

22

ET

05348

1987

SERIE RESEARCH MEMORANDA

TESTS FOR MODEL MISSPECIFICATION

Herman J. Bierens

Researchmemorandum 1987-22 April '87



VRIJE UNIVERSITEIT
Faculteit der Economische Wetenschappen en Econometrie
A M S T E R D A M

TESTS FOR MODEL MISSPECIFICATION *

Herman J. Bierens
Department of Econometrics
Free University, Amsterdam

December 1988

* Chapter 5 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

PART 1: PRELIMINARY MATHEMATICS

1. BASIC PROBABILITY THEORY
 - 1.1 Measure-theoretical foundation of probability theory
 - 1.2 Independence
 - 1.3 Borel measurable functions
 - 1.4 Mathematical expectation
 - 1.5 Characteristic functions
 - 1.6 Random functions
2. CONVERGENCE
 - 2.1 Weak and strong convergence of random variables
 - 2.2 Convergence of mathematical expectations
 - 2.3 Convergence of distributions
 - 2.4 Central limit theorems
 - 2.5 Further results on convergence of distributions and mathematical expectations, and laws of large numbers
 - 2.6 Convergence of random functions
 - 2.7 Uniform strong and weak laws of large numbers
3. INTRODUCTION TO CONDITIONING
 - 3.1 Definition of conditional expectation
 - 3.2. Basic properties of conditional expectations
 - 3.3 Identification of conditional expectations

PART 2: REGRESSION ANALYSIS UNDER INDEPENDENCE

4. NONLINEAR PARAMETRIC REGRESSION ANALYSIS
 - 4.1 Nonlinear regression models and the nonlinear least squares estimator
 - 4.2 Consistency and asymptotic normality: General theory
 - 4.2.1 Consistency
 - 4.2.2 Asymptotic normality
 - 4.3 Consistency and asymptotic normality of nonlinear least squares estimators in the i.i.d. case
 - 4.3.1 Consistency

- 4.3.2 Asymptotic normality
 - 4.3.3 Consistent estimation of the asymptotic variance matrix
 - 4.4 Consistency and asymptotic normality of the nonlinear least squares estimator under data heterogeneity
 - 4.4.1 Data heterogeneity
 - 4.4.2 Strong and weak consistency
 - 4.4.3 Asymptotic normality
 - 4.5 Testing parameter restrictions: The Wald test
5. TESTS FOR MODEL MISSPECIFICATION
 - 5.1 White's version of Hausman's test
 - 5.2 Newey's M-test
 - 5.2.1 Introduction
 - 5.2.2 The conditional M-test
 - 5.3 A consistent conditional M-test
 - 5.4 The integrated M-test
6. THE NADARAYA-WATSON KERNEL REGRESSION FUNCTION ESTIMATOR
 - 6.1 Introduction
 - 6.2 Asymptotic normality in the continuous case
 - 6.3 Uniform consistency in the continuous case
 - 6.4 Discrete and mixed continuous-discrete regressors
 - 6.4.1 The discrete case
 - 6.4.2 The mixed continuous-discrete case
 - 6.5 The choice of the kernel
 - 6.6 The choice of the window width
 - 6.7 An empirical application to specification of household expenditure systems and equivalence scales
 - 6.7.1 Introduction
 - 6.7.2 Model and data
 - 6.7.3 The results
 - 6.7.4 Sample selection
7. SAMPLE MOMENTS INTEGRATING NORMAL KERNEL ESTIMATORS
 - 7.1 Kernel density estimators
 - 7.1.1 Integral conditions
 - 7.1.2 Sample moments integrating kernel density estimators
 - 7.1.3 The Dirac-catastrophe and the modified SMINK density estimator

- 7.1.4 How to choose the window width of the modified SMINK density estimator
- 7.2 Regression
 - 7.2.1 SMINK estimators of a regression function
 - 7.2.2 How to choose the window width parameters
- 7.3 A numerical example
- 7.4 Proofs

- 8. NONLINEAR REGRESSION WITH DISCRETE EXPLANATORY VARIABLES
 - 8.1. The earnings function
 - 8.2. The functional form of a regression model with discrete explanatory variables
 - 8.3. The choice of the linear separator
 - 8.4. Estimating and testing the regression function
 - 8.4.1 Estimation
 - 8.4.2 Model specification testing
 - 8.4.3 The selection of the polynomial order
 - 8.5 Proofs
 - 8.5.1 Proof of theorem 8.4.2
 - 8.5.2 Proof of theorem 8.4.3

PART 3: TIME SERIES

- 9. CONDITIONING AND DEPENDENCE
 - 9.1 Conditional expectations relative to a Borel field
 - 9.1.1 Definition and basic properties
 - 9.1.2 Martingales
 - 9.1.3 Martingale convergence theorems
 - 9.1.4 A martingale difference central limit theorem
 - 9.2 Measures of dependence
 - 9.2.1 Mixingales
 - 9.2.2 Uniform and strong mixing
 - 9.2.3 ν -Stability
 - 9.3 Weak laws of large numbers for dependent random variables
 - 9.4 Proper heterogeneity and uniform laws for functions of infinitely many random variables

