SERIE RESEARCH MEMORANDA

ARMAX MODELS: ESTIMATION AND TESTING

Herman J. Bierens

Research Memorandum 1988-64 Dec. '88

VRIJE UNIVERSITEIT
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN
EN ECONOMETRIE
AMSTERDAM

ARMAX MODELS: ESTIMATION AND TESTING *

Herman J.Bierens
Department of Econometrics
Free University, Amsterdam

December 1988

* Chapter 11 of:
PRINCIPLES OF NONLINEAR AND NONPARAMETRIC REGRESSION ANALYSIS
To be published by Marcel Dekker Inc., New York.

PRINCIPLES OF NONLINEAR AND NONPARAMETRIC REGRESSION ANALYSIS

TABLE OF CONTENTS:

PREFACE

This book deals with statistical inference of nonlinear regression models from two opposite points of view, namely the case where the functional form of the model is completely specified as a known function of regressors and unknown parameters, and the opposite case where the functional form of the model is completely unknown. First it is assumed that the response function of the regression model under review belongs to a certain well-specified parametric family of functional forms, by which estimation of the model merely amounts to estimation of the unknown parameters. For this class of models we review the asymptotic properties of the nonlinear least squares estimator for independent data as well as for time series.

In practice assumptions on the functional form are often made on the basis of computational convenience rather than on the basis of precise a priori knowledge of the empirical phenomenon under review. Therefore the linear regression model is still the most popular model specification in applied research. However, even if the specification of the functional form is based on sound theoretical considerations there is quite often a large range of functional forms that are theoretically admissible, so that there is no guarantee that the actually chosen functional form is true. Functional specification of a parametric nonlinear regression model should therefore always be verified by conducting model misspecification tests. Various model misspecification tests will therefore be discussed, in particular consistent tests which have asymptotic power 1 against all deviations from the null hypothesis that the model is correct.

The opposite case of parametric regression is nonparametric regression. Nonparametric regression analysis is concerned with estimation of a regression model without specifying in advance its functional form. Thus the only source of information about the functional form of the model is the data set itself. In this book we shall review various nonparametric regression approaches, with special emphasis on the kernel method, under various distributional assumptions.

This book is divided into three parts. In the first part we review the elements of abstract probability theory we need in part 2. Part 2 is devoted to the asymptotic theory of parametric and nonparametric regression analysis in the case of independent data generating processes. In part 3 we extend the analysis involved to time series.

The selection of the topics mainly reflexes my own interest in the subject. Instead of providing an encyclopedic survey of the literature, I have chosen for a setup which aims to fill the gap between intermediate statistics (including linear time series analysis) and the level necessary to get access to the recent literature on nonlinear and nonparametric regression analysis, with emphasis on my own contributions. The ultimate goal is to provide the student with the tools for his own independent research in this area, by showing what tools I and others have used and what they have been used for. Thus, this book may be viewed as an account of my own struggle with the material involved. I think this book is particularly suitable for self-tuition (at least it aims to be), and may prove useful in a graduate course in mathematical statistics and advanced econometrics.

Acknowledgements:

The first five chapters of this book have been disseminated in draft form as working papers. I am grateful to Anil Bera, Alexander Georgiev and Jan Magnus for suggesting additional references, and in particular to Lourens Broersma, Johan Smits and Ton Steerneman who suggested various improvements.

A large body of the material in chapter 6 has been published earlier in Truman F.Bewley (ed.), *Advances in Econometrics, Fifth World Congress*, Cambridge University Press. I am indebted to Cambridge University Press for granting permission to reprint it.

11. ARMAX MODELS: ESTIMATION AND TESTING

In this chapter we first consider the asymptotic properties of least squares estimators of the parameters of a linear ARMAX models, and then we extend the results involved to nonlinear ARMAX models. A new feature of our approach is that we allow the X-variables to be stochastic time series themselves, possibly depending on lagged Y-values. Moreover, we allow the data generating process to be heterogeneous. Furthermore, we propose consistent tests of the null hypothesis that the errors are martingale differences, and a less general but easier test of the null hypothesis that the errors are uncorrelated. Most of these results are obtained by a further elaboration of the results in Bierens (198?).

11.1 *Estimation of linear ARMAX models*

11.1.1 *Introduction*

We recall that, given a k-variate time series process $\{(Y_t,X_t)\}$, where Y_t and the k-1 components of X_t are real-valued random variables, the linear ARMAX model assumes the form:

$$(1-\Sigma_{s=1}^{p}\alpha_s L^s)Y_t = \mu + \Sigma_{s=1}^{r}\beta_s{}'L^s X_t + (1+\Sigma_{s=1}^{q}\gamma_s L^s)U_t, \qquad (11.1.1)$$

where L is the usual lag operator, $\mu \in R$, $\alpha_s \in R$, $\beta_s \in R^{k-1}$ and $\gamma_s \in R$ are unknown parameters, the U_t's are the errors and p, q and r are natural numbers specified in advance. The exclusion of X_t in this model is no loss of generality, as we may replace X_t by $X_t^* = X_{t+1}$.

The correctness of this linear ARMAX model specification now corresponds to the null hypothesis

$$H_0: E(U_t|(Y_{t-1},X_{t-1}),(Y_{t-2},X_{t-2}),.....) = 0 \text{ a.s.} \qquad (11.1.2)$$

for each t. Assuming that the lag polynomial $1+\Sigma_{s=1}^{r}\gamma_s L^s$ is invertible, this hypothesis implies that the ARMAX model (11.1.1) represents the mathematical expectation of Y_t conditional on the entire past of the process.

The ARMAX model specification is particularly suitable for macroeconomic vector time series modeling without imposing a priori restrictions prescribed by macroeconomic theory. Such macroeconomic analysis has been advocated and conducted by Sims

(1980, 1981) and Doan, Litterman and Sims (1984) in the framework of unrestricted vector autoregressions and observable index models. Cf. Sargent and Sims(1977) for the latter models. The advantage of ARMAX models over the VAR models used by Sims (1980) is that ARMAX models allow an infinite lag structure with a parsimonious parametrisation, by which we get a tractable model that may better reflect the strong dependence of macroeconomic time series.

Estimation of the parameters of a linear ARMAX model for the case that the X_t's are exogenous (in the sense that the X_t's are either nonstochastic or independent of the U_t's) has been considered by Hannan, Duinsmuir and Deistler (1980), among others. This estimation theory, however, is not straightforwardly applicable to the model under review. The condition that the X_t's are exogenous is too strong a condition, as then feedback from Y_t to X_t is excluded. Also, we do not assume that the errors U_t are Gaussian or independent, but merely that (U_t) is a martingale difference sequence.

We recall that the ARMAX model (11.1.1) represents the conditional expectation of Y_t given the entire past of the process $\{(Y_t,X_t)\}$, provided condition (11.1.2) holds and the MA lag polynomial $1+\Sigma_{s=1}^{q}\gamma_s L^s$ is invertible. We then have

$$E(Y_t|(Y_{t-1},X_{t-1}),(Y_{t-2},\ X_{t-2}),\ldots\ldots)$$

$$= \mu\ /\ (1 + \Sigma_{s=1}^{q}\gamma_s)$$

$$+ \{(\Sigma_{s=1}^{p}\alpha_s L^s + \Sigma_{s=1}^{q}\gamma_s L^s)/(1 + \Sigma_{s=1}^{q}\gamma_s L^s)\}Y_t$$

$$+ \{(\Sigma_{s=1}^{r}\beta_s' L^s)/(1 + \Sigma_{s=1}^{q}\gamma_s L^s)\}X_t \text{ a.s.} \qquad (11.1.3)$$

and

$$U_t = Y_t - E(Y_t|(Y_{t-1},X_{t-1}),(Y_{t-2},X_{t-2}),\ldots) \text{ a.s.} \qquad (11.1.4)$$

Since the MA lag polynomial can be written as

$$1 + \Sigma_{s=1}^{q}\gamma_s L^s = \Pi_{s=1}^{q}(1 - \lambda_s L),$$

where $\lambda_1^{-1},\ldots\ldots,\lambda_q^{-1}$ are its possibly complex-valued roots, invertibility requires $|\lambda_s| < 1$ for s=1,...,r. In particular, for $0 < \delta < 1$ the set

$$\Gamma_\delta = \{(\gamma_1,\ldots,\gamma_q)' \in R^q\colon |\lambda_s| \le 1-\delta \text{ for } s=1,\ldots,q\} \quad (11.1.5)$$

is a compact set of vectors $(\gamma_1,\ldots,\gamma_q)'$ with this property. The compactness of Γ_δ follows from the fact that the γ_s's are continuous functions of $\lambda_1,\ldots,\lambda_q$, hence Γ_δ is the continuous image of a compact set and therefore compact itself. Cf. Royden (1968, Proposition 4, p. 158).

From now on we assume that there are known compact subsets M of R, A of R^p, B of $R^{(k-1)r}$ and Γ_δ of R^q such that, if (11.1.2) is true,

$$\mu \in M,\ (\alpha_1,\ldots,\alpha_p)' \in A,\ (\beta_1',\ \ldots,\beta_r')' \in B \text{ and } (\gamma_1,\ldots,\gamma_q)' \in \Gamma_\delta \quad (11.1.6)$$

Stacking all these parameters in a vector θ_0:

$$\theta_0 = (\mu,\alpha_1,\ldots,\alpha_p,\beta_1',\ldots,\beta_r',\gamma_1,\ldots,\gamma_q)' \quad (11.1.7)$$

and denoting the parameter space by

$$\Theta = M\times A\times B\times\Gamma_\delta \subset R^m \text{ with } m = 1 + p + (k-1)r + q, \quad (11.1.8)$$

which is a compact set, we thus have $\theta_0 \in \Theta$ if (11.1.2) is true.

Denoting $Z_t = (Y_t,X_t')'$, the conditional expectation (11.1.3) can now be written as

$$E(Y_t|Z_{t-1},Z_{t-2},\ldots) = \varphi(\theta_0) + \Sigma_{s=1}^{\infty}\eta_s(\theta_0)'Z_{t-s}, \quad (11.1.9)$$

where

$$\varphi(\theta_0) = \mu/(1 + \Sigma_{s=1}^{q}\gamma_s)$$

and the $\eta_s(.)$ are continuously differentiable vector-valued functions defined by:

$$(1 + \Sigma_{s=1}^{q}\gamma_s L^s)(\Sigma_{s=1}^{\infty}\eta_s(\theta_0)L^s) = (\Sigma_{s=1}^{p}\alpha_s L^s + \Sigma_{s=1}^{q}\gamma_s L^s,\ \Sigma_{s=1}^{r}\beta_s' L^s)' \quad (11.1.10)$$

It is not too hard to verify that each component $\eta_{i,s}(\theta)$ of $\eta_s(\theta)$ satisfies

$$\Sigma_{s=1}^{\infty}\sup_{\theta\in\Theta}|\eta_{i,s}(\theta)| < \infty, \tag{11.1.11}$$

$$\Sigma_{s=1}^{\infty}\sup_{\theta\in\Theta}|(\partial/\partial\theta_j)\eta_{i,s}(\theta)| < \infty \ (j=1,2,\ldots,m) \tag{11.1.12}$$

$$\Sigma_{s=1}^{\infty}\sup_{\theta\in\Theta}|(\partial/\partial\theta_j)(\partial/\partial\theta_\ell)\eta_{i,s}(\theta)| < \infty \ (j,\ell=1,2,\ldots\ldots,m) \tag{11.1.13}$$

etc. Cf. exercise 1.These properties will play a crucial role in our estimation theory. In particular, the model (11.1.3) can now be written as a nonlinear regression model:

$$Y_t = g_t(\theta_0) + U_t. \tag{11.1.14}$$

where the response function

$$g_t(\theta) = \varphi(\theta) + \Sigma_{s=1}^{\infty}\eta_s(\theta)'Z_{t-s} \tag{11.1.15}$$

and its first and second partial derivatives are well-defined random functions.

