or current endogenous variables are implicitly considered to be contained in *X*.

In fact, as we have shown:

1. If X only comprises exogenous variables, asymptotic inference based upon either the G.N.R., its simplified version or its restricted simplified version (evaluated either at N.L.S. or at any other consistent estimates) is perfectly valid.

2. If X contains either lagged dependent or current endogenous variables, then the restricted simplified version of the G.N.R. (evaluated either at N.L.S., or N.L.I.V., or at any other consistent estimates) should not be used for inference because the estimation of β and ρ will not be independent:

2.1. In the presence of lagged dependent variables, those variables will be correlated with u_{-1}, violating the necessary orthogonality condition for independence;

2.2. In the presence of current endogenous variables, the G.N.R. must be estimated by I.V. and some of the instruments recommend (on the grounds of efficiency) will be correlated with u_{-1}, violating the necessary orthogonality condition for independence;

3. On the other hand, if X contains current endogenous variables and we apply a two-stage L.S. procedure, not even the simplified (non restricted) version of the G.N.R. will be valid for inference because w^2 will not be estimated consistently: the residuals produced by this version will not be I.V. residuals;

4. Nevertheless, it is valid to use the simplified (non restricted) version of the G.N.R. if we use I.V. directly, but then we cannot perform any test basing the statistic of the test on the *E.S.S.* from that regression. This issue deserves further attention in the following chapter.

CHAPTER 4

NESTED TESTING PROCEDURES

4.1 Introduction

In the previous chapter we have been assuming that the original model exhibits nonspherical disturbances which follow an AR(1) process. This is such a fundamental assumption that underlies our previous discussion of appropriate nonlinear estimation methods.

Those methods have also been adapted in response to other assumptions. We have in fact dealt with the problem of the model containing lagged dependent variables or any current endogenous variable so that an I.V. approach is recommended in the former case and strictly necessary in the latter case.

Strictly speaking, propositions about serial correlation and/or the presence of lagged dependent variables as regressors and/or correlation of regressors with disturbance terms should be tested in advance. Otherwise, one may end up making misleading inference, namely if it is based on a seriously misspecified model.

It is well known that not rejecting the null hypothesis of serial correlation can be and usually is interpreted as evidence of misspecification of the main part of the model. Using an estimation process appropriate to serially correlated disturbances, which implies transforming the model first, is not always the right way to proceed.

Apparent serial correlation might happen if a current exogenous or endogenous variable that was itself serially correlated or any lagged variables (either dependent, exogenous or endogenous) were incorrectly excluded from the regression function.

In fact, dependence between economic variables is rarely fully worked through within the unit time period. Respecifying the model by including the incorrectly excluded lagged variables should account for that.

However, in practice, not all of those incorrectly excluded variables will be recognised as such and it is sometimes unavoidable to have lagged effects spilling over to the disturbance term, which acts as a summary representation of what is still missing.

Fortunately there is a family of tests, the so called 'Common Factor Restrictions Tests', which can be used to detect misspecification in models with apparent autoregressive errors.

4.2 Testing for Serial Correlation

As explained in chapter 2 (section 2.3), any restrictions on the parameters of a nonlinear regression function can be tested by estimating the corresponding G.N.R. evaluated at restricted estimates, provided the estimation methods involved produce consistent root-n estimates under the null hypothesis under testing.

Take then the regression function given by (3.4)

$$x_t'(\beta,\rho) = x_t\ \beta + \rho\ (y_{t-1} - x_{t-1}\ \beta\ )\ ,$$

with corresponding G.N.R.

$$y_t - \rho\, y_{t-1} - \left(x_t - \rho\, x_{t-1}\right)\beta = \left(x_t - \rho\, x_{t-1}\right)b + r\left(y_{t-1} - x_{t-1}\beta\right) + error\ term\,. \qquad (4.1)$$

Consider $\left(\dot{\beta}_R, 0\right)$ any consistent root-n restricted estimate of the parameter vector $\left(\beta,\rho\right)$ under $H_0\colon \rho = 0$.

Evaluated at these estimates, the G.N.R. becomes

$$y_t - x_t\,\dot{\beta}_R = x_t b + r\left(y_{t-1} - x_{t-1}\,\dot{\beta}_R\right) + error\ term\,. \qquad (4.2)$$

Under the null, however, the regression function itself is simply given by

$$x_t'(\beta,\rho) = x_t'(\beta,0) = x_t\beta \tag{4.3}$$

with corresponding G.N.R. evaluated at $\overset{*}{\beta}_R$

$$y_t - x_t\overset{*}{\beta}_R = x_t b + error\ term\,. \tag{4.4}$$

Hence, to test the restriction $\rho = 0$ one simply has to test the significance of the extra term $\overset{*}{u}_{-1R}$ with typical element $y_{t-1} - x_{t-1}\overset{*}{\beta}_R$ in (4.2).

If X does not contain any current endogenous variable, G.N.R. (4.2) (also G.N.R. (4.4)) can simply be estimated by O.L.S. (as one concluded in section 3.3.1). On the other hand, the restricted estimation of β can also be obtained by O.L.S. estimation of model (3.1) regardless the fact that X may contain lagged dependent variables since, under the null, $u_t = \xi_t$.

Consider now the *E.S.S.* of regression (4.2) which is given by

$$\overset{*}{u}_R{}^T P_X \overset{*}{u}_R \ + \ \overset{*}{u}_R{}^T M_X \overset{*}{u}_{-1R}\left(\overset{*}{u}_{-1R}{}^T M_X \overset{*}{u}_{-1R}\right)^{-1} \overset{*}{u}_{-1R}{}^T M_X \overset{*}{u}_R \ ^{1} \tag{4.5}$$

[1] Deduced in Annex 1 (also cf.(2.93)).

where $\dot{u}_R$ has typical element $y_t - x_t \dot{\beta}_R$, and the matrices P_X and M_X have the usual definitions.

Provided one has used $t = 2, ..., N$ to initially estimate β as well as to run (4.2) (or alternatively, one has used $t = 1, ..., N$ to estimate β as well as to run (4.2) setting then $\dot{u}_{t-1,R} = 0$ (t=1)) the regressand $\dot{u}_R$ will be orthogonal to X.[2] In these circumstances,

$$\dot{u}_R^T M_X \dot{u}_{-1R} \left(\dot{u}_{-1R}^T M_X \dot{u}_{-1R} \right)^{-1} \dot{u}_{-1R}^T M_X \dot{u}_R \tag{4.6}$$

represents itself the *E.S.S.* of regression (4.2) since the first term in (4.5) would be equal to zero. Expression (4.6) multiplied by (n-k-1) and divided by the *S.S.R.* taken from regression (4.2) is easily seen to be the square of the t-statistic to test the significance of the extra term $\dot{u}_{-1R}$.[3] Then, the $(n-k)R_u^2$ statistic[4] taken from the same regression which would be given by

$$(n-k)\frac{\dot{u}_R^T M_X \dot{u}_{-1R} \left(\dot{u}_{-1R}^T M_X \dot{u}_{-1R} \right)^{-1} \dot{u}_{-1R}^T M_X \dot{u}_R}{\dot{u}_R^T \dot{u}_R} \tag{4.7}$$

---

[2] For the rest of this work, unless otherwise explicitly assumed, we will denote by n the number of observations effectively used on the estimation process.

[3] See Annex 1.

[4] $R_u^2 = R^2$ if the restricted model contains a contains term (see chapter 2, section 2.3.1).

and is distributed as $\chi^2(1)$ is asymptotically equivalent to the pseudo-t test, both based on the L.M. principle.

Obviously the $(n-k)R_u^2$ statistic uses the *T.S.S.* of regression (4.2) to estimate w^2 whereas the pseudo-t statistic uses the O.L.S. *S.S.R.*. Clearly, the *T.S.S.* divided by $n-k$ will provide a consistent estimate of w^2.[5] It is however also valid to use the *S.S.R.* divided by $n-k-1$ because the correction for the degrees of freedom is negligible asymptotically and as X is orthogonal to $\overset{*}{u}_R$, the G.N.R. (4.2) should have no explanatory power (asymptotically) under the null $\rho = 0$. In other words, both statistics use the same consistent estimate of w^2.

On the contrary, if we have used the entire sample to obtain the restricted estimate of β and only the last N-1 observations to estimate (4.2), the orthogonality condition between $\overset{*}{u}_R$ and X would not be satisfied for finite samples, though it would be satisfied asymptotically.[6] For finite samples, expression (4.6) should now rather be interpreted as the contribution of the extra term for the *E.S.S.* of regression (4.2), or alternatively, as the reduction in the *S.S.R.* brought about by the inclusion of $\overset{*}{u}_{-1R}$ in the regression. Therefore, a pseudo-t test now based on the

---

[5] Note that this is s_R^2, the restricted estimate of w^2: the restricted model has k parameters and n denotes the number of observations effectively used (cf. footnote 2 in this section).

[6] Asymptotically does not matter which time period we have used to estimate β.

$C(\alpha)$ principle [7] would still be strictly valid, and should be preferred to the L.M. $(n-k)R_u^2$ version which will tend to reject the null too often in finite samples.[8]

Let us turn now our attention to the I.V. case.

If X does contain any current endogenous variable, G.N.R. (4.2) has to be estimated by I.V. (as one concluded in section 3.3.2). If the instrument set includes, as it should, all the non-redundant columns of the pre-determined variables contained in X, [9] the regressors contained in X_{-1}, the first lag of y and eventually other exogenous variables that we suspect might explain the offending regressors (and also their first lag, as one concluded in section 3.3.3), then the I.V. estimates of b and r that we would obtain running G.N.R. (4.2) by I.V., are exactly the same as the O.L.S estimates that we obtain running by O.L.S. the following G.N.R.

$$y_t - x_t \dot{\beta}_R = \mathring{x}_t b + r\left(y_{t-1} - x_{t-1}\dot{\beta}_R\right) + error\ term \tag{4.8}$$

---

[7] In this case $\dot{\beta}_R$ is viewed as an arbitrary root-n consistent estimate not satisfying the first-order condition $\left(y - X\dot{\beta}_R\right)^T X = 0$.

[8] The $C(\alpha)$ test can also be computed as $(n-k-1)R_u^2$ from (4.2) minus $(n-k-1)R_u^2$ from (4.4) (see subsection 2.3.2.2). Therefore, $(n-k)R_u^2$ from (2.42) will be obviously greater than the value computed by the $C(\alpha)$ test. Similarly, when testing for serial correlation of order p, the statistic $F(p, n-k-p)$, given by $\dfrac{(R.S.S.R.-U.S.S.R.)/p}{U.S.S.R./(n-k-p)}$ should be preferred to the $(n-k)R_u^2(p)$ statistic.

[9] That is, all the exogenous, lagged dependent or other lagged endogenous variables contained in X.

where $\overset{\circ}{x}_t$ denotes row t of matrix $\overset{\circ}{X} = P_W X$.

As far as the restricted estimation of β is concerned, it should also be obtained by I.V. estimation of model (3.1), eventually using the same set of instruments. We say eventually for two reasons: first, there is no need to include lagged dependent or other lagged exogenous variables in the instrument set as at this stage there is no lagged residual to form predicted values for ; secondly, all those variables are nevertheless valid instruments under the null and therefore may be included.

Whatever the choice we make when estimating the initial consistent root-n estimate of β, the I.V. analog version of equation (4.4) can still be written as

$$y_t - x_t \dot{\beta}_R = \overset{\circ}{x}_t b + error\ term \tag{4.9}$$

where the instrument set implicitly considered to form the instrumental variables for the offending regressors is the same as the one considered in equation (4.8). [10]

The O.L.S. *E.S.S.* of regression (4.8) is given by

$$\dot{u}_R^T P_{\overset{\circ}{X}} \dot{u}_R + \dot{u}_R^T M_{\overset{\circ}{X}} \dot{u}_{-1R} \left(\dot{u}_{-1R}^T M_{\overset{\circ}{X}} \dot{u}_{-1R} \right)^{-1} \dot{u}_{-1R}^T M_{\overset{\circ}{X}} \dot{u}_R \,.^{11} \tag{4.10}$$

---

[10] That is the comprehensive set of instruments.

[11] See Annex 1 (also cf. (2.106), where now $X_{1R} = P_W X = \overset{\circ}{X}$ and $X_{2R} = P_W \dot{u}_{-1R} = \dot{u}_{-1R}$.

All the considerations concerning the time periods used for the estimations involved and previously emphasised are also valid in the I.V. case. However, in this case another requirement is needed to ensure that $\overset{\bullet}{u}_R$ is orthogonal to $\overset{\circ}{X}$: one should use the same set of instruments in both initially estimating β and forming the instrumental variables for the offending regressors. In these circumstances,

$$\overset{\bullet}{u}_R{}^T M_{\overset{\circ}{X}} \overset{\bullet}{u}_{-1R}\left(\overset{\bullet}{u}_{-1R}{}^T M_{\overset{\circ}{X}} \overset{\bullet}{u}_{-1R}\right)^{-1} \overset{\bullet}{u}_{-1R}{}^T M_{\overset{\circ}{X}} \overset{\bullet}{u}_R \tag{4.11}$$

represents itself the *E.S.S.* of regression (4.8) since the first term in (4.10) would be equal to zero. Then, the nR_u^2 statistic is asymptotically equivalent to the pseudo-t test, both based on the L.M. principle. [12]

On the contrary, if we have chosen to use different sets of instruments to initially estimate β and subsequently to form the instrumental variables for the offending regressors to be used in the G.N.R., the orthogonality condition between $\overset{\circ}{X}$ and $\overset{\circ}{u}_R$ would not be satisfied for finite samples (cf. (2.102)), even if we had been careful and had used the proper time periods in both estimation processes. Then, one would have rather to think of expression (4.11) as being the contribution of the extra term for the *E.S.S.* of regression (4.8); or alternatively, as in the previous case, as the

[12] As emphasised in subsection 2.3.1, there is now some doubt about the appropriateness of the degrees-of-freedom adjustment since the *S.S.R.* is not the value of the objective function in the I.V. context. On the other hand, for the conditions under which $R_u^2 = R^2$, cf. footnote 12 in that subsection.

reduction in the *S.S.R.* brought about by the inclusion of $\overset{*}{u}_{-1R}$ (now an I.V. residual) in the regression.

However, as we have shown in chapter 2 (subsection 2.3.2.2), a pseudo-t test based on the $C(\alpha)$ principle will still be valid since $\mathring{X}$ will have no asymptotic explanatory power for $\overset{*}{u}_R$ (cf. (2.111)). That is to say, under the null, the G.N.R. will have no asymptotic explanatory power and, therefore, w^2 will still be estimated consistently.

Nevertheless, one might wonder why not to regress (4.2) by an I.V. procedure, using W as matrix of instruments. This would certainly avoid the initial stage of regressing the offending regressors on W. More fundamental than that, the right I.V. residuals would be used to estimate w^2 as the package would substitute $\mathring{X}$ by X when forming the residuals.

There is however a problem doing this, as the *E.S.S.* from an I.V. regression would rather be given by

$$\overset{*}{u}_R{}^T \mathring{X}\left(\mathring{X}^T \mathring{X}\right)^{-1} X^T X\left(\mathring{X}^T \mathring{X}\right)^{-1} \mathring{X}^T \overset{*}{u}_R$$
$$+2\overset{*}{u}_R{}^T \mathring{X}\left(\mathring{X}^T \mathring{X}\right)^{-1} X^T\left(I - X\left(\mathring{X}^T \mathring{X}\right)^{-1} \mathring{X}^T\right)\overset{*}{u}_{-1R}\left(\overset{*}{u}_{-1R}{}^T M_{\mathring{X}}\, \overset{*}{u}_{-1R}\right)^{-1} \overset{*}{u}_{-1R}{}^T M_{\mathring{X}}\, \overset{*}{u}_R$$

$$+\overset{\bullet}{u}_R^T M_{\mathring{X}} \overset{\bullet}{u}_{-1R}\left(\overset{\bullet}{u}_{-1R}^T M_{\mathring{X}} \overset{\bullet}{u}_{-1R}\right)^{-1} \overset{\bullet}{u}_{-1R}^T\left(I-\mathring{X}\left(\mathring{X}^T \mathring{X}\right)^{-1} X^T\right)\left(I-X\left(\mathring{X}^T \mathring{X}\right)^{-1} \mathring{X}^T\right)$$

$$\overset{\bullet}{u}_{-1R}\left(\overset{\bullet}{u}_{-1R}^T M_{\mathring{X}} \overset{\bullet}{u}_{-1R}\right)^{-1} \overset{\bullet}{u}_{-1R}^T M_{\mathring{X}} \overset{\bullet}{u}_R \,. \text{ [13]} \qquad (4.12)$$

Comparing this expression with the O.L.S. *E.S.S.* of G.N.R. (4.8), which is given by (4.10) one may conclude that the two *E.S.S.*'s are in general different, even if the orthogonality condition $\overset{\bullet}{u}_R^T \mathring{X} = 0$ is satisfied. In fact as $X \neq \mathring{X}$,

$$\left(I-\mathring{X}\left(\mathring{X}^T \mathring{X}\right)^{-1} X^T\right)\left(I-X\left(\mathring{X}^T \mathring{X}\right)^{-1} \mathring{X}^T\right), \qquad (4.13)$$

does not collapse to $M_{\mathring{X}}$.

In other words, a $nR_u^2(1)$ test based on the L.M. principle would never be valid if one uses an I.V. package (no matter if one has eventually chosen the same set of instruments to initially estimate β and to run the G.N.R.). [14]

Nevertheless, the *t*-statistic reported by the I.V. package would still be the correct one as it would still be based on the right estimate of *r* and on the right estimate of its variance but also on the right I.V. residuals that estimate w^2 consistently. The pseudo *t*-test based on the $C(\alpha)$ principle is therefore also valid if based on I.V. estimation rather than on a two-stage O.L.S. procedure.

---

[13] See Annex 1.

[14] Neither it would be a $nR_u^2(p)$ test to test for serial correlation of order p (cf. last paragraph of subsection 2.3.1).

Unfortunately, this conclusion cannot be generalised for higher-order serial correlation processes, unless the I.V. package reports not only the *S.S.R.* of the restricted and unrestricted versions but also the transformed *S.S.R.* of both regressions,[15] whose difference should be used to compute the numerator of the pseudo-*F* test. That is to say, for the test to be valid, the pseudo-*F* test should not be based upon the last two terms of expression (4.12), which represent the difference between the *E.S.S.*'s one would obtain running both the G.N.R.'s (4.2) and (4.4) by I.V., using the same set of instruments. Rather it should be based upon the difference between the transformed *E.S.S.*'s that one would obtain including P_W in the middle of the products of the corresponding predicted values. In fact, looking at Annex 1, it is straightforward to conclude that using that correction, expression (A.1.13) would become expression (A.1.8) and therefore expression (A.1.17) would simply be expression (A.1.12). Hence, the difference between the transformed *E.S.S.*'s obtained from I.V. regressions would finally be given by (4.11), which is the difference between the explained *E.S.S.*'s obtained from second-stage O.L.S. regressions.

Unfortunately, very few I.V. regression packages make it easy to compute an *F*-test in this way.[16] This means that calculating the proper test statistic is frequently harder than it should be. Going back and using a two-stage O.L.S. procedure to obtain the right *E.S.S.*'s or, running instead two extra regressions on W using the

[15] That is the I.V. minimised criterion functions.

[16] Although this is not a problem for T.S.P., for example.

I.V. residuals as regressands and taking the *E.S.S.*'s of those regressions, both procedures are alternatives to be considered. Simplicity has however been one of the motivations to use I.V. rather than O.L.S. twice. Simplicity is nevertheless at stake with all these necessary artificial regressions. The simpler way, therefore, to perform the test is either:

i) To use the same set of instruments as well as the proper time periods both to initially estimate β and to run G.N.R.(4.8) by O.L.S. and base the test on the L.M. principle, either using the pseudo-t statistic or the $nR_u^2(1)$ version to test for first-order serial correlation; Then, when testing for higher-order serial correlation the $nR_u^2(p)$ (p being the order considered) or the pseudo-F statistic $F(p,n)$ are still valid and very easy to implement.

ii) To use instead the 'natural' reduced set of instruments to initially estimate β and base the test on the $C(\alpha)$ principle using the pseudo-t statistic to test for first-order serial correlation in the O.L.S. regression (4.8); Then when testing for higher-order serial correlation the pseudo-F statistic $F(p,n)$ is still valid and very easy to implement.

Finally there is still something that we can do to simplify further the test procedure if X does not contain current endogenous variables.

Consider regressing y rather than $\overset{*}{u}_R$ on X and $\overset{*}{u}_{-1R}$. That is, consider instead the S.G.N.R.

$$y_t = x_t b + r\,\overset{*}{u}_{t-1,R} + error\ term\,. \tag{4.14}$$

Comparing this S.G.N.R. with the previous version one may easily conclude that the O.L.S. estimate of r is the same in both, since $M_X X = 0$ provided we have used the same time period to form the residual $\overset{*}{u}_R$. If that critical condition has been fulfilled, $\overset{\circ}{b}_s$ in this version equals $\overset{\circ}{b}$ in the previous version (which is identically zero) plus the value of the initial restricted estimate of β. Hence the O.L.S. residuals of both versions will be exactly the same and given by

$$y_t - x_t \overset{\circ}{b}_s - \overset{\circ}{r}\,\overset{*}{u}_{t-1,R}\,. \tag{4.15}$$

As $\overset{\circ}{r}$ is the same and both regressions would report the same estimate for the standard error, one may easily conclude that a valid pseudo-t test can be conducted by running the simpler regression (4.14) by O.L.S..

Note that even under the null, this G.N.R. will have asymptotically explanatory power as the regressand is far from being orthogonal to X. The test just described is therefore based on the $C(\alpha)$ principle, that is the $(n-k)R_u^2(1)$ version is not valid.

This result, as all the others we have been obtaining in this chapter can be seen, in fact, as specialisations of the general results obtained in chapters 2 and 3.

However, if X contains current endogenous variables, a two-stage O.L.S. procedure applied to the S.G.N.R. version would never produce a consistent estimate of w^2 (as one concluded in subsection 3.3.2). Therefore, regression (4.14) should instead be run by I.V. with all the complications we have already referred to if we want to generalise the test for higher-order serial correlation. In that case we suggest that this procedure should be avoided and we should stick with the previous version of the G.N.R. to implement either a $nR_u^2(p)$ test based on a two-stage O.L.S. procedure (using the same set of instruments as well as the same time period in both stages), or a pseudo-$F(p,n)$ test based on a two-stage O.L.S. procedure (if using the 'natural' reduced set of instruments to initially estimate β).

Before moving to the next section, a final remark. All the discussion until now has been conducted using the term 'serial correlation' rather than 'autoregressive errors'. In Annex 2, we consider a model containing Moving Average (M.A.) disturbances of first-order.