- 10. FUNCTIONAL SPECIFICATION OF TIME SERIES MODELS
 - 10.1 Linear times series regression models
 - 10.1.1 Introduction

- 10.1.2 The Wold decomposition
- 10.1.3 Linear vector time series models
- 10.2 ARMA memory index models
 - 10.2.1 Introduction
 - 10.2.2 Finite conditioning of univariate rational-valued time series
 - 10.2.3 Infinite conditioning of univariate rational-valued time series
 - 10.2.3 The multivariate case
 - 10.2.4 The nature of the ARMA memory index parameters and the response functions
 - 10.2.5 Discussion
- 10.3 Nonlinear ARMAX models

- 11. ARMAX MODELS: ESTIMATION AND TESTING
 - 11.1 Estimation of linear ARMAX models
 - 11.1.1 Introduction
 - 11.1.2 Consistency and asymptotic normality
 - 11.2 Estimation of nonlinear ARMAX models
 - 11.3 A consistent $N(0,1)$ model specification test
 - 11.4 A consistent Hausman-type model specification test
 - 11.5 An autocorrelation test

- 12. NONPARAMETRIC TIME SERIES REGRESSION
 - 12.1 Assumptions and preliminary lemmas
 - 12.2 Consistency and asymptotic normality of time series kernel regression function estimators

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 para-

metric 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.

5. TESTS FOR MODEL MISSPECIFICATION

In the literature on model specification testing two trends can be distinguished. One trend consists of tests using one or more well-specified nonnested alternative specifications. See Cox (1961, 1962), Atkinson (1969, 1970), Quandt (1974), Pereira (1977, 1978), Pesaran and Deaton (1978), Davidson and MacKinnon (1981), among others. The other trend consists of tests of the orthogonality condition, i.e. the condition that the conditional expectation of the error relative to the regressors equals zero a.s., without employing a well-specified alternative. Notable work on this problem has been done by Ramsey (1969, 1970), Hausman (1978), White (1981), Holly (1982), Bierens (1982) and Newey (1985), among others.

A pair of models is called nonnested if it is not possible to construct one model out of the other by fixing some parameters. The nonnested models considered in the literature usually have different vectors of regressors, for testing nonnested models with common regressors makes no sense. In the latter case one may simply choose the model with the minimum estimated error variance, and this choice will be consistent in the sense that the probability that we pick the wrong model converges to zero. A serious point overlooked by virtually all authors is that nonnested models with different sets of regressors may be all correct. This is obvious if the dependent variable and the all regressors involved are jointly normally distributed and the nonnested models are all linear, for conditional expectations on the basis of jointly normally distributed random variables are always linear functions of the conditioning variables. Moreover, in each model involved the errors are independent of the regressors. In particular, in this case the tests of Davidson and MacKinnon (1981) will likely reject each of these true models, as these tests are based on combining linearly the nonnested models into a compound regression model. Since other tests of nonnested hypotheses are basically in the same spirit one may expect this flaw to be a pervasive phenomenon. Thus, these tests are only valid if either the null or only one of the alternatives is true. Moreover, tests of nonnested hypotheses may have low power against nonspecified alternatives, as pointed out by Bierens (1982). Therefore we shall not review these tests further.

In this chapter we only consider tests of the orthogo-

nality condition, without employing a specific alternative. First we discuss White's version of Hausman's test in section 5.1 and then, in section 5.2, the more general M-test of Newey. In section 5.3 we modify the M-test to a consistent test and in section 5.4 we consider a further elaboration of Bierens' integrated M-test.

5.1 *White's version of Hausman's test*

In an influential paper, Hausman (1978) proposed to test for model misspecification by comparing an efficient estimator with a consistent but inefficient estimator. Under the null hypothesis that the model is correctly specified the difference of these estimators times the square root of the sample size, will converge in distribution to the normal with zero mean, whereas under the alternative that the model is misspecified it is likely that these two estimators have different probability limits. White (1981) has extended Hausman's test to nonlinear models, using the nonlinear least squares estimator as the efficient estimator and a weighted nonlinear least squares estimator as the nonefficient consistent estimator.

The null hypothesis to be tested is that assumption 4.1.1. holds:

$$H_0: E(Y_j | X_j) = f(X_j, \theta_0) \text{ a.s. for some } \theta_0 \in \Theta,$$

where $f(x, \theta)$ is a given Borel measurable real function on $R^k \times \Theta$ which for each $x \in R^k$ is continuous on the compact Borel set $\Theta \subset R^m$.