Assuming that only $Z_1,\ldots,Z_n$ have been observed, we now propose to estimate θ_0 by nonlinear least squares, as follows. Let

$$\tilde{g}_t(\theta) = \varphi(\theta) + \Sigma_{s=1}^{t-1}\eta_s(\theta)'Z_{t-s} \text{ if } t \geq 2,$$
$$\tilde{g}_t(\theta) = \varphi(\theta) \text{ if } t \leq 1. \tag{11.1.16}$$

Thus (11.1.16) is $g_t(\theta)$ with Z_t set equal to the zero vector for $t < 1$. Alternatively, we may set $Z_t = Z_1$ for $t < 1$, but for convenience the analysis below will be conducted for the case (11.1.16) only. Moreover, denote

$$\hat{Q}(\theta) = (1/n)\Sigma_{t=1}^{n}(Y_t - \tilde{g}_t(\theta))^2. \tag{11.1.17}$$

Then the proposed least squares estimator $\hat{\theta}$ of θ_0 is a (measurable) solution of

$$\hat{\theta} \in \Theta : \quad \hat{Q}(\hat{\theta}) = \inf_{\theta \in \Theta} \hat{Q}(\theta). \tag{11.1.18}$$

Similar to the results in chapter 4 we can set forth conditions such that under the null hypothesis (11.1.2),

$$\sqrt{n}(\hat{\theta} - \theta_0) \to N_m(0, \Omega_1^{-1}\Omega_2\Omega_1^{-1}) \text{ in distr.}, \tag{11.1.19}$$

where Ω_1 is the probability limit of

$$\hat{\Omega}_1 = (1/n)\Sigma_{t=1}^{n}\{(\partial/\partial\theta')g_t(\hat{\theta})\}\{(\partial/\partial\theta)g_t(\hat{\theta})\} \tag{11.1.20}$$

and Ω_2 is the probability limit of

$$\hat{\Omega}_2 = (1/n)\Sigma_{t=1}^{n}\{Y_t - g_t(\hat{\theta})\}^2\{(\partial/\partial\theta')g_t(\hat{\theta})\}\{(\partial/\partial\theta)g_t(\hat{\theta})\}. \tag{11.1.21}$$

Moreover, when the null hypothesis (11.1.2) is false we show that there exists a $\theta^* \in \Theta$ such that

$$\text{plim}_{n\to\infty}\hat{\theta} = \theta^*. \tag{11.1.22}$$

11.1.2 *Consistency and asymptotic normality*

In this section we set forth conditions such that the results in section 11.1.1 hold.

Assumption 11.1.1. The data generating process (Z_t) in R^k, with $Z_t = (Y_t, X_t')$, is ν-stable in L^1 with respect to an α or φ-mixing base, where either $\Sigma_{j=0}^{\infty}\varphi(j) < \infty$ or $\Sigma_{j=0}^{\infty}\alpha(j) < \infty$, and is properly heterogeneous. Moreover, $\sup_t E|Z_t|^{4+\delta} < \infty$ for some $\delta > 0$.

(Cf. Definitions 9.2.2, 9.2.3 and 9.4.1 and theorem 9.4.1).

In the sequel we shall denote the base involved by (v_t) where $v_t \in V$ with V a Euclidean space, and the mean process of (Z_t) (cf. definition 9.4.1) will be denoted by (Z_t^*). It should be noted that the error U_t of the ARMAX model (11.1.1) need not be a component of v_t, as it is possible that the U_t's themselves are generated by a one-sided infinite sequence of v_t's.

If we would make the strict stationarity assumption then assumption 11.1.1 simplifies to:

*Assumption 11.1.1**. There exist a strictly stationary α or φ-mixing process (v_t) in a Euclidean space V, with α and φ the same as in assumption 11.1.1, and a Borel measurable mapping G from the space of one-sided infinite sequences in V into R^k such that

$$Z_t = (Y_t, X_t')' = G(v_t, v_{t-1}, v_{t-2}, \ldots\ldots) \text{ a.s.}$$

Moreover, $E|Z_t|^{4+\delta} < \infty$ for some $\delta > 0$.

Thus assumption 11.1.1* implies assumption 11.1.1. The proof of this proposition follows straightforwardly from theorem 9.2.2 and the fact that by the strict stationarity assumption the proper heterogeneity condition automatically holds with mean process (Z_t).

Next consider the function $\hat{Q}(\theta)$ defined by (11.1.17). Let Y_0^* be the first component of Z_0^* and let

$$\overline{Q}(\theta) = E(Y_0^* - \varphi(\theta) - \Sigma_{s=1}^{\infty}\eta_s(\theta)'Z_{-s}^*)^2. \qquad (11.1.23)$$

Then it follows from theorem 9.4.1:

Theorem 11.1.1. Under assumption 11.1.1,

$$\text{plim}_{n\to\infty}\sup_{\theta\in\Theta}|\hat{Q}(\theta) - \overline{Q}(\theta)| = 0.$$

Proof: Condition (9.4.1) is implied by (11.1.11). Since the function ψ in theorem 9.4.1 is now $\psi(.) = (.)^2$, condition (9.4.2) holds with $\mu = 1$. The other conditions of theorem 9.4.1 follow now from assumption 11.1.1. Q.E.D.

Next, we assume

Assumption 11.1.2. There exists a unique $\theta_* \in \Theta$ such that

$$\overline{Q}(\theta_*) = \inf_{\theta\in\Theta}\overline{Q}(\theta).$$

Since Θ is compact and $\overline{Q}(\theta)$ is continuous there is always a θ_* in Θ which minimizes $\overline{Q}(\theta)$ over Θ. Thus the actual contents of this assumption is the uniqueness of θ_*. If the null hypothesis (11.1.2) is true then $\theta_* = \theta_0$, so that assumption 11.1.2 then identifies the parameters of model (11.1.1). However, this assumption is also supposed to hold in the case that the null hypothesis (11.1.2) is false.

Applying theorem 4.2.1 it follows from theorem 11.1.1 and assumption 11.1.2:

Theorem 11.1.2. Under assumptions 11.1.1 and 11.1.2 the least squares estimator $\hat{\theta}$ defined by (11.1.18) satisfies

$$\text{plim}_{n\to\infty}\hat{\theta} = \theta_*.$$

This proves (11.1.22).

Next we show the consistency and asymptotic normality of $\hat{\theta}$ under the assumption that (11.1.2) holds for each t. Since (11.1.1) and (11.1.2) are equivalent to (11.1.9), we now assume:

Assumption 11.1.3. There exists a point θ_0 in an open convex subset Θ_0 of Θ, such that (11.1.9) holds for each t.

This assumption is hardly a condition. The sets M, A, and B [cf. (11.1.6) and (11.1.8)] can be chosen to be the closures of open convex sets M_0, A_0 and B_0, respectively, whereas the set

$$\Gamma_{0,\delta} = \{(\gamma_1,\ldots,\gamma_q)' \in R^q\colon \lambda_s < 1-\delta \text{ for } s=1,2,\ldots,q\}$$

(cf. (11.1.5)) is for $\delta \in (0,1)$ the continuous image of an open set and therefore open itself, with closure Γ_δ. Assuming that γ_0 corresponding to θ_0 is an interior point of Γ_δ for some δ, hence $\gamma_0 \in \Gamma_{0,\delta}$, there exists an open convex neighborhood Γ_0 of γ_0 contained in Γ_δ. Thus $\Theta_0 = M_0 \times A_0 \times B_0 \times \Gamma_0$ (cf. (11.1.7)) is then an open convex subset of Θ.

In order to establish the consistency of the least squares estimator $\hat{\theta}$ it suffices to show that θ_0 minimizes $\overline{Q}(\theta)$, as then by the uniqueness condition in assumption 11.1.2, θ_* must be equal to θ_0. To show this, let

$$\tilde{Q}(\theta) = (1/n)\Sigma_{t=1}^{n}\{Y_t - \varphi(\theta) - \Sigma_{s=1}^{\infty}\eta_s(\theta)'Z_{t-s}\}^2. \quad (11.1.24)$$

It follows from lemmas 9.4.2 and 9.4.3 that

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}|E\ \tilde{Q}(\theta) - \bar{Q}(\theta)| = 0 \qquad (11.1.25)$$

and from (11.1.4) and (11.1.9) it follows that under assumption 11.1.3

$$E\ \tilde{Q}(\theta) = (1/n)\Sigma_{t=1}^{n} E\ U_t^2$$
$$+ (1/n)\Sigma_{t=1}^{n} E\{\varphi(\theta) - \varphi(\theta_0) + \Sigma_{s=1}^{\infty}(\eta_s(\theta) - \eta_s(\theta_0))'Z_{t-s}\}^2. \qquad (11.1.26)$$

Consequently, under assumption 11.1.3 we have:

$$\bar{Q}(\theta) = \lim_{n\to\infty}\Sigma_{t=1}^{n} E\ U_t^2$$
$$+ E\{\varphi(\theta) - \varphi(\theta_0) + \Sigma_{s=1}^{\infty}(\eta_s(\theta) - \eta_s(\theta_0))'Z_{-s}^{*}\}^2. \qquad (11.1.27)$$

Clearly θ_0 minimizes $\bar{Q}(\theta)$ and hence $\theta_0 = \theta_*$. This proves:

Theorem 11.1.3. Under assumptions 11.1.1, 11.1.2 and 11.1.3,

$$\text{plim}_{n\to\infty}\hat{\theta} = \theta_0.$$

The asymptotic normality proof follows the classical lines. Cf. chapter 4. Thus we first apply the mean value theorem to $(\partial/\partial\theta_i)\hat{Q}(\hat{\theta})$, where θ_i is the i-th component of θ. This yields

$$(\partial/\partial\theta_i)\hat{Q}(\hat{\theta}) = (\partial/\partial\theta_i)\hat{Q}(\theta_0)$$
$$+ (\hat{\theta} - \theta_0)'(\partial/\partial\theta')(\partial/\partial\theta_i)\hat{Q}(\tilde{\theta}^{(i)}), \qquad (11.1.28)$$

with $\tilde{\theta}^{(i)}$ a mean value satisfying

$$|\tilde{\theta}^{(i)} - \theta_0| \leq |\hat{\theta} - \theta_0| \text{ a.s.} \qquad (11.1.29)$$

Theorem 11.1.3 and assumption 11.1.3 imply that $\hat{\theta}$ is an interior point of Θ with probability converging to 1. Thus $\hat{\theta}$

and $\tilde{\theta}^{(i)}$ are with probablity converging to 1 contained in the open convex subset Θ_0 of Θ. Cf. assumption 11.1.3. Consequently, we have similarly to (4.2.7),

$$\text{plim}_{n\to\infty}\sqrt{n}(\partial/\partial\theta_i)\hat{Q}(\hat{\theta}) = 0. \qquad (11.1.30)$$

The next step is to establish

$$\text{plim}_{n\to\infty}\{(\partial/\partial\theta')(\partial/\partial\theta_1)\hat{Q}(\tilde{\theta}^{(1)}),\ldots,(\partial/\partial\theta')(\partial/\partial\theta_m)\hat{Q}(\tilde{\theta}^{(m)})\}$$
$$= 2\Omega_1, \qquad (11.1.31)$$

where Ω_1 is a nonsingular matrix, and the last step is to show

$$\sqrt{n}(\partial/\partial\theta')\hat{Q}(\theta_0) \to N_m(0,4\Omega_2). \qquad (11.1.32)$$

Combining (11.1.28) through (11.1.32) then yields

$$\sqrt{n}(\hat{\theta} - \theta_0) \to N_m(0,\Omega_1^{-1}\Omega_2\Omega_1^{-1}) \text{ in distr.} \qquad (11.1.33)$$

For proving (11.1.31) and (11.1.32) the following lemma is convenient. Let

$$g^*(\theta) = \varphi(\theta) + \Sigma_{s=1}^{\infty}\eta_s(\theta)'Z_{-s}^*, \qquad (11.1.34)$$

where we recall that (Z_t^*) is the mean process of (Z_t).

Lemma 11.1.1: Under assumptions 11.1.1 and 11.1.3,

$$E(Y_0^* - g^*(\theta_0))Z_{-s}^* = 0 \text{ for } s = 1,2,\ldots,$$

where Y_0^* is the first component of Z_0^*.

Proof: Since $E(Y_t - g_t(\theta_0))Z_{t-s} = E\,U_t Z_{t-s} = 0$ under assumption 11.1.3 and since by theorem 9.4.1,

$$\lim_{n\to\infty}E(Y_t - g_t(\theta_0))Z_{t-s} = E(Y_0^* - g^*(\theta_0))Z_{-s}^*,$$

the lemma follows. Q.E.D.