Taking into consideration the transformed model in (A.2.7) it is straightforward to conclude that under the null $\alpha = 0$, and when evaluated at restricted estimates, the G.N.R. given in (A.2.13) would simply be the G.N.R. in (4.2) looking at (A.2.5). Thus, the G.N.R. for testing against the alternative of M.A.(1) errors is identical to the G.N.R. for testing against A.R.(1) errors. In other words, under the null hypothesis of no serial correlation, regression models with A.R.(1) and M.A.(1) errors are locally equivalent alternatives. Strictly speaking, models with A.R.$(p+q)$, models with M.A.$(p+q)$ and models with A.R.M.A.$(p\ ,\ q)$ errors are all locally equivalent alternatives. Therefore, the same G.N.R. which would include $p+q$ lags of $\dot{u}_R$ would be appropriate for testing against any of these processes.

4.3 Testing for Common Factor Restrictions

Return to model (3.3),

$$y_t = x_t\ \beta + \rho\ (y_{t-1} - x_{t-1}\ \beta\ ) + \xi_t \quad ; \quad \xi_t \sim \text{IID}\ (0, w^2)$$

and consider as alternative, model (3.40)

$$y_t = \rho\ y_{t-1} + x_t\ \beta + x_{t-1}\gamma + \xi_t \quad ; \ \xi_t \sim \text{IID}\ (0, w^2)\,.$$

The former can be rewritten as

$$\left(1-\rho L\right)y_t = \left(1-\rho L\right)x_t\ \beta + \xi_t \quad ; \ \xi_t \sim \text{IID}\ (0, w^2) \tag{4.16}$$

where L is the lag operator, and the latter as

$$(1-\rho L)y_t = x_t\,\beta + L\,x_t\,\gamma + \xi_t \quad ; \; \xi_t \sim \text{IID } (0, w^2) \tag{4.17}$$

It is easily seen that model (4.16) is the restricted version of model (4.17) when imposing the nonlinear restrictions

$\gamma = -\rho\beta$. [17]

Consider now the regression function of the alternative model (4.17) (the regression function of model (3.40)),

$$x_t'(\beta, \rho, \gamma) = x_t\,\beta + \rho\, y_{t-1} + x_{t-1}\gamma \tag{4.18}$$

which is a linear function. Nevertheless, the corresponding G.N.R. will be given by

$$(y_t - \rho\, y_{t-1}) - (x_t\,\beta + x_{t-1}\gamma) = x_t\, b + c\, y_{t-1} + x_{t-1} d + \textit{error term}\,. \text{ [18]} \tag{4.19}$$

Evaluating this G.N.R. at the restricted consistent root-n estimates under the null $\left(\dot{\beta}_R, \dot{\rho}_R, -\dot{\rho}_R\,\dot{\beta}_R\right)$, the G.N.R. becomes

---

[17] The fact that $(1-\rho L)$ appears in both sides of equation (4.16) but not in (4.17) explains the name given to the test.

[18] The regressors are simply the derivatives in order to the parameters (as for the nonlinear case).

$$\left(y_t - \dot{\rho}_R\, y_{t-1}\right) - \left(x_t - \dot{\rho}_R\, x_{t-1}\right)\dot{\beta}_R = x_t\, b + c\, y_{t-1} + x_{t-1} d + error\ term\,. \tag{4.20}$$

To test the restrictions we simply have to confront the *S.S.R.* of this G.N.R. with the *S.S.R.* of the following G.N.R.

$$\left(y_t - \dot{\rho}_R\, y_{t-1}\right) - \left(x_t - \dot{\rho}_R\, x_{t-1}\right)\dot{\beta}_R = \left(x_t - \dot{\rho}_R\, x_{t-1}\right)b + r\left(y_{t-1} - x_{t-1}\,\dot{\beta}_R\right) + error\ term \tag{4.21}$$

which is the G.N.R. corresponding to the null model when evaluated at the same estimates.

If X does not contain any current endogenous variable the initial consistent estimates $\dot{\beta}_R$ and $\dot{\rho}_R$ could be obtained either by N.L.S. applied to model (3.3) or, as also described in section 3.3.1, using I.V. estimation to first obtain a consistent estimate of β in regression (3.20) and then using the I.V. residuals in the usual way to estimate ρ.

If we opt for N.L.S., then G.N.R. (4.21) will have no asymptotic explanatory power. In these circumstances all we have to do is to run G.N.R. (4.20) by O.L.S. (with $t = 2, \ldots, N$), considering only the non-redundant regressors and to use the $(n-k-1)R_u^2(l)$ statistic to test their significance. Obviously, the value so obtained will be exactly the same as the $(n-k-1)R_u^2(l)$ one we would obtain extending G.N.R. (4.21) with the non-redundant regressors in (4.20) and running then by O.L.S. (with $t = 2, \ldots, N$) the extended regression.

Using $(n-k-1)R_u^2(l)$ one is implicitly using s_R^2 to estimate w^2. Note that G.N.R.(4.21) has $k+1$ parameters and would be estimated over n observations. The finite-simple adjustment is therefore given by $n-k-1$.

Note also that the number of degrees of freedom (which we have denoted by l) is not equal to the number of non-redundant regressors in (4.20). It is rather given by the difference between that number and $k+1$ (the number of parameters in equation (4.21)). For example consider X containing a constant, two exogenous variables (z_1 and z_2) and the first lag of y. Then, the number of non-redundant regressors in (4.20) would be equal to 7. As in that case (4.21) would contain 5 parameters (4 b's plus 1 r), the number of degrees of freedom would be equal to 2 (l = 2).

If instead we have opted for the second alternative, that is, by initial I.V. estimation of β and subsequent estimation of ρ using the I.V. residuals, then, as we have emphasised in section 3.3.1, the k regressors in $\left(X-\dot{\rho}_R X_{-1}\right)$ would still have some explanatory power in G.N.R. (4.21). In that case, the test should be based on the $C(\alpha)$ principle rather than on the L.M. principle (as above). Therefore, we should run both regressions (4.20) and (4.21) by O.L.S. and compute a pseudo-F test using the statistic

$$\frac{(R.S.S.R.-U.S.S.R.)/l}{U.S.S.R./(n-k-l-1)} \tag{4.22}$$

where *R.S.S.R.* and *U.S.S.R.* are taken from G.N.R. (4.21) and (4.20), respectively.[19]

Finally if *X* does contain any current endogenous variable, the initial estimates $\overset{*}{\beta}_R$ and $\overset{*}{\rho}_R$ must be obtained either by N.L.I.V. applied to model (2.3), or, as also described in subsection 3.3.2, using the same procedure as in the previous simple dynamic case.

If we opt for N.L.I.V. estimation, the following G.N.R.

$$\left(y_t-\tilde{\rho}_R\,y_{t-1}\right)-\left(x_t-\tilde{\rho}_R\,x_{t-1}\right)\tilde{\beta}_R=\left(\overset{\circ}{x}_t-\tilde{\rho}_R\,x_{t-1}\right)b+r\left(y_{t-1}-x_{t-1}\tilde{\beta}_R\right)+error\ term \tag{4.23}$$

will have no asymptotic explanatory power provided we use the same set of instruments to form the instrumental variables for the offending regressors in *X* [20] and the proper time period.[21] Then, we simply have to run the following G.N.R.

$$\left(y_t-\tilde{\rho}_R\,y_{t-1}\right)-\left(x_t-\tilde{\rho}_R\,x_{t-1}\right)\tilde{\beta}_R=\overset{\circ}{x}_t\,b+c\,y_{t-1}+x_{t-1}d+error\ term \tag{4.24}$$

---

[19] The denominator is $n-k-l-1$ because (4.20) is estimated over n observations and has $k+1+l$ parameters corresponding to k b's, c, and l additional parameters.

[20] The choice of the set of instruments has been clarified earlier (in subsection 3.3.3).

[21] By 'proper' we mean the same time period as used to obtain $\tilde{\beta}_R$ and $\tilde{\rho}_R$.

by O.L.S., using the proper time period and only the non-redundant regressors, and then compute the $nR_u^2(l)$ statistic to make a decision about their significance.

On the contrary, if we have not used either the same set of instruments or the same time period in both regressions or if we have opted for an alternative to N.L.I.V. to initially estimate β and ρ, the L.M. test is no longer valid and we will have to base our decision on a pseudo-F test based on the $C(\alpha)$ principle. That is, using the statistic in (4.22), where the denominator $n-k-l-1$ should now be replaced by n and *R.S.S.R.* and *U.S.S.R.* are taken from G.N.R.(4.23) and (4.24), respectively.

Running by I.V. both regressions (4.20) and (4.21) is still an alternative but then one should be careful and to make sure that we use $(T.R.S.S.R - T.U.S.S.R.)$[22] in the numerator of the F-statistic.

Finally, in the introduction to this chapter we have referred to the 'Common Factor Restrictions' as a family of tests. To complete this section there is still another test that should be mentioned.

Consider the following model as alternative to model (3.3)

$$y_t = x_t\,\beta + \rho\, y_{t-1} - \delta\, x_{t-1}\,\beta + \xi_t \quad ; \; \xi_t \sim \text{IID }(0, w^2) \tag{4.25}$$

---

[22] As on footnote 15 in the previous section.

which can be rewritten as

$$(1-\rho L)y_t = (1-\delta L)x_t\ \beta + \xi_t \quad ; \xi_t \sim \text{IID } (0, w^2) \tag{4.26}$$

Now, model (3.3), which is (4.16), it is easily seen to be the restricted version of model (4.26) when imposing the linear restriction $\delta = \rho$.

The regression function of this alternative model is

$$x_t'(\beta, \rho, \delta) = x_t\ \beta + \rho\, y_{t-1} - \delta\, x_{t-1}\beta \tag{4.27}$$

which is nonlinear.

The corresponding G.N.R. will be given by

$$(y_t - \rho\, y_{t-1}) - (x_t - \delta\, x_{t-1})\ \beta = (x_t - \delta\, x_{t-1})\ b + r\, y_{t-1} - d\, x_{t-1}\beta + \textit{error term}. \tag{4.28}$$

When evaluated at the restricted consistent root-n estimates under the null $\left(\dot{\beta}_R, \dot{\rho}_R\right)$, the G.N.R. becomes

$$\left(y_t - \dot{\rho}_R\, y_{t-1}\right) - \left(x_t - \dot{\rho}_R\, x_{t-1}\right)\dot{\beta}_R = \left(x_t - \dot{\rho}_R\, x_{t-1}\right)\, b + r\, y_{t-1} - d\, x_{t-1}\,\dot{\beta}_R + \textit{error term} \tag{4.29}$$

or, after subtracting and adding up the term $r\,x_{t-1}\,\dot{\beta}_R$ in the right hand side of the equation, and collecting terms,

$$\left(y_t - \dot{\rho}_R\, y_{t-1}\right) - \left(x_t - \dot{\rho}_R\, x_{t-1}\right)\dot{\beta}_R = \left(x_t - \dot{\rho}_R\, x_{t-1}\right) b + r\left(y_{t-1} - x_{t-1}\,\dot{\beta}_R\right) + (r-d)x_{t-1}\,\dot{\beta}_R + \textit{error term}. \quad (4.30)$$

To test the restrictions we still could confront the *S.S.R.* of this G.N.R. with the *S.S.R.* of G.N.R. given in (4.21), which is still the G.N.R. corresponding to the null model, when evaluated at the same restricted estimates. In this case, however, that would be equivalent to test the significance of the extra term $(r-d)x_{t-1}\,\dot{\beta}_R$ in the extended regression (4.30).

It is straightforward to conclude that all the previous considerations about the type of test we would use are still valid in this case. The only difference is the dimension of the test, that is the number of degrees of freedom. Now $l = 1$, and therefore the $(n-k-1)R_u^2(l)$ (the $nR_u^2(l)$) becomes either a pseudo-t test or a $(n-k-1)R_u^2(1)$ (a $nR_u^2(1)$) test and the pseudo-F test can be conducted as a pseudo-t test. It is worth noting that the two tests will be equivalent if the regression function is such that l was equal to 1 in the previous case as well. Otherwise, the test with only 1 degree of freedom is testing against a less general alternative. That is, model (4.25) is in general more restrictive than model (3.40).

As either test might perform better than the other, depending on how the data were actually generated, both tests should be implemented.

4.4 Conclusions

Before moving to the next chapter, it will be useful to summarise the main results so far obtained.

As the so called 'Common Factor Restrictions' tests have not been extended to higher-order autoregressive processes,[23] there is no need to include the obtained results in the following summary. Hence, we will restrict ourselves to the results on the serial correlation testing which are in fact as illuminating and of great help for the discussion of nonnested testing.

1. To test for serial correlation of order p, one simply has to test the significance of the extra terms (the first p lags of $\overset{*}{u}_R$) in the maintained G.N.R. (the G.N.R. associated with the alternative model) evaluated at restricted consistent root-n estimates under the null.

2. If X does not contain current endogenous variables and one uses the same time period to restrictly estimate β as well as to run the maintained G.N.R., the tests (of the significance of the extra terms, which will be O.L.S. restricted

---

[23] Two interesting cases might be considered for further research: the AR(2) and a combination of first and fourth-order autoregressive disturbances (like $\left(1-\rho_1 L\right)\left(1-\rho_4 L^4\right) u_t = \xi_t$). Note that this would correspond to include the first, the fourth and also the fifth lag of $\overset{*}{u}_R$ in the analysis.

residuals) can be based on the L.M. principle. Therefore, one may simply use the $(n-k)R_u^2(p)$ statistic from the G.N.R., which is the easiest to compute.

However, taking into consideration the results of Kiviet (1986) (mentioned as a final comment in chapter 2), the pseudo-F test version $\dfrac{U.E.S.S./p}{U.S.S.R./(n-k-p)}$ may be a better choice to avoid rejecting the null too often in finite samples. To see this, not that $(n-k)R_u^2 = \dfrac{U.E.S.S.}{T.S.S./(n-k)}$ (if the restricted model contains a constant term) and, under the null, $T.S.S. = U.S.S.R.$ since the G.N.R. has no explanatory power under the null.

Similarly, when testing for AR (1), the pseudo-$t(n-k-1)$ test associated with the lagged O.L.S. restricted residual should be preferred to the corresponding $(n-k)R_u^2(1)$ version.

3. If X does contain any current endogenous variable and one uses not only the same time period but also the same comprehensive set of instruments to restrictly estimate β as well as to form the instrumental variable(s) for the offending regressor(s) in the maintained G.N.R., the tests (of significance of the extra terms, which will be I.V. restricted residuals) can still be based on the L.M. principle and the statistics above, now computed from the second-stage O.L.S. maintained G.N.R., may still be used (in this context there is

however some doubt about the appropriateness of using finite-sample adjustments).

4. If any of the crucial conditions in 3. is not fulfilled, the tests must be based on the $C(\alpha)$ principle, rather than on the L.M. principle. The R_u^2 version in the maintained G.N.R. is never valid in these circumstances. Instead (after substituting the offending regressors by the appropriate instrumentals variables if X contains any current endogenous variable(s)), one should run by O.L.S. both the maintained and the null G.N.R.'s to compute the numerator of the pseudo-F test (the difference $(R.S.S.R.–U.S.S.R.)$ from those regressions).

   Alternatively, if $p = 1$, then a pseudo-t test on the significance of the lagged restricted residual (a restricted I.V. residual if X contains any current endogenous variable(s)) in the maintained G.N.R. is the easiest to compute.

5. As an alternative, wherever I.V. estimation is required, all the tests described in 3. and 4. can also be conducted using an I.V. package. As we said before, this would avoid unnecessary O.L.S. regressions to form the instrumental variable(s) for the offending regressor(s) at first-stage.

   The t-statistics produced by this procedure would be exactly the same as those obtained from the second-stage O.L.S. regressions. However, as we also stressed before, one should be careful and to make sure that the package also prints the 'transformed' quantities required to be used in the numerator of the R_u^2 and F statistics (when testing for higher-order serial correlation processes). Otherwise, extra O.L.S. auxiliary regressions are needed to compute those quantities.

6. The maintained and the null G.N.R.'s can always be substituted by their simplified versions (that is, the regressand in both regressions can simply be replaced by y), but then the tests must necessarily be based on the $C(\alpha)$ principle. More fundamental than that, if X contains any current endogenous variable(s), then a two-stage L.S. procedure must not be used since w^2 would not be estimated consistently. In other words, the use of the S.G.N.R. implies the use of an I.V. package so that the residuals used to estimate w^2 are (proper) I.V. residuals.

7. Finally, as explained in the final comments on chapter 2, the use of testing procedures based on the $C(\alpha)$ principle is most interesting when testing restrictions in the I.V. context because there is no need to enforce a unique set of instruments both to restrictly estimate β and to run the maintained G.N.R.. The harmful effect of such enforcement is however most relevant when the 'extra' instruments are too many (increased bias) and have little ability to explain the offending regressors (increased inefficiency). This argument against the L.M. tests is therefore much more relevant in the context of nonnested models than in the context of nested ones.

CHAPTER 5

NONNESTED TESTING PROCEDURES

5.1 Introduction

Diagnostic checks as those discussed in the last chapter, or any other nested testing procedures, cannot be expected to detect inappropriate specifications with high probability in every application.

On the other hand, it is very unlikely that one of the models under consideration will be the true model.

It follows that quite often a collection of nonnested econometric models, usually based upon different economic theories will be declared 'data consistent' after a nested testing routine. Then, a sensible thing to do is to use these competing models to provide additional checks of each other's specification. In other words, rather than choosing among 'data consistent' alternatives, nonnested testing is valuable in providing additional specification testing.

There is nothing strange therefore, if the outcome of the nonnested testing, for example between a pair of models, may be either a rejection or a non-rejection of both models. When both are rejected, each model provides evidence that the other is misspecified; when neither is rejected, either the two models are very similar or we need to use a more informative data set to distinguish between them. Last but not the least, if only one is rejected, the other model should be considered as the preferred model.

5.2 The PA Tests

Consider the two alternative models

$$M_1 \colon y_t = x_{1,t}\beta_1 + \rho_1\left(y_{t-1} - x_{1,t-1}\beta_1\right) + \xi_{1,t} \quad ; \quad \xi_{1,t} \sim \text{IID}\left(0, w_1^2\right); \ \left|\rho_1\right| < 1 \tag{5.1}$$

$$M_2 \colon y_t = x_{2,t}\beta_2 + \rho_2\left(y_{t-1} - x_{2,t-1}\beta_2\right) + \xi_{2,t} \quad ; \quad \xi_{2,t} \sim \text{IID}\left(0, w_2^2\right); \ \left|\rho_2\right| < 1 \tag{5.2}$$

where β_1 and β_2 are k_1 and k_2 dimensional vectors, respectively.

These models are said to be nonnested if it is in general not possible to find restrictions on $\left(\beta_1, \rho_1\right)$ such that, for arbitrary $\left(\beta_2, \rho_2\right)$, the regression functions of M_1 and M_2 are exactly the same and vice versa.

Consider also the linear combination of the two models which have been put forward after passing the 'Common Factor Restrictions Tests' described in the previous chapter,

$$M_c\colon y_t = (1-\alpha)\left[x_{1,t}\beta_1 + \rho_1\left(y_{t-1} - x_{1,t-1}\beta_1\right)\right] + \alpha\left[x_{2,t}\beta_2 + \rho_2\left(y_{t-1} - x_{2,t-1}\beta_2\right)\right] + \xi_t\,. \quad (5.3)$$

The regression function of M_c is given by

$$\begin{aligned} x'_{c,t}\left(\beta_1,\beta_2,\rho_1,\rho_2,\alpha\right) = (1-\alpha)&\left[x_{1,t}\beta_1 + \rho_1\left(y_{t-1} - x_{1,t-1}\beta_1\right)\right] \\ +\alpha&\left[x_{2,t}\beta_2 + \rho_2\left(y_{t-1} - x_{2,t-1}\beta_2\right)\right] \end{aligned} \quad (5.4)$$

with corresponding G.N.R.

$$\begin{aligned} y_t - x'_{c,t}\left(\beta_1,\beta_2,\rho_1,\rho_2,\alpha\right) &= (1-\alpha)\left(x_{1,t} - \rho_1 x_{1,t-1}\right)b_1 + \alpha\left(x_{2,t} - \rho_2 x_{2,t-1}\right)b_2 \\ &\quad + (1-\alpha)\left(y_{t-1} - x_{1,t-1}\beta_1\right)r_1 + \alpha\left(y_{t-1} - x_{2,t-1}\beta_2\right)r_2 \\ + a\Big\{x_{2,t}\beta_2 + \rho_2\left(y_{t-1} - x_{2,t-1}\beta_2\right) &- \left[x_{1,t}\beta_1 + \rho_1\left(y_{t-1} - x_{1,t-1}\beta_1\right)\right]\Big\} + error\ term \end{aligned} \quad (5.5)$$

where the regressors have been obtained, as usual, by differentiation of the regression function with respect to the parameters.

Testing $\alpha = 0$ in M_c corresponds to test M_1 (as the null model) against M_2.
Testing $\alpha = 1$ in M_c corresponds to test M_2 (as the null model) against M_1.

Consider the former case.

When evaluated at restricted root-n consistent estimates under the null

$$\left(\beta_1, \beta_2, \rho_1, \rho_2, \alpha\right)=\left(\overset{*}{\beta}_1, \overset{*}{\beta}_2, \overset{*}{\rho}_1, \overset{*}{\rho}_2, 0\right)$$ [1]

the G.N.R. (associated with the comprehensive model) simplifies to

$$\begin{aligned} & y_t-\overset{*}{\rho}_1 y_{t-1}-\left(x_{1,t}-\overset{*}{\rho}_1 x_{1,t-1}\right) \overset{*}{\beta}_1=\left(x_{1,t}-\overset{*}{\rho}_1 x_{1,t-1}\right) b_1+r_1\left(y_{t-1}-x_{1,t-1} \overset{*}{\beta}_1\right) \\ & +a\left\{x_{2,t} \overset{*}{\beta}_2+\overset{*}{\rho}_2\left(y_{t-1}-x_{2,t-1} \overset{*}{\beta}_2\right)-\left[x_{1,t} \overset{*}{\beta}_1+\overset{*}{\rho}_1\left(y_{t-1}-x_{1,t-1} \overset{*}{\beta}_1\right)\right]\right\}+\textit{error term} \end{aligned} \tag{5.6}$$

On the other hand, the G.N.R. associated with the null model will simply be given by

$$y_t-\overset{*}{\rho}_1 y_{t-1}-\left(x_{1,t}-\overset{*}{\rho}_1 x_{1,t-1}\right) \overset{*}{\beta}_1=\left(x_{1,t}-\overset{*}{\rho}_1 x_{1,t-1}\right) b_1+r_1\left(y_{t-1}-x_{1,t-1} \overset{*}{\beta}_1\right)+\textit{error term}\,. \tag{5.7}$$

Therefore, to test $\alpha=0$ in the comprehensive model, one simply has to test $a=0$ in G.N.R.(5.6).

[1] We deliberately drop the subscript R, which stands for restricted estimation, to avoid awkward notation.