The weighted nonlinear least squares estimator is a measurable solution $\hat{\theta}^*$ of:

$$\hat{\theta}^* \in \Theta \text{ a.s., } \hat{Q}^*(\hat{\theta}^*) = \inf_{\theta \in \Theta} \hat{Q}^*(\theta),$$

where

$$\hat{Q}^*(\theta) = (1/n) \sum_{j=1}^n [Y_j - f(X_j, \theta)]^2 w(x_j),$$

with $w(\cdot)$ a positive Borel measurable real weight function on R^k . Following White (1981), we shall now set forth conditions such that under the null hypothesis,

$$\sqrt{n}(\hat{\theta} - \hat{\theta}^*) \rightarrow N_m[0, \Omega] \text{ in distr.},$$

with $\hat{\theta}$ the nonlinear least squares estimator, whereas if H_0 is false,

$$\text{plim}_{n \rightarrow \infty} \hat{\theta} \neq \text{plim}_{n \rightarrow \infty} \hat{\theta}^*.$$

Given a consistent estimator $\hat{\Omega}$ of the asymptotic variance matrix Ω the test statistic of White's version of Hausman's test is now

$$\hat{W}^* = n(\hat{\theta} - \hat{\theta}^*)' \hat{\Omega}^{-1} (\hat{\theta} - \hat{\theta}^*),$$

which is asymptotically χ_m^2 distributed under H_0 and converges in probability to infinity if H_0 is false.

Let us now list the maintained hypotheses which are assumed to hold regardless whether or not the model is correctly specified.

Assumption 5.1.1. Assumption 4.3.1 holds and $E Y_j^2 w(X_j) < \infty$.

Assumption 5.1.2. Assumption 4.3.2 holds and

$$E \sup_{\theta \in \Theta} f(X_1, \theta)^2 w(X_1) < \infty.$$

Assumption 5.1.3. There are unique vectors θ_* and θ_{**} in Θ such that

$$E[E(Y_1 | X_1) - f(X_1, \theta_*)]^2 = \inf_{\theta \in \Theta} E[E(Y_1 | X_1) - f(X_1, \theta)]^2$$

and

$$\begin{aligned} E[E(Y_1 | X_1) - f(X_1, \theta_{**})]^2 w(X_1) \\ = \inf_{\theta \in \Theta} E[E(Y_1 | X_1) - f(X_1, \theta)]^2 w(X_1). \end{aligned}$$

If H_0 is false then $\theta_* \neq \theta_{**}$.

Assumption 5.1.4. The parameter space Θ is convex and $f(x, \theta)$ is for each $x \in \mathbb{R}^k$ twice continuously differentiable on Θ . If H_0 is true then θ_0 is an interior point of Θ .

Assumption 5.1.5. Let assumption 4.3.5 hold. Moreover, let for $i, i_1, i_2 = 1, \dots, m$,

$$E \sup_{\theta \in \Theta} [(\partial/\partial \theta_{i_1}) f(X_1, \theta)]^2 w(X_1) < \infty,$$

$$E \sup_{\theta \in \Theta} [(\partial/\partial \theta_{i_1})(\partial/\partial \theta_{i_2}) f(X_1, \theta)]^2 w(X_1) < \infty.$$

Assumption 5.1.6. The matrices

$$\Omega_2 = E [(\partial/\partial \theta') f(X_1, \theta_*)] [(\partial/\partial \theta) f(X_1, \theta_*)],$$

$$\Omega_2^* = E w(X_1) [(\partial/\partial \theta') f(X_1, \theta_*)] [(\partial/\partial \theta) f(X_1, \theta_*)]$$

are nonsingular.

Assumption 5.1.7. Let assumption 4.3.7 hold and let for $i_1, i_2 = 1, \dots, m$,

$$E \sup_{\theta \in \Theta} [Y_1 - f(X_1, \theta)]^2 w(X_1)^2 |(\partial/\partial \theta_{i_1}) f(X_1, \theta)| \\ \times |(\partial/\partial \theta_{i_2}) f(X_1, \theta)| < \infty.$$

Finally, denoting

$$\Omega_1 = E [Y_1 - f(X_1, \theta_*)]^2 [(\partial/\partial \theta') f(X_1, \theta_*)] [(\partial/\partial \theta) f(X_1, \theta_*)],$$

$$\Omega_1^* = E [Y_1 - f(X_1, \theta_*)]^2 w(X_1) [(\partial/\partial \theta') f(X_1, \theta_*)] [(\partial/\partial \theta) f(X_1, \theta_*)],$$

$$\Omega_1^{**} = E [Y_1 - f(X_1, \theta_*)]^2 w(X_1)^2 [(\partial/\partial \theta') f(X_1, \theta_*)] [(\partial/\partial \theta) f(X_1, \theta_*)],$$

$$\Omega = \Omega_2^{-1} \Omega_1 \Omega_2^{-1} - \Omega_2^{-1} \Omega_1^* (\Omega_2^*)^{-1} \\ - (\Omega_2^*)^{-1} \Omega_1^{**} \Omega_2^{-1} + (\Omega_2^*)^{-1} \Omega_1^{**} (\Omega_2^*)^{-1}$$

we assume:

Assumption 5.1.8. The matrix Ω is nonsingular.