Now consider the derivatives

$$(\partial/\partial\theta_i)\hat{Q}(\theta) = -2(1/n)\Sigma_{t=1}^{n}(Y_t - \tilde{g}_t(\theta))(\partial/\partial\theta_i)\tilde{g}_t(\theta),$$

$$(\partial/\partial\theta_i)(\partial/\partial\theta_j)\hat{Q}(\theta)$$

$$= 2(1/n)\Sigma_{t=1}^{n}\{(\partial/\partial\theta_i)\tilde{g}_t(\theta)\}\{(\partial/\partial\theta_j)\tilde{g}_t(\theta)\}$$

$$- 2(1/n)\Sigma_{t=1}^{n}\{Y_t - \tilde{g}_t(\theta)\}(\partial/\partial\theta_i)(\partial/\partial\theta_j)\tilde{g}_t(\theta).$$

It follows from theorem 9.4.1 and (11.1.11), (11.1.12) and (11.1.13) that

Lemma 11.1.2. Under assumptions 11.1.1 and 11.1.2,

$$\text{plim}_{n\to\infty}\sup_{\theta\in\Theta}\left|(1/n)\Sigma_{t=1}^{n}\{(\partial/\partial\theta_i)\tilde{g}_t(\theta)\}\{\partial/\partial\theta_j)\tilde{g}_t(\theta)\}\right.$$

$$\left. - E\{(\partial/\partial\theta_i)g^*(\theta)\}\{(\partial/\partial\theta_j)g^*(\theta)\}\right| = 0 \quad (11.1.35)$$

and

$$\text{plim}_{n\to\infty}\sup_{\theta\in\Theta}\left|(1/n)\Sigma_{t=1}^{n}(Y_t - \tilde{g}_t(\theta))(\partial/\partial\theta_i)(\partial/\partial\theta_j)\tilde{g}_t(\theta)\right.$$

$$\left. - E(Y_0^* - g^*(\theta))(\partial/\partial\theta_i)(\partial/\partial\theta_j)g^*(\theta)\right| = 0, \quad (11.1.36)$$

Proof: We only prove (11.1.35). The proof of (11.1.36) is left as exercise 2. We verify the conditions of theorem 9.4.1. For t ≥ 2 we have

$$(\partial/\partial\theta_i)\tilde{g}_t(\theta) = (\partial/\partial\theta_i)\varphi(\theta) + \Sigma_{s=1}^{t-1}(\partial/\partial\theta_i)\eta_s(\theta)'Z_{t-s},$$

hence

$$\{(\partial/\partial\theta_i)\tilde{g}_t(\theta)\}\{(\partial/\partial\theta_j)\tilde{g}_t(\theta)\} = \{(\partial/\partial\theta_i)\varphi(\theta)\}\{(\partial/\partial\theta_j)\varphi(\theta)\}$$

$$+ \psi_1(\Sigma_{s=1}^{t-1}\Gamma_s^{(1)}(\theta)'Z_{t-s}) + \psi_1(\Sigma_{s=1}^{t-1}\Gamma_s^{(2)}(\theta)'Z_{t-s})$$

$$+ \psi_2(\Sigma_{s=1}^{t-1}\Gamma_s^{(3)}(\theta)'Z_{t-s}),$$

where

$$\Gamma_s^{(1)}(\theta)' = \{(\partial/\partial\theta_i)\varphi(\theta)\}\{(\partial/\partial\theta_j)\eta_s(\theta)'\},$$

$$\Gamma_s^{(2)}(\theta)' = \{(\partial/\partial\theta_j)\varphi(\theta)\}\{(\partial/\partial\theta_i)\eta_s(\theta)'\},$$

$$\Gamma_s^{(3)}(\theta)' = \begin{pmatrix} (\partial/\partial\theta_i)\eta_s(\theta)' \\ (\partial/\partial\theta_j)\eta_s(\theta)' \end{pmatrix},$$

and for ξ, ξ_1, $\xi_2 \in R$, $\psi_1(\xi) = \xi$, $\psi_2(\xi_1,\xi_2) = \xi_1\xi_2$. Moreover, the parameter μ in (9.4.2) is

$$\mu = 0 \text{ for } \psi_1, \ \mu = 1 \text{ for } \psi_2.$$

Now part (11.1.35) of the lemma follows easily from (11.1.11), (11.1.12), assumptions 11.1.1 and 11.1.2 and theorem 9.4.1. The proof of (11.1.36) is similar. Q.E.D.

Next, observe that lemma 11.1.1 implies

$$E(Y_0^* - g^*(\theta_0))(\partial/\partial\theta_i)(\partial/\partial\theta_j)g^*(\theta_0) = 0. \qquad (11.1.37)$$

Since

$$\text{plim}_{n\to\infty}\tilde{\theta}_i = \theta_0$$

because of (11.1.29) and theorem 11.1.3, (11.1.31) follows from (11.1.35), (11.1.36) and (11.1.37) with

$$\Omega_1 = E\{(\partial/\partial\theta')g^*(\theta_0)\}\{(\partial/\partial\theta)g^*(\theta_0)\}. \qquad (11.1.38)$$

The non-singularity of Ω_1 cannot be derived, but has to be assumed as part of the identification assumption. Moreover, in order that this matrix is also defined in the case that the null hypothesis (11.1.2) fails to hold we redefine Ω_1 as

$$\Omega_1 = E\{(\partial/\partial\theta')g^*(\theta_*)\}\{(\partial/\partial\theta)g^*(\theta_*)\}. \qquad (11.1.39)$$

Since $\theta_* = \theta_0$ under the null hypothesis, there is no loss of generality in doing so.

Assumption 11.1.4. The matrix Ω_1 defined by (11.1.39) is non-singular.

This assumption is part of the set of maintained hypotheses which are assumed to hold regardless whether or not the null hypothesis is true.

Using lemma 11.1.2 it easily follows that the matrix $\hat{\Omega}_1$ defined by (11.1.20) is a consistent estimator of Ω_1:

Lemma 11.1.3. Under assumptions 11.1.1 and 11.1.2,

$$\text{plim}_{n\to\infty}\hat{\Omega}_1 = \Omega_1 .$$

Note that assumption 11.1.3 is not needed for this result, due to the more general definition (11.1.39) of Ω_1.

For proving (11.1.32), we observe first that

$$\sqrt{n}(\partial/\partial\theta_i)\hat{Q}(\theta_0)$$

$$= -2(1/\sqrt{n})\Sigma_{t=1}^{n}(U_t + g_t(\theta_0) - \tilde{g}_t(\theta_0))(\partial/\partial\theta_i)\bar{g}_t(\theta_0)$$

$$= -2(1/\sqrt{n})\Sigma_{t=1}^{n}U_t(\partial/\partial\theta_i)\bar{g}_t(\theta_0)$$

$$+ 2(1/\sqrt{n})\Sigma_{t=1}^{n}(g_t(\theta_0) - \bar{g}_t(\theta_0))(\partial/\partial\theta_i)\bar{g}_t(\theta_0). \tag{11.1.40}$$

Secondly, we shall prove:

$$\text{plim}_{n\to\infty}(1/\sqrt{n})\Sigma_{t=1}^{n}(g_t(\theta_0) - \bar{g}_t(\theta_0)(\partial/\partial\theta_i)\bar{g}_t(\theta_0) = 0, \tag{11.1.41}$$

so that

$$\text{plim}_{n\to\infty}(\sqrt{n}(\partial/\partial\theta_i)\hat{Q}(\theta_0) + 2(1/\sqrt{n})\Sigma_{t=1}^{n}U_t(\partial/\partial\theta_i)\bar{g}_t(\theta_0)) = 0 \tag{11.1.42}$$

Thirdly, denoting

$$X_{n,t} = U_t(\partial/\partial\theta)\bar{g}_t(\theta_0)\xi, \tag{11.1.43}$$

cf. theorem 9.1.7, where ξ is an arbitrary non-random vector in R^m, we show that $(X_{n,t})$ is a sequence of martingale differences for which the martingale central limit theorem 9.1.7 applies:

$$(1/\sqrt{n})\Sigma_{t=1}^{n} X_{n,t} \rightarrow N(0,\sigma^2) \text{ in distr.}, \qquad (11.1.44)$$

where

$$\sigma^2 = \text{plim}_{n\rightarrow\infty}(1/n)\Sigma_{t=1}^{n} X_{n,t}^{2} = \xi'\Omega_2\xi \qquad (11.1.45)$$

with

$$\Omega_2 = E(Y_0^* - g^*(\theta_*))^2\{(\partial/\partial\theta')g^*(\theta_*)\}\{(\partial/\partial\theta)g^*(\theta_*)\}. \qquad (11.1.46)$$

From these results it then follows

$$\sqrt{n}(\partial/\partial\theta)\hat{Q}(\theta_0)\xi \rightarrow N(0,\xi'\Omega_2\xi) \text{ in distr. for every } \xi \in R^m, \qquad (11.1.47)$$

which implies (11.1.32).

For proving (11.1.41) we need the following extension of (11.1.11).

Lemma 11.1.4. For $s\rightarrow\infty$ and $i=1,\ldots\ldots,m$,

$$\sup_{\theta\in\Theta}|\eta_{i,s}(\theta)| = O(s^q(1-\delta)^s),$$

where δ is defined by (11.1.5).

Proof: It follows from (11.1.5) that

$$(1 + \Sigma_{s=1}^{q}\gamma_s L^s)^{-1} = \prod_{s=1}^{q}(1 - \lambda_s L)^{-1} = \prod_{s=1}^{q}(\Sigma_{\ell=0}^{\infty}\lambda_s^{\ell}L^{\ell})$$

$$= \Sigma_{s=0}^{\infty}(\Sigma_{\ell_1+\ell_2+\ldots+\ell_r=s,\ell_j=0,1,2,..,s}\lambda_1^{\ell_1}\lambda_2^{\ell_2}\ldots\lambda_r^{\ell_r})L^s$$

and

$$\Sigma_{\ell_1+\ell_2+\ldots+\ell_r=s,\ell_j=0,1,2,..,s}\lambda_1^{\ell_1}\lambda_2^{\ell_2}\ldots\lambda_r^{\ell_r}$$

$$\leq \Sigma_{\ell_1+\ell_2+\ldots+\ell_r=s,\ell_j=0,1,2,..,s}(1 - \delta)^s \leq s^q(1-\delta)^s.$$

Combining these results with (11.1.10), the lemma easily follows. Q.E.D.

Using this lemma, Cauchy-Schwarz inequality and the fact that

$$\Sigma_{s=t}^{\infty} s^q (1-\delta)^s = O(t^q (1-\delta)^t)$$

we see that there are constants C_* and C_{**} such that

$$E|(1/\sqrt{n})\Sigma_{t=1}^{n}(g_t(\theta_0) - \bar{g}_t(\theta_0))(\partial/\partial\theta_i)\bar{g}_t(\theta_0)|$$

$$\leq (1/\sqrt{n})\Sigma_{t=1}^{n}\Sigma_{s=t}^{\infty}|\eta_s(\vartheta_0)|E|Z_{t-s}||(\partial/\partial\theta_i)\bar{g}_t(\theta_0)|$$

$$\leq (1/\sqrt{n})\Sigma_{t=1}^{n}\Sigma_{s=t}^{\infty}C_* s^q(1-\delta)^s(E|Z_{t-s}|^2)^{\frac{1}{2}}(E|(\partial/\partial\theta_i)\bar{g}_t(\theta_0)|^2)^{\frac{1}{2}}$$

$$\leq \sup_t(E|Z_t|^2)^{\frac{1}{2}}(1/\sqrt{n})\Sigma_{t=1}^{n}(\Sigma_{s=t}^{\infty}C_* s^q(1-\delta)^s)(E|(\partial/\partial\theta_i)\bar{g}_t(\theta_0)|^2)^{\frac{1}{2}}$$

$$\leq C_{**}(1/\sqrt{n})\Sigma_{t=1}^{n}t^q(1-\delta)^t(E|(\partial/\partial\theta_i)\bar{g}_t(\theta_0)|^2)^{\frac{1}{2}}. \qquad (11.1.48)$$

Moreover, denoting.

$$\rho_s = \max_i|(\partial/\partial\theta_i)\eta_s(\theta_0)|, \ \rho_0 = \max_i|(\partial/\partial\theta_i)\varphi(\theta_0)|$$

we have from (11.1.12) and Liapounov's inequality,

$$E|(\partial/\partial\theta_i)\tilde{g}_t(\theta_0)|^2 \leq E|\rho_0 + \Sigma_{s=1}^{\infty}\rho_s|Z_{t-s}||^2$$

$$\leq 2\rho_0^2 + 2(\Sigma_{s=1}^{\infty}\rho_s)\Sigma_{s=1}^{\infty}\rho_s E|Z_{t-s}|^2$$

$$\leq 2\rho_0^2 + 2(\Sigma_{s=1}^{\infty}\rho_s)^2\cdot\sup_t E|Z_t|^2. \qquad (11.1.49)$$

Thus the right hand side of (11.1.48) is of order

$$O((1/\sqrt{n})\Sigma_{t=1}^{n}t^q(1-\delta)^t) = O(1/\sqrt{n})$$

and converges therefore to zero. Now (11.1.41) follows from Chebishev's inequality.