Now, as we said before, we want to evaluate G.N.R.(5.6) at restricted consistent root-n estimates under the null. As under the null M_c collapses to M_1, any root-n consistent estimates $\overset{*}{\beta}_1$ and $\overset{*}{\rho}_1$ that we may obtain to substitute for the parameters in M_1 can be used in the G.N.R..

M_1 does not contain any current endogenous explanatory variable

If $x_{1,t}$ does not contain current endogenous variables not only the parameters in M_1 but also the G.N.R.(5.6) could simply be estimated by L.S.: eventually N.L.S. to estimate M_1 if we do not consider initial root-n consistent estimates of β_1 and ρ_1 to be used in the G.N.R..

Then, for example the term

$$x'_{1,t}\left(\hat{\beta}_1, \hat{\rho}_1\right) = x_{1,t}\,\hat{\beta}_1 + \hat{\rho}_1\left(y_{t-1} - x_{1,t-1}\,\hat{\beta}_1\right) \tag{5.8}$$

would simply be the N.L.S. predicted value in period t given by M_1.

The problem is how should we estimate β_2 and ρ_2 (consistently), taking into consideration that M_1 is the null model.

Provided $x_{2,t}$ does not contain any current endogenous variable either, one argument for the estimation of β_2 and ρ_2 will be to minimise the squared differences between the predicted values of the two models, taking the predicted values of M_1 as given.

To achieve that, we simply have to obtain consistent estimates of β_2 and ρ_2 so that they minimise the criterion function

$$\left(x_1'\left(\hat{\beta}_1,\hat{\rho}_1\right)-x_2'\left(\beta_2,\rho_2\right)\right)^T\left(x_1'\left(\hat{\beta}_1,\hat{\rho}_1\right)-x_2'\left(\beta_2,\rho_2\right)\right). \tag{5.9}$$

That is to say, we simply have to estimate by N.L.S the auxiliary regression

$$x_{1,t}'\left(\hat{\beta}_1,\hat{\rho}_1\right)=x_{2,t}\beta_2+\rho_2\left(y_{t-1}-x_{2,t-1}\beta_2\right)+error\ term \tag{5.10}$$

which is equivalent to the substitution of the actual value of y_t in M_2 by the N.L.S. predicted value of y_t given by M_1.

Obviously, the *S.S.R.* of this auxiliary regression will be the minimal value of the criterion function in (5.9) which can be represented by

$$\left(x_1'\left(\hat{\beta}_1,\hat{\rho}_1\right)-x_2'\left(\hat{\beta}_{2/1},\hat{\rho}_{2/1}\right)\right)^T\left(x_1'\left(\hat{\beta}_1,\hat{\rho}_1\right)-x_2'\left(\hat{\beta}_{2/1},\hat{\rho}_{2/1}\right)\right) \tag{5.11}$$

that is the minimal sum-of-squares of the differences between the predicted values given by the two models.

If either of the models contain any current endogenous explanatory variable, this desirable principle of minimising a function of the differences between the predicted values of the two models in direction of M_1 is still valid.

For example, if M_2 is the only to be estimated by N.L.I.V., the auxiliary regression in (5.10) should rather be estimated by N.L.I.V. to obtain consistent estimates of β_2 and ρ_2.

After forming W_2 for M_2 in the usual way (as in section 3.3.3), regressing (5.10) by N.L.I.V. can be reinterpreted as regressing by N.L.S.

$$x'_{1,t}\left(\hat{\beta}_1, \hat{\rho}_1\right) = \overset{\circ}{x}_{2,t}\beta_2 + \rho_2\left(y_{t-1} - x_{2,t-1}\beta_2\right) + error\ term \qquad (5.12)$$

where only the offending regressors in $x_{2,t}$ have been substituted by instrumental variables.

Note that this two-stage L.S. procedure will minimise

$$\left(x_1'\left(\hat{\beta}_1,\hat{\rho}_1\right)-\mathring{x}_2'\left(\beta_2,\rho_2\right)\right)^T\left(x_1'\left(\hat{\beta}_1,\hat{\rho}_1\right)-\mathring{x}_2'\left(\beta_2,\rho_2\right)\right) \tag{5.13}$$

rather than the criterion function in (5.9).

Alternatively, the same I.V. estimates $\tilde{\beta}_{2/1}$ and $\tilde{\rho}_{2/1}$ can be obtained by minimising the criterion function

$$\left(x_1'\left(\hat{\beta}_1,\hat{\rho}_1\right)-x_2'\left(\beta_2,\rho_2\right)\right)^T P_{W_2}\left(x_1'\left(\hat{\beta}_1,\hat{\rho}_1\right)-x_2'\left(\beta_2,\rho_2\right)\right). \tag{5.14}$$

which corresponds to run (directly) by I.V. the auxiliary regression in (5.10) using W_2 as the set of instruments.

That is to say, as it is critical to obtain consistent estimates of all the parameters involved, a weighted sum-of-squares (rather than an 'ordinary' sum-of-squares) of the differences between the predicted values given by the two models must be considered in this case.

Finally, replacing β_2 and ρ_2 in (5.14) by the consistent estimates $\tilde{\beta}_{2/1}$ and $\tilde{\rho}_{2/1}$ one obtains

$$\left(x_1'\left(\hat{\beta}_1,\hat{\rho}_1\right)-x_2'\left(\tilde{\beta}_{2/1},\tilde{\rho}_{2/1}\right)\right)^T P_{W_2}\left(x_1'\left(\hat{\beta}_1,\hat{\rho}_1\right)-x_2'\left(\tilde{\beta}_{2/1},\tilde{\rho}_{2/1}\right)\right) \tag{5.15}$$

which is simply the $T.S.S.R.$[2] that one would obtain running by I.V. the auxiliary regression in (5.10).

Note also that there is no reason to consider including in W_2 any of the non-redundant columns of M_1 since for the moment we are assuming that $x_{2,t}$ and $x_{1,t}$ do not have any current endogenous variable in common.[3]

Now, provided M_1 has been estimated by N.L.S. using the same time period as the one used to run the G.N.R., the orthogonality condition between the regressand and each of the regressors in G.N.R.(5.6) (in matrix form) will be satisfied, with one exception: the extra regressor

$$X_2 \dot{\beta}_{2/1} + \dot{\rho}_{2/1}\left(y_{-1} - X_{2,-1}\dot{\beta}_{2/1}\right) - \left[X_1 \hat{\beta}_1 + \hat{\rho}_1\left(y_{-1} - X_{1,-1}\hat{\beta}_1\right)\right]^4 \qquad (5.16)$$

where $\dot{\beta}_{2/1}$ and $\dot{\rho}_{2/1}$ denote either N.L.S. or N.L.I.V. estimates from regression (5.10), as appropriate. [5]

---

[2] $T.S.S.R.$ stands for 'Transformed' $S.S.R.$.

[3] Actually, we are assuming that $x_{1,t}$ does not contain any current endogenous variable at all.

[4] When evaluated at N.L.S. estimates, the first two regressors will be orthogonal to the regressand because $\hat{\beta}_1$ and $\hat{\rho}_1$ satisfy the N.L.S. first-order conditions. Also, when evaluated at N.L.S. estimates, the last term in (5.16), which is the N.L.S. predicted values of M_1, it is orthogonal to the N.L.S. residuals; that, however, does not mean that this last term can simply be dropped from the G.N.R. since that would affect (the estimation of) the variance of the estimate of a.

[5] Or any other consistent root-n estimates counterparts (like, for example, the appropriate O.S. estimates of the parameters of that regression).

Therefore, this extra regressor simply represents either the N.L.S. or the N.L.I.V. residual vector from the auxiliary regression (5.10) (apart from the sign).

In these circumstances, as we learned from previous chapters, the test of the significance of this residual vector (the extra regressor) in G.N.R.(5.6) can be based on the L.M. principle. That is to say, either the $(n-k_1-1)R_u^2(1)$ statistic or the pseudo-$t(n-k_1-2)$ statistic from the G.N.R. can be used to perform the test.

On the contrary, if either O.S. estimation has been used to estimate M_1 or we do not have used the same time period to both estimate M_1 and to run the G.N.R., then only the pseudo-t statistic will be valid since, as the G.N.R. associated with the null model (G.N.R.(5.7)) will still have some explanatory power in general, the test must be based on the $C(\alpha)$ principle.

Whichever the case, as we also have learned from previous chapters, a S.G.N.R. version can be used instead to perform the test. Replacing the regressand simply by $y_t - \hat{\rho}_1\, y_{t-1}$ one would still obtain the same estimates of r_1 and a and also the same O.L.S. residuals (cf. for example equations (3.8) and (3.11) and also equations (3.14), (3.15) and (3.19)). As the O.L.S. estimate of a is the same and both the G.N.R. and its simplified version would report the same estimate for the standard error, a valid pseudo-t test can be conducted by running the S.G.N.R. by O.L.S..

M_1 is the only containing current endogenous explanatory variables

If $x_{1,t}$ is the only containing any current endogenous variable, then not only the parameters in M_1 but also the G.N.R. (5.6) must be estimated by I.V.: eventually N.L.I.V. to estimate M_1 if we do not consider initial root-n consistent estimates of β_1 and ρ_1 to be used in the G.N.R..

In that case all the non-redundant columns of the non offending regressors in M_1 should be included in W_1 not only to run the G.N.R. but also to estimate M_1. [6]

Moreover, the non-redundant columns of M_2 should also be considered to extend the instrument set used to run the G.N.R. (say, the instrument set W_{1E}). [7]

Then, if one uses not only the same time period but also the same extended set of instruments W_{1E} to reestimate M_1 as well as to form the instrumental variable(s) for the offending regressor(s) in the G.N.R., the orthogonality condition between the regressand and each of the regressors in the following G.N.R.

---

[6] It is true that if we opt instead for initial estimation of β_1 and ρ_1, some of those instruments will not be valid instruments when obtaining the I.V. residuals at that stage (as one concluded in section 3.3.3). They are, nevertheless, valid instruments both to estimate M_1 and to run the G.N.R..

[7] Where $W_{1E} = W_1 \cup X_2 \cup X_{2,-1}$.

$$y_t - \tilde{\rho}_1 y_{t-1} - \left(x_{1,t} - \tilde{\rho}_1 x_{1,t-1}\right)\tilde{\beta}_1 = \left(\mathring{x}_{1,t} - \tilde{\rho}_1 x_{1,t-1}\right)b_1 + r_1\left(y_{t-1} - x_{1,t-1}\tilde{\beta}_1\right)$$
$$+ a\left\{x_{2,t}\hat{\beta}_{2/1} + \hat{\rho}_{2/1}\left(y_{t-1} - x_{2,t-1}\hat{\beta}_{2/1}\right) - \left[\mathring{x}_{1,t}\tilde{\beta}_1 + \tilde{\rho}_1\left(y_{t-1} - x_{1,t-1}\tilde{\beta}_1\right)\right]\right\} + error\ term \tag{5.17}$$

will be satisfied, with one exception: the extra regressor

$$X_2\hat{\beta}_{2/1} + \hat{\rho}_{2/1}\left(y_{-1} - X_{2,-1}\hat{\beta}_{2/1}\right) - \left[\mathring{X}_1\tilde{\beta}_1 + \tilde{\rho}_1\left(y_{-1} - X_{1,-1}\tilde{\beta}_1\right)\right] \tag{5.18}$$ [8]

where $\hat{\beta}_{2/1}$ and $\hat{\rho}_{2/1}$ still denote the N.L.S. estimates of β_2 and ρ_2 but now from the auxiliary regression

$$\mathring{x}'_{1,t}\left(\tilde{\beta}_1, \tilde{\rho}_1\right) = x_{2,t}\beta_2 + \rho_2\left(y_{t-1} - x_{2,t-1}\beta_2\right) + error\ term \tag{5.19}$$

which is equivalent to the substitution of the actual value of y_t in M_2 by the two-stage N.L.S. predicted value of y_t given by M_1.

That is to say, the N.L.S. estimates of β_2 and ρ_2 minimise now the criterion function

---

[8] When evaluated at N.L.I.V. estimates, the first two regressors will be orthogonal to the regressand because $\tilde{\beta}_1$ and $\tilde{\rho}_1$ satisfy the N.L.I.V. first-order conditions. Also, when evaluated at N.L.I.V. estimates, the last term in (5.18), which is the two-stages N.L.S. predicted values of M_1, it is orthogonal to the N.L.I.V. residuals; that, however, does nor mean that this last term can simply be dropped from the G.N.R. since that would affect (the estimation of) the variance of the estimate of a.

$$\left(\overset{\circ}{x}_1'\left(\tilde{\beta}_1, \tilde{\rho}_1\right) - x_2'(\beta_2, \rho_2)\right)^T \left(\overset{\circ}{x}_1'\left(\tilde{\beta}_1, \tilde{\rho}_1\right) - x_2'(\beta_2, \rho_2)\right) \tag{5.20}$$

whose minimal value can be represented by

$$\left(\overset{\circ}{x}_1'\left(\tilde{\beta}_1, \tilde{\rho}_1\right) - x_2'\left(\hat{\beta}_{2/1}, \hat{\rho}_{2/1}\right)\right)^T \left(\overset{\circ}{x}_1'\left(\tilde{\beta}_1, \tilde{\rho}_1\right) - x_2'\left(\hat{\beta}_{2/1}, \hat{\rho}_{2/1}\right)\right) \tag{5.21}$$

The *S.S.R.* of the auxiliary regression in (5.19).

Under the circumstances described above, once again the test of the significance of the extra regressor, once again a N.L.S. residual vector (apart from the sign) but now from the auxiliary regression (5.19), can be based on the L.M. principle. Either the statistic $nR_u^2(1)$ or the pseudo-$t(n)$ statistic may be used to perform the test. Moreover, there is now an option between obtaining these statistics either from G.N.R.(5.17) or (directly) from the following G.N.R.

$$\begin{aligned} y_t - \tilde{\rho}_1 y_{t-1} - \left(x_{1,t} - \tilde{\rho}_1 x_{1,t-1}\right)\tilde{\beta}_1 &= \left(x_{1,t} - \tilde{\rho}_1 x_{1,t-1}\right) b_1 + r_1\left(y_{t-1} - x_{1,t-1}\tilde{\beta}_1\right) \\ &+ a\left\{x_{2,t}\hat{\beta}_{2/1} + \hat{\rho}_{2/1}\left(y_{t-1} - x_{2,t-1}\hat{\beta}_{2/1}\right) - \left[x_{1,t}\tilde{\beta}_1 + \tilde{\rho}_1\left(y_{t-1} - x_{1,t-1}\tilde{\beta}_1\right)\right]\right\} + \mathit{error\ term} \end{aligned} \tag{5.22}$$

to be run by I.V. using the 'proper' time period and the extended set of instrument W_{1E} mentioned above. In this case, the extra regressor should be reinterpreted as the N.L.S. residual vector (apart from the sign) from the auxiliary regression

$$x'_{1,t}\left(\tilde{\beta}_1, \tilde{\rho}_1\right) = x_{2,t}\beta_2 + \rho_2\left(y_{t-1} - x_{2,t-1}\beta_2\right) + error\ term \tag{5.23}$$

which is equivalent to the substitution of the actual value of y_t in M_2 by the N.L.I.V. predicted value of y_t given by M_1. [9]

However, if we opt for the R_u^2 test based on this second alternative, we must be careful and to make sure that the package prints the correct R_u^2, that is, the one that makes use of the *T.E.S.S.* (see conclusion 5 in the previous chapter).

On the contrary, still using the extended set of instrument to run the G.N.R. but using W_1 to estimate M_1, will certainly have implications on the choice of the type of test to perform when testing $a = 0$. In other words, the test must now be based on the $C(\alpha)$ principle and therefore the R_u^2 version is not a valid alternative, no matter the choice between the G.N.R.'s (5.17) and (5.22) mentioned above (see conclusion 4 in the previous chapter).

Whichever the case, as we have learned from previous chapters, as long as one opts for direct I.V. estimation, the S.G.N.R. version can be used instead to perform the pseudo-*t* test based on the $C(\alpha)$ principle. Replacing the regressand simply by

---

[9] Note that we are assuming that the non-redundant columns of M_2 have been taken into consideration to extend the set of instruments. Therefore, both auxiliary regressions (5.19) and (5.23) will produce the same estimates $\hat{\beta}_{2/1}$ and $\hat{\rho}_{2/1}$.

$y_t - \tilde{\rho}_1 y_{t-1}$ one would still obtain the same estimates of r_1 and a and also the same I.V. residuals (cf. for example equations (3.51), (3.55), (3.58) and (3.59)) which will estimate w^2 consistently.

Both M_1 and M_2 contain current endogenous explanatory variables

If both $x_{1,t}$ and $x_{2,t}$ contain current endogenous variables all the comments on the preceding two cases are still valid and should be combined adequately.

In this case, however, the extended instrument set used to run the G.N.R. should include all the non-redundant columns of the non-offending regressors either in M_1 or in M_2 (say, the instrument set W). [10] The G.N.R to be run by O.L.S. should now be rewritten as

$$\begin{aligned} y_t - \tilde{\rho}_1 y_{t-1} - \left(x_{1,t} - \tilde{\rho}_1 x_{1,t-1}\right)\tilde{\beta}_1 &= \left(\mathring{x}_{1,t} - \tilde{\rho}_1 x_{1,t-1}\right)b_1 + r_1\left(y_{t-1} - x_{1,t-1}\tilde{\beta}_1\right) \\ &+ a\left\{\mathring{x}_{2,t}\tilde{\beta}_{2/1} + \tilde{\rho}_{2/1}\left(y_{t-1} - x_{2,t-1}\tilde{\beta}_{2/1}\right) - \left[\mathring{x}_{1,t}\tilde{\beta}_1 + \tilde{\rho}_1\left(y_{t-1} - x_{1,t-1}\tilde{\beta}_1\right)\right]\right\} + error\ term \end{aligned} \tag{5.24}$$

where the extra regressor

$$\mathring{X}_2\tilde{\beta}_{2/1} + \tilde{\rho}_{2/1}\left(y_{-1} - X_{2,-1}\tilde{\beta}_{2/1}\right) - \left[\mathring{X}_1\tilde{\beta}_1 + \tilde{\rho}_1\left(y_{-1} - X_{1,-1}\tilde{\beta}_1\right)\right] \tag{5.25}$$

---

[10] Where $W = W_1 \cup W_2$.

can be obtained as the N.L.S. residual vector (apart from the sign) from the auxiliary regression

$$\overset{\circ}{x}{}'_{1,t}\left(\tilde{\beta}_1, \tilde{\rho}_1\right) = \overset{\circ}{x}_{2,t}\,\beta_2 + \rho_2\left(y_{t-1} - x_{2,t-1}\beta_2\right) + error\ term\,. \tag{5.26}$$

In this case, the N.L.I.V estimates of β_2 and ρ_2 minimise the criterion function

$$\left(x_1'\left(\tilde{\beta}_1, \tilde{\rho}_1\right) - x_2'\left(\beta_2, \rho_2\right)\right)^T P_W \left(x_1'\left(\tilde{\beta}_1, \tilde{\rho}_1\right) - x_2'\left(\beta_2, \rho_2\right)\right) \tag{5.27}$$

whose minimal value can be represented by

$$\left(x_1'\left(\tilde{\beta}_1, \tilde{\rho}_1\right) - x_2'\left(\tilde{\beta}_{2/1}, \tilde{\rho}_{2/1}\right)\right)^T P_W \left(x_1'\left(\tilde{\beta}_1, \tilde{\rho}_1\right) - x_2'\left(\tilde{\beta}_{2/1}, \tilde{\rho}_{2/1}\right)\right) \tag{5.28}$$

the *S.S.R.* of the auxiliary regression in (5.26).

Alternatively, the G.N.R. to be run (directly) by I.V. should now be rewritten as

$$\begin{aligned} y_t - \tilde{\rho}_1\, y_{t-1} - \left(x_{1,t} - \tilde{\rho}_1\, x_{1,t-1}\right)\tilde{\beta}_1 &= \left(x_{1,t} - \tilde{\rho}_1\, x_{1,t-1}\right)b_1 + r_1\left(y_{t-1} - x_{1,t-1}\,\tilde{\beta}_1\right) \\ + a\left\{x_{2,t}\,\tilde{\beta}_{2/1} + \tilde{\rho}_{2/1}\left(y_{t-1} - x_{2,t-1}\,\tilde{\beta}_{2/1}\right) - \left[x_{1,t}\,\tilde{\beta}_1 + \tilde{\rho}_1\left(y_{t-1} - x_{1,t-1}\,\tilde{\beta}_1\right)\right]\right\} &+ error\ term \end{aligned} \tag{5.29}$$

where the extra regressor should be interpreted as the N.L.S. residual vector (apart from the sign) from the auxiliary regression in (5.23).

Then, if one uses not only the same time period but also the same extended set of instruments W to reestimate M_1 as well as to form the instrumental variables for the offending regressors in the G.N.R.(5.24), the orthogonality condition between the regressand and each of the regressors will be satisfied with the exception of the extra regressor under testing. The test of the significance of the extra regressor can therefore be based on the L.M. principle. The statistics to be used are exactly the same as in the previous case and, in a similar fashion, one might also prefer instead to run by I.V. the G.N.R.(5.29) to conduct the test.

Also, as in the previous case, still using the extended set of instruments W to run the G.N.R. but using W_1 to estimate M_1 will have the same implication: the test must be based on the $C(\alpha)$ principle and therefore the R_u^2 version is not a valid statistic.

Finally, according to our previous comments on the interest of $C(\alpha)$ tests in the context of I.V. estimation (see conclusion 7. in the previous chapter), it will not make much sense to reestimate M_1 using the extended set of instruments when both models do not have the same current endogenous explanatory variables in common.

In other words, whereas in the previous case there is never a good reason to base the test on the L.M. principle (M_2 has no current endogenous explanatory variables), in this latter case that reason may exist: it will exist if M_1 and M_2 contain the same offending regressors.

Nevertheless, no matter the choice one has made about the set of instruments to be used, one might also prefer instead to run by I.V. the G.N.R.(5.29) to conduct the test. As in the previous case, this preference will have the advantage that a simplified version, where the regressand would simply be $y_t - \bar{\rho}_1 y_{t-1}$, can be used instead to perform the pseudo-*t* test based on the $C(\alpha)$ principle.

The tests so far proposed should be considered as PA tests because, as Fisher and McAleer (1981) firstly suggested and Godfrey (1983) further studied, the basic idea of implicitly minimising a criterion function like the one in (5.9) was firstly due to Atkinson (1970). The papers by Atkinson (1969, 1970) together with the papers by Cox (1961,1962) are considered pioneer works on nonnested models. Their basic ideas have been adapted to linear regression models by Pesaran (1974) and to nonlinear regression models by Pesaran and Deaton (1978). Following the artificial regression approach firstly advocated by Davidson and MacKinnon (1981) [11] (but also making use of what we have learned from previous chapters in this work) has made it possible, so far, to handle the particular (nonlinear) case of models like those in (5.1) and (5.2). We have by now discussed most of the points of interest

[11] Only in the context of L.S.. The initial work in the I.V. context is due to Godfrey (1983), Ericson (1983) and MacKinnon et al (1983).

concerning the nonnested testing of such models by proposing adequate alternative criterion functions (to the one in (5.9)) whenever an I.V. approach is required. The issues related to the choice of different sets of instruments have also been clarified. The validity of $C(\alpha)$ tests also in this context has proved to be of great relevance.