Now observe that under assumptions 5.1.1 and 5.1.2,

$$\hat{Q}(\theta) - Q(\theta) \text{ a.s. uniformly on } \Theta$$

$$\hat{Q}^*(\theta) \rightarrow Q^*(\theta) \text{ a.s. uniformly on } \Theta,$$

where $\hat{Q}(\theta)$ and $Q(\theta)$ are defined in (4.1.9) and (4.3.3), respectively, and

$$Q^*(\theta) = E[Y_1 - f(X_1, \theta)]^2 w(X_1).$$

Together with assumption 5.1.3. these results now imply:

Theorem 5.1.1. Under assumptions 5.1.1-5.1.3,

$$\hat{\theta} \rightarrow \theta_{**} \text{ a.s. and } \hat{\theta}^* \rightarrow \theta_{**} \text{ a.s.}$$

(cf. Theorem 4.2.1). Moreover, if H_0 is true then clearly

$$\theta_{**} = \theta_{**} = \theta_0.$$

Now assume that H_j is true, and denote

$$U_j = Y_j - f(X_j, \theta_0).$$

Then it follows from assumptions 5.1.1-5.1.8, similarly to (4.2.12)

$$\text{plim}_{n \rightarrow \infty} \{ \sqrt{n}(\hat{\theta} - \theta_0) \cdot \Omega_2^{-1} (1/\sqrt{n}) \sum_{j=1}^n U_j (\partial/\partial \theta') f(X_j, \theta_0) \} = 0 \quad (5.1.1)$$

$$\text{plim}_{n \rightarrow \infty} \{ \sqrt{n}(\hat{\theta}^* - \theta_0) \cdot (\Omega_2^*)^{-1} (1/\sqrt{n}) \sum_{j=1}^n U_j w(X_j) (\partial/\partial \theta') f(X_j, \theta_0) \} = 0 \quad (5.1.2)$$

hence,

$$\text{plim}_{n \rightarrow \infty} \{ (\sqrt{n}(\hat{\theta} - \theta_0))', \sqrt{n}(\hat{\theta}^* - \theta_0)' \}' \cdot (1/\sqrt{n}) \sum_{j=1}^n Z_j \} = 0,$$

where

$$Z_j = U_j [\Omega_2^{-1} (\partial/\partial\theta) f(X_j, \theta_0), (\Omega_2^*)^{-1} w(X_j) (\partial/\partial\theta) f(X_j, \theta_0)]'$$

Moreover, from the central limit theorem it follows

$$(1/\sqrt{n}) \sum_{j=1}^n Z_j \rightarrow N_{2m}[0, A],$$

where

$$A = E Z_j Z_j' = [A_{i_1 i_2}], \quad i_1, i_2 = 1, 2,$$

is a $2m \times 2m$ matrix with $m \times m$ blocks

$$A_{11} = \Omega_2^{-1} \Omega_1 \Omega_2^{-1}, \quad A_{12} = \Omega_2^{-1} \Omega_1^* (\Omega_2^*)^{-1}$$

$$A_{21} = (\Omega_2^*)^{-1} \Omega_1^* \Omega_2^{-1}, \quad A_{22} = (\Omega_2^*)^{-1} \Omega_1^{**} (\Omega_2^*)^{-1}$$

From these results it easily follows now:

Theorem 5.1.2. Under H_0 and the assumption 5.1.1-5.1.8,

$$\sqrt{n}(\hat{\theta} - \hat{\theta}^*) \rightarrow N_m(0, \Omega) \text{ in distr.}$$

A consistent estimator of Ω can be constructed as follows. Let

$$\hat{\Omega}_1^{(1)} = (1/n) \sum_{j=1}^n w(X_j) \dot{I} [Y_j - f(X_j, \hat{\theta})]^2 \times [(\partial/\partial\theta') f(X_j, \hat{\theta})][(\partial/\partial\theta) f(X_j, \hat{\theta})]$$

$$\hat{\Omega}_2^{(1)} = (1/n) \sum_{j=1}^n w(X_j) \dot{I} [(\partial/\partial\theta') f(X_j, \hat{\theta})][(\partial/\partial\theta) f(X_j, \hat{\theta})]$$

$$\hat{\Omega}_1 = \hat{\Omega}_1^{(0)}, \quad \hat{\Omega}_1^* = \hat{\Omega}_1^{(1)}, \quad \hat{\Omega}_1^{**} = \hat{\Omega}_1^{(2)},$$

$$\hat{\Omega}_2 = \hat{\Omega}_2^{(0)}, \quad \hat{\Omega}_2^* = \hat{\Omega}_2^{(1)}$$

and define $\hat{\Omega}$ analogously to Ω . Then

Theorem 5.1.3. Under assumptions 5.1.1 - 5.1.7,

$$\hat{\Omega} \rightarrow \Omega \text{ a.s.}$$

regardless whether or not the null is true.