With $X_{n,t}$ defined by (11.1.43), condition (9.1.14) and thus condition (9.1.13) of theorem 9.1.7 follows from

$$E|X_{n,t}|^{2+\delta^*} \leq E(|U_t|^{2+\delta^*}.\max_i|(\partial/\partial\theta_i)\tilde{g}_t(\theta_0)|^{2+\delta^*})|\xi|^{2+\delta^*}$$

$$\leq (E|U_t|^{4+2\delta^*})^{\frac{1}{2}}\max_i(E|(\partial/\partial\theta_i)\bar{g}_t(\theta_0)|^{4+2\delta^*})^{\frac{1}{2}}|\xi|^{2+\delta^*}$$

$$\leq (2^{4+2\delta^*}E|Y_t|^{4+2\delta^*}E|\overline{g}_t(\theta_0)|^{4+2\delta^*})^{½}$$

$$\times \max_i (E|(\partial/\partial\theta_i)\overline{g}_t(\theta_0)|^{4+2\delta^*})^{½}|\xi|^{2+\delta^*}$$

and the fact that $\sup_t E|Z_t|^{4+2\delta^*} < \infty$ implies

$$\sup_t E|Y_t|^{4+2\delta^*} < \infty$$

and, similarly to (11.1.49),

$$\sup_t E|g_t(\theta_0)|^{4+2\delta^*} < \infty \text{ and } \sup_t E|(\partial/\partial\theta_i)\overline{g}_t(\theta_0)|^{4+2\delta^*} < \infty.$$

The δ^* for which this holds must therefore be smaller than $½\delta$, with δ as in assumption 11.1.1.

Condition (9.1.12) of theorem 9.1.7 follows easily from theorem 9.4.1, where σ^2 is given by (11.1.45) and (11.1.46). So (11.1.47) is proved.

Furthermore, we note that similarly to lemma 11.1.3 we have:

Lemma 11.1.5. Under assumptions 11.1.1 en 11.1.2,

$$\text{plim}_{n\to\infty}\hat{\Omega}_2 = \Omega_2 .$$

Proof: Exercise 3.

Summarizing, we have:

Theorem 11.1.4. Under assumptions 11.1.1 through 11.1.4,

$$\sqrt{n}(\hat{\theta} - \theta_0) \to N_m(0,\Omega_1^{-1}\Omega_2\Omega_1^{-1}) \text{ in distr.} \tag{11.1.50}$$

and under assumptions 11.1.1, 11.1.2 and 11.1.4,

$$\text{plim}_{n\to\infty}\hat{\Omega}_1^{-1}\hat{\Omega}_2\hat{\Omega}_1^{-1} = \Omega_1^{-1}\Omega_2\Omega_1^{-1}. \tag{11.1.51}$$

Note that assumption 11.1.3 is not needed for (11.1.51). However, if assumption 11.1.3 does not hold then $\Omega_1^{-1}\Omega_2\Omega_1^{-1}$ is no longer the variance matrix of the limiting distribution of $\hat{\theta}$. Moreover, note that non-singularity of Ω_2 is not required: if

Ω_2 is singular the limiting normal distribution (11.1.50) is singular too.

Exercises:

1. Prove (11.1.11), (11.1.12) and (11.1.13)
2. Prove (11.1.36)
3. Prove lemma 11.1.5.

11.2 *Estimation of nonlinear ARMAX models*

In this section we consider the asymptotic properties of the least squares parameter estimators of model (10.4.1):

$$Y_t = g(Z_{t-1},\ldots,Z_{t-p},\beta_0) + U_t + \Sigma_{j=1}^{q}\gamma_{0,j}U_{t-j},\ \beta_0 \in B, \tag{11.2.1}$$

with $B \subset R^r$ a parameter space. Let again

$$\gamma_0 = (\gamma_{0,1},\ldots,\gamma_{0,q})' \in \Gamma_\delta,$$

where Γ_δ is defined by (11.1.5) and let

$$\theta_0 = (\beta_0{}',\gamma_0{}')',\ \Theta = B \times \Gamma_\delta \subset R^m,\ \text{with } m = r+q.$$

Since the lag polynomial $1 + \Sigma_{s=1}^{q}\gamma_s L^s$ is invertible, we can write

$$\Sigma_{s=0}^{\infty}\rho_s(\gamma)L^s = (1 + \Sigma_{s=1}^{q}\gamma_s L^s)^{-1},$$

where the $\rho_s(\gamma)$'s are continuously differentiable functions on Γ_δ such that

$$\rho_0(\gamma) = 1,$$

$$\Sigma_{s=1}^{\infty}\sup_{\gamma\in\Gamma_\delta}|\rho_s(\gamma)| < \infty, \tag{11.2.2}$$

$$\Sigma_{s=1}^{\infty}\sup_{\gamma\in\Gamma_\delta}|(\partial/\partial\gamma_i)\rho_s(\gamma)| < \infty,\ i=1,\ldots,q, \tag{11.2.3}$$

$$\Sigma_{s=1}^{\infty}\sup_{\gamma\in\Gamma_\delta}|(\partial/\partial\gamma_i)(\partial/\partial\gamma_j)\rho_s(\gamma)| < \infty,\ i,j=1,..,q. \tag{11.2.4}$$

Cf. (11.1.11), (11.1.12), (11.1.13). Similarly to (11.1.15) we can write the model in nonlinear ARX(∞) form as

$$Y_t = g_t(\theta) + U_t,$$

where now

$$g_t(\theta) = \Sigma_{s=0}^{\infty}\rho_s(\gamma)g(Z_{t-1},\ldots,Z_{t-p},\beta). \tag{11.2.5}$$

Moreover, similarly to (11.1.16) we truncate the response function $g_t(\theta)$ to

$$\tilde{g}_t(\theta) = \Sigma_{s=0}^{t-p-1}\rho_s(\gamma)g(Z_{t-1},\ldots,Z_{t-p},\beta) \text{ for } t \geq p+1,$$
$$\tilde{g}_t(\theta) = 0 \text{ for } t < p+1, \tag{11.2.6}$$

and we define the least squares estimator $\hat{\theta}$ of θ_0 as a (measurable) solution of

$$\hat{\theta} \in \Theta : \hat{Q}(\hat{\theta}) = \inf_{\theta\in\Theta}\hat{Q}(\theta),$$

where

$$\hat{Q}(\theta) = (1/(n-p))\Sigma_{t=p+1}^{n}\{Y_t - \tilde{g}_t(\theta)\}^2.$$

Now assume:

Assumption 11.2.1.
(a) The process (Z_t) is ν-stable in L^1 with respect to a φ- or α-mixing base, where either $\Sigma_{j=0}^{\infty}\varphi(j) < \infty$ or $\Sigma_{j=0}^{\infty}\alpha(j) < \infty$.
(b) The function $g(w,\beta)$ is for each $w \in R^{pk}$ a continuous real function on B, and for each $\beta \in B$ a Borel measurable real function on R^{pk}, where B is a compact subset of R^r.
(c) The process (Z_t) is setwise or pointwise properly heterogeneous.
(d) If (Z_t) is pointwise properly heterogeneous then $g(w,\beta)$ is continuous on $R^{pk}\times B$.
(e) For some $\delta > 0$,

$$\sup_t E|Z_t|^{4+\delta} < \infty, \quad \sup_t E \sup_{\beta\in B}|g(Z_{t-1},\ldots,Z_{t-p},\beta)|^{4+\delta} < \infty.$$

Let similarly to (11.1.23)

$$\bar{Q}(\theta) = E\{Y_0^* - \Sigma_{s=0}^{\infty}\rho_s(\gamma)g(Z_{-1-s}^*,\ldots,Z_{-p-s}^*,\beta)\}^2.$$

where Y_0^* and Z_{-j}^* are the same as before. Then

Theorem 11.2.1. Under assumption 11.2.1,

$$\text{plim}_{n\to\infty}\sup_{\theta\in\Theta}|\hat{Q}(\theta) - \bar{Q}(\theta)| = 0.$$

Proof: We verify the conditions of theorems 9.4.2 and 9.4.3. Let for $w = (w_1,\ldots,w_{1+pk})' \in R^{1+pk}$,

$$\gamma_s(\theta,w) = w_1 - g(w_2,\ldots,w_{1+pk},\beta) \text{ if } s = 0,$$

$$\gamma_s(\theta,w) = -\rho_s(\gamma)g(w_2,\ldots,w_{1+pk},\beta) \text{ if } s \geq 1,$$

and let

$$\rho_s = \sup_{\gamma\in\Gamma_\delta}|\rho_s(\gamma)|,$$

$$\bar{b}(w) = |w_1| + \sup_{\beta\in B}|g(w_2,\ldots,w_{1+pk},\beta)|.$$

Note that $\bar{b}(w)$ is continuous on R^{1+pk} if $g(w_2,\ldots,w_{1+pk},\beta)$ is continuous in all its arguments. Cf. theorem 1.6.1. Then

$$\sup_{\theta\in\Theta}|\gamma_s(\theta,w)| \leq \rho_s\bar{b}(w), \quad \Sigma_{s=0}^{\infty}\rho_s < \infty.$$

The latter result follows from (11.2.2). Moreover, denoting

$$W_t = (Y_t, Z_{t-1}', \ldots, Z_{t-p}')'$$

it follows from assumption 11.2.1 that (W_t) is ν-stable in L^1 with respect to a φ- or α-mixing base and that

$$\sup_t E|\bar{b}(W_t)|^{2+\delta} < \infty.$$

The theorem under review now easily follows from theorems 9.4.2 and 9.4.3. Q.E.D.

Next, assume:

Assumption 11.2.2. There exists a unique $\theta_* \in \Theta$ such that

$$\bar{Q}(\theta_*) = \inf_{\theta\in\Theta}\bar{Q}(\theta .$$

Assumption 11.2.3. There exists a point θ_0 in an open convex subset Θ_0 of Θ, such that for each t,

$$E(Y_t | Z_{t-1}, Z_{t-2}, \ldots) = g_t(\theta) \text{ a.s.}$$

Then similarly to theorems 11.1.2 and 11.1.3 we have:

Theorem 11.2.2. Under assumptions 11.2.1 and 11.2.2,

$$\text{plim}_{n\to\infty}\hat{\theta} = \theta_*.$$

Theorem 11.2.3. Under assumptions 11.2.1, 11.2.2 and 11.2.3,

$$\text{plim}_{n\to\infty}\hat{\theta} = \theta_0 .$$

The proof of the asymptotic normality of $\hat{\theta}$ is left as an exercise. We only give here the additional assumptions involved.

Assumption 11.2.4. The function $g(w,\beta)$ is for each $w \in R^{pk}$ twice continuously differentiable on B. If (Z_t) is pointwise properly heterogeneous then for $i,i_1,i_2=1,..,r$, $(\partial/\partial\beta_i)g(w,\beta)$ and $(\partial/\partial\beta_{i_1})(\partial/\partial\beta_{i_2})g(w,\beta)$ are continuous functions on $R^{pk}\times B$. Moreover, for $i,i_1,i_2=1,..,r$ and some $\delta > 0$,

$$\sup_t E|(\partial/\partial\beta_i)g(Z_{t-1},\ldots,Z_{t-p},\beta)|^{4+\delta} < \infty,$$

$$\sup_t E|(\partial/\partial\beta_{i_1})(\partial/\partial\beta_{i_2})g(Z_{t-1},\ldots,Z_{t-p},\beta)|^{2+\delta} < \infty.$$

Let $\hat{\Omega}_1$, $\hat{\Omega}_2$, Ω_1 and Ω_2 be defined as in section 11.1.2, with n replaced by n-p and $g^*(\theta)$ replaced by

$$g^*(\theta) = \Sigma_{s=0}^{\infty}\rho_s(\gamma)g(Z^*_{-1-s},\ldots,Z^*_{-p-s},\beta), \qquad (11.2.7)$$

and assume:

Assumption 11.2.5. The matrix Ω_1 is nonsingular.