5.3 The P Test

In the previous section we have proposed a variety of auxiliary regressions (either (5.10), (5.12), (5.19), (5.23) or (5.26)) to estimate β_2 and ρ_2 consistently, taking into consideration that M_1 is the null model. The PA tests so far proposed are in fact valid since the first term of the extra regressor

$$x_2'\left(\dot{\beta}_{2/1}, \dot{\rho}_{2/1}\right) = X_2\,\dot{\beta}_{2/1} + \dot{\rho}_{2/1}\left(y_{-1} - X_{2,-1}\,\dot{\beta}_{2/1}\right) \quad (5.30)$$

[12]

which appears on the right-hand side of thè G.N.R., depends on y through $x_1'\left(\dot{\beta}_1, \dot{\rho}_1\right)$, the predicted values of y given by the null model. Therefore, the extra

[12] Cf. equation (5.16), for example.

regressor can be treated as if it were a predetermined variable (provided $x_{1,t}$ does not contain any current endogenous explanatory variable).[13]

Now, following Davidson and MacKinnon (1981) suggestion (in the L.S. context), one can also consider to use $x_2'\left(\dot{\beta}_2, \dot{\rho}_2\right)$ rather than $x_2'\left(\dot{\beta}_{2/1}, \dot{\rho}_{2/1}\right)$ as the first term of the extra regressor. In this case, however, it is not so obvious that the test of the significance of the extra regressor is in fact valid since $x_2'\left(\dot{\beta}_2, \dot{\rho}_2\right)$ will depend directly on y. Nevertheless, provide that, under the null (model), the vector $\left(\dot{\beta}_2, \dot{\rho}_2\right)$ converges asymptotically to some constant vector, so it will be the vector $x_2'\left(\dot{\beta}_2, \dot{\rho}_2\right)$.[14] Once again, it is therefore asymptotically valid to treat the extra regressor as if it were a predetermined variable.[15]

Using $x_2'\left(\dot{\beta}_2, \dot{\rho}_2\right)$ rather than $x_2'\left(\dot{\beta}_{2/1}, \dot{\rho}_{2/1}\right)$ as first term of the extra regressor (under testing) has clearly the advantage of avoiding the auxiliary regressions mentioned above. That is to say, the extra regressor in equation (5.16) would simply become, either

---

[13] If it does, $x_1'\left(\dot{\beta}_1, \dot{\rho}_1\right)$ should be replaced by $\overset{\circ}{x}_1'\left(\dot{\beta}_1, \dot{\rho}_1\right)$ as suggested in (5.19) and (5.26).

[14] And so it will be the vector $\overset{\circ}{x}_2'\left(\dot{\beta}_2, \dot{\rho}_2\right)$ to be used, instead, if I.V. estimation of M_2 is required.

[15] This type of argument can also reinforce the treatment of the extra regressor in the PA tests as a predetermined variable.

$$x_2'\left(\hat{\beta}_2, \hat{\rho}_2\right) - x_1'\left(\hat{\beta}_1, \hat{\rho}_1\right) \tag{5.31}$$

or

$$\overset{\circ}{x}{}_2'\left(\tilde{\beta}_2, \tilde{\rho}_2\right) - x_1'\left(\hat{\beta}_1, \hat{\rho}_1\right). \tag{5.32}$$

On the other hand, the extra regressor in equations (5.18) and (5.25) would simply become, respectively

$$x_2'\left(\hat{\beta}_2, \hat{\rho}_2\right) - \overset{\circ}{x}{}_1'\left(\tilde{\beta}_1, \tilde{\rho}_1\right) \tag{5.33}$$

and

$$\overset{\circ}{x}{}_2'\left(\tilde{\beta}_2, \tilde{\rho}_2\right) - \overset{\circ}{x}{}_1'\left(\tilde{\beta}_1, \tilde{\rho}_1\right) \tag{5.34}$$

where $\left(\hat{\beta}_i, \hat{\rho}_i\right)$ and $\left(\tilde{\beta}_i, \tilde{\rho}_i\right)$ denote N.L.S. and N.L.I.V. estimates from M_i $(i = 1,2)$, respectively.

All the equations above are simply the difference vector between the appropriate predicted values given by the two models: the extra regressor in (5.31) to be tested when **neither M_1 nor M_2 contain any current endogenous explanatory variable**; the extra regressor in (5.32) to be tested when M_2 **is the only to contain any current endogenous explanatory variable**; the extra regressor in (5.33) to be tested when M_1 **is the only to contain any current endogenous explanatory variable**; and the extra regressor in (5.34) when **both M_1 and M_2 contain current endogenous explanatory variables**. All these tests to be conducted in a G.N.R. to be run by O.L.S..

If, instead, one prefer to run the G.N.R. directly by I.V., then the extra regressor in the last two equations would rather be

$$x_2'\left(\hat{\beta}_2, \hat{\rho}_2\right) - x_1'\left(\tilde{\beta}_1, \tilde{\rho}_1\right) \tag{5.35}$$

and

$$x_2'\left(\tilde{\beta}_2, \tilde{\rho}_2\right) - x_1'\left(\tilde{\beta}_1, \tilde{\rho}_1\right). \tag{5.36}$$

In the former case, the G.N.R. should be run using W_{1E} as the set of instruments. In the latter case, W should be used instead.

This option, as for the PA tests, would have the advantage that a simplified version, where the regressand would simply be $y_t - \tilde{\rho}_1 y_{t-1}$ can also be used to perform the pseudo-t test based on the $C(\alpha)$ principle.

Finally, this alternative is obviously always valid when $x_{1,t}$ does not contain any current endogenous variable and therefore the simplified version of the G.N.R. can be run by O.L.S. to test the significance of the extra regressor either in (5.31) or (5.32).

This type of tests is named in the literature as a P test. As mentioned before, Davidson and MacKinnon (1981) were the first to suggest such a test in the context of N.L.S.. MacKinnon et al (1983) further studied the P test (in the same context) when the models contain lagged dependent variables and nonnormal disturbances. These authors, in the same paper, also have considered the I.V. version of the test but only in the linear context.

Like for linear models, the PA test and the P test so far proposed are obviously asymptotically equivalent. In small samples and in the spherical context, the PA test can be shown to be less biased than the P test (Fisher and McAleer (1981) and Godfrey (1983)). Unfortunately, the PA test is in many circumstances much less powerful than the P test (Davidson and MacKinnon (1982), Godfrey and Pesaran (1983) and MacKinnon (1983)).

Bernanke et al (1988) and McAleer et al (1990) have extended the nonnested criteria of Pesaran (1974), Fisher and McAleer (1981) and Davidson and MacKinnon (1981) to situations involving first-order serially correlated errors. In both papers, a variety of asymptotically equivalent test statistics are derived.

In the former paper, the computational and statistical tradeoffs of several statistics are evaluated. Bernanke et al (1988, p.320) conclude that: i) there is a distinct size and power advantage to using a Generalised Least Squares (G.L.S.) test when the serial correlation in the residuals of both models is high; ii) the G.L.S. nonnested tests have desirable power properties when the two models have an approximate number of regressors; iii) there is, however, an unacceptably large bias toward rejection of a true model when the alternative model has a large number of regressors and the residuals are highly serially correlated; iv) the P test version appears to be the best of the alternatives presented.[16]

In turn, McAleer et al (1990, p.3631) generalise G.L.S. results previously obtained to the case where lagged dependent variables are permitted and the disturbances follow a stationary autoregressive process of order p`. In this paper, the authors end up suggesting to test the significance of $-\left(\hat{\xi}_{2,t}-\hat{\xi}_{1,t}\right)$ in a regression where $y_t-\hat{\rho}_1\, y_{t-1}$ is the regressand, and $x_{1,t}-\hat{\rho}_1\, x_{1,t-1}$ and $y_{t-1}-x_{1,t-1}\,\hat{\beta}_1$ are the other regressors (cf. the authors equation 12, on p. 3628). As they emphasise, this regression is simply «Hatanaka's (1974) two-step (or residual-adjusted Aitken) estimator for the dynamic

16 See below.

adjustment model with autoregressive errors», and therefore «didactically more appealing» than an alternative regression they also propose (cf. the authors equation 10, on p. 3627).

It happens that the alternative regression of McAleer et al (1990) is simply the G.N.R. we have been proposing, but evaluated at M.L. estimates and where $\hat{\xi}_{1,t} - \hat{\xi}_{2,t}$, rather than $x'_{2,t}\left(\hat{\beta}_2, \hat{\rho}_2\right) - x'_{1,t}\left(\hat{\beta}_1, \hat{\rho}_1\right)$, is under testing. On the other hand, McAleer et al (1990) preferred regression is simply the S.G.N.R. we have been proposing, but where $-\left(\hat{\xi}_{2,t} - \hat{\xi}_{1,t}\right)$, rather than $x'_{2,t}\left(\hat{\beta}_2, \hat{\rho}_2\right) - x'_{1,t}\left(\hat{\beta}_1, \hat{\rho}_1\right)$, is under testing. However, taking into consideration that $y_t = x'_{1,t}\left(\hat{\beta}_1, \hat{\rho}_1\right) + \hat{\xi}_{1,t} = x'_{2,t}\left(\hat{\beta}_2, \hat{\rho}_2\right) + \hat{\xi}_{2,t}$, it follows that $\hat{\xi}_{1,t} - \hat{\xi}_{2,t} = x'_{2,t}\left(\hat{\beta}_2, \hat{\rho}_2\right) - x'_{1,t}\left(\hat{\beta}_1, \hat{\rho}_1\right)$ and therefore the alternative P tests proposed by McAleer et al (1990) are precisely the tests we have been able to derive using the G.N.R. approach. Moreover, the preferred regression of McAleer et al (1990) is also the preferred regression of Bernanke et al (1988) (cf. the authors equation 16, on p. 302), on the basis of Monte Carlo experiments on nonnested models of investment subject to serial correlation. Finally, it is important to emphasise that McAleer et al (1990, pp. 3628-30) also show that similar results can be obtained through the application of the Cox test of Pesaran and Deaton (1978) directly to models like (5.1) and (5.2).

Neither Bernanke et al (1988) nor McAleer et al (1990), however, handle the situation where either of the models contain current endogenous explanatory variables and therefore the relative performance of the G.L.S. tests in an I.V. context are completely unknown and deserves further investigation.

5.4 The J and JA Tests

An alternative and (apparently) simpler approach which does not make use of the G.N.R. was also originally suggested by Davidson and Mackinnon (1981) and further studied by Godfrey (1983) and MacKinnon et al (1983).

Consider again M_c given by (5.3). This model cannot be estimated because not all the parameters will be separately identifiable. One solution to this problem is simply to replace the regression function of M_2 by a consistent estimate in (5.3). M_c would become

$$y_t = (1-\alpha)\left[x_{1,t}\beta_1 + \rho_1\left(y_{t-1} - x_{1,t-1}\beta_1\right)\right] + \alpha\, x'_{2,t}\left(\dot{\beta}_2, \dot{\rho}_2\right) + error\ term. \qquad (5.37)$$

These resulting model has only $k_1 + 2$ parameters,[17] therefore all the parameters will be asymptotically identified because the two original models are nonnested and M_1 is assumed to be asymptotically identified.

For the sake of simplicity consider for a moment the case where neither $x_{1,t}$ nor $x_{2,t}$ contain current endogenous variables.

Estimating model (5.37) by N.L.S., one can then test M_1 by simply testing the significance of the extra term using a pseudo-t test based on the $C(\alpha)$ principle.[18]

Alternatively, the JA test suggested by Fisher and McAleer (1981) uses $x'_{2,t}\left(\dot{\beta}_{2/1}, \dot{\rho}_{2/1}\right)$ instead of $x'_{2,t}\left(\dot{\beta}_2, \dot{\rho}_2\right)$ as the extra term whose significance is under testing.

However, as we said before, only apparently this is a simpler approach. What is then deceptive about such approach?

First, our discussion in chapter 3 allows us to conclude that due care and attention is needed to guarantee that one obtains the correct estimate of the covariance matrix of

---

[17] $k_1 + 1$ parameters in $x'_{1,t}\left(\beta_1, \rho_1\right)$, plus the parameter α.

[18] The test is called J because both $\left(\beta_1, \rho_1\right)$ and α are estimated jointly.

the N.L.S. estimates when $x_{1,t}$ does contain lagged dependent variables. In fact, if M_1 contains any lagged dependent variable, the estimation of β_1 and ρ_1 will not be independent and therefore the inference based upon model (5.37) may be misleading if one uses for example a Cochrane-Orcutt procedure to estimate the model. The fact that the extra term can be treated as if it were a vector of observations on a predetermined variable (asymptotically uncorrelated with the error term) does not validate the inference. For example, $(x_t - \rho x_{t-1})$ so widely used in chapter 3, also contains some predetermined variables if evaluated at initial consistent estimates $\acute{\rho}$. Nevertheless, the absence of the first lag residual in the corresponding G.N.R., as one concluded in chapter 3, would invalidate the inference with respect to all the parameters and not only to those associated with the lagged dependent variable(s). Moreover, even if $x_{1,t}$ only contained strictly exogenous variables, a similar requirement must be fulfilled by $x_{2,t}$ because nonnested testing always implies reversing the roles of the two models.

Secondly, estimation of α and testing $\alpha = 0$ in (5.37) involves nonlinear restrictions between the parameters $(1-\alpha)\beta_1$, $(1-\alpha)\rho_1$, $-\rho_1(1-\alpha)\beta_1$ and α, which is computationally demanding if nonlinear estimation is adopted, as it must be, to guarantee a correct estimate of the covariance matrix mentioned above.

Obviously, the inconvenience of such approach can only be reinforced if $x_{1,t}$ contains any current endogenous variable, so that I.V. estimation of M_1 is also required.

In that case, regression (5.37) must also be estimated directly by I.V. (to obtain a consistent estimate of w^2) using a proper set of instruments: either W_{1E} **if $x_{2,t}$ does not contain current endogenous variables**, or W **if $x_{2,t}$ also contains any current endogenous variable**.

The inconvenience of such approach either in the L.S. or in the I.V. context, can be avoided if a G.N.R. approach is used to guarantee valid inference about the parameter estimates of model (5.37). But this was in first instance the motivation for the P and PA tests previously deduced. Applying a G.N.R. approach to model (5.37) will not produce therefore new results.

To see this, note that the regression function of model (5.37) is given by

$$x_t'(\beta_1, \rho_1, \alpha) = (1-\alpha)\left[x_{1,t}\beta_1 + \rho_1\left(y_{t-1} - x_{1,t-1}\beta_1\right)\right] + \alpha\, x_{2,t}'\left(\dot{\beta}_2, \dot{\rho}_2\right) \tag{5.38}$$

with corresponding G.N.R.

$$y_t - x_t'(\beta_1, \rho_1, \alpha) = (1-\alpha)(x_{1,t} - \rho_1 x_{1,t-1})b_1 + (1-\alpha)(y_{t-1} - x_{1,t-1}\beta_1)r_1 \\ + a\left\{x_{2,t}'\left(\dot{\beta}_2, \dot{\rho}_2\right) - \left[x_{1,t}\beta_1 + \rho_1\left(y_{t-1} - x_{1,t-1}\beta_1\right)\right]\right\} + error\ term \quad (5.39)$$

When evaluated at restricted root-n consistent estimates under the null

$$(\beta_1, \rho_1, \alpha) = \left(\dot{\beta}_1, \dot{\rho}_1, 0\right)$$

the G.N.R. associated with the maintained model is now easily seen to be the G.N.R. in (5.6), and the G.N.R. associated with the null model will still be the G.N.R. in (5.7). In other words, the J and JA tests so obtained would be exactly the same as the P and PA tests deduced before. This comparison also clarifies the asymptotic equivalence between the J and JA tests directly based upon model (5.37) and the P and PA tests based upon the G.N.R.: as we learned from previous chapters, the use of a G.N.R. is mostly valuable to produce a consistent estimate of the covariance matrix of the parameter estimates of model (5.37).

Given the drawbacks mentioned before, it is useless therefore to treat the J and JA tests in more detail (as we did for the P and PA tests). The J and JA versions are not really worthwhile when the models contain either lagged dependent variables or current endogenous explanatory variables. Nevertheless, as their specialisations are still interesting, we will consider those versions in the following section.

5.5. Summary of Results

The generalisation for serial correlation processes of higher-order is straightforward. Therefore, the corresponding results will be considered in this summary. The summary will be useful not only for the propose of synthesis, but also to show that the general results obtained in this work 'encompass' the well known specialised results for the linear case (absence of serial correlation). New specialised results for Spherical Univariate Linear Models estimated by I.V. under different conditioning sets of instruments will also come through as a by-product of this work.

5.5.1. Generalised PA, P , JA and J Tests for Higher-Order Autoregressive Processes

5.5.1.1. Univariate Linear Models with Autoregressive Disturbances Estimated by N.L.S.

PA test: Test $a = 0$, running by O.L.S. the following G.N.R.:

$$\dot{y}_t - \dot{x}_{1,t}\hat{\beta}_1 = \dot{x}_{1,t} b_1 + \sum_{j=1}^{p_1} r_j \hat{u}_{1,t-j} + a\left[x'_{2,t}\left(\hat{\beta}_{2/1}, \hat{\rho}_{21/1}, \ldots, \hat{\rho}_{2p_2/1}\right) - x'_{1,t}\left(\hat{\beta}_1, \hat{\rho}_{11}, \ldots, \hat{\rho}_{1p_1}\right)\right] \quad (5.40)$$
$$+ \, error \; term$$

where

$$\dot{y}_t = y_t - \sum_{j=1}^{p_1} \hat{\rho}_{1j}\, y_{t-j} \quad (5.41)$$

$$\dot{x}_{1,t} = x_{1,t} - \sum_{j=1}^{p_1} \hat{\rho}_{1j}\, x_{1,t-j} \quad (5.42)$$

$$\hat{u}_{1,t-j} = y_{t-j} - x_{1,t-j}\hat{\beta}_1 \quad (5.43)$$

$t = p+1, p+2, \ldots, N$ and $p = max(p_1, p_2)$, where p_1 and p_2 are the autoregressive orders of M_1 and M_2, respectively, and

$$x'_{2,t}\left(\hat{\beta}_{2/1}, \hat{\rho}_{21/1}, \ldots, \hat{\rho}_{2p_2/1}\right) - x'_{1,t}\left(\hat{\beta}_1, \hat{\rho}_{11}, \ldots, \hat{\rho}_{1p_1}\right) \quad (5.44)$$

is the (negative of the) N.L.S. residual value taken from the auxiliary regression

$$x'_{1,t}\left(\hat{\beta}_1, \hat{\rho}_{11}, \ldots, \hat{\rho}_{1p_1}\right) = x_{2,t}\beta_2 + \sum_{j=1}^{p_2} \rho_{2j} u_{2,t-j} + error\; term \quad (5.45)$$

where

$$x'_{1,t}\left(\hat{\beta}_1, \hat{\rho}_{11}, \ldots, \hat{\rho}_{1p_1}\right) = x_{1,t}\,\hat{\beta}_1 + \sum_{j=1}^{p_1} \hat{\rho}_{1j}\,\hat{u}_{1,t-j} \tag{5.46}$$

and

$$u_{2,t-j} = y_{t-j} - x_{2,t-j}\beta_2\,. \tag{5.47}$$

P test: Test $a = 0$, running by O.L.S. the following G.N.R.:

$$\overset{*}{y}_t - \overset{*}{x}_{1,t}\,\hat{\beta}_1 = \overset{*}{x}_{1,t}\,b_1 + \sum_{j=1}^{p_1} r_j\,\hat{u}_{1,t-j} + a\left[x'_{2,t}\left(\hat{\beta}_2, \hat{\rho}_{21}, \ldots, \hat{\rho}_{2p_2}\right) - x'_{1,t}\left(\hat{\beta}_1, \hat{\rho}_{11}, \ldots, \hat{\rho}_{1p_1}\right)\right] + \textit{error term} \tag{5.48}$$

where

$$x'_{2,t}\left(\hat{\beta}_2, \hat{\rho}_{21}, \ldots, \hat{\rho}_{2p_2}\right) = x_{2,t}\,\hat{\beta}_2 + \sum_{j=1}^{p_2} \hat{\rho}_{2j}\,\hat{u}_{2,t-j} \tag{5.49}$$

and

$$\hat{u}_{2,t-j} = y_{t-j} - x_{2,t-j}\,\hat{\beta}_2\,; \tag{5.50}$$

All the remaining expressions have the same definitions as for the PA test.

Comments:

1) Use either the $(n-k_1-p_1)R_u^2(1)$ statistic or the pseudo-t $(n-k_1-p_1-1)$ statistic to conduct the test;

2) Alternatively, $\overset{*}{x}_{1,t}\hat{\beta}_1$ can be dropped in both G.N.R.'s, but then only the pseudo-t test will be valid;

3) If O.S., rather than N.L.S. estimation has been used to estimate M_1, only the pseudo-t test will be valid, no matter the choice one has made about the regressand.

JA test: Test $\alpha = 0$, running by N.L.S. the following regression:

$$y_t = (1-\alpha)\left[x_{1,t}\beta_1 + \sum_{j=1}^{p_1}\rho_{1j}u_{1,t-j}\right] + \alpha\, x'_{2,t}\left(\hat{\beta}_{2/1}, \hat{\rho}_{21/1}, \ldots, \hat{\rho}_{2p_2/1}\right) + error\ term \tag{5.51}$$

where $x'_{2,t}\left(\hat{\beta}_{2/1}, \hat{\rho}_{21/1}, \ldots, \hat{\rho}_{2p_2/1}\right)$ is the N.L.S. predicted value obtained from the auxiliary regression in (5.45).

J test: Test $\alpha = 0$, running by N.L.S. the following regression:

$$y_t = (1-\alpha)\left[x_{1,t}\beta_1 + \sum_{j=1}^{p_1}\rho_{1j}u_{1,t-j}\right] + \alpha\, x'_{2,t}\left(\hat{\beta}_2, \hat{\rho}_{21}, \ldots, \hat{\rho}_{2p_2}\right) + error\ term \tag{5.52}$$

where $x'_{2,t}\left(\hat{\beta}_2, \hat{\rho}_{21}, \ldots, \hat{\rho}_{2p_2}\right)$ has the same definition as for the P test.