Combining theorems 5.1.1-5.1.3 we now have

Theorem 5.1.4. Under assumptions 5.1.1-5.1.8,

$$\hat{W}^* \rightarrow \chi_m^2 \text{ if } H_0 \text{ is true and}$$

$$\hat{W}^*/n \rightarrow (\theta_* - \theta_{**})' \Omega^{-1} (\theta_* - \theta_{**}) > 0 \text{ a.s. if } H_0 \text{ is false.}$$

The latter implies, of course, that $\text{plim}_{n \rightarrow \infty} \hat{W}^* = \infty$.

The power of this test heavily depends on the condition that under misspecification $\theta_* \neq \theta_{**}$, and for that the choice of the weight function $w(\cdot)$ is crucial. Take for example the true model

$$Y_j = X_{1j} + X_{2j} + X_{1j}X_{2j} + U_j$$

where the X_{1j} 's, X_{2j} 's and U_j 's are independent $N(0,1)$ distributed, and let $f(x, \theta) = \theta_1 x_1 + \theta_2 x_2$, $w(x) = x_1^2 + x_2^2$.

Then

$$E [Y_1 - f(X_1, \theta)]^2 = (1 - \theta_1)^2 + (1 - \theta_2)^2 + 2 \quad (5.1.3)$$

and

$$E [Y_1 - f(X_1, \theta)]^2 w(X_1) = 4(1 - \theta_1)^2 + 4(1 - \theta_2)^2 + 8, \quad (5.1.4)$$

hence $\theta_* - \theta_{**} = (1, 1)'$. Moreover, in this case we still have

$$\hat{W}^* \rightarrow \chi_2^2 \text{ in distr.,} \quad (5.1.5)$$

although the model is misspecified. Thus Hausman's test is not consistent against all alternatives, a result also confirmed by Holly (1982).

Exercises:

1. Prove (5.1.2)
2. Prove theorem 5.1.2
3. Prove theorem 5.1.3
4. Prove (5.1.3) and (5.1.4), using the fact that the fourth moment of a standard normally distributed random variable equals 3.
5. Prove (5.1.5).

5.2 *Newey's M-test*

5.2.1 *Introduction*

Newey (1985) argues that testing model correctness usually amounts to testing a null hypothesis of the form

$$H_0: E M(Y_j, X_j, \theta_0) = 0, \quad (5.2.1)$$

where $M(y, x, \theta) = (M_1(y, x, \theta), \dots, M_p(y, x, \theta))'$ is a vector-valued function on $R \times R^k \times \Theta$ (with Borel measurable components). A specification test can then be based on the sample moment vector

$$\hat{M}(\hat{\theta}) = (1/n) \sum_{j=1}^n M(Y_j, X_j, \hat{\theta}), \quad (5.2.2)$$

where $\hat{\theta}$ is, under H_0 , a consistent and asymptotically normally distributed estimator of θ_0 .

We show now that the Hausman-White test is indeed asymptotically equivalent under H_0 with a particular M-test. Let

$$M(y, x, \theta) = (y - f(x, \theta)) (\partial / \partial \theta') f(x, \theta) w(x) \quad (5.2.3)$$

and let $\hat{\theta}$ be the nonlinear least squares estimator. Under H_0 we have for the i -th component of $M(\hat{\theta})$,

$$\begin{aligned} \sqrt{n} \hat{M}_i(\hat{\theta}) &= (1/\sqrt{n}) \sum_{j=1}^n [Y_j - f(X_j, \hat{\theta})] (\partial / \partial \theta_i) f(X_j, \hat{\theta}) w(X_j) \\ &= (1/\sqrt{n}) \sum_{j=1}^n U_j (\partial / \partial \theta_i) f(X_j, \hat{\theta}) w(X_j) \\ &= (1/\sqrt{n}) \sum_{j=1}^n [f(X_j, \hat{\theta}) - f(X_j, \theta_0)] (\partial / \partial \theta_i) f(X_j, \hat{\theta}) w(X_j). \end{aligned}$$

Using the mean value theorem we see that there exists a mean value $\hat{\theta}^{(i)}$ satisfying $|\hat{\theta}^{(i)} - \theta_0| \leq |\hat{\theta} - \theta_0|$ a.s. such that

$$\sqrt{n} \hat{M}_1(\hat{\theta}) = \sqrt{n} \hat{M}_1(\theta_0) + [(\partial/\partial\theta)\hat{M}_1(\hat{\theta}^{(1)})]\sqrt{n}(\hat{\theta}-\theta_0).$$