Then

Theorem 11.2.4. Under assumptions 11.2.1 through 11.2.5,

$$\sqrt{(n-p)}\ (\hat{\theta} - \theta_0) \rightarrow N_{r+q}(0,\Omega_1^{-1}\Omega_2\Omega_1^{-1}) \text{ in distr.}$$

and under assumptions 11.2.1, 11.2.2, 11.2.4 and 11.2.5,

$$\text{plim}_{n\rightarrow\infty}\hat{\Omega}_1^{-1}\hat{\Omega}_2\hat{\Omega}_1^{-1} = \Omega_1^{-1}\Omega_2\Omega_1^{-1}.$$

Exercise:

1. Prove theorem 11.2.4 along the lines of the proof of theorem 11.1.4.

11.3 *A consistent N(0,1) model specification test*

In Bierens (1984) we have proposed a consistent model specification test for nonlinear time series regressions. This test tests the null hypothesis that the errors of a nonlinear ARX(p) model obey a condition of the type (11.1.2). This model specification test is in principle also applicable to the ARMAX case considered in sections 11.1 and 11.2. A disadvantage of this test, however, is that the distribution of the test statistic under the null hypothesis is of an unknown type, so that the critical region of the test involved had to be derived on the basis of Chebishev's inequality for first absolute moments. This approach will lead, of course, to overestimating the effective type I error of the test, as Chebishev's inequality is not very sharp. Moreover, this test is quite laborious for relatively large data sets and models. On the other hand, the test involved is consistent in the sense that any model misspecification will be detected as the sample size grows to infinity, provided the data generating process is strictly stationary. To the best of our knowledge no other model specification test for time series regressions has this consistency property.

We shall now propose a new test which has a known limiting distribution under the null hypothesis and is consistent in the above sense. In particular, in this section we shall construct a test statistic $\sqrt{n}\hat{T}$, say, with the property that under the null hypothesis (11.1.2), $\sqrt{n}\hat{T} \to N(0,1)$ in distribution as $n \to \infty$, whereas under the alternative hypothesis that (11.1.2) does not hold and under the stationarity hypothesis, $\text{plim}_{n\to\infty}\sqrt{n}\hat{T} = \infty$. This test is a further elaboration of the test of Bierens (1987), and is reminiscent of the consistent conditional moment tests in chapter 5.

The consistency of our test requires that assumption 11.1.1* holds. Thus, strict stationarity of (Z_t) and thus of $\{(U_t, Z_t)\}$ is part of our maintained hypothesis. Some of our results below also hold under data heterogeneity. This will be indicated by not explicitly referring to assumption 11.1.1*. The reason for imposing the stationarity assumption is that under data heterogeneity the null hypothesis (11.1.2) may be false for only finitely many t's. Clearly, no test based on asymptotic arguments can detect this.

Let $U_t = Y_t - g_t(\theta^*)$, where g_t is defined by (11.2.5) and θ^* is defined in assumption 11.2.2. The null hypothesis of a correct model specification can now be restated as

$$H_0: P\{E(U_t|Z_{t-1}, Z_{t-2}, Z_{t-3}, \ldots)=0\} = 1. \quad (11.3.1)$$

The alternative hypothesis we consider is that H_0 is false. Under stationarity this general alternative hypothesis becomes

$$H_1: P\{E(U_t|Z_{t-1}, Z_{t-2}, Z_{t-3}, \ldots)=0\} < 1. \quad (11.3.2)$$

Let us assume for the moment that H_1 is true. Then theorem 9.1.5 implies that there exists an integer ℓ_0 such that

$$\sup_{\ell\geq\ell_0} P\{E(U_t|Z_{t-1}^{(\ell)}, Z_{t-2}^{(\ell)}, \ldots\ldots) = 0\} < 1, \quad (11.3.3)$$

where $Z_t^{(\ell)}$ is the vector of components of Z_t rounded off to ℓ decimal digits. Cf. exercise 1. Since $\{(U_t, Z_t)\}$ is strictly stationary, ℓ_0 is independent of t. Because the process $(Z_t^{(\ell)})$ is rational-valued, we may now apply theorem 10.3.3:

Lemma 11.3.1. Let assumptions 10.3.2, 11.1.1* and 11.2.1 hold. There exists a subset N of R^{k+1} with Lebesgue measure zero such that (11.3.3) implies

$$\sup_{\ell\geq\ell_0} P\{E(U_t|\Sigma_{s=1}^{\infty} \tau^{s-1}\xi' Z_{t-s}^{(\ell)}) = 0\} < 1$$

for all $(\xi,\tau) \in R^k\times(-1,1)\backslash N$.

Proof: Let $\theta^* = (\beta_*',\gamma_*')'$ and

$$U_t^{(j)} = Y_t - \Sigma_{s=0}^{j-p}\rho_s(\gamma_*)g(Z_{t-1-s},\ldots,Z_{t-p-s},\beta_*)$$

Cf. (11.2.5). Then $U_t^{(j)} \in F_{t-j}^{\infty}$ [cf. assumption 10.3.2] and

$$E|U_t - U_t^{(j)}| \le \Sigma_{s=j+p+1}^{\infty}|\rho_s(\gamma_*)|E|g(Z_{t-1-s},\ldots,Z_{t-p-s},\beta_*)|$$

$$\to 0 \text{ as } j \to \infty.$$

Consequently,

$$E|E(U_t|\Sigma_{s=1}^{\infty}\tau^{s-1}\xi'Z_{t-s}^{(\ell)}) - E(U_t^{(j)}|\Sigma_{s=1}^{\infty}\tau^{s-1}\xi'Z_{t-s}^{(\ell)})|$$

$$\le E|U_t - U_t^{(j)}| \to 0 \text{ as } j \to \infty \qquad (11.3.4)$$

and

$$E|E(U_t|Z_{t-1}^{(\ell)},Z_{t-2}^{(\ell)},\ldots) - E(U_t^{(j)}|Z_{t-1}^{(\ell)},Z_{t-2}^{(\ell)},\ldots)|$$

$$\le E|U_t - U_t^{(j)}| \to 0 \text{ as } j \to \infty. \qquad (11.3.5)$$

Moreover, from theorem 10.3.3 and assumption 10.3.2 it follows that there exists a subset $N_{j,\ell}$ of R^{k+1} with Lebesgue measure zero such that for all $(\xi,\tau) \in R^k\times(-1,1)\backslash N_{j,\ell}$,

$$E(U_t^{(j)}|Z_{t-1}^{(\ell)},Z_{t-2}^{(\ell)},\ldots) = E(U_t^{(j)}|\Sigma_{s=1}^{\infty}\tau^{s-1}\xi'Z_{t-s}^{(\ell)}). \qquad (11.3.6)$$

Combining (11.3.4), (11.3.5) and (11.3.6) and taking for N the union of all the sets $N_{j,\ell}$, the lemma follows. Q.E.D.

Next, we combine lemma 11.3.1 with lemma 3.3.1:

Lemma 11.3.2. Let ψ be an arbitrary bounded Borel measurable one-to-one mapping from R into R. Let the conditions of lemma 11.3.1 hold. There exists a subset S of R^{k+2} with Lebesgue measure zero such that (11.3.3) implies

$$E\ U_t \cdot \exp(\rho \cdot \psi(\Sigma_{s=1}^{\infty} \tau^{s-1} \xi' Z_{t-s}^{(\ell)})) \neq 0. \tag{11.3.7}$$

for all $\ell \geq \ell_0$ and all $(\rho,\xi,\tau) \in R \times R^k \times (-1,1) \backslash S$.

Proof: Lemmas 3.3.1 and 11.3.1 imply that for each $(\xi,\tau) \in R^k \times (-1,1) \backslash N$ there exists a countable subset $C_\ell(\xi,\tau)$ of R such that (11.3.7) holds for $\rho \notin C_\ell(\xi,\tau)$. The proof can now easily be completed along the lines of the proof of theorem 3.3.4. Cf. exercise 2. Q.E.D.

Let

$$\hat{U}_t = Y_t - \tilde{g}_t(\hat{\theta}),$$

where $\tilde{g}_t$ is defined by (11.2.6). Denoting

$$w_{t,\ell}(\rho,\xi,\tau) = \exp(\rho \cdot \psi(\Sigma_{j=1}^{\infty} \tau^{j-1} \xi' Z_{t-j}^{(\ell)})); \tag{11.3.8}$$

$$\hat{w}_{t,\ell}(\rho,\xi,\tau) = \exp(\rho \cdot \psi(\Sigma_{j=1}^{t-1} \tau^{j-1} \xi' Z_{t-j}^{(\ell)})) \text{ if } t \geq 2,$$
$$\hat{w}_{t,\ell}(\rho,\xi,\tau) = 1 \text{ if } t \leq 1; \tag{11.3.9}$$

$$c_{n,\ell}(\rho,\xi,\tau) = (1/(n-p))\Sigma_{t=p+1}^{n} U_t w_{t,\ell}(\rho,\xi,\tau); \tag{11.3.10}$$

$$\hat{c}_{n,\ell}(\rho,\xi,\tau) = (1/(n-p))\Sigma_{t=p+1}^{n} \hat{U}_t \hat{w}_{t,\ell}(\rho,\xi,\tau), \tag{11.3.11}$$

$$\bar{c}_\ell(\rho,\xi,\tau) = \lim_{n\to\infty} E\ c_{n,\ell}(\rho,\xi,\tau), \tag{11.3.12}$$

where ψ is now a bounded *uniformly continuous* one-to-one mapping from R into R, it follows:

Theorem 11.3.1. Let the conditions of theorem 11.2.2 be satisfied. Then

$$\text{plim}_{n\to\infty} \hat{c}_{n,\ell}(\rho,\xi,\tau) = \bar{c}_\ell(\rho,\xi,\tau).$$

Under H_1 and assumptions 10.3.2 and 11.1.1* there exists an integer ℓ_0 and a subset S of R^{k+2} with Lebesgue measure zero such that for all $\ell \geq \ell_0$ and all $(\rho,\xi,\tau) \in R \times R^k \times (-1,1) \backslash S$,

$$\bar{c}_\ell(\rho,\xi,\tau) \neq 0,$$

whereas under H_0,

$$\bar{c}_\ell(\rho,\xi,\tau) = 0$$

for all ℓ and all $(\rho,\xi,\tau) \in R\times R^k\times(-1,1)$.

Proof: This theorem follows easily from lemma 11.3.2 and theorems 9.4.3 and 11.2.2. Cf. exercise 3. Q.E.D.

Theorem 11.3.1 suggests to use $\hat{c}_{n,\ell}(\rho,\xi,\tau)$ as a basis for a consistent test of the null hypothesis (11.3.1) versus the alternative hypothesis (11.3.2). The next two lemmas establish the asymptotic normality of $\hat{c}_{n,\ell}(\rho,\xi,\tau)$ under H_0.

Lemma 11.3.3. Under the null hypothesis (11.3.1) and the conditions of theorem 11.2.4,

$$\text{plim}_{n\to\infty}(\surd(n-p)\hat{c}_{n,\ell}(\rho,\xi,\tau) - (1/\surd(n-p))\Sigma_{t=1}^n U_t[w_{t,\ell}(\rho,\xi,\tau) - \bar{b}_\ell(\rho,\xi,\tau)'\Omega_1^{-1}(\partial/\partial\theta')g_t(\theta^*)])=0, \quad (11.3.13)$$

where Ω_1 is defined by (11.1.38) with g^* defined by (11.2.7), and

$$\bar{b}_\ell(\rho,\xi,\tau) = \lim_{n\to\infty}(1/n)\Sigma_{t=1}^n E\, w_{t,\ell}(\rho,\xi,\tau)(\partial/\partial\theta')g_t(\theta^*). \quad (11.3.14)$$

Proof: Exercise 4.