Comments:

1) Use the pseudo-t $\left(n - k_1 - p_1 - 1\right)$ statistic to conduct the test;

2) These tests have the inconvenience that we emphasised when dealing with the AR(1) case: Attention and due care is needed to insure that valid inference is obtained when $x_{1,t}$ contains any lagged dependent variable; O.S. estimation (to avoid nonlinearity) must not be adopted;

5.5.1.2. Univariate Linear Models with Autoregressive Disturbances Estimated by N.L.I.V.

5.5.1.2.1. Under the Same Conditioning Set of Instruments

PA test: Test $a = 0$, running by I.V. the following G.N.R.:

$$\dot{y}_t - \dot{x}_{1,t}\,\tilde{\beta}_1 = \dot{x}_{1,t}\,b_1 + \sum_{j=1}^{p_1} r_j\,\tilde{u}_{1,t-j} + a\left[x'_{2,t}\left(\tilde{\beta}_{2/1}, \tilde{\rho}_{21/1}, \ldots, \tilde{\rho}_{2p_2/1}\right) - x'_{1,t}\left(\tilde{\beta}_1, \tilde{\rho}_{11}, \ldots, \tilde{\rho}_{1p_1}\right)\right] \quad (5.53)$$
$$+ \, error\ term$$

where

$$\overset{*}{y}_t = y_t - \sum_{j=1}^{p_1} \tilde{\rho}_{1j}\, y_{t-j} \tag{5.54}$$

$$\overset{*}{x}_{1,t} = x_{1,t} - \sum_{j=1}^{p_1} \tilde{\rho}_{1j}\, x_{1,t-j} \tag{5.55}$$

$$\tilde{u}_{1,t-j} = y_{t-j} - x_{1,t-j}\, \tilde{\beta}_1 \tag{5.56}$$

$t = p' + 1, p' + 2, \ldots, N$ and p' is the maximum lag of the lagged variables considered in the common (extended) set of instruments W, that is the maximum lag induced by the model with higher autoregressive process; and

$$x'_{2,t}\left(\tilde{\beta}_{2/1}, \tilde{\rho}_{21/1}, \ldots, \tilde{\rho}_{2p_2/1}\right) - x'_{1,t}\left(\tilde{\beta}_1, \tilde{\rho}_{11}, \ldots, \tilde{\rho}_{1p_1}\right) \tag{5.57}$$

is the (negative of the) N.L.I.V. residual value taken from the auxiliary regression

$$x'_{1,t}\left(\tilde{\beta}_1, \tilde{\rho}_{11}, \ldots, \tilde{\rho}_{1p_1}\right) = x_{2,t}\beta_2 + \sum_{j=1}^{p_2} \rho_{2j} u_{2,t-j} + error\ term \tag{5.58}$$

where

$$x'_{1,t}\left(\tilde{\beta}_1, \tilde{\rho}_{11}, \ldots, \tilde{\rho}_{1p_1}\right) = x_{1,t}\, \tilde{\beta}_1 + \sum_{j=1}^{p_1} \tilde{\rho}_{1j}\, \tilde{u}_{1,t-j} \tag{5.59}$$

and $u_{2,t-j}$ has the same definition as in (5.47).

P test: Test $a = 0$, running by I.V. the following G.N.R.:

$$\dot{y}_t - \dot{x}_{1,t}\tilde{\beta}_1 = \dot{x}_{1,t} b_1 + \sum_{j=1}^{p_1} r_j \tilde{u}_{1,t-j} + a\left[x'_{2,t}\left(\tilde{\beta}_2, \tilde{\rho}_{21}, \ldots, \tilde{\rho}_{2p_2}\right) - x'_{1,t}\left(\tilde{\beta}_1, \tilde{\rho}_{11}, \ldots, \tilde{\rho}_{1p_1}\right)\right] \qquad (5.60)$$

$$+ \, error \; term$$

where

$$x'_{2,t}\left(\tilde{\beta}_2, \tilde{\rho}_{21}, \ldots, \tilde{\rho}_{2p_2}\right) = x_{2,t}\tilde{\beta}_2 + \sum_{j=1}^{p_2} \tilde{\rho}_{2j}\tilde{u}_{2,t-j} \qquad (5.61)$$

and

$$\tilde{u}_{2,t-j} = y_{t-j} - x_{2,t-j}\tilde{\beta}_2 \, ; \qquad (5.62)$$

All the remaining expressions have the same definitions as for the PA test.

Comments:

1) Use either the $nR_u^2(1)$ statistic (with due care) or the pseudo-$t(n)$ statistic to conduct the test;

2) Alternatively, $\overset{*}{x}_{1,t}\tilde{\beta}_1$ can be dropped in both G.N.R.'s, but then only the pseudo-t test will be valid;

3) If O.S., rather than N.L.I.V. estimation has been used to estimate M_1, only the pseudo-t test will be valid, no matter the choice one has made about the regressand;

4) The use of a common (extended) set of instruments W to estimate both M_1 and M_2 as well as to run the G.N.R. is recommended when both models have the same current endogenous explanatory variables;

5) All the regressions above (including the G.N.R.'s) can be run instead by L.S. (eventually N.L.S. to estimate both M_1 and M_2 as well as the extra regressor under testing) if $x_{1,t}$ and $x_{2,t}$ are prior replaced by $\overset{\circ}{x}_{1,t}$ and $\overset{\circ}{x}_{2,t}$, respectively. In that case, the extra regressor should be rewritten as $\overset{\circ}{x}{}'_{2,t}(\,.\,) - \overset{\circ}{x}{}'_{1,t}(\,.\,)$; However, this option would imply that $\overset{*}{x}_{1,t}\tilde{\beta}_1$ could not be dropped from the G.N.R.'s: otherwise w^2 would not be estimated consistently.

JA test: Test $\alpha = 0$, running by N.L.I.V. the following regression:

$$y_t = (1-\alpha)\left[x_{1,t}\beta_1 + \sum_{j=1}^{p_1}\rho_{1j}u_{1,t-j}\right] + \alpha\, x'_{2,t}\left(\tilde{\beta}_{2/1}, \tilde{\rho}_{21/1}, \ldots, \tilde{\rho}_{2p_2/1}\right) + error\ term \tag{5.63}$$

where $x'_{2,t}\left(\tilde{\beta}_{2/1}, \tilde{\rho}_{21/1}, \ldots, \tilde{\rho}_{2p_2/1}\right)$ is the N.L.I.V. predicted value obtained from the auxiliary regression in (5.58).

J test: Test $\alpha = 0$, running by N.L.I.V. the following regression:

$$y_t = (1-\alpha)\left[x_{1,t}\beta_1 + \sum_{j=1}^{p_1}\rho_{1j}u_{1,t-j}\right] + \alpha\, x'_{2,t}\left(\tilde{\beta}_2, \tilde{\rho}_{21}, \ldots, \tilde{\rho}_{2p_2}\right) + error\ term \tag{5.64}$$

where $x'_{2,t}\left(\tilde{\beta}_2, \tilde{\rho}_{21}, \ldots, \tilde{\rho}_{2p_2}\right)$ has the same definition as for the P test.

Comments:

1) Use the pseudo-t (n) statistic to conduct the test;

2) These tests have the inconvenience that we emphasised before: Attention and due care is needed to insure that valid inference is obtained when $x_{1,t}$ contains at least a current endogenous variable; O.S. estimation (to avoid nonlinearity) must not be adopted;

3) The use of a common (extended) set of instruments W to estimate both M_1 and M_2 as well as to run regressions (5.63-64) is recommended when both models have the same current endogenous explanatory variables;

4) A two-stages L.S. procedure cannot be adopted in this case: otherwise w^2 would not be estimated consistently.

5.5.1.2.2. Under Different Conditioning Sets of Instruments

5.5.1.2.2.1. Only M_2 Requires I.V. Estimation

PA test: Test $a = 0$, running by O.L.S. the following G.N.R.:

$$\dot{y}_t - \dot{x}_{1,t}\,\hat{\beta}_1 = \dot{x}_{1,t}\,b_1 + \sum_{j=1}^{p_1} r_j\,\hat{u}_{1,t-j} + a\left[x'_{2,t}\left(\tilde{\beta}_{2/1}, \tilde{\rho}_{21/1}, \ldots, \tilde{\rho}_{2p_2/1}\right) - x'_{1,t}\left(\hat{\beta}_1, \hat{\rho}_{11}, \ldots, \hat{\rho}_{1p_1}\right)\right] \quad (5.65)$$

$$+\, error\ term$$

where $\dot{y}_t$, $\dot{x}_{1,t}$, $\hat{u}_{1,t-j}$ have the same definitions as for the PA test in 5.5.1.1.; $t = p' + 1, p' + 2, \ldots, N$ and p' is the maximum lag of the lagged variables considered in the set of instruments W_2, that is the maximum lag induced by the autoregressive process in M_2; [19] and

$$x'_{2,t}\left(\tilde{\beta}_{2/1}, \tilde{\rho}_{21/1}, \ldots, \tilde{\rho}_{2p_2/1}\right) - x'_{1,t}\left(\hat{\beta}_1, \hat{\rho}_{11}, \ldots, \hat{\rho}_{1p_1}\right) \quad (5.66)$$

is the (negative of the) N.L.I.V. residual value taken from the auxiliary regression

---

[19] If the autoregressive process in M_1 is of higher order than the maximum lag induced by the autoregressive process in M_2, then p' is the order of the autoregressive process in M_1.

$$x'_{1,t}\left(\hat{\beta}_1, \hat{\rho}_{11}, \ldots, \hat{\rho}_{1p_1}\right) = x_{2,t}\beta_2 + \sum_{j=1}^{p_2} \rho_{2j} u_{2,t-j} + error\ term \tag{5.67}$$

where $x'_{1,t}\left(\hat{\beta}_1, \hat{\rho}_{11}, \ldots, \hat{\rho}_{1p_1}\right)$ and $u_{2,t-j}$ have the same definitions as for the PA test in 5.5.1.1..

P test: Test $a = 0$, running by O.L.S. the following G.N.R.:

$$\overset{*}{y}_t - \overset{*}{x}_{1,t} \hat{\beta}_1 = \overset{*}{x}_{1,t} b_1 + \sum_{j=1}^{p_1} r_j \hat{u}_{1,t-j} + a\left[\overset{\circ}{x}'_{2,t}\left(\tilde{\beta}_2, \tilde{\rho}_{21}, \ldots, \tilde{\rho}_{2p_2}\right) - x'_{1,t}\left(\hat{\beta}_1, \hat{\rho}_{11}, \ldots, \hat{\rho}_{1p_1}\right)\right] \tag{5.68}$$

$$+ error\ term$$

where

$$\overset{\circ}{x}'_{2,t}\left(\tilde{\beta}_2, \tilde{\rho}_{21}, \ldots, \tilde{\rho}_{2p_2}\right) = \overset{\circ}{x}_{2,t} \tilde{\beta}_2 + \sum_{j=1}^{p_2} \tilde{\rho}_{2j} \tilde{u}_{2,t-j} \tag{5.69}$$

and

$$\tilde{u}_{2,t-j} = y_{t-j} - x_{2,t-j} \tilde{\beta}_2 \; ; \tag{5.70}$$

All the remaining expressions have the same definitions as for the PA test.

Comments:

1) The same comments as for the PA and P tests in 5.5.1.1.;

2) The relevant difference is that in the P test, $\mathring{x}'_{2,t}\left(\tilde{\beta}_2, \tilde{\rho}_{21}, \ldots, \tilde{\rho}_{2p_2}\right)$ rather than $x'_{2,t}\left(\tilde{\beta}_2, \tilde{\rho}_{21}, \ldots, \tilde{\rho}_{2p_2}\right)$ must be considered as first term of the extra regressor: otherwise the extra regressor cannot be treated as a predetermined variable.

JA test: Test $\alpha = 0$, running by N.L.S. the following regression:

$$y_t = (1-\alpha)\left[x_{1,t}\beta_1 + \sum_{j=1}^{p_1} \rho_{1j} u_{1,t-j}\right] + \alpha\, x'_{2,t}\left(\tilde{\beta}_{2/1}, \tilde{\rho}_{21/1}, \ldots, \tilde{\rho}_{2p_2/1}\right) + error\ term \qquad (5.71)$$

where $x'_{2,t}\left(\tilde{\beta}_{2/1}, \tilde{\rho}_{21/1}, \ldots, \tilde{\rho}_{2p_2/1}\right)$ is the N.L.I.V. predicted value obtained from the auxiliary regression in (5.67).

J test: Test $\alpha = 0$, running by N.L.S. the following regression:

$$y_t = (1-\alpha)\left[x_{1,t}\beta_1 + \sum_{j=1}^{p_1} \rho_{1j} u_{1,t-j}\right] + \alpha\ \mathring{x}'_{2,t}\left(\tilde{\beta}_2, \tilde{\rho}_{21}, \ldots, \tilde{\rho}_{2p_2}\right) + error\ term \qquad (5.72)$$

where $\mathring{x}'_{2,t}\left(\tilde{\beta}_2, \tilde{\rho}_{21}, \ldots, \tilde{\rho}_{2p_2}\right)$ has the same definition as for the P test.

Comments:

1) The same comments as for the JA and J tests in 5.5.1.1.;

2) For the J test, also see the extra comment on the P test in this subsection.

5.5.1.2.2.2. Only M_1 Requires I.V. Estimation

PA test: Test $a = 0$, running by I.V. the following G.N.R.:

$$\overset{*}{y}_t - \overset{*}{x}_{1,t}\tilde{\beta}_1 = \overset{*}{x}_{1,t} b_1 + \sum_{j=1}^{p_1} r_j \tilde{u}_{1,t-j} + a\left[x'_{2,t}\left(\hat{\beta}_{2/1}, \hat{\rho}_{21/1}, \ldots, \hat{\rho}_{2p_2/1}\right) - x'_{1,t}\left(\tilde{\beta}_1, \tilde{\rho}_{11}, \ldots, \tilde{\rho}_{1p_1}\right)\right] + error\ term \quad (5.73)$$

where $\overset{*}{y}_t$, $\overset{*}{x}_{1,t}$, $\tilde{u}_{1,t-j}$ have the same definitions as for the PA test in 5.5.1.2.1.; $t = p'+1, p'+2, \ldots, N$ and p' is the maximum lag of the lagged variables considered in the (extended) set of instruments W_{1E}, that is the maximum lag induced by the autoregressive process in M_1;[20] and

$$x'_{2,t}\left(\hat{\beta}_{2/1}, \hat{\rho}_{21/1}, \ldots, \hat{\rho}_{2p_2/1}\right) - x'_{1,t}\left(\tilde{\beta}_1, \tilde{\rho}_{11}, \ldots, \tilde{\rho}_{1p_1}\right) \quad (5.74)$$

[20] If the autoregressive process in M_2 is of higher order than the maximum lag induced by the autoregressive process in M_1, then p' is the order of the autoregressive process in M_2.

is the (negative of the) N.L.S. residual value taken from the auxiliary regression (5.58), where $x'_{1,t}\left(\tilde{\beta}_1, \tilde{\rho}_{11}, \ldots, \tilde{\rho}_{1p_1}\right)$ has the same definition as in (5.59) and $u_{2,t-j}$ as the same definition as in (5.47).

P test: Test $a = 0$, running by I.V. the following G.N.R.:

$$\overset{\bullet}{y}_t - \overset{\bullet}{x}_{1,t}\tilde{\beta}_1 = \overset{\bullet}{x}_{1,t} b_1 + \sum_{j=1}^{p_1} r_j \tilde{u}_{1,t-j} + a\left[x'_{2,t}\left(\hat{\beta}_2, \hat{\rho}_{21}, \ldots, \hat{\rho}_{2p_2}\right) - x'_{1,t}\left(\tilde{\beta}_1, \tilde{\rho}_{11}, \ldots, \tilde{\rho}_{1p_1}\right)\right] \tag{5.75}$$
$$+ \, error \; term$$

where $x'_{2,t}\left(\hat{\beta}_2, \hat{\rho}_{21}, \ldots, \hat{\rho}_{2p_2}\right)$ has the same definition as in (5.49-50) and all the remaining expressions have the same definitions as for the PA test.

Comments:

1) The same three first comments as in 5.5.1.2.1. ;

2) Any I.V. regression above (including the G.N.R.'s) can be run instead by L.S. (eventually N.L.S. to estimate M_1) if $x_{1,t}$ is prior replaced by $\overset{\circ}{x}_{1,t}$. In that case, the extra regressor should be rewritten as $x'_{2,t}(\,.\,) - \overset{\circ}{x}'_{1,t}(\,.\,)$; However, this option would imply that $\overset{\bullet}{x}_{1,t}\tilde{\beta}_1$ could not be dropped from the G.N.R.'s: otherwise w^2 would not be estimated consistently.

JA test: Test $\alpha = 0$, running by N.L.I.V. the following regression:

$$y_t = (1-\alpha)\left[x_{1,t}\beta_1 + \sum_{j=1}^{p_1}\rho_{1j}u_{1,t-j}\right] + \alpha\, x'_{2,t}\left(\hat{\beta}_{2/1}, \hat{\rho}_{21/1}, \ldots, \hat{\rho}_{2p_2/1}\right) + error\ term \qquad (5.76)$$

where $x'_{2,t}\left(\hat{\beta}_{2/1}, \hat{\rho}_{21/1}, \ldots, \hat{\rho}_{2p_2/1}\right)$ is the N.L.S. predicted value obtained from the auxiliary regression in (5.58).

J test: Test $\alpha = 0$, running by N.L.I.V. the following regression:

$$y_t = (1-\alpha)\left[x_{1,t}\beta_1 + \sum_{j=1}^{p_1}\rho_{1j}u_{1,t-j}\right] + \alpha\, x'_{2,t}\left(\hat{\beta}_2, \hat{\rho}_{21}, \ldots, \hat{\rho}_{2p_2}\right) + error\ term \qquad (5.77)$$

where $x'_{2,t}\left(\hat{\beta}_2, \hat{\rho}_{21}, \ldots, \hat{\rho}_{2p_2}\right)$ has the same definition as for the P test.

Comments:

1) The same two first comments as in 5.5.1.2.1.;

2) Also, the same fourth comment as in 5.5.1.2.1..

5.5.1.2.2.3. Both M_1 and M_2 Require I.V. Estimation

PA test: Test $a = 0$, running the G.N.R. in (5.53) by I.V., using the (extended) set of instruments W.

Previously estimate both M_1 and M_2 by N.L.I.V. using W_1 and W_2 as set of instruments, respectively.

Also estimate by N.L.I.V. the auxiliary regression in (5.58) using W_2 as set of instruments, to obtain the (negative of the) N.L.I.V. residual values to be tested in the G.N.R..

All the regressions may be run using the same value of t as in 5.5.1.2.1.. [21]

P test: As for the PA test but running instead the G.N.R. in (5.60).

Comments:

1) Only the pseudo-$t(n)$ statistic will be valid since the test must be based on the $C(\alpha)$ principle: The G.N.R.'s associated with the null model would still have some explanatory power in general;

[21] Strictly speaking, a greater number of observations can be used either to estimate M_1 or to estimate M_2.

2) Alternatively $\overset{*}{x}_{1,t}\,\tilde{\beta}_1$ can be dropped in both G.N.R.'s if one uses the same time period to both estimate M_1 and to run the G.N.R.'s;

3) O.S. estimation, rather than N.L.I.V. estimation, can be used to both estimate M_1 and M_2. If using the same time period to both estimate M_1 and to run the G.N.R.'s, $\overset{*}{x}_{1,t}\,\dot{\beta}_1$ can still be dropped;

4) The use of different sets of instruments to estimate the G.N.R., M_1 and M_2 is recommended if the two models do not have the same current endogenous explanatory variables (see the final comments on chapter 2 and also conclusion 7 on chapter 4);

5) The same comment as in 5.5.1.2.1..

JA test: Test $\alpha = 0$, running the regression in (5.63) by N.L.I.V., using the extended set of instruments W.

Obtain $x_{2,t}'\left(\tilde{\beta}_{2/1}, \tilde{\rho}_{21/1}, \ldots, \tilde{\rho}_{2p_2/1}\right)$ as for the PA test.

J test: Test $\alpha = 0$, running the regression in (5.64) by N.L.I.V., using the extended set of instruments W.

Obtain $x_{2,t}'\left(\tilde{\beta}_2, \tilde{\rho}_{21}, \ldots, \tilde{\rho}_{2p_2}\right)$ as for the P test.

Comments:

1) The first two comments as in 5.5.1.2.1.;

2) The use of different sets of instruments to estimate M_1, M_2 and the regressions in (5.63-64) is recommended if M_1 and M_2 do not have the same current endogenous explanatory variables;

3) Also the last comment as in 5.5.1.2.1..

5.5.2. Specialisations of the PA, P , JA and J Tests

5.5.2.1. Spherical Univariate Linear Models Estimated by O.L.S.

PA test: Test $a = 0$, running by O.L.S. the following G.N.R.:

$$y_t - x_{1,t}\hat{\beta}_1 = x_{1,t}b_1 + a\left[x_{2,t}\hat{\beta}_{2/1} - x_{1,t}\hat{\beta}_1\right] + error\ term \tag{5.78}$$

where $t = 1,\ 2,\ \ldots,\ N$ and

$$x_{2,t}\hat{\beta}_{2/1} - x_{1,t}\hat{\beta}_1 \tag{5.79}$$

is the (negative of the) O.L.S. residual value taken from the auxiliary regression

$$x_{1,t}\hat{\beta}_1 = x_{2,t}\beta_2 + error\ term \tag{5.80}$$

where $x_{1,t}\hat{\beta}_1$ is the O.L.S. predicted value of y_t given by M_1.

P test: Test $a = 0$, running by O.L.S. the following G.N.R.:

$$y_t - x_{1,t}\hat{\beta}_1 = x_{1,t}b_1 + a\left[x_{2,t}\hat{\beta}_2 - x_{1,t}\hat{\beta}_1\right] + error\ term \tag{5.81}$$

where $x_{2,t}\hat{\beta}_2$ is the O.L.S. predicted value of y_t given by M_2.

Comments:

1) Use either the $(n-k_1)R_u^2(1)$ statistic or the $t(n-k_1-1)$ statistic to conduct the test;

2) Alternatively, $x_{1,t}\hat{\beta}_1$ can be dropped in both G.N.R.'s, but then only the t-test will be valid.

JA test: Test $\alpha = 0$, running by O.L.S. the following regression:

$$y_t = (1-\alpha)\, x_{1,t}\beta_1 + \alpha\, x_{2,t}\hat{\beta}_{2/1} + error\ term \tag{5.82}$$

where $x_{2,t}\hat{\beta}_{2/1}$ is the O.L.S. predicted value of the auxiliary regression (5.80).

J test: Test $\alpha = 0$, running by O.L.S. the following regression:

$$y_t = (1-\alpha)\, x_{1,t}\beta_1 + \alpha\, x_{2,t}\hat{\beta}_2 + \textit{error term} \tag{5.83}$$

where $x_{2,t}\hat{\beta}_2$ has the same definition as for the P test.