We leave it as an exercise (cf. exercise 1) to show that under the conditions of theorem 5.1.2,

$$\text{plim}_{n \rightarrow \infty} [(\partial/\partial\theta')\hat{M}_1(\hat{\theta}^{(1)}), \dots, (\partial/\partial\theta')\hat{M}_m(\hat{\theta}^{(m)})]' = \Omega_2^*,$$

hence

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} (\sqrt{n} \hat{M}(\hat{\theta}) - (1/\sqrt{n}) \sum_{j=1}^n U_j (\partial/\partial\theta') f(X_j, \theta_0) w(X_j) \\ + \Omega_2^* \sqrt{n}(\hat{\theta} - \theta_0)) = 0. \end{aligned} \quad (5.2.4)$$

Comparing this result with (5.1.2) we now see that

$$\text{plim}_{n \rightarrow \infty} (\sqrt{n} \hat{M}(\hat{\theta}) - \Omega_2^* \sqrt{n}(\hat{\theta}^* - \hat{\theta})) = 0,$$

hence

$$(\Omega_2^*)^{-1} \sqrt{n} \hat{M}(\hat{\theta})$$

has the same asymptotic normal distribution as

$$\sqrt{n}(\hat{\theta} - \hat{\theta}^*).$$

This result demonstrates the asymptotic equivalence under H_0 of this special case of the M-test and the Hausman-White test.

Next, consider the case that H_0 is false. Under the conditions of theorem 5.1.4 we have

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \hat{M}(\hat{\theta}) \\ = E[E(Y_1 | X_1) - f(X_1, \theta_*)] (\partial/\partial\theta') f(X_1, \theta_*) w(X_1) \end{aligned} \quad (5.2.5)$$

as is not hard to verify. Cf. exercise 2. Now assume that the function

$$[E(Y_1 | X_1) - f(X_1, \theta)]^2 w(X_1)$$

has no local extremum at $\theta = \theta_*$. This condition is only a slight augmentation of assumption 5.1.3. Then the right hand

side of (5.2.5) is unequal to the zero vector, i.e.,

$$\text{plim}_{n \rightarrow \infty} \hat{M}(\hat{\theta}) \neq 0.$$

This establishes the asymptotic power of the M-test under review. However, also the M-test is not generally watertight. It is not hard to verify that for the example at the end of section 5.1 this version of the M-test has also low power.

Another example of an M-test is the Wald test in section 4.5, with $M(y, x, \theta) = \eta(\theta)$. Also, Ramsey's (1969, 1970) RESET test may be considered as a special case of the M-test.

5.2.2 The conditional M-test

In regression analysis, where we deal with conditional expectations, model correctness usually corresponds to a null hypothesis of the form

$$H_0: E[r(Y_j, X_j, \theta) | X_j] = 0 \text{ a.s. if and only if } \theta = \theta_0. \quad (5.2.6)$$

For example in the regression case considered in section 4.3 an obvious candidate for this function r is $r(y, x, \theta) = y - f(x, \theta)$. Also, we may choose for $r(Y_j, X_j, \theta)$ the j -th term of

$$-\frac{1}{2}(\partial/\partial\theta')\hat{Q}(\theta),$$

where $\hat{Q}(\theta)$ is defined in (4.1.9), i.e.,

$$r(y, x, \theta) = [y - f(x, \theta)](\partial/\partial\theta')f(x, \theta). \quad (5.2.7)$$

Clearly,

$$E[r(Y_j, X_j, \theta) | X_j] = [E(Y_j | X_j) - f(X_j, \theta)](\partial/\partial\theta')f(X_j, \theta) = 0 \text{ a.s.} \\ \text{if and only if } E(Y_j | X_j) = f(X_j, \theta_0) \text{ a.s. and } \theta = \theta_0. \quad (5.2.8)$$

Furthermore, observe that in the case (5.2.7) the estimator $\hat{\theta}$ of θ_0 is such that

$$P\{(1/n)\sum_{j=1}^n r(Y_j, X_j, \hat{\theta}) = 0\} \rightarrow 1. \quad (5.2.9)$$

Cf. (4.2.7). This is true even if the model is misspecified, provided that $\hat{\theta}$ converges in probability to an interior point

of the parameter space Θ . Consequently we cannot choose $r=M$, for then

$$P(\hat{M}(\hat{\theta})=0) \rightarrow 1$$

anyhow and thus any test based on $\hat{M}(\hat{\theta})$ will have no power at all. What we need is a weight function, similarly to (5.2.3), i.e., let $r_i(y,x,\theta)$ be the i -th component of the vector $r(y,x,\theta)$, and let $w_i(x,\theta)$ be a weight function. Then

$$M_i(y,x,\theta) = r_i(y,x,\theta)w_i(x,\theta), \quad i=1,2,\dots,m. \quad (5.2.10)$$

Note that in the case (5.2.3),

$$w_i(x,\theta) = w(x), \quad r_i(y,x,\theta) = (y-f(x,\theta))(\partial/\partial\theta_i)f(x,\theta), \\ i=1,2,\dots,k.$$

In view of the above argument, we can now state the basic ingredients of the conditional M-test. First, let us assume that the data generating process is i.i.d.:

Assumption 5.2.1. The data generating process $\{(Y_j, X_j)\}$ with $Y_j \in \mathbb{R}$, $X_j \in \mathbb{R}^k$ is i.i.d.