Denoting

$$\bar{s}_\ell^2(\rho,\xi,\tau) = \lim_{n\to\infty}(1/n)\Sigma_{t=1}^n E\, U_t^2\{w_{t,\ell}(\rho,\xi,\tau) - \bar{b}_\ell(\rho,\xi,\tau)'\Omega_1^{-1}(\partial/\partial\theta')g_t(\theta^*)\}^2. \quad (11.3.15)$$

it follows now from lemma 11.3.3 and the martingale central limit theorem 9.1.7 that

Lemma 11.3.4. Under the null hypothesis (11.3.1) and the conditions of theorem 11.2 4,

$$\surd(n-p)\ \hat{c}_{n,\ell}(\rho,\xi,\tau) \to N(0,\bar{s}_\ell^2(\rho,\xi,\tau)) \text{ in distr.} \quad (11.3.16)$$

for each $(\rho,\xi,\tau) \in R\times R^k\times(-1,1)$ and each ℓ.

Proof: Exercise 5.

Moreover, denoting

$$\hat{s}_{n,\ell}^2(\rho,\xi,\tau) = (1/(n-p))\Sigma_{t=p+1}^n \hat{U}_t{}^2(\hat{w}_{t,\ell}(\rho,\xi,\tau)$$

$$- \hat{b}_{n,\ell}(\rho,\xi,\tau)'\hat{\Omega}_1{}^{-1}(\partial/\partial\theta')g_t(\hat{\theta}))^2 \quad (11.3.17)$$

with

$$\hat{b}_{n,\ell}(\rho,\xi,\tau)$$

$$= (1/(n-p))\Sigma_{t=p+1}^n \hat{w}_{t,\ell}(\rho,\xi,\tau)(\partial/\partial\theta')g_t(\hat{\theta}) \quad (11.3.18)$$

we have:

Lemma 11.3.5. Under assumptions 11.2.1, 11.2.2, 11.2.4 and 11.2.5,

$$\text{plim}_{n\to\infty}\hat{s}_{n,\ell}^2(\rho,\xi,\tau) = \bar{s}_\ell^2(\rho,\xi,\tau),$$

regardless whether or not the null is true.

Proof: Exercise 6.

Our test statistic is now

$$\hat{T}_\ell(\rho,\xi,\tau) = \hat{c}_{n,\ell}(\rho,\xi,\tau)\ /\ \surd(\hat{s}_{n,\ell}^2(\rho,\xi,\tau)). \quad (11.3.19)$$

However, before we can draw conclusions about the limiting behavior of $\hat{T}_\ell(\rho,\xi,\tau)$ under H_0 and H_1 we have to address the question whether $\bar{s}_\ell^2(\rho,\xi,\tau) > 0$ for all (ρ,ξ,τ). The answer is no, i.e., if the model contains a constant term then at least

$$\bar{s}_\ell^2(0,\xi,\tau) = 0 \text{ and } \bar{s}_\ell^2(\rho,0,\tau) = 0,$$

where in the latter case 0 is the zero vector. If $\xi = 0$ then clearly

$$\hat{w}_{t,\ell}(\rho,\xi,\tau) = w_{t,\ell}(\rho,\xi,\tau) = \exp(\rho\cdot\psi(0))$$

and consequently

$$\hat{c}_{n,\ell}(\rho,\xi,\tau) = \exp(\rho\cdot\psi(0))(1/(n-p))\Sigma_{t=p+1}^{n}\hat{U}_t .$$

Similarly, if $\rho = 0$ then

$$\hat{c}_{n,\ell}(\rho,\xi,\tau) = (1/(n-p))\Sigma_{t=p+1}^{n}\hat{U}_t .$$

But if model (11.2.1) contains a constant term then

$$\Sigma_{t=p+1}^{n}\hat{U}_t = 0 \quad \text{a.s.},$$

hence

$$\bar{s}_\ell^2(\rho,\xi,\tau) = 0.$$

We may exclude this case by including $\rho = 0$ and/or $\xi = 0$ in the set S. Because the new set S is then the union of the original set S with another set with Lebesgue measure zero, it still has Lebesgue measure zero. The following assumption guaranties that the points $(\rho,\xi,\tau) \in R\times R^k\times(-1,1)$ for which $\bar{s}_\ell^2(\rho,\xi,\tau) = 0$ are indeed contained in a set with Lebesgue measure zero.

Assumption 11.3.1. The process (Z_t) is strictly stationary. There exists a stationary process (f_t), where f_t is defined on the Borel field generated by $Z_{t-1}, Z_{t-2}, \ldots,$ such that the random vector $\kappa_t = (f_t, (\partial/\partial\theta)g_t(\theta^*))'$ has positive definite second moment matrix $E\,\kappa_t\kappa_t'$.

Lemma 11.3.6. Under assumption 11.3.1 and the conditions of lemma 11.3.5 there exists a natural number ℓ_1 such that for all $\ell > \ell_1$ the set

$$S_\ell^* = \{(\rho,\xi,\tau) \in R\times R^k\times(-1,1): \bar{s}_\ell^2(\rho,\xi,\tau) = 0\}$$

has Lebesgue measure zero.

Proof: Without loss of generality we may assume $P(U_t = 0) < 1$. Let $(\rho,\xi,\tau) \in S_\ell^*$. Then

$$w_{t,\ell}(\rho,\xi,\tau) = \bar{b}_\ell(\rho,\xi,\tau)'\Omega_1^{-1}(\partial/\partial\theta')g_t(\theta^*) \text{ a.s.}$$

Hence

$$E\{f_t \cdot w_{t,\ell}(\rho,\xi,\tau)\} = \bar{b}_\ell(\rho,\xi,\tau)'\Omega_1^{-1}E\{(\partial/\partial\theta')g_t(\theta^*)f_t\}$$

$$= E\{(\partial/\partial\theta')g_t(\theta^*)\lambda\cdot w_{t,\ell}(\rho,\xi,\tau)\},$$

where

$$\lambda = \Omega_1^{-1}E\{(\partial/\partial\theta')g_t(\theta^*)f_t\}.$$

Since $P\{f_t=(\partial/\partial\theta')g_t(\theta^*)\lambda\} = 1$ would imply that $E\,\kappa_t\kappa_t'$ is singular, it follows from assumption 11.3.1 that

$$P\{f_t=(\partial/\partial\theta')g_t(\theta^*)\lambda\} < 1.$$

Similarly to lemma 11.3.2 it follows now that for sufficiently large ℓ, S_ℓ^* has Lebesgue measure zero. Q.E.D.

Taking the union of the former set S with the union of the S_ℓ^* over $\ell \geq \ell_1$ and denoting the new set again by S we now have:

Theorem 11.3.2. There exists a natural number ℓ_0 and a subset S of R^{k+2} with Lebesgue measure zero such that for all $(\rho,\xi,\tau) \in R\times R^k\times(-1,1)\backslash S$ and all $\ell \geq \ell_0$ the following hold.
(a) Under the null hypothesis, the conditions of theorem 11.2.4 and assumption 11.3.1, we have

$$\surd(n-p)\hat{T}_\ell(\rho,\xi,\tau) \rightarrow N(0,1) \text{ in distr.} \qquad (11.3.20)$$

(b) If the null is false then under assumptions 10.3.2, 11.1.1*, 11.3.1, 11.2.1, 11.2.2, 11.2.4 and 11.2.5,

$$\text{plim}_{n\rightarrow\infty}\surd(n-p)\hat{T}_\ell(\rho,\xi,\tau) = \infty. \qquad (11.3.21)$$

In practice the exceptional set S is unknown. However, similarly to theorems 5.3.1, 8.4.3 and 8.4.5 we have:

Theorem 11.3.3. Draw ρ and the components of ξ randomly from continuous distributions and draw τ randomly from the uniform (-1,1) distribution. Let $\ell \geq \ell_0$. Then the conclusions of theorem 11.3.2 carry over.

Proof: Exercise 7.

Next we propose a slightly modified version of the test under review, for the very same reason as in section 5.3: Suppose we have chosen

$$\psi(.) = tg^{-1}(\delta(.)), \text{ where } \delta > 0 \text{ is some constant.} \qquad (11.3.22)$$

Moreover, suppose that

$$\hat{x}_{t,\ell}(\tau,\xi) = \Sigma_{j=1}^{t-1}\tau^{j-1}\xi' Z_{t-j}^{(\ell)} \qquad (11.3.23)$$

takes with high probability only large positive values. Then $\psi(\hat{x}_{t,\ell}(\tau,\xi))$ takes with high probability only values close to $\frac{1}{2}\pi$, which is the upperbound of the function ψ. Hence the function

$$\hat{w}_{t,\ell}(\rho,\xi,\tau) = \exp(\rho\cdot\psi(\hat{x}_{t,\ell}(\tau,\xi))) \qquad (11.3.24)$$

takes values close to $\exp(\rho\frac{1}{2}\pi)$ and consequently

$$\hat{c}_{n,\ell}(\rho,\xi,\tau) = \{(1/n)\Sigma_{t=1}^{n}\hat{U}_t\}\exp(\rho^{\frac{1}{2}}\pi).$$

But if the $\hat{U}_t$ are least squares residuals of a model with a constant term they sum up to zero and hence $\hat{c}_{n,\ell}(\rho,\xi,\tau)$ will then be close to zero. Clearly this will destroy the power of the test. Therefore we propose to standardize the argument of ψ, i.e., we propose to replace $\hat{x}_{t,\ell}(\tau,\xi)$ in (11.3.24) by $\hat{\hat{x}}_{t,\ell}(\tau,\xi)$ defined as follows:

$$\hat{\hat{x}}_{t,\ell}(\tau,\xi) = \{\hat{x}_{t,\ell}(\tau,\xi) - (1/t)\Sigma_{s=1}^{t}\hat{x}_{s,\ell}(\tau,\xi)\}$$
$$/\{(1/t)\Sigma_{s=1}^{t}\hat{x}_{s,\ell}(\tau,\xi)^2 - ((1/t)\Sigma_{s=1}^{t}\hat{x}_{s,\ell}(\tau,\xi))^2\}^{\frac{1}{2}} \text{ if } t \geq 3, \tag{11.3.25}$$
$$\hat{\hat{x}}_{t,\ell}(\tau,\xi) = 0 \quad \text{if } t \leq 2,$$

where $\hat{x}_{t,\ell}(\tau,\xi) = 0$ if $t \leq 1$ and defined by (11.3.24) if $t \geq 2$. The function $\hat{w}_{t,\ell}(\rho,\xi,\tau)$ is thus redefined as

$$\hat{w}_{t,\ell}(\rho,\xi,\tau) = \exp(\rho\cdot\psi(\hat{\hat{x}}_{t,\ell}(\tau,\xi))) \text{ if } t \geq 3, \tag{11.3.26}$$
$$\hat{w}_{t,\ell}(\rho,\xi,\tau) = 1 \text{ if } t \leq 2.$$

Redefine the function $w_{t,\ell}(\rho,\xi,\tau)$ accordingly as

$$w_{t,\ell}(\rho,\xi,\tau) = \exp(\rho\cdot\psi(\overline{x}_{t,\ell}(\tau,\xi))), \tag{11.3.27}$$

where

$$\overline{x}_{t,\ell}(t,\ell) = \{x_{t,\ell}(\tau,\xi) - E\, x_{t,\ell}(\tau,\xi)\}/\{\mathrm{var}(x_{t,\ell}(\tau,\xi))\}^{\frac{1}{2}} \tag{11.3.28}$$

with

$$x_{t,\ell}(\tau,\xi) = \Sigma_{j=1}^{\infty}\tau^{j-1}\xi' Z_{t-j}^{(\ell)}. \tag{11.3.29}$$

Then

Theorem 11.3.4. With (11.3.26) instead of (11.3.9) and (11.3.27) instead of (11.3.8), all the previous results in this section go through.

Proof: Exercise 8.