Comments:

1) Use the $t(n-k_1-1)$ statistic to conduct the test.

These are the well known results for the spherical linear case estimated by O.L.S.. Davidson and MacKinnon (1981) derived both the P and the J tests which can very easily be shown to be exactly the same in this simpler context. Similarly for the PA and JA tests due to Fisher and McAleer (1981) and that we have also obtained. Godfrey (1983) derived the *t*-test which is an equivalent test. In fact, in the t-test, the extra regressor is the negative of the extra regressor obtained in the JA and PA tests above. That is, in Godfrey (1983), the regressor under testing is precisely the O.L.S. residual vector taken from the auxiliary regression (5.80). However, because the alternative is two-sided, does not really matter the sign one affects to the parameter under testing.

5.5.2.2. Spherical Univariate Linear Models Estimate by I.V.

5.5.2.2.1.Under the Same Conditioning Set of Instruments.

PA test: Test $a=0$, running by I.V. the following G.N.R.:

$$y_t - x_{1,t}\tilde{\beta}_1 = x_{1,t} b_1 + a\left[x_{2,t}\tilde{\beta}_{2/1} - x_{1,t}\tilde{\beta}_1\right] + error\ term \tag{5.84}$$

where $t = p'+1,\ p'+2,\ \ldots,\ N$ and p' is the maximum lag of the lagged variables eventually considered in the common (extended) set of instruments W;[22] and

$$x_{2,t}\tilde{\beta}_{2/1} - x_{1,t}\tilde{\beta}_1 \tag{5.85}$$

is the (negative of the) I.V. residual value taken from the auxiliary regression

$$x_{1,t}\tilde{\beta}_1 = x_{2,t}\beta_2 + error\ term \tag{5.86}$$

where $x_{1,t}\tilde{\beta}_1$ is the I.V. predicted value of y_t given by M_1.

---

[22] If lagged variables are not considered as instruments, then $p'=0$.

P test: Test $a = 0$, running by I.V. the following G.N.R.:

$$y_t - x_{1,t}\tilde{\beta}_1 = x_{1,t}\, b_1 + a\left[x_{2,t}\tilde{\beta}_2 - x_{1,t}\tilde{\beta}_1\right] + error\ term \tag{5.87}$$

where $x_{2,t}\tilde{\beta}_2$ is the I.V. predicted value of y_t given by M_2.

Comments:

1) Use either the $(n)R_u^2(1)$ statistic (with due care) or the $t(n)$ statistic to conduct the test;

2) Alternatively, $x_{1,t}\tilde{\beta}_1$ can be dropped in both G.N.R.'s, but then only the t-test will be valid;

3) The same comment as before (in 5.5.1.2.1.) concerning the use of a common (extended) set of instruments;

4) All the regressions above (including the G.N.R.'s) can be run instead by O.L.S. if $x_{1,t}$ and $x_{2,t}$ are prior replaced by $\mathring{x}_{1,t}$ and $\mathring{x}_{2,t}$, respectively; In that case, the extra regressor should be rewritten as $\mathring{x}_{2,t}\tilde{\beta}_{2/1} - \mathring{x}_{1,t}\tilde{\beta}_1$ and $\mathring{x}_{2,t}\tilde{\beta}_2 - \mathring{x}_{1,t}\tilde{\beta}_1$ in the PA and P tests, respectively; However, as for the nonlinear case, this option would imply that $x_{1,t}\tilde{\beta}_1$ could not be dropped from the G.N.R.'s: otherwise w^2 would not be estimated consistently.

JA test: Test $\alpha = 0$, running by I.V. the following regression:

$$y_t = (1-\alpha)\, x_{1,t}\beta_1 + \alpha\, x_{2,t}\,\tilde{\beta}_{2/1} + error\ term \tag{5.88}$$

where $x_{2,t}\,\tilde{\beta}_{2/1}$ is the I.V. predicted value of the auxiliary regression in (5.86).

J test: Test $\alpha = 0$, running by I.V. the following regression:

$$y_t = (1-\alpha)\, x_{1,t}\beta_1 + \alpha\, x_{2,t}\,\tilde{\beta}_2 + error\ term \tag{5.89}$$

where $x_{2,t}\,\tilde{\beta}_2$ has the same definition as for the P test.

Comments:

1) Use the $t(n)$ statistic to conduct the test;

2) The same comment about the use of a common (extended) set of instruments as for the PA and P tests;

3) As in the nonlinear case, a two-stage O.L.S. procedure cannot be adopted: otherwise w^2 would not be estimated consistently.

These are the well known results for the spherical linear case estimated by I.V., using a common (extended) set of instruments. Godfrey (1983) derived the t-test in I.V. which is equivalent to the PA and JA tests above. The J and P tests in I.V.

are due to MacKinnon et al (1983). Also in this context, the P and the J tests can very easily be shown to be exactly the same. Similarly for the PA and JA tests.

5.5.2.2.2.Under Different Conditioning Sets of Instruments

5.5.2.2.2.1. Only M_2 Requires I.V. Estimation

PA test: Test $a = 0$, running by O.L.S. the following G.N.R.:

$$y_t - x_{1,t}\,\hat{\beta}_1 = x_{1,t}\,b_1 + a\left[x_{2,t}\,\tilde{\beta}_{2/1} - x_{1,t}\,\hat{\beta}_1\right] + error\ term \tag{5.90}$$

where $t = p'+1,\ p'+2,\ \ldots,\ N$ and p' is the maximum lag of the lagged variables eventually considered in W_2;[23] and

$$x_{2,t}\,\tilde{\beta}_{2/1} - x_{1,t}\,\hat{\beta}_1 \tag{5.91}$$

is the (negative of the) I.V. residual value taken from the auxiliary regression

[23] See footnote 22.

$$x_{1,t}\hat{\beta}_1 = x_{2,t}\beta_2 + error\ term \tag{5.92}$$

where $x_{1,t}\hat{\beta}_1$ is the O.L.S. predicted value of y_t given by M_1.

P test: Test $a = 0$, running by O.L.S. the following G.N.R.:

$$y_t - x_{1,t}\hat{\beta}_1 = x_{1,t}b_1 + a\left[\overset{\circ}{x}_{2,t}\tilde{\beta}_2 - x_{1,t}\hat{\beta}_1\right] + error\ term \tag{5.93}$$

where $\overset{\circ}{x}_{2,t}\tilde{\beta}_2$ is the two-stage O.L.S. predicted value of y_t given by M_2.

Comments:

1) The same comments as for the PA and P tests in 5.5.2.1.;

2) As for the nonlinear case, the relevant difference is that in the P test, $\overset{\circ}{x}_{2,t}\tilde{\beta}_2$ rather than $x_{2,t}\beta_2$, must be considered as first term of the extra regressor: otherwise the extra regressor cannot be treated as a predetermined variable.

JA test: Test $\alpha = 0$, running by O.L.S. the following regression:

$$y_t = (1-\alpha)\, x_{1,t}\beta_1 + \alpha\, x_{2,t}\tilde{\beta}_{2/1} + error\ term \tag{5.94}$$

where $x_{2,t}\tilde{\beta}_{2/1}$ is the I.V. predicted value obtained from the auxiliary regression in (5.92).

J test: Test $\alpha = 0$, running by O.L.S. the following regression:

$$y_t = (1-\alpha)\, x_{1,t}\beta_1 + \alpha\ \overset{\circ}{x}_{2,t}\,\tilde{\beta}_2 + error\ term \tag{5.95}$$

where $\overset{\circ}{x}_{2,t}\,\bar{\beta}_2$ has the same definition as for the P test.

Comments:

1) The same comment as for the JA and J tests in 5.5.2.1.;

2) Also see the extra comment on the P test in this subsection.

5.5.2.2.2.2. Only M_l Requires I.V. Estimation

PA test: Test $a = 0$, running by I.V. the following G.N.R.:

$$y_t - x_{1,t}\,\tilde{\beta}_1 = x_{1,t}\, b_1 + a\left[x_{2,t}\,\hat{\beta}_{2/1} - x_{1,t}\,\tilde{\beta}_1\right] + error\ term \tag{5.96}$$

where $t = p'+1,\ p'+2,\ \ldots,\ N$ and p' is the maximum lag of the lagged variables eventually considered in W_{1E};[24] and

$$x_{2,t}\,\hat{\beta}_{2/1} - x_{1,t}\,\tilde{\beta}_1 \tag{5.97}$$

is the (negative of the) O.L.S. residual value taken from the auxiliary regression in (5.86), where $x_{1,t}\,\tilde{\beta}_1$ is the I.V. predicted value of y_t given by M_1.

P test: Test $a = 0$, running by I.V. the following G.N.R.:

$$y_t - x_{1,t}\,\tilde{\beta}_1 = x_{1,t}\,b_1 + a\left[x_{2,t}\,\hat{\beta}_2 - x_{1,t}\,\tilde{\beta}_1\right] + \mathit{error\ term} \tag{5.98}$$

where $x_{2,t}\,\hat{\beta}_2$ is the O.L.S. predicted value of y_t given by M_2.

Comments:

1) The same two first comments as in 5:5.2.2.1.;

2) Any I.V. regression above (including the G.N.R.'s) can be run instead by O.L.S. if $x_{1,t}$ is prior replaced by $\mathring{x}_{1,t}$; In that case, the extra regressor should be rewritten as $x_{2,t}\,\hat{\beta}_{2/1} - \mathring{x}_{1,t}\,\tilde{\beta}_1$ and $x_{2,t}\,\hat{\beta}_2 - \mathring{x}_{1,t}\,\tilde{\beta}_1$ in the PA and P tests, respectively; However, as for the nonlinear case, this option would

[24] See footnote 22.

imply that $x_{1,t}\tilde{\beta}_1$ could not be dropped from the G.N.R.'s: otherwise w^2 would not be estimated consistently.

JA test: Test $\alpha = 0$, running by I.V. the following regression:

$$y_t = (1-\alpha)\, x_{1,t}\beta_1 + \alpha\, x_{2,t}\, \hat{\beta}_{2/1} + error\ term \tag{5.99}$$

where $x_{2,t}\,\hat{\beta}_{2/1}$ is the O.L.S. predicted value obtained from the auxiliary regression in (5.86).

J test: Test $\alpha = 0$, running by I.V. the following regression:

$$y_t = (1-\alpha)\, x_{1,t}\beta_1 + \alpha\, x_{2,t}\, \hat{\beta}_2 + error\ term \tag{5.100}$$

where $x_{2,t}\,\hat{\beta}_2$ has the same definition as for the P test.

Comments:

1) Use the $t(n)$ statistic to conduct the test;

2) Also the same third comment as in 5.5.2.2.1..

5.5.2.2.2.3. Both M_1 and M_2 Require I.V. Estimation

PA test: Test $a = 0$, running the G.N.R. in (5.84) by I.V., using the (extended) set of instruments W.

Previously, estimate both M_1 and M_2 by I.V. using W_1 and W_2 as the set of instruments respectively.

Also estimate by I.V. the auxiliary regression in (5.86), using W_2 as the set of instruments, to obtain the (negative of the) I.V. residual values to be tested in the G.N.R..

All the regressions may be run using the same value of t as in 5.5.2.2.1.. [25]

P test: As for the PA test but running instead the G.N.R. in (5.87).

Comments:

1) Only the $t(n)$ statistic will be valid since the test must be based on the $C(\alpha)$ principle: the G.N.R.'s associated with the null model would still have some explanatory power in general;

[25] See footnotes 21 and 22.

2) Alternatively, $x_{1,t}\tilde{\beta}_1$ can be dropped in both G.N.R.'s if one uses the same time period to both estimate M_1 and to run the G.N.R.'s;

3) The use of different sets of instruments to estimate M_1, M_2 and the G.N.R.'s is recommended if M_1 and M_2 do not have the same current endogenous explanatory variables (see comment 4 in 5.5.1.2.2.3.);

4) Also the same comment as in 5.5.2.2.1..

JA test: Test $\alpha = 0$ running the regression in (5.88) by I.V. using the extended set of instruments W.

Obtain $x_{2,t}\tilde{\beta}_{2/1}$ as for the PA test.

J test: Test $\alpha = 0$ running the regression in (5.89) by I.V. using the extended set of instruments W.

Obtain $x_{2,t}\tilde{\beta}_2$ as for the P test.

Comments:

1) Use the $t(n)$ statistic to conduct the test;

2) The use of different sets of instruments to estimate M_1, M_2 and the regressions in (5.88-89) is recommended if M_1 and M_2 do not have the same current endogenous explanatory variables;

3) Also the last comment in 5.5.2.2.1..

One should be able to show that, also in this (new) context, the JA and the PA tests are exactly the same. Similarly for the J and P tests.

Consider then the following regressions:

$$y - X_1 \tilde{\beta}_1 = X_1\, b_1 + a\left[I - X_2\left(X_2{}^T P_{W_2} X_2\right)^{-1} X_2{}^T P_{W_2}\right] X_1\left(X_1{}^T P_{W_1} X_1\right)^{-1} X_1{}^T P_{W_1}\, y \quad (5.101)$$
$$+\, error\ vector$$

$$y - X_1 \tilde{\beta}_1 = X_1\, b_1 + a\left[X_2\left(X_2{}^T P_{W_2} X_2\right)^{-1} X_2{}^T P_{W_2} - X_1\left(X_1{}^T P_{W_1} X_1\right)^{-1} X_1{}^T P_{W_1}\right] y \quad (5.102)$$
$$+\, error\ vector$$

$$y = (1-\alpha) X_1\, \beta_1 + \alpha\, X_2\left(X_2{}^T P_{W_2} X_2\right)^{-1} X_2{}^T P_{W_2}\, X_1\left(X_1{}^T P_{W_1} X_1\right)^{-1} X_1{}^T P_{W_1}\, y \quad (5.103)$$
$$+\, error\ vector$$

$$y = (1-\alpha) X_1\, \beta_1 + \alpha\, X_2\left(X_2{}^T P_{W_2} X_2\right)^{-1} X_2{}^T P_{W_2}\ \ y\ +\ error\ vector. \quad (5.104)$$

The above regressions, are regressions (5.84), (5.87), (5.88) and (5.89), respectively, in matrix form, after estimation of M_1 and M_2 by I.V., using W_1 and W_2 as set of instruments, respectively.

As the above regressions should be run by I.V. using W as set of instruments, define

$$M = I - X_1\left(X_1{}^T P_W X_1\right)^{-1} X_1{}^T P_W \ .$$

Now consider, for example, the JA test which corresponds to test $\alpha = 0$ in regression (5.103).

Since, under the null

$$y - X_1 \check{\beta}_1 = y - X_1\left(X_1{}^T P_W X_1\right)^{-1} X_1{}^T P_W \ y = M\, y$$

where $\check{\beta}_1$ is the I.V. estimate from regression (5.103) using W as set of instruments, it follows that one wants to test the significance of

$$\begin{aligned} & y^T M^T P_W\, M X_2\left(X_2{}^T P_{W_2} X_2\right)^{-1} X_2{}^T P_{W_2} X_1\left(X_1{}^T P_{W_1} X_1\right)^{-1} X_1{}^T P_{W_1}\, y \\ & = y^T M^T P_W\, X_2\left(X_2{}^T P_{W_2} X_2\right)^{-1} X_2{}^T P_{W_2} X_1\left(X_1{}^T P_{W_1} X_1\right)^{-1} X_1{}^T P_{W_1}\, y, \text{ since } M^T P_W\, M = M^T P_W \\ & = -\, y^T M^T P_W\left[I - X_2\left(X_2{}^T P_{W_2} X_2\right)^{-1} X_2{}^T P_{W_2}\right] X_1\left(X_1{}^T P_{W_1} X_1\right)^{-1} X_1{}^T P_{W_1}\, y \qquad (5.105) \\ & \text{since } \; -\, y^T M^T P_W\, X_1\left(X_1{}^T P_{W_1} X_1\right)^{-1} X_1{}^T P_{W_1}\, y = 0, \quad \text{as} \quad M^T P_W\, X_1 = 0 \,. \end{aligned}$$

Consider now the corresponding PA test. Under the null, the I.V. estimate of b_1 in regression (5.101) is equal to the difference between $\check{\beta}_1$ (the I.V. estimate from regression (5.103)) and $\tilde{\beta}_1$, the I.V. estimate from M_1, using W_1 as set of instruments. In fact,

$$\begin{aligned}\check{b}_1 &= \left(X_1^T P_W X_1\right)^{-1} X_1^T P_W \left(y - X_1\right)\tilde{\beta}_1 \\ &= \left(X_1^T P_W X_1\right)^{-1} X_1^T P_W\, y - \tilde{\beta}_1 \\ &= \check{\beta}_1 - \tilde{\beta}_1 \ .\end{aligned}$$

Note that the I.V. estimate $\check{b}_1$ (in regression (5.101)) is not identically zero because

$$\left(y - X_1 \tilde{\beta}_1\right)^T P_W \neq 0$$

as $W_1 \neq W$. That is to say, the G.N.R. (5.101) as still some explanatory power in general, even under the null. Therefore, the corresponding expression being tested in the PA test must be (the negative of) expression (5.105) [26] because

$$\begin{aligned}y - X_1 \tilde{\beta}_1 - X_1 \check{b}_1 &= y - X_1 \tilde{\beta}_1 - X_1\left(\check{\beta}_1 - \tilde{\beta}_1\right) = y - X_1 \check{\beta}_1 \\ &= Y - X_1\left(X_1^T P_W X_1\right)^{-1} X_1^T P_W\, y = M\, y\,.\end{aligned}$$

On the same lines would be possible to prove that the J and P tests are exactly the same in this (new) context. Note finally that the cases where either M_1 or M_2 (but

[26] Obviously it does not really matter the sign one affects to the parameter under testing, because the alternative is two-sided.

not both) require I.V. estimation, can only be seen as particular cases of this more general case for which we have provided a formal proof.

5.6. Final Comments.

1. The roles of M_1 and M_2 can simply be interchanged to perform the corresponding nonnested tests of M_2 against M_1. Looking at equation (5.3) is easily seen that testing $\alpha = 1$ in M_c corresponds to interchange subscripts in equations (5.6-7). As the next comment will clarify, that does not necessarily mean that one should always treat M_1 and M_2 symmetrically: in the I.V. context, that will be the case if and only if the same extended set of instruments is recommended to estimate both models as well as to run either the appropriate G.N.R.'s (PA and P tests) or the extended regressions (JA and J tests).

2. To the best of our knowledge, nonnested tests for competing models estimated under differing instrument validity assumptions as been addressed only by Smith (1992). However, the author's results are presented for nonnested linear regression models with heteroskedasticity and serial correlation of unknown form estimated by Generalised Method of Moments.

In this work we have also relaxed the assumption of using a common extended instrument set for both specifications. The results summarised in subsection 5.5.2.2.2. (for spherical univariate linear models), as a specialisation of the general results summarised in subsection 5.5.1.2.2. (for univariate linear models with autoregressive disturbances of order p) also generalise Godfrey's (1983,1985) results but in a simpler and most attractive way.

Quoting MacKinnon et al (1983, pp. 63-4) helps to understand why nonnested tests under differing instrument validity assumptions have not been widely addressed in the literature:

«We are explicitly assuming that both competing hypotheses specify the same matrix of instruments. This assumption is somewhat restrictive, but is, we believe, a good one, even though it is entirely possible to devise tests based on more general assumptions. Such tests would have the undesirable property that their results might depend on which instruments were associated with each hypotheses, rather than on the specifications of H_0 and H_1 themselves. Moreover, our assumption makes it impossible for the applied worker to treat the same variables as exogenous in one model and endogenous in the other, an error which could easily cause non-nested tests to yield misleading results».

In our view, whereas the second argument justifies the use of the same (extended) set of instruments when both models contain the same current endogenous explanatory variables, the first argument is somehow strange and deserves debate.

If economic theory suggests the use of different conditioning sets of instruments to estimate M_1 and M_2 (and the validity of those sets can and must always be tested in advance) that information should be taken into account to 'complete' both the null and the alternative specifications. For the model under testing not to be rejected, its 'complete' specification should be able to encompass the 'complete' alternative. This is, in our view, what really should be under testing. Of course, some degree of arbitrariness still remains when choosing instrument sets in a Limited Information context and this is the reason why we have used inverted commas in the word 'complete'. However, using a common (extended) set of instruments to estimate both models is not a neutral strategy to either the rejection or non-rejection of either model as MacKinnon et al (1983) seem to suggest: an excessive number of instruments and/or the little ability of some of them to explain the current endogenous explanatory variable(s) contained in either of the two models will increase bias; Also the I.V. estimators may be extremely inefficient (see the final comments on chapter 2 and also conclusion 7 on chapter 4). This harmful effects may lead to the wrong conclusion. In other words, and using MacKinnon et al (1983) own words, such tests would have the undesirable property that their results might depend on which instruments were **not** associated with each hypotheses, rather than on the specifications of H_0 and H_1 themselves.