The model is implicitly specified by the functions $r_i(y,x,\theta)$:

Assumption 5.2.2. For $i=1,2,\dots,m$ the functions $r_i(y,x,\theta)$ are Borel measurable real functions on $\mathbb{R} \times \mathbb{R}^k \times \Theta$, where Θ is compact Borel subset of \mathbb{R}^m , such that

$$E \sup_{\theta \in \Theta} |r_i(Y_j, X_j, \theta)| < \infty.$$

Moreover, for each $(y,x) \in \mathbb{R} \times \mathbb{R}^k$ the functions $r_i(y,x,\theta)$ are continuously differentiable on Θ . Let

$$r(y,x,\theta) = (r_1(y,x,\theta), \dots, r_m(y,x,\theta))'.$$

There exists a unique interior point θ_0 of Θ such that

$$E r(Y_j, X_j, \theta_0) = 0 \quad (\in \mathbb{R}^m).$$

Note that the latter condition does not say anything about model correctness. For example in the case of the regression model in section 4.3 this condition merely says that the function

$$E[Y_j - f(X_j, \theta)]^2 = E[Y_j - E(Y_j | X_j)]^2 + E[E(Y_j | X_j) - f(X_j, \theta)]^2$$

has a unique extremum on Θ at an interior point θ_0 of Θ , without saying that

$$E[E(Y_j | X_j) - f(X_j, \theta_0)]^2 = 0.$$

Next, we consider an estimator θ_n of θ_0 , satisfying (5.2.9):

Assumption 5.2.3. Let (θ_n) be a sequence of random vectors in Θ such that

$$\lim_{n \rightarrow \infty} P\{(1/n) \sum_{j=1}^n r(Y_j, X_j, \theta_n) = 0\} = 1.$$

We may think of θ_n as an estimator obtained by minimizing an objective function of the form

$$D_n(\theta) = \left\| (1/n) \sum_{j=1}^n r(Y_j, X_j, \theta) \right\| \quad (5.2.11)$$

over Θ , where $\|\cdot\|$ is the Euclidean norm. Then under assumptions 5.2.1-5.2.3,

$$D_n(\theta) \rightarrow D(\theta) = |E r(Y_1, X_1, \theta)| \text{ a.s., uniformly on } \Theta. \quad (5.2.12)$$

Cf. exercise 3. This result, together with assumption 5.2.3, implies $\theta_n \rightarrow \theta_0$ in prob., i.e.,

Theorem 5.2.1. Under assumptions 5.2.1-5.2.3, $\text{plim}_{n \rightarrow \infty} \theta_n = \theta_0$.

Proof: We have

$$\begin{aligned} 0 \leq D(\theta_n) &= D(\theta_n) - D(\theta_0) = D(\theta_n) - D_n(\theta_n) + D_n(\theta_n) \\ &\leq \sup_{\theta \in \Theta} |D_n(\theta) - D(\theta)| + D_n(\theta_n) \rightarrow 0 \text{ in prob.} \end{aligned} \quad (5.2.13)$$

by (5.2.12) and assumption 5.2.3, hence

$$\text{plim}_{n \rightarrow \infty} D(\theta_n) - D(\theta_0) = 0. \quad (5.2.14)$$

The theorem follows now similarly to the proof of theorem 4.2.1. Q.E.D.

Also asymptotic normality applies. By the mean value theorem we have

$$\begin{aligned} & (1/\sqrt{n}) \sum_{j=1}^n r_i(Y_j, X_j, \theta_n) - (1/\sqrt{n}) \sum_{j=1}^n r_i(Y_j, X_j, \theta_0) \\ & + [(1/n) \sum_{j=1}^n (\partial/\partial \theta') r_i(Y_j, X_j, \theta_n^{(i)})]' \sqrt{n}(\theta_n - \theta_0), \end{aligned} \quad (5.2.15)$$

where $\theta_n^{(i)}$ is a mean value satisfying $|\theta_n^{(i)} - \theta_0| \leq |\theta_n - \theta_0|$.
Now assume

Assumption 5.2.4. For $i, \ell=1, 2, \dots, m$, let

$$E \sup_{\theta \in \Theta} |(\partial/\partial \theta_\ell) r_i(Y_1, X_1, \theta)| < \infty$$