Remark 1: It is easy to verify that the results in this section also go throught if we use more general ARMA memory indices than (11.3.29). In particular, we may replace (11.3.29) by say

$$x_{t,\ell}(\tau,\xi) = (1+\tau_1 L^1+\ldots+\tau_{q_1} L^{q_1})^{-1}(\xi_1 L+\ldots+\xi_{p_1} L^{p_1})' Z_t^{(\ell)}$$

and replace (11.3.23) by

$$\hat{x}_{t,\ell}(\tau,\xi) = (1+\tau_1 L^1+\ldots+\tau_{q_1} L^{q_1})^{-1}(\xi_1 L+\ldots+\xi_{p_1} L^{p_1})' \hat{Z}_t^{(\ell)},$$

where now

$$\tau = (\tau_1,\ldots,\tau_{q_1})' \in \Delta,\ \xi = (\xi_1',\ldots,\xi_{p_1}')' \in R^{k+p_1}$$

with Δ such that lag polynomial $1+\tau_1 L^1+\ldots+\tau_{q_1} L^{q_1}$ has roots all outside the complex unit circle, and

$$\hat{Z}_t^{(\ell)} = Z_t^{(\ell)} \text{ for } t \geq 1, \qquad \hat{Z}_t^{(\ell)} = 0 \text{ for } t < 1.$$

Remark 2: In chapter 10 we argued that the parameters τ and ξ for which

$$E(U_t | Z_{t-1}^{(\ell)}, Z_{t-2}^{(\ell)},\ldots) = E(U_t | \Sigma_{j=1}^{\infty} \tau^{j-1} \xi' Z_{t-j}^{(\ell)}) \text{ a.s.}$$

are likely to be irrational. Since this result plays a key-role in the proof of theorem 11.3.1, one might therefore think that the consistency of our test can not hold in practice as consistency requires irrational τ and ξ, whereas in practice it is impossible to deal with irrational numbers. However, the functions $\bar{c}_\ell(\rho,\xi,\tau)$ and $\bar{s}_\ell^2(\rho,\xi,\tau)$ are continuous, and so is

$$\bar{T}_\ell(\rho,\xi,\tau) = \text{plim}_{n\to\infty} \hat{T}_\ell(\rho,\xi,\tau) = \bar{c}_\ell(\rho,\xi,\tau)/\sqrt{(\bar{s}_\ell^2(\rho,\xi,\tau))},$$

provided $\bar{s}_\ell^2(\rho,\xi,\tau) > 0$, hence if

$$\bar{T}_\ell(\rho,\xi,\tau) \neq 0$$

holds then it holds too on an open neighborhood of (ρ,ξ,τ) and thus also for all rational (ρ^*,ξ^*,τ^*) in this neighborhood.

Exercises:

1. Prove (11.3.3).
2. Complete the proof of lemma 11.3.2.
3. Complete the proof of theorem 11.3.1.
4. Prove lemma 11.3.3.
5. Prove lemma 11.3.4.
6. Prove lemma 11.3.5
7. Prove theorem 11.3.3.
8. Prove theorem 11.3.4.

11.4 *A consistent Hausman-type model specification test*

The test in section 11.3 is a special case of the consistent conditional moment test in section 5.3 generalized to ARMAX models. This suggests that it is possible to construct more general consistent tests along the lines of section 5.3. In the present section we shall now generalize the Hausman-White test [cf. section 5.1], which is also a conditional moment test, to testing (non)linear ARMAX model specifications, on the basis of the model and asymptotic results of sections 11.2 and 11.3. All the proofs will be left to the reader as exercises.

Let

$$\tilde{Q}_\ell(\theta|\rho,\xi,\tau) = (1/(n-p)\Sigma_{t=p+1}^{n}[Y_t-\tilde{g}_t(\theta)]^2\hat{w}_{t,\ell}(\rho,\xi,\tau)$$

where $\hat{w}_{t,\ell}(\rho,\xi,\tau)$ is defined in (11.3.26). Then the weighted least squares estimator of θ_0 is defined by

$$\tilde{\theta}_\ell(\rho,\xi,\tau) \in \Theta\text{: } \tilde{Q}_\ell(\tilde{\theta}_\ell(\rho,\xi,\tau)|\rho,\xi,\tau) = \inf_{\theta\in\Theta}\tilde{Q}_\ell(\theta|\rho,\xi,\tau)$$

and similarly to section 5.1 we can construct a χ^2 test on the basis of $\tilde{\theta}_\ell(\rho,\xi,\tau) - \hat{\theta}$. This approach, however, requires conducting nonlinear least squares twice, which is a computational burden. Therefore we propose to approximate $\tilde{\theta}_\ell(\rho,\xi,\tau)$ by a single Newton-type iteration step starting from $\hat{\theta}$, i.e., we propose to approximate $\tilde{\theta}_\ell(\rho,\xi,\tau)$ by

$$\hat{\theta}_\ell(\rho,\xi,\tau) = \hat{\theta} + \hat{B}_\ell(\rho,\xi,\tau)^{-1}\hat{d}_\ell(\rho,\xi,\tau),$$

where

$$\hat{d}_\ell(\rho,\xi,\tau) = (1/(n-p))\Sigma_{t=p+1}^{n}\hat{U}_t(\partial/\partial\theta')\bar{g}_t(\hat{\theta})\hat{w}_{t,\ell}(\rho,\xi,\tau)$$

and

$$\hat{B}_\ell(\rho,\xi,\tau)$$

$$=(1/(n-p))\Sigma_{t=p+1}^{n}[(\partial/\partial\theta')\bar{g}_t(\hat{\theta})][(\partial/\partial\theta)\bar{g}_t(\hat{\theta})]\hat{w}_{t,\ell}(\rho,\xi,\tau).$$

The estimator $\hat{\theta}_\ell(\rho,\xi,\tau)$ is motivated by the following lemma.

Lemma 11.4.1. Under the conditions of theorem 11.2.4,

$$\sqrt{(n-p)}[\tilde{\theta}_\ell(\rho,\xi,\tau)-\theta_0] \text{ and } \sqrt{(n-p)}[\hat{\theta}_\ell(\rho,\xi,\tau)-\theta_0]$$

have the same limiting normal distribution.

Proof: Exercise 1.

Consequently, under the null hypothesis (11.3.1) a Hausman-White test based on $\tilde{\theta}_\ell(\rho,\xi,\tau) - \hat{\theta}$ is asymptotically equivalent to a Hausman-White test based on $\hat{\theta}_\ell(\rho,\xi,\tau) - \hat{\theta}$.

Similarly to lemma 11.3.3 it can be shown that under H_0,

$$\text{plim}_{n\to\infty}(\sqrt{(n-p)}\hat{d}_\ell(\rho,\xi,\tau)$$

$$- (1/\sqrt{(n-p)})\Sigma_{t=p+1}^{n}U_t[(\partial/\partial\theta')g_t(\theta^*)w_{t,\ell}(\rho,\xi,\tau)$$

$$- B_\ell(\rho,\xi,\tau)\Omega_1^{-1}(\partial/\partial\theta')g_t(\theta^*)]) = 0,$$

where

$$B_\ell(\rho,\xi,\tau)$$

$$= \lim_{n\to\infty}(1/n))\Sigma_{t=1}^{n}E\{[(\partial/\partial\theta')g_t(\theta)][(\partial/\partial\theta)g_t(\theta)]w_{t,\ell}(\rho,\xi,\tau)\}.$$

Thus, under H_0, $\sqrt{(n-p)}\hat{d}_\ell(\rho,\xi,\tau)$ is asymptotically distributed as

$$(1/\surd(n-p))\Sigma_{t=p+1}^{n} U_t[(\partial/\partial\theta')g_t(\theta^*)w_{t,\ell}(\rho,\xi,\tau)$$

$$- B_\ell(\rho,\xi,\tau)\Omega_1^{-1}(\partial/\partial\theta')g_t(\theta^*)],$$

The latter asymptotic distribution is multivariate normal with zero mean vector and asymptotic variance matrix

$$\Sigma_\ell(\rho,\xi,\tau) = A_\ell^{(2)}(\rho,\xi,\tau) - B_\ell(\rho,\xi,\tau)\Omega_1^{-1}A_\ell^{(1)}(\rho,\xi,\tau)$$

$$- A_\ell^{(1)}(\rho,\xi,\tau)\Omega_1^{-1}B_\ell(\rho,\xi,\tau) + B_\ell(\rho,\xi,\tau)\Omega_1^{-1}\Omega_2\Omega_1^{-1}B_\ell(\rho,\xi,\tau),$$

where for j=1,2,

$$A_\ell^{(j)}(\rho,\xi,\tau)$$

$$=\lim_{n\to\infty}(1/n))\Sigma_{t=1}^{n} E\{U_t^2[(\partial/\partial\theta')g_t(\theta)][(\partial/\partial\theta)g_t(\theta)]w_{t,\ell}(\rho,\xi,\tau)^j\}.$$

Since

$$\text{plim}_{n\to\infty}\hat{B}_\ell(\rho,\xi,\tau) = B_\ell(\rho,\xi,\tau)$$

and since assumption 11.2.5 implies that $B_\ell(\rho,\xi,\tau)$ is nonsingular [cf. exercise 2], it follows:

Theorem 11.4.1. Under the conditions of theorem 11.2.4,

$$\surd(n-p)[\hat{\theta}_\ell(\rho,\xi,\tau)-\theta_0] \to N_m[0,\Delta_\ell(\rho,\xi,\tau)] \text{ in distr.},$$

for all $(\rho,\xi,\tau) \in R\times R^k\times(-1,1)$, where m is the dimension of θ_0 and

$$\Delta_\ell(\rho,\xi,\tau) = B_\ell(\rho,\xi,\tau)^{-1}A_\ell^{(2)}(\rho,\xi,\tau)B_\ell(\rho,\xi,\tau)^{-1}$$

$$- \Omega_1^{-1}A_\ell^{(1)}(\rho,\xi,\tau)B_\ell(\rho,\xi,\tau)^{-1} - B_\ell(\rho,\xi,\tau)^{-1}A_\ell^{(1)}(\rho,\xi,\tau)\Omega_1^{-1}$$

$$+ \Omega_1^{-1}\Omega_2\Omega_1^{-1}.$$

Proof: Exercise 3.

Next, let

$$\hat{\Delta}_\ell(\rho,\xi,\tau) = \hat{B}_\ell(\rho,\xi,\tau)^{-1}\hat{A}_\ell^{(2)}(\rho,\xi,\tau)\hat{B}_\ell(\rho,\xi,\tau)^{-1}$$

$$- \hat{\Omega}_1^{-1}\hat{A}_\ell^{(1)}(\rho,\xi,\tau)\hat{B}_\ell(\rho,\xi,\tau)^{-1} - \hat{B}_\ell(\rho,\xi,\tau)^{-1}\hat{A}_\ell^{(1)}(\rho,\xi,\tau)\hat{\Omega}_1^{-1}$$

$$+ \hat{\Omega}_1^{-1}\hat{\Omega}_2\hat{\Omega}_1^{-1},$$

where for j=1,2,

$$\hat{A}_\ell^{(j)}(\rho,\xi,\tau)$$

$$=(1/(n-p))\Sigma_{t=p+1}^{n}E(\hat{U}_t^2[(\partial/\partial\theta')g_t(\hat{\theta})][(\partial/\partial\theta)g_t(\hat{\theta})]\hat{w}_{t,\ell}(\rho,\xi,\tau)^j).$$

Then

Theorem 11.4.2. Under assumptions 11.2.1, 11.2.2, 11.2.4 and 11.2.5,

$$\text{plim}_{n\to\infty}\hat{\Delta}_\ell(\rho,\xi,\tau) = \Delta_\ell(\rho,\xi,\tau)$$

Moreover, under the additional assumption 11.3.1 there exists a natural number ℓ_1 such that for all $\ell \geq \ell_1$ the set

$$S_\ell^* = \{(\rho,\xi,\tau) \in R\times R^k\times(-1,1): \det[\Delta_\ell(\rho,\xi,\tau)]=0\}$$

has Lebesgue measure zero.

Proof: Exercise 4.

The latter result is analogous to lemma 11.3.6 and also it proof is almost the same. Again, note that theorem 11.4.2 does not require that H_0 is true. Of course, in that case $\Delta_\ell(\rho,\xi,\tau)$ is no longer the variance matrix of the limiting distribution of $\sqrt{(n-p)}[\hat{\theta}_\ell(\rho,\xi,\tau)-\theta_0]$.