1. *E.S.S.* of O.L.S. regression (4.2):

$$E.S.S. = \left(X\mathring{b}+\mathring{r}\overset{*}{u}_{-1R}\right)^{T}\left(X\mathring{b}+\mathring{r}\overset{*}{u}_{-1R}\right)$$

$$= \mathring{b}^{\,T}X^{T}X\mathring{b}+2\mathring{b}^{\,T}X^{T}\overset{*}{u}_{-1R}\mathring{r}+\mathring{r}\overset{*}{u}_{-1R}^{\;T}\overset{*}{u}_{-1R}\mathring{r} \tag{A.1.1}$$

where

$$\mathring{r} = \left(\overset{*}{u}_{-1R}^{\;T}M_{X}\overset{*}{u}_{-1R}\right)^{-1}\overset{*}{u}_{-1R}^{\;T}M_{X}\overset{*}{u}_{R} \tag{A.1.2}$$

$$\mathring{b} = \left(X^{T}X\right)^{-1}X^{T}\left(\overset{*}{u}_{R}-\overset{*}{u}_{-1R}\mathring{r}\right). \tag{A.1.3}$$

Hence,

$$\mathring{b}^{\,T}X^{T}X\mathring{b} = \left(\overset{*}{u}_{R}-\overset{*}{u}_{-1R}\mathring{r}\right)^{T}X\left(X^{T}X\right)^{-1}X^{T}X\left(X^{T}X\right)^{-1}X^{T}\left(\overset{*}{u}_{R}-\overset{*}{u}_{-1R}\mathring{r}\right)$$

$$= \left(\overset{*}{u}_{R}-\overset{*}{u}_{-1R}\mathring{r}\right)^{T}P_{X}\left(\overset{*}{u}_{R}-\overset{*}{u}_{-1R}\mathring{r}\right)$$

$$2\mathring{b}^{\,T}X^{T}\overset{*}{u}_{-1R}\mathring{r} = 2\left(\overset{*}{u}_{R}-\overset{*}{u}_{-1R}\mathring{r}\right)^{T}X\left(X^{T}X\right)^{-1}X^{T}\overset{*}{u}_{-1R}\mathring{r}$$

$$= 2\left(\overset{*}{u}_{R}-\overset{*}{u}_{-1R}\mathring{r}\right)^{T}P_{X}\overset{*}{u}_{-1R}\mathring{r}$$

Therefore,

$$E.S.S. = \left(\overset{*}{u}_R - \overset{*}{u}_{-1R}\overset{\circ}{r}\right)^T P_X \overset{*}{u}_R - \left(\overset{*}{u}_R - \overset{*}{u}_{-1R}\overset{\circ}{r}\right)^T P_X \overset{*}{u}_{-1R}\overset{\circ}{r}$$

$$+2\left(\overset{*}{u}_R - \overset{*}{u}_{-1R}\overset{\circ}{r}\right)^T P_X \overset{*}{u}_{-1R}\overset{\circ}{r} + \overset{\circ}{r}\overset{*}{u}_{-1R}{}^T \overset{*}{u}_{-1R}\overset{\circ}{r}$$

$$= \overset{*}{u}_R{}^T P_X \overset{*}{u}_R - \overset{\circ}{r}\overset{*}{u}_{-1R}{}^T P_X \overset{*}{u}_R - \overset{*}{u}_R{}^T P_X \overset{*}{u}_{-1R}\overset{\circ}{r}$$
$$+\overset{\circ}{r}\overset{*}{u}_{-1R}{}^T P_X \overset{*}{u}_{-1R}\overset{\circ}{r} + 2\overset{*}{u}_R{}^T P_X \overset{*}{u}_{-1R}\overset{\circ}{r} - 2\overset{\circ}{r}\overset{*}{u}_{-1R}{}^T P_X \overset{*}{u}_{-1R}\overset{\circ}{r}$$
$$+\overset{\circ}{r}\overset{*}{u}_{-1R}{}^T \overset{*}{u}_{-1R}\overset{\circ}{r}$$

The 2nd, 3rd and 5th terms add up zero and collecting the 4th and the 6th terms,

$$= \overset{*}{u}_R{}^T P_X \overset{*}{u}_R - \overset{\circ}{r}\overset{*}{u}_{-1R}{}^T P_X \overset{*}{u}_{-1R}\overset{\circ}{r} + \overset{\circ}{r}\overset{*}{u}_{-1R}{}^T \overset{*}{u}_{-1R}\overset{\circ}{r}.$$

Finally collecting the last two terms,

$$= \overset{*}{u}_R{}^T P_X \overset{*}{u}_R + \overset{\circ}{r}\overset{*}{u}_{-1R}{}^T M_X \overset{*}{u}_{-1R}\overset{\circ}{r} \tag{A.1.4}$$

and substituting for

$$\overset{\circ}{r} = \left(\overset{*}{u}_{-1R}{}^T M_X \overset{*}{u}_{-1R}\right)^{-1} \overset{*}{u}_{-1R}{}^T M_X \overset{*}{u}_R$$

$$E.S.S. = \overset{*}{u}_R^T P_X \overset{*}{u}_R + \overset{*}{u}_R^T M_X \overset{*}{u}_{-1R} \left(\overset{*}{u}_{-1R}^T M_X \overset{*}{u}_{-1R} \right)^{-1} \overset{*}{u}_{-1R}^T M_X \overset{*}{u}_R \,.. \tag{A.1.5}$$

2. The Square of the t Statistic in regression (4.2):

$$\frac{(n-k-1)\overset{*}{u}_R^T M_X \overset{*}{u}_{-1R} \left(\overset{*}{u}_{-1R}^T M_X \overset{*}{u}_{-1R} \right)^{-1} \overset{*}{u}_{-1R}^T M_X \overset{*}{u}_R}{S.S.R.} \tag{A.1.6}$$

$$= \left[\frac{\left(\overset{*}{u}_{-1R}^T M_X \overset{*}{u}_{-1R} \right)^{-1/2} \overset{*}{u}_{-1R}^T M_X \overset{*}{u}_R}{\left[S.S.R./(n-k-1) \right]^{1/2}} \right]^2$$

$$= \left[\frac{\left(\overset{*}{u}_{-1R}^T M_X \overset{*}{u}_{-1R} \right)^{-1} \overset{*}{u}_{-1R}^T M_X \overset{*}{u}_R}{\left[S.S.R./n-k-1 \right]^{1/2} \left(\overset{*}{u}_R^T M_X \overset{*}{u}_{-1R} \right)^{-1/2}} \right]^2$$

$$= \left[\frac{\overset{\circ}{r}}{s.e.\left(\overset{\circ}{r}\right)} \right]^2 = \overset{\circ}{t}^2 \,.. \tag{A.1.7}$$

3. E.S.S. of O.L.S. regression (4.8):

$$E.S.S.=\left(\mathring{X}\mathring{b}+\mathring{r}\dot{u}_{-1R}\right)^T\left(\mathring{X}\mathring{b}+\mathring{r}\dot{u}_{-1R}\right)$$

$$=\mathring{b}^{\,T}\mathring{X}^{\,T}\mathring{X}\mathring{b}+2\mathring{b}^{\,T}\mathring{X}^{\,T}\dot{u}_{-1R}\,r+\mathring{r}\dot{u}_{-1R}^{\,T}\dot{u}_{-1R}\mathring{r} \qquad \text{(A.1.8)}$$

where

$$\mathring{r}=\left(\dot{u}_{-1R}^{\,T}M_{\mathring{X}}\dot{u}_{-1R}\right)^{-1}\dot{u}_{-1R}^{\,T}M_{\mathring{X}}\dot{u}_R \qquad \text{(A.1.9)}$$

$$\mathring{b}=\left(\mathring{X}^{\,T}\mathring{X}\right)^{-1}\mathring{X}^{\,T}\left(\dot{u}_R-\dot{u}_{-1R}\mathring{r}\right). \qquad \text{(A.1.10)}$$

Therefore, simply substituting X by $\mathring{X}$ in (A.1.4)

$$E.S.S.=\dot{u}_R^{\,T}P_{\mathring{X}}\dot{u}_R+\mathring{r}\dot{u}_{-1R}^{\,T}M_{\mathring{X}}\dot{u}_{-1R}\mathring{r} \qquad \text{(A.1.11)}$$

or finally, substituting for

$$\mathring{r}=\left(\dot{u}_{-1R}^{\,T}M_{\mathring{X}}\dot{u}_{-1R}\right)^{-1}\dot{u}_{-1R}^{\,T}M_{\mathring{X}}\dot{u}_R$$

$$E.S.S. = \dot{u}_R^T P_{\mathring{X}} \dot{u}_R + \dot{u}_R^T M_{\mathring{X}} \dot{u}_{-1R} \left(\dot{u}_{-1R}^T M_{\mathring{X}} \dot{u}_{-1R} \right)^{-1} \dot{u}_{-1R}^T M_{\mathring{X}} \dot{u}_R \tag{A.1.12}$$

4. E.S.S. of I.V. regression (4.2):

$$E.S.S. = \left(X\check{b} + \check{r}\dot{u}_{-1R} \right)^T \left(X\check{b} + \check{r}\dot{u}_{-1R} \right)$$

$$= \check{b}^T X^T X \check{b} + 2\check{b}^T X^T \dot{u}_{-1R} \check{r} + \check{r}\dot{u}_{-1R}^T \dot{u}_{-1R} \check{r} \tag{A.1.13}$$

where

$$\check{r} = \left(\dot{u}_{-1R}^T M_{\mathring{X}} \dot{u}_{-1R} \right)^{-1} \dot{u}_{-1R}^T M_{\mathring{X}} \dot{u}_R \tag{A.1.14}$$

$$\check{b} = \left(\mathring{X}^T \mathring{X} \right)^{-1} \mathring{X}^T \left(\dot{u}_R - \dot{u}_{-1R} \check{r} \right). \tag{A.1.15}$$

Hence,

$$\check{b}^T X^T X \check{b} = \left(\dot{u}_R - \dot{u}_{-1R} \check{r} \right)^T \mathring{X} \left(\mathring{X}^T \mathring{X} \right)^{-1} X^T X \left(\mathring{X}^T \mathring{X} \right)^{-1} \mathring{X}^T \left(\dot{u}_R - \dot{u}_{-1R} \check{r} \right)$$

$$2\check{b}^T X^T \dot{u}_{-1R} \check{r} = 2\left(\dot{u}_R - \dot{u}_{-1R} \check{r} \right)^T \mathring{X} \left(\mathring{X}^T \mathring{X} \right)^{-1} X^T \dot{u}_{-1R} \check{r}$$

Therefore,

$$
\begin{aligned}
E.S.S. = {} & \left(\dot{u}_R - \dot{u}_{-1R}\,\check{r}\right)^T \mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1} X^T X\left(\mathring{X}^T\mathring{X}\right)^{-1}\mathring{X}^T\dot{u}_R \\
& -\left(\dot{u}_R - \dot{u}_{-1R}\,\check{r}\right)^T \mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1} X^T X\left(\mathring{X}^T\mathring{X}\right)^{-1}\mathring{X}^T\dot{u}_{-1R}\,\check{r} \\
& +2\left(\dot{u}_R - \dot{u}_{-1R}\,\check{r}\right)^T \mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1} X^T\dot{u}_{-1R}\,\check{r} + \check{r}\,\dot{u}_{-1R}^T\,\dot{u}_{-1R}\,\check{r}
\end{aligned}
$$

$$
\begin{aligned}
= {} & \dot{u}_R^T\mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1} X^T X\left(\mathring{X}^T\mathring{X}\right)^{-1}\mathring{X}^T\dot{u}_R \\
& -\check{r}\,\dot{u}_{-1R}^T\mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1} X^T X\left(\mathring{X}^T\mathring{X}\right)^{-1}\mathring{X}^T\dot{u}_R \\
& -\dot{u}_R^T\mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1} X^T X\left(\mathring{X}^T\mathring{X}\right)^{-1}\mathring{X}^T\dot{u}_{-1R}\,\check{r} \\
& +\check{r}\,\dot{u}_{-1R}^T\mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1} X^T X\left(\mathring{X}^T\mathring{X}\right)^{-1}\mathring{X}^T\dot{u}_{-1R}\,\check{r} \\
& +2\dot{u}_R^T\mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1} X^T\dot{u}_{-1R}\,\check{r} \\
& -2\check{r}\,\dot{u}_{-1R}^T\mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1} X^T\dot{u}_{-1R}\,\check{r} \\
& +\check{r}\,\dot{u}_{-1R}^T\,\dot{u}_{-1R}\,\check{r}
\end{aligned}
$$

Adding the 2nd and the 3rd terms,

$$
\begin{aligned}
&= \dot{u}_R^{\,T}\,\mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1}X^TX\left(\mathring{X}^T\mathring{X}\right)^{-1}\mathring{X}^T\dot{u}_R \\
&\quad -2\dot{u}_R^{\,T}\,\mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1}X^TX\left(\mathring{X}^T\mathring{X}\right)^{-1}\mathring{X}^T\dot{u}_{-1R}\,\check{r} \\
&\quad +\check{r}\,\dot{u}_{-1R}^{\,T}\,\mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1}X^TX\left(\mathring{X}^T\mathring{X}\right)^{-1}\mathring{X}^T\dot{u}_{-1R}\,\check{r} \\
&\quad +2\dot{u}_R^{\,T}\,\mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1}X^T\dot{u}_{-1R}\,\check{r} \\
&\quad -2\check{r}\,\dot{u}_{-1R}^{\,T}\,\mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1}X^T\dot{u}_{-1R}\,\check{r} \\
&\quad +\check{r}\,\dot{u}_{-1R}^{\,T}\,\dot{u}_{-1R}\,\check{r}
\end{aligned}
$$

Collecting the 4th and the 2nd terms and splitting the 5th term,

$$
\begin{aligned}
&= \dot{u}_R^{\,T}\,\mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1}X^TX\left(\mathring{X}^T\mathring{X}\right)^{-1}\mathring{X}^T\dot{u}_R \\
&\quad +2\dot{u}_R^{\,T}\,\mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1}X^T\left(I-X\left(\mathring{X}^T\mathring{X}\right)^{-1}\mathring{X}^T\right)\dot{u}_{-1R}\,\check{r} \\
&\quad +\check{r}\,\dot{u}_{-1R}^{\,T}\,\mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1}X^TX\left(\mathring{X}^T\mathring{X}\right)^{-1}\mathring{X}^T\dot{u}_{-1R}\,\check{r} \\
&\quad -\check{r}\,\dot{u}_{-1R}^{\,T}\,\mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1}X^T\dot{u}_{-1R}\,\check{r} \\
&\quad -\check{r}\,\dot{u}_{-1R}^{\,T}\,\mathring{X}\left(\mathring{X}^T\mathring{X}\right)^{-1}X^T\dot{u}_{-1R}\,\check{r} \\
&\quad +\check{r}\,\dot{u}_{-1R}^{\,T}\,\dot{u}_{-1R}\,\check{r}
\end{aligned}
$$

Collecting the 4th and the 3rd terms, and collecting the 6th and the transpose of the 5th term,

$$
\begin{aligned}
= \ & \dot{u}_R^T \mathring{X}\left(\mathring{X}^T \mathring{X}\right)^{-1} X^T X\left(\mathring{X}^T \mathring{X}\right)^{-1} \mathring{X}^T \dot{u}_R \\
& + 2\dot{u}_R^T \mathring{X}\left(\mathring{X}^T \mathring{X}\right)^{-1} X^T\left(I - X\left(\mathring{X}^T \mathring{X}\right)^{-1} \mathring{X}^T\right)\dot{u}_{-1R}\check{r} \\
& - \check{r}\dot{u}_{-1R}^T \mathring{X}\left(\mathring{X}^T \mathring{X}\right)^{-1} X^T\left(I - X\left(\mathring{X}^T \mathring{X}\right)^{-1} \mathring{X}^T\right)\dot{u}_{-1R}\check{r} \\
& + \check{r}\dot{u}_{-1R}^T\left(I - X\left(\mathring{X}^T \mathring{X}\right)^{-1} \mathring{X}^T\right)\dot{u}_{-1R}\check{r}
\end{aligned}
$$

Finally, collecting the 3rd and the 4th terms,

$$
\begin{aligned}
= \ & \dot{u}_R^T \mathring{X}\left(\mathring{X}^T \mathring{X}\right)^{-1} X^T X\left(\mathring{X}^T \mathring{X}\right)^{-1} \mathring{X}^T \dot{u}_R \\
& + 2\dot{u}_R^T \mathring{X}\left(\mathring{X}^T \mathring{X}\right)^{-1} X^T\left(I - X\left(\mathring{X}^T \mathring{X}\right)^{-1} \mathring{X}^T\right)\dot{u}_{-1R}\check{r} \\
& + \check{r}\dot{u}_{-1R}^T\left(I - \mathring{X}\left(\mathring{X}^T \mathring{X}\right)^{-1} X^T\right)\left(I - X\left(\mathring{X}^T \mathring{X}\right)^{-1} \mathring{X}^T\right)\dot{u}_{-1R}\check{r}
\end{aligned}
\tag{A.1.16}
$$

or, substituting for

$$
\check{r} = \left(\dot{u}_{-1R}^T M_{\mathring{X}} \dot{u}_{-1R}\right)^{-1} \dot{u}_{-1R}^T M_{\mathring{X}} \dot{u}_R \ .
$$

$$
\begin{aligned}
&= \dot{u}_R^{\,T} \mathring{X}\left(\mathring{X}^{T}\mathring{X}\right)^{-1} X^T X \left(\mathring{X}^{T}\mathring{X}\right)^{-1} \mathring{X}^{T} \dot{u}_R \\
&\quad + 2\dot{u}_R^{\,T} \mathring{X}\left(\mathring{X}^{T}\mathring{X}\right)^{-1} X^T \left(I - X\left(\mathring{X}^{T}\mathring{X}\right)^{-1}\mathring{X}^{T}\right) \dot{u}_{-1R}\left(\dot{u}_{-1R}^{\,T} M_{\mathring{X}}\, \dot{u}_{-1R}\right)^{-1} \dot{u}_{-1R}^{\,T} M_{\mathring{X}}\, \dot{u}_R \\
&\quad + \dot{u}_R^{\,T} M_{\mathring{X}}\, \dot{u}_{-1R}\left(\dot{u}_{-1R}^{\,T} M_{\mathring{X}}\, \dot{u}_{-1R}\right)^{-1} \dot{u}_{-1R}^{\,T}\left(I - \mathring{X}\left(\mathring{X}^{T}\mathring{X}\right)^{-1} X^T\right) \\
&\qquad\qquad \left(I - X\left(\mathring{X}^{T}\mathring{X}\right)^{-1}\mathring{X}^{T}\right) \dot{u}_{-1R}\left(\dot{u}_{-1R}^{\,T} M_{\mathring{X}}\, \dot{u}_{-1R}\right)^{-1} \dot{u}_{-1R}^{\,T} M_{\mathring{X}}\, \dot{u}_R
\end{aligned}
\tag{A.1.17}
$$

It is now easily seen that this *E.S.S.* would collapse to the *E.S.S.* in (A.1.12) if $X = \mathring{X}$, since:

in the 1st term $P_{\mathring{X}} P_{\mathring{X}} = P_{\mathring{X}}$;
in the 2nd term $P_{\mathring{X}} M_{\mathring{X}} = 0$;
in the 3rd term $M_{\mathring{X}} M_{\mathring{X}} = M_{\mathring{X}}$.

However that is not supposed to be the case, otherwise the I.V. estimation would be redundant.

Estimating regression models with MA(1) errors

Consider the model

$$y_t = x_t\beta + u_t \quad ; \quad u_t = \xi_t + \alpha\xi_{t-1} \quad ; \quad \xi_t \sim \mathrm{IID}\left(0, w^2\right) \tag{A.2.1}$$

where x_t is row t of matrix X, $t = 1, 2, ..., n$.

Rewrite the disturbance process as

$$\xi_t = u_t - \alpha\xi_{t-1} \tag{A.2.2}$$

and assume $\xi_0 = 0$ to obtain

$$\begin{aligned}
\xi_1 &= u_1 - \alpha\xi_0 = u_1 \\
\xi_2 &= u_2 - \alpha\xi_1 = u_2 - \alpha u_1 \\
\xi_3 &= u_3 - \alpha\xi_2 = u_3 - \alpha\left(u_2 - \alpha u_1\right) = u_3 - \alpha u_2 + \alpha^2 u_1 \\
\xi_4 &= u_4 - \alpha\xi_3 = u_4 - \alpha\left(u_3 - \alpha u_2 + \alpha^2 u_1\right) = u_4 - \alpha u_3 + \alpha^2 u_2 - \alpha^3 u_1
\end{aligned} \tag{A.2.3}$$

(...)

Therefore,

$$
\begin{aligned}
y_1 &= x_1\beta + u_1 = x_1\beta + \xi_1 + \alpha\xi_0 = x_1\beta + \xi_1 \\
y_2 &= x_2\beta + u_2 = x_2\beta + \xi_2 + \alpha\xi_1 = x_2\beta + \alpha u_1 + \xi_2 \\
y_3 &= x_3\beta + u_3 = x_3\beta + \xi_3 + \alpha\xi_2 = x_3\beta + \alpha\left(u_2 - \alpha u_1\right) + \xi_3 \\
&= x_3\beta + \alpha u_2 - \alpha^2 u_1 + \xi_3 \\
y_4 &= x_4\beta + u_4 = x_4\beta + \xi_4 + \alpha\xi_3 = x_4\beta + \alpha\left(u_3 - \alpha u_2 + \alpha^2 u_1\right) + \xi_4 \\
&= x_4\beta + \alpha u_3 - \alpha^2 u_2 + \alpha^3 u_1 + \xi_4
\end{aligned}
\qquad \text{(A.2.4)}
$$

(...)

or, by making the definitions

$\overset{*}{u}_0 = 0$ and

$$\overset{*}{u}_{t-1} = u_{t-1} - \alpha \overset{*}{u}_{t-2} \quad ; \quad t = 2, \ldots, n \qquad \text{(A.2.5)}$$

We can write equations (A.2.4) in the form

$$y_t = x_t\beta + \alpha \overset{*}{u}_{t-1} + \xi_t \quad ; \quad t = 1, \ldots, n \qquad \text{(A.2.6)}$$

or

$$y_t = x_t\beta + \alpha\left(y_{t-1} - x_{t-1}\right)^{*} + \xi_t \quad ; \quad t = 1, \ldots, n \quad ; \quad \xi_t \sim \mathrm{IID}\left(0, w^2\right) \qquad \text{(A.2.7)}$$

with regression function

$$x_t'(\beta,\alpha) = x_t\beta + \alpha\left(y_{t-1} - x_{t-1}\right)^{*} \tag{A.2.8}$$

Comparing this model with model (3.4) one can conclude that the two models seem very similar. However, the star in the lagged disturbance makes it clear that the regression function in (A.2.8) depends on the entire sample up to period t. Moreover, since $\overset{*}{y}_{t-1}$ and $\overset{*}{x}_{t-1}$ have similar definitions as $\overset{*}{u}_{t-1}$ in (A.2.5), both are functions of the parameter α.

A specialised N.L.S. program should therefore allows us to define the regression function recursively and also take into account the higher nonlinearity in it.

Those N.L.S. estimates may be obtained by minimising, with respect to (β,α) the criterion function

$$\left(y - x'(\beta,\alpha)\right)^T\left(y - x'(\beta,\alpha)\right)$$

$$\equiv \left\{y - \left[X\beta + \alpha\left(\overset{*}{y}_{-1}(\alpha) - \overset{*}{X}_{-1}(\alpha)\beta\right)\right]\right\}^T\left\{y - \left[X\beta + \alpha\left(\overset{*}{y}_{-1}(\alpha) - \overset{*}{X}_{-1}(\alpha)\beta\right)\right]\right\}. \tag{A.2.9}$$

The F.O.C.'s are given by

$$\left[X-\hat{\alpha}\dot{X}_{-1}\left(\hat{\alpha}\right) \quad \left(\dot{y}_{-1}\left(\hat{\alpha}\right)-\dot{X}_{-1}\left(\hat{\alpha}\right)\hat{\beta}\right)+\hat{\alpha}\left.\frac{\partial\left(\dot{y}_{-1}(\alpha)-\dot{X}_{-1}(\alpha)\hat{\beta}\right)}{\partial\alpha}\right|_{\alpha=\hat{\alpha}} \right] \tag{A.2.10}$$

$$\left\{ y-\left[X\hat{\beta}+\hat{\alpha}\left(\dot{y}_{-1}\left(\hat{\alpha}\right)-\dot{X}_{-1}\left(\hat{\alpha}\right)\hat{\beta}\right)\right]\right\}=0$$

where the matrix

$$\dot{X}_{-1}\left(\hat{\alpha}\right)=$$

$$=\begin{bmatrix} 0 \\ x_1 \\ x_2-\hat{\alpha}x_1 \\ x_3-\hat{\alpha}x_2+\hat{\alpha}^2x_1 \\ \cdots \\ x_{n-1}-\hat{\alpha}x_{n-2}+\hat{\alpha}^2x_{n-3}-\ldots+(-1)^{n-4}\hat{\alpha}^{n-4}x_3+(-1)^{n-3}\hat{\alpha}^{n-3}x_2+(-1)^{n-2}\hat{\alpha}^{n-2}x_1 \end{bmatrix} \tag{A.2.11}$$

with x_t as row t of matrix X, and the vector $\dot{y}_{-1}\left(\hat{\alpha}\right)$ as a similar definition with y_t as its component of period t.