Next we consider the case where H_0 is false. Denote

$$\bar{d}_\ell(\rho,\xi,\tau) = \lim_{n\to\infty}(1/n))\Sigma_{t=1}^{n}E[U_t(\partial/\partial\theta')g_t(\theta)w_{t,\ell}(\rho,\xi,\tau)]$$

Similarly to theorem 11.3.1 it follows that under H_1 and the stationarity condition there exists a natural number ℓ_0 and a subset S of R^{k+2} with Lebesgue measure zero such that for all $(\rho,\xi,\tau) \in R\times R^k\times(-1,1)\backslash S$ and all $\ell \geq \ell_0$,

$$\bar{d}_\ell(\rho,\xi,\tau) \neq 0.$$

Since

$$\bar{\theta}_\ell(\rho,\xi,\tau) = \text{plim}_{n\to\infty}\hat{\theta}_\ell(\rho,\xi,\tau) = \theta^* + B_\ell(\rho,\xi,\tau)^{-1}\bar{d}_\ell(\rho,\xi,\tau)$$

and $B_\ell(\rho,\xi,\tau)$ is nonsingular for all $(\rho,\xi,\tau) \in R\times R^k\times(-1,1)$ and all ℓ, it therefore follows:

Theorem 11.4.3. Under H_1, assumptions 10.3.2 and 11.1.1* and the conditions of theorem 11.2.2 there exists a natural number ℓ_0 such that for all $\ell \geq \ell_2$ the sets

$$S_\ell = \{(\rho,\xi,\tau) \in R\times R^k\times(-1,1) : \bar{\theta}_\ell(\rho,\xi,\tau) = \theta^*\}$$

have Lebesgue measure zero.

Proof: Exercise 5.

Denoting

$$\ell_0 = \max(\ell_1,\ell_2),\ S = \cup_{\ell=\ell_0}^{\infty}(S_\ell \cup S_\ell^*),$$

$$\hat{H}_\ell(\rho,\xi,\tau) = (n-p)[\hat{\theta}_\ell(\rho,\xi,\tau)-\hat{\theta}]'\hat{\Delta}_\ell(\rho,\xi,\tau)^{-1}[\hat{\theta}_\ell(\rho,\xi,\tau)-\hat{\theta}],$$

it follows now from theorems 11.4.1, 11.4.2 and 11.4.3,

Theorem 11.4.4. There exists a natural number ℓ_0 and a subset S of R^{k+2} with Lebesgue measure zero such that for all $(\rho,\xi,\tau) \in R\times R^k\times(-1,1)\backslash S$ and all $\ell \geq \ell_0$ the following hold.

(a) Under the null hypothesis, the conditions of theorem 11.2.4 and assumption 11.3.1, we have

$$\hat{H}_\ell(\rho,\xi,\tau) \to \chi_m^2 \text{ in distr.}$$

(b) If the null is false then under assumptions 10.3.2, 11.1.1*, 11.3.1, 11.2.1, 11.2.2, 11.2.4 and 11.2.5,

$$\text{plim}_{n\to\infty}\hat{H}_\ell(\rho,\xi,\tau) = \infty.$$

Finally, we note that remarks 1 and 2 in section 11.3

also apply to the present test, and that similarly to theorem 11.3.3 we have:

Theorem 11.4.5. Draw ρ and the components of ξ randomly from continuous distributions and draw τ randomly from the uniform (-1,1) distribution. Let $\ell \geq \ell_0$. Then the conclusions of theorem 11.4.4 carry over.

Exercises:

1. Prove lemma 11.4.1.
2. Why is the nonsingularity of $B_\ell(\rho,\xi,\tau)$ implied by assumption 11.2.5 ?
3. Prove theorem 11.4.1
4. Prove theorem 11.4.2
5. Prove theorem 11.4.3

11.5 *An autocorrelation test*

In this section we briefly discuss a test for first or higher order autocorrelation of the errors U_t of model (11.2.1). The null hypothesis is still H_0 defined by (11.3.1), but instead of (11.3.2) we consider the less general alternative

$$H_1^{(r)}: \operatorname{cov}(U_t, U_{t-j}) \neq 0 \text{ for some } j \in \{1,2,..,r\}.$$

The reason for considering the problem of testing H_0 against $H_1^{(r)}$ is threefold. First, in traditional times analysis most tests for model specification test the null hypothesis of white noise errors against an alternative of the type $H_1^{(r)}$. Second, such a test is rather easy to construct, and its construction is a very useful exercise that highlights the essence of the approach in the previous sections. Third, severe model misspecification will likely be covered by $H_1^{(r)}$ for r sufficiently large. Therefore we advocate to conduct the test below first, as a pretest of model misspecification. If H_0 is rejected in favor of $H_1^{(r)}$ there is no need to conduct a consistent test. However, since $H_1^{(r)}$ may be false while H_0 is false, not rejecting H_0 in favor of $H_1^{(r)}$ does not provide sufficient evidence that H_0 is true In that case the consistent tests in sections 11.3 and 11.4 should be used in order to verify whether H_0 is true or not.

The test involved can simply be based on the statistic

$$\hat{c} = (1/(n-r-p))\Sigma_{t=r+p+1}^{n}\hat{U}_t\hat{V}_t .$$

where

$$\hat{U}_t = Y_t - \bar{g}_t(\hat{\theta}),$$

with $\bar{g}_t$ defined by (11.2.6), and

$$\hat{V}_t = (\hat{U}_{t-1},\ldots,\hat{U}_{t-r})'.$$

Let

$$U_t = Y_t - g_t(\theta^*),$$

where g_t is defined by (11.2.5) and θ^* is defined in assumption 11.2.2, and let

$$V_t = (U_{t-1},\ldots,U_{t-r})'.$$

We recall that under H_0 . $\theta^*=\theta_0$. Denoting

$$\Delta = A - B_1\Omega_1^{-1}B_2' - B_2\Omega_1^{-1}B_1' + B_1\Omega_1^{-1}\Omega_2\Omega_1^{-1}B_1',$$

$$\hat{\Delta} = \hat{A} - \hat{B}_1\hat{\Omega}_1^{-1}\hat{B}_2' - \hat{B}_2\hat{\Omega}_1^{-1}\hat{B}_1' + \hat{B}_1\hat{\Omega}_1^{-1}\hat{\Omega}_2\hat{\Omega}_1^{-1}\hat{B}_1',$$

with

$$B_1 = \lim_{n\to\infty}(1/(n-r-p))\Sigma_{t=r+p+1}^{n}E[V_t(\partial/\partial\theta)g_t(\theta^*)],$$

$$B_2 = \lim_{n\to\infty}(1/(n-r-p))\Sigma_{t=r+p+1}^{n}E[U_t^2V_t(\partial/\partial\theta)g_t(\theta^*)],$$

$$A = \lim_{n\to\infty}(1/(n-r-p))\Sigma_{t=r+p+1}^{n}E[U_t^2V_tV_t']$$

$$\hat{B}_1 = (1/(n-r-p))\Sigma_{t=r+p+1}^{n}E[\hat{V}_t(\partial/\partial\theta)\bar{g}_t(\hat{\theta})],$$

$$\hat{B}_2 = (1/(n-r-p))\Sigma_{t=r+p+1}^{n}E[\hat{U}_t^2\hat{V}_t(\partial/\partial\theta)\bar{g}_t(\hat{\theta})],$$

$$\hat{A} = (1/(n-r-p))\Sigma_{t=r+p+1}^{n}E[\hat{U}_t^2\hat{V}_t\hat{V}_t'],$$

the test statistic involved is now

$$\hat{a}_r = (n-r-p)\hat{c}'\hat{\Delta}^{-1}\hat{c}.$$

Theorem 11.5.1. Let $\det(\Delta) \neq 0$.
(a) Under the null hypothesis (11.3.1) and assumptions 11.2.1 through 11.2.5 we have

$$\hat{a}_r \to \chi^2_r \quad \text{in distr}$$

(b) Under $H_1^{(r)}$ and the assumptions 11.1.1*, 11.2.1, 11.2.2, 11.2.4 and 11.2.5 we have

$$\text{plim}_{n\to\infty}\hat{a}_r = \infty.$$

Proof: The details of the proof are left as exercises. Below we only give the main steps of the argument.
First, under the conditions of part (b) of the theorem we have

$$\text{plim}_{n\to\infty}\hat{c} \neq 0. \tag{11.5.1}$$

Next, let the conditions of part (a) hold. Let $\hat{c}_i$ be the i-th component of $\hat{c}$. By the mean value theorem there exists a mean value $\hat{\theta}^{(i)}$ satisfying $|\hat{\theta}^{(i)}-\theta_0| \leq |\hat{\theta}-\theta_0|$ such that

$$\hat{c}_i = (1/(n-r-p))\Sigma^n_{t=r+p+1}[Y_t-\bar{g}_t(\theta_0)][Y_{t-i}-\bar{g}_{t-i}(\theta_0)]$$
$$-\{(1/(n-r-p))\Sigma^n_{t=r+p+1}[Y_t-\bar{g}_t(\hat{\theta}^{(i)})][(\partial/\partial\theta)\bar{g}_{t-i}(\hat{\theta}^{(i)})]$$
$$+(1/(n-r-p))\Sigma^n_{t=r+p+1}[(\partial/\partial\theta)\bar{g}_t(\hat{\theta}^{(i)})][Y_{t-i}-\bar{g}_{t-i}(\hat{\theta}^{(i)})]\}$$
$$\times(\hat{\theta}-\theta_0)$$

Since

$$\text{plim}_{n\to\infty}\{(1/\sqrt{(n-r-p)})\Sigma^n_{t=r+p+1}[Y_t-\bar{g}_t(\theta_0)][Y_{t-i}-\bar{g}_{t-i}(\theta_0)]$$
$$-(1/\sqrt{(n-r-p)})\Sigma^n_{t=r+p+1}U_tU_{t-i}\} = 0,$$

$$\text{plim}_{n\to\infty}(1/(n-r-p))\Sigma_{t=r+p+1}^{n}[Y_t-\bar{g}_t(\hat{\theta}^{(i)})][(\partial/\partial\theta)\bar{g}_{t-i}(\hat{\theta}^{(i)})]$$

$$= \text{plim}_{n\to\infty}(1/(n-r-p))\Sigma_{t=r+p+1}^{n}U_t(\partial/\partial\theta)g_{t-i}(\theta_0) = 0,$$

$$\text{plim}_{n\to\infty}(1/(n-r-p))\Sigma_{t=r+p+1}^{n}[(\partial/\partial\theta)\bar{g}_t(\hat{\theta}^{(i)})][Y_{t-i}-\bar{g}_{t-i}(\hat{\theta}^{(i)})]$$

$$= \text{plim}_{n\to\infty}(1/(n-r-p))\Sigma_{t=r+p+1}^{n}U_{t-i}(\partial/\partial\theta)g_t(\theta_0)$$

and

$$\text{plim}_{n\to\infty}\{\surd(n-r-p)(\hat{\theta}-\theta_0)$$

$$- (1/\surd(n-r-p))\Sigma_{t=r+p+1}^{n}U_t\Omega_1^{-1}(\partial/\partial\theta')g_t(\theta_0)\} = 0$$

it follows now that

$$\text{plim}_{n\to\infty}\{\surd(n-r-p)\hat{c}$$

$$- (1/\surd(n-r-p))\Sigma_{t=r+p+1}^{n}U_t[V_t-B_1\Omega_1^{-1}(\partial/\partial\theta')g_t(\theta_0)]\} = 0,$$

hence by theorem 9.1.7,

$$\surd(n-r-p)\hat{c} \to N_r[0,\Delta] \text{ in distr.} \qquad (11.5.2)$$

Finally, we have

$$\text{plim}_{n\to\infty}\hat{\Delta} = \Delta. \qquad (11.5.3)$$

Combining (11.5.1), (11.5.2) and (11.5.3), the theorem follows.
Q.E.D.

References

Bierens, H.J. (1984), "Model Specification Testing of Time Series Regressions", *Journal of Econometrics* 26, 323-353.

Bierens, H.J. (1987), "ARMAX Model Specification Testing, with an Application to Unemployment in the Netherlands", *Journal of Econometrics* 35, 161-190.

Doan,T., P.Litterman and C.Sims (1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions", *Econometric Reviews* 3, 1-100.

Hannan,E.J., W.T.M.Dunsmuir and M.Deistler (1980), "Estimation of Vector ARMAX Models", *Journal of Multivariate Analysis* 10, 275-295.

Royden,H.L. (1968), *Real Analysis*, London: Macmillan.

Sargent,T.J. and C.A.Sims (1977), "Business Cycle Modeling Without Pretending to Have Too Much A Priori Economic Theory", in: C.A.Sims (ed.), *New Methods in Business Cycle Research: Proceedings from a Conference*, Minneapolis: Federal Reserve Bank of Minneapolis.

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

Sims,C.A. (1981), "An autogressive Index Model for the U.S., 1948-1975", in: J.Kmenta and J.B.Ramsey (eds.), *Large-Scale Macro-Econometric Models*, Amsterdam: North-Holland.