The corresponding G.N.R. is

$$y-\left[X\hat{\beta}+\hat{\alpha}\left(\dot{y}_{-1}\left(\hat{\alpha}\right)-\dot{X}_{-1}\left(\hat{\alpha}\right)\hat{\beta}\right)\right]=$$

$$=\left[X-\hat{\alpha}\,\dot{X}_{-1}\left(\hat{\alpha}\right)\quad\left(\dot{y}_{-1}\left(\hat{\alpha}\right)-\dot{X}_{-1}\left(\hat{\alpha}\right)\hat{\beta}\right)+\left(\hat{\alpha}\right)\left.\frac{\partial\left(\dot{y}_{-1}(\alpha)-\dot{X}_{-1}(\alpha)\hat{\beta}\right)}{\partial\alpha}\right|_{\alpha=\hat{\alpha}}\right]\begin{bmatrix}b\\ r\end{bmatrix}\tag{A.2.12}$$

$$+\,error\ vector$$

or

$$y_t-\hat{\alpha}\,\dot{y}_{t-1}\left(\hat{\alpha}\right)-\left(x_t\hat{\beta}-\hat{\alpha}\,\dot{x}_{t-1}\left(\hat{\alpha}\right)\hat{\beta}\right)=$$

$$=\left(x_t-\hat{\alpha}\,\dot{x}_{t-1}\left(\hat{\alpha}\right)\right)b+r\left[\left(\dot{y}_{t-1}\left(\hat{\alpha}\right)-\dot{x}_{t-1}\left(\hat{\alpha}\right)\hat{\beta}\right)+\hat{\alpha}\left.\frac{\partial\left(\dot{y}_{t-1}(\alpha)-\dot{x}_{t-1}(\alpha)\hat{\beta}\right)}{\partial\alpha}\right|_{\alpha=\hat{\alpha}}\right]+error\ term\tag{A.2.13}$$

where

$$y_t-\hat{\alpha}\,\dot{y}_{t-1}\left(\hat{\alpha}\right)-\left(x_t\hat{\beta}-\hat{\alpha}\,\dot{x}_{t-1}\left(\hat{\alpha}\right)\hat{\beta}\right)=\hat{\xi}_t$$

$$x_t-\hat{\alpha}\,\dot{x}_{t-1}\left(\hat{\alpha}\right)=\left.\frac{\partial x_t'(\beta,\alpha)}{\partial\beta}\right|_{(\beta,\alpha)=\left(\hat{\beta},\hat{\alpha}\right)}$$

$$\left(\dot{y}_{t-1}\left(\hat{\alpha}\right)-\dot{x}_{t-1}\left(\hat{\alpha}\right)\hat{\beta}\right)+\hat{\alpha}\left.\frac{\partial\left(\dot{y}_{t-1}(\alpha)-\dot{x}_{t-1}\left(\hat{\alpha}\right)\hat{\beta}\right)}{\partial\alpha}\right|_{\alpha=\hat{\alpha}}=\left.\frac{\partial x_t'(\beta,\alpha)}{\partial\alpha}\right|_{(\beta,\alpha)=\left(\hat{\beta},\hat{\alpha}\right)}$$

$$=\hat{\dot{u}}_{t-1}+\hat{\alpha}\left.\frac{\partial\hat{\dot{u}}_{t-1}(\alpha)}{\partial\alpha}\right|_{\alpha=\hat{\alpha}}$$

and ^ denotes N.L.S. estimates.

Summarising, given suitable regularity condition on $x_t'(\beta, \alpha)$, making the asymptotically innocuous assumption that $\xi_0 = 0$ and also assuming the invertibility condition $|\alpha| \leq 1$ so that α will be identified by the data, the N.L.S. estimators for β and α will be consistent, asymptotically efficient and asymptotically normal and with a covariance matrix consistently estimated by the estimated covariance matrix of the O.L.S. estimators for b and r in the G.N.R. above.

References

Amemiya, T. (1974). "The nonlinear two-stage least-squares estimator," *Journal of Econometrics*, **2**, 105-10.

Amemiya, T. (1983). " Non-linear regression models," Ch. 6 in *Handbook of Econometrics*, Vol. I, eds. Z. Griliches and M. D. Intriligator, Amesterdam, North-Holland.

Amemiya, T. (1985). *Advanced Econometrics*, Cambridge, Mass.,Harvard University Press.

Atkinson, A. C. (1969). "A test for discriminating between models," *Biometrika*. **56**, 337-47.

Atkinson, A. C. (1970). "A method for discriminating between models," *Journal of the Royal Statistical Society*, Series B, **32**, 323-53.

Backus, D. (1984). "Empirical Models of the exchange rate: separating the wheat from the chaff ," *Canadian Journal of Economics*, **17**, 824-46.

Bates, D. M., and D. G. Watts (1988). *Nonlinear Regression Analysis and its Applications*, New York, John Wiley & Sons.

Bera, A. K., & M. McAleer (1987): "On exact and asymptotic tests of nonnested models", *Statistics & Probability Letters*, **5**(1), 19-22.

Bernanke, B., H. Bohn, and P.C. Reiss (1988). "Alternative non-nested specification tests of time-series investment models," *Journal of Econometrics*, **37**, 293-326.

Berndt, E. R., and N. E. Savin (1975). "Estimation and hypothesis testing in singular equation systems with autoregressive disturbances," *Econometrica*, **43**, 937-57.

Betancourt, R., and H. Kelejian (1981). "Lagged endogenous variables and the Cochrane-Orcutt procedure," *Econometrica*, **49**, 1073-78.

Breusch, T. S. (1978). "Testing for autocorrelation in dynamic linear models," *Australian Economic Papers*, **17**, 334-55.

Breusch, T. S., and A. R. Pagan (1980). "The Lagrange Multiplier test and its applications to model specification in econometrics," *Review of Economic Studies*, **47**, 239-53.

Cochrane, D., and G. H. Orcutt (1949). "Application of least squares regression to relationships containing autocorrelated error terms," *Journal of the American Statistical Association*, **44**, 32-61.

Cooper, J. P. (1972). "Asymptotic covariance matrix of procedures for linear regression in the presence of first-order autoregressive disturbances," *Econometrica*, **40**, 305-10.

Cox, D. R. (1961). "Tests of separate families of hypotheses," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 105-23.

Cox, D. R. (1962). "Further results on tests of separate families of hypotheses," *Journal of the Royal Statistical Society*, Series B, **24**, 406-**24**

Dastoor, N. K. (1981): "A note on the interpretation of the Cox procedure for non-nested hypotheses," *Economics Letters*, **8**, 113-19.

Dastoor, N. K. (1983). "Some aspects of testing non-nested hypotheses," *Journal of Econometrics*, **21**, 213-28.

Dastoor, N. K. (1985): "A classical approach to Cox: test for non-nested hypotheses," *Journal of Econometrics*, **27**, 363-70.

Dastoor, N. K., and G.R. Fisher (1988). "On point-optimal Cox tests," *Econometric Theory*, **4**, 97-107.

Dastoor, N. K., & M. McAleer (1984): "Testing separate models with stochastic regressors," *Economic Modelling*, **2**, 331-8.

Dastoor, N. K., & M. McAlleer (1989): "Some power comparisons of joint and paired tests for nonnested models under local hypotheses," *Econometric Theory*, **5**, 83-94.

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

Davidson, R., R. J. G. Mackinnon (1980a): "On a simple procedure for testing non-nested regressions models," *Economics Letters*, **5**; 45-48.

Davidson, R., and J. G. Mackinnon (1980b). "Estimating the covariance matrix for regression models with AR (1) errors and lagged dependent variables," *Economic Letters*, **6**, 119-23.

Davidson, R., and J. G. Mackinnon (1981). "Several tests for model specification in the presence of alternative hypotheses," *Econometrica*, **49**, 781-93.

Davidson, R., and J. G. Mackinnon (1982). "Some non-nested hypothesis tests and the relations among them," *Review of Economic Studies*, **49**, 551-65.

Davidson, R., and J. G. Mackinnon (1984). "Model specification tests based on artificial linear regressions," *International Economic Review*, **25**, 485-502.

Davidson, R., and J. G. Mackinnon (1990). "Specification tests based on artificial regressions," *Journal of the American Statistical Association*, **85**, 220-27.

Davidson, R., and J. G. Mackinnon (1993). *Estimation and Inference in Econometrics*, Oxford, Oxford University Press.

Durbin, J. (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.

Durbin, J., and G. S. Watson (1951). "Testing for serial correlation in least squares regression II," *Biometrika*, **38**, 159-78.

Durbin, J., and G.S. Watson (1971). "Testing for serial correlation in least squares regression III," *Biometrika*, **58**, 1-19.

Dyer, A. R. (1973). "Discrimination procedures for separate families of hypotheses," *Journal of the American Statistical Association*, **68**, 970-74.

Dyer, A. R. (1974). "Hypothesis testing procedures for separate families of hypotheses," *Journal of the American Statistical Association*, **69**, 140-145.

Efron, B. (1983). "Comparing non-nested linear models," *Journal of the American Statistical Association*, **79**, 791-803.

Engle, R. F. (1982). "A general approach to Lagrange Multiplier model diagnostics," *Journal of Econometrics*, **20**, 83-104.

Engle, R. F. (1984). "Wald, likelihood ratio and Lagrange multiplier tests in econometrics," Ch. 13 in *Handbook of Econometrics*, Vol.II, eds. Z. Griliches and M. D. Intriligator, Amesterdam, North-Holland.

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

Ericsson, N. R. (1983). "Asymptotic properties of instrumental variables statistics for testing non-nested hypotheses," *Review of Economic Studies*, **50**, 287-304.

Ericsson, N. R. (1986). "Post-simulation analysis of Monte-Carlo experiments-interpreting pesaran's (1974) study of non-nested hypothesis test statistics," *Review of Economic Studies*, **53**, 691-707.

Ericsson, N. R. (1991). "Monte Carlo methodology and the finite sample properties of instrumental variables statistics for testing nested and non-nested hypotheses," *Econometrica*, **59**, 1249-77.

Fisher, G. R., & M. McAleer (1979). "On the interpretation of the cox test in Econometrics," *Economics Letters*, **4**, 145-150.

Fisher, G. R., and M. McAleer (1981). "Alternative procedures and associated tests of significance for non-nested hypotheses," *Journal of Econometrics*, **16**,103-19.

Gallant, A. R. (1987). *Nonlinear Statistical Models*, New York, John Wiley & Sons.

Gallant, A. R., and H. White (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Oxford, Basil Blackwell.

Godfrey, L. G. (1976). "Testing for serial correlation in dynamic simultaneous equation models," *Econometrica*, **44**, 1077-84.

Godfrey, L. G. (1978a). "Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables," *Econometrica*, **46**, 1293-1301.

Godfrey, L. G. (1978b). "Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables," *Econometrica*, **46**, 1303-10.

Godfrey, L. G. (1981). "On the invariance of the Lagrange multiplier test with respect to changes in the alternative hypothesis," *Econometrica*, **49**, 1443-55.

Godfrey, L. G. (1983). "Testing non-nested models after estimation by instrumental variables or least squares," *Econometrica*, **51**, 355-65.

Godfrey, L. G. (1985). "On the uses of misspecification checks and tests of non-nested hypotheses in empirical econometrics," *The Economic Journal*, **94**, 69-81.

Godfrey, L. G. (1988). *Misspecification Tests in Econometrics*, Cambridge, Cambridge University Press.

Godfrey, L. G., and M. H. Pesaran (1983). "Tests of non-nested regression models: small sample adjustments and Monte Carlo evidence," *Journal of Econometrics*, **21**, 133-54.

Godfrey, L. G., and A. R. Tremayne (1988). "Chekcs of model adequacy for univariate time series models and their application to econometric relationships," (Discussion and Reply) *Econometric Reviews*, **7**, 1-64.

Godfrey, L. G., and M. R. Wickens (1982). "Tests of misspecification using locally equivalent alternative models," in *Evaluating the Reliability of Econometric Models*, eds. G. C. Chow and P. Corsi, New York, John Wiley & Sons, 71-99.

Gourieroux, C., A. Monfort, & A. Trognon (1983). "Testing nested or non-nested hypotheses," *Journal of Econometrics*, **21**, 83-115.

Greene, W. H. (1990). *Econometric Analysis*, New York, Macmillan.

Hall, A. D. (1983). "Confidence contours for two test statistics for non-nested regression models," *Journal of Econometrics*, **21**, 155-160.

Hall, A. (1985). "A simplified method of calculating the distribution free Cox test," *Economics Letters*, **18**, 149-51.

Hall, A. (1986). "Corrijendum: a simplified method of calculating the distribution free Cox test," *Economic Letters*, **22**, 321.

Hatanaka, M. (1974). "An efficient two-step estimator for the dynamic adjustment model with autoregressive errors," *Journal of Econometrics*, **2**, 199-220.

Hatanaka, M. (1976). "Several efficient two-step estimators for the dynamic simultaneous equations model with autoregressive disturbances," *Journal of Econometrics*, **4**, 189-204.

Hausman, J. A. (1978). "Specification tests in econometrics," *Econometrica*, **46**, 1251-72.

Hausman, J. A., & M. H. Pesaran (1983). "The j-test as a hausman specification test," *Economics Letters*, **12**, 277-81.

Hendry, D. F., and G. E. Mizon (1978). "Serial correlation as a convenient simplification not a nuisance: a comment on a study of the demand for money by the Bank of England," *Economic Journal*, **88**, 549-63.

Hendry, D. F., and J. F. Richard (1983). "On the formulation of empirical models in dynamic econometrics," *Journal of Econometrics*, **20**, 3-33.

Hendry, D. F., A. R. Pagan, and J. D. Sargan (1984). "Dynamic specification," Ch. 18 in *Handbook of Econometrics*, Vol.II, eds. Z. Griliches and M. D. Intriligator, Amsterdam, North-Holland.

Holly, A. (1982). "A remark on Hausman`s specification test," *Econometrica*, **50**, 749-59.

Holly, A., and A. Monfort (1986). "Some useful equivalence properties of Hausman`s test," *Economics Letters*, **20**, 39-43.

Johannes, J. M. and A. R. Nasseh (1985). "Income or wealth in money demand: An application of non-nested hypothesis tests," *Southern Economic Journal*, **51**, 1099-106.

Judge, G. G., R. C. Hill, W. E. Griffiths, H. Lütkepohl, and T.-C. Lee (1985). *The Theory and Practice of Econometrics*, Second edition, New York, John Wiley & Sons.

Kent, J. T. (1986). "The underlying structure of nonnested hypothesis tests," *Biometrika*, **73**, 333-343.

King, M. L. (1985). "A point-optimal test for autoregressive disturbances," *Journal of Econometrics*, **27**, 21-37.

King, M. L. (1987). "Towards a theory of point optimal testing,"(with discussions and reply), *Econometric Reviews*, **6**, 169-255.

King, M. L., and M. J. McAleer (1987). "Further results on testing AR(1) against MA(1) disturbances in the linear regression model," *Review of Economic Studies*, **54**, 649-63.

Kiviet, J. F. (1986). "On the rigour of some misspecification tests for modelling dynamic relationships," *Review of Economic Studies*, **53**, 241-61.

MacKinnon, J. G. (1983). "Model specification tests against non-nested alternatives" (with discussion and reply), *Econometric Reviews*, **2**, 85-158.

MacKinnon, J. G., H. White, and R. Davidson (1983). "Tests for model specification in the presence of alternative hypotheses: some further results," *Journal of Econometrics*, **21**, 53-70.

Maddala, G. S., and A. S. Rao (1973). "Tests for serial correlation in regression models with lagged dependent variables and serially correlated errors," *Econometrica*, **41**, 761-74.

Magee, L. (1987). "A note on Cochrane-Orcutt estimation," *Journal of Econometrics*, **35**, 211-18.

McAleer, M. (1981). " A small sample test for non-nested regression models," *Economics Letters*, **7**, 335-338.

McAleer, M. (1983). "Exact tests of a model against nonnested alternatives," *Biometrika*, **70**, 285-8.

McAleer, M. J. (1987). "Specification tests for separate models: a survey," in *Specification Analysis in the Linear Model*, eds. M. L. King and D. E. A. Giles, London, Routledge and Kegan Paul, 146-95.

McAleer, M., & G. R. Fisher (1982). "Testing separate regression models subject to specification error," *Journal of Econometrics*, **19**, 125-45.

McAleer, M., C. R. McKenzie and A. D. Hall (1988). "Testing separate time series models," *Journal of Time Series Analysis*, **9**, 169-189.

McAller. M., & M. H. Pesaran (1986). "Statistical inference in non-nested econometric models," *Applied Mathematics and Computation*, **20**, 271-311.

McAleer, M. J., M. H. Pesaran, and A. K. Bera (1990). "Alternative approaches to testing non-nested models with autocorrelated disturbances: an application to models of U. S. unemployment," *Communications in Statistics*, Series A, **19**, 3619-44.

Milbourne, R. (1985). "Distinguishing between Australian demand for money models," *Australian Economic Papers*, **24**, 154-68.

Miliken, G. A., and F. A. Graybill (1970). "Extensions of the general linear hypothesis model," *Journal of the American Statistical Association*, **65**, 797-807.

Mizon, G. E. (1984). "The encompassing approach in econometrics," in *Quantitative Economics and Econometric Analysis*, eds. K. F. Wallis and D.F. Hendry, Oxford, Basil Blackwell, 135-72.

Mizon, G. E., and D. F. Hendry (1980). "An empirical and Monte Carlo analysis of tests of dynamic specification," *Review of Economic Studies*, **47**, 21-45.

Mizon, G. E., and J.-F. Richard (1986). "The encompassing principle and its application to testing non-nested hypotheses," *Econometrica*, **54**, 657-78.

Pagan, A. R. (1984). "Model evaluation by variable addition," Ch. 5 in *Quantitative Economics and Econometrica Analysis*, eds. K. F. Wallis and D. F. Hendry, Oxford, Basil Blackwell.

Pesaran, M. H. (1974). "On the general problem of model selection," *Review of Economic Studies*, **41**, 153-71.

Pesaran, M. H. (1981). "Pitfalls of testing non-nested hypotheses by the Lagrange multiplier method," *Journal of Econometrics*, **17**, 323-31.

Pesaran, M. H. (1982a). "On the Comprehensive method of testing non-nested regression models," *Journal of Econometrics*, **18**, 263-274.

Pesaran, M. H. (1982b). "Comparison of local power of alternative tests of non-nested regression models," *Econometrica*, **50**, 1287-1305.

Pesaran, M. H. (1987). "Global and partial non-nested hypotheses and asymptotic local power," *Econometric Theory*, **3**, 69-97.

Pesaran, M. H., and A. S. Deaton, (1978). "Testing non-nested nonlinear regression models," *Econometrica*, **46**, 677-94.

Pesaran, H., and A. Hall (1989). "A test of non-nested linear regression models subject to linear restrictions," *Economics Letters*, **2**, pp 341-348.

Pesaran, M. H. and B. Pesaran (1993). "A simulation approach to the problem of computing cox's statistics for testing nonnested models," *Journal of Econometrics*, **57**, 377-392.

Quandt, R. E. (1974). "A comparison of methods for testing nonnested hypotheses," *Review of Economics and Statistics*, **56**, 92-99.

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

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

Sargan J. D. (1980). "Some tests of dynamic specification for a single equation," *Econometrica*, **48,** 879-97.

Sargan, J. D., and F. Mehta (1983). "A generalization of the Durbin significance test and its application to dynamic specification," *Econometrica*, **51**, 1551-67.

Savin, N. E. (1976). "Conflict among testing procedures in a linear regression model with autoregressive disturbances," *Econometrica*, **44**, 1303-15.

Sawyer, K. R. (1983). "Testing separate families of hypotheses: an information criterion," *Journal of the Royal Statistical Society, B,* **45**, 89-99.

Seber, G. A. F., and C. J. Wild (1989). *Nonlinear Regression*, New York, John Wiley & Sons.

Silvapulle, P., and M. L. King (1993). "Non-nested testing for autocorrelation in the linear regression model," *Journal of Econometrics*, **58**, 295-314.

Smith, R. J. (1987). "Alternative asymptotically optimal tests and their application to dynamic specification," *Review of Economic Studies*, **54**, 665-80.

Smith, R. J. (1992). "Non-nested tests for competing models estimated by generalized method of moments," *Econometrica*, **60**, 973- 80.

Spanos, A. (1986). *Statistical Foundations of Econometric Modelling*. Cambridge, Cambridge University Press.

Spencer, B. G. (1975). "The small sample bias of Durbin`s tests for serial correlation when one of the regressors is the lagged dependent variable," *Journal of Econometrics*, **3**, 249-54.

Stewart, J. (1991). *Econometrics. London, Philip Allan.*

Thornton, D. L. (1985). "The appropriate interest rate and scale variable in money demand: Results from non-nested tests," *Applied Economics*, **17**, 735-44.

Thornton, D. L. (1987). "A note on the effect of the Cochrane-Orcutt estimator of the AR(1) regression model," *Journal of Econometrics*, **36**, 369-76.

Vuong, Q. H. (1989). "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica*, **57**, 307-334.

Wallis, K. F. (1972). "Testing for fourth-order autocorrelation in quarterly regression equations," *Econometrica*, **40**, 617-36.

White, H. (1982). "Regularity conditions for cox's test of non-nested hypotheses," *Journal of Econometrics*, **19**, 301-318.

White, H. (1984). *Asymptotic Theory for Econometricians*, Orlando, Academic Press.

White, H. (1987). "Specification Testing in Dynamic models," Ch. 1 in *Advances in Econometrics-Fifth World Congress*, Vol. 1, ed. T. Bewley, Cambridge, Cambridge University Press.

White, H., and I. Domowitz (1984). "Nonlinear regression with dependent observations," *Econometrica*, **52**, 143-61.

Wooldridge, J. M. (1990a). "A note on the Lagrange multiplier and *F* statistics for two stage least squares regressions," *Economics Letters*, **34**, 151-55.

Wooldridge, J. M. (1990b). "A unified approach to robust, regression-based specification tests," *Econometric Theory*, **6**, 17-43.

Wooldridge, J. M. (1990c). "An encompassing approach to conditional mean tests with applications to testing non-nested hypotheses," Journal of *Econometrics*, **45**, 331-50.

Wu, D.-M. (1973). "Alternative tests of independence between stochastic regressors and disturbances," *Econometrica*, **41**, 733-50.