Then by theorem 2.7.5,

$$(1/n) \sum_{j=1}^n (\partial/\partial \theta') r_i(Y_j, X_j, \theta) \rightarrow E(\partial/\partial \theta') r_i(Y_1, X_1, \theta) \text{ a.s.} \quad (5.2.16)$$

uniformly on Θ , hence by theorem 2.6.1 and the fact that by theorem 5.2.1,

$$\text{plim}_{n \rightarrow \infty} \theta_n^{(i)} = \theta_0.$$

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} (1/n) \sum_{j=1}^n (\partial/\partial \theta') r_i(Y_j, X_j, \theta_n^{(i)}) \\ = E(\partial/\partial \theta') r_i(Y_1, X_1, \theta_0) \end{aligned}$$

Denoting

$$\begin{aligned} \hat{\Gamma}^* = & [(1/n) \sum_{j=1}^n (\partial/\partial \theta') r_1(Y_j, X_j, \theta_n^{(1)}), \dots \\ & \dots, (1/n) \sum_{j=1}^n (\partial/\partial \theta') r_m(Y_j, X_j, \theta_n^{(m)})]' \end{aligned}$$

and

$$\Gamma = [E(\partial/\partial\theta')r_1(Y_1, X_1, \theta_0), \dots, E(\partial/\partial\theta')r_m(Y_1, X_1, \theta_0)]' \quad (5.2.17)$$

we thus have

$$\text{plim}_{n \rightarrow \infty} \hat{\Gamma}^{*k} = \Gamma \quad (5.2.18)$$

Next assume:

Assumption 5.2.5. The $(m \times m)$ matrix Γ is nonsingular.

Then

$$\text{plim}_{n \rightarrow \infty} \hat{\Gamma}^{*-1} = \Gamma^{-1} \quad (5.2.19)$$

(Cf. exercise 4), whereas by (5.2.15) and assumption 5.2.3

$$\text{plim}_{n \rightarrow \infty} \{ (1/\sqrt{n}) \sum_{j=1}^n \hat{\Gamma}^{*-1} r(Y_j, X_j, \theta_0) + \sqrt{n}(\theta_n - \theta_0) \} = 0. \quad (5.2.20)$$

This result, together with (5.2.19), implies

$$\text{plim}_{n \rightarrow \infty} \{ (1/\sqrt{n}) \sum_{j=1}^n \Gamma^{-1} r(Y_j, X_j, \theta_0) + \sqrt{n}(\theta_n - \theta_0) \} = 0 \quad (5.2.21)$$

provided that

$$(1/\sqrt{n}) \sum_{j=1}^n r(Y_j, X_j, \theta_0) \text{ converges in distribution.}$$

Cf. exercise 5. A sufficient additional condition for the latter is:

Assumption 5.2.6. For $i=1, 2, \dots, m$, $E \sup_{\theta \in \Theta} [r_i(Y_1, X_1, \theta)]^2 < \infty$,

as then the $(m \times m)$ variance matrix

$$\Delta = E r(Y_1, X_1, \theta_0) r(Y_1, X_1, \theta_0)' \quad (5.2.22)$$

has finite elements. Since the random vectors $r(Y_j, X_j, \theta_0)$ are i.i.d. with zero mean vectors and finite variance matrix Δ it follows now from the central limit theorem

$$(1/\sqrt{n})\sum_{j=1}^n r(Y_j, X_j, \theta_0) \rightarrow N_m[0, \Delta] \text{ in distr.} \quad (5.2.23)$$

Combining (5.2.21) and (5.2.23) yields:

Theorem 5.2.2. Under assumptions 5.2.1-5.2.6,

$$\sqrt{n}(\theta_n - \theta_0) \rightarrow N_m[0, \Omega] \text{ in distr., where } \Omega = (\Gamma)^{-1}\Delta(\Gamma')^{-1}.$$

Note that this result holds regardless whether or not the underlying model is correctly specified. A similar result has been obtained by White (1980, 1982) for misspecified linear models and maximum likelihood under misspecification. Moreover, if the underlying model is correctly specified, r defined by (5.2.7) and if θ_n is the nonlinear least squares estimator then Ω reduces to $\Omega_2^{-1}\Omega_1\Omega_2^{-1}$. Cf. theorem 4.3.2 and exercise 6.

A consistent estimator of Ω can be obtained as follows.

Let

$$\hat{\Gamma} = \left[\begin{array}{c} (1/n)\sum_{j=1}^n (\partial/\partial\theta')r_1(Y_j, X_j, \theta_n), \dots, \dots, \\ \dots, (1/n)\sum_{j=1}^n (\partial/\partial\theta')r_m(Y_j, X_j, \theta_n) \end{array} \right]' \quad (5.2.24)$$

$$\hat{\Delta} = (1/n)\sum_{j=1}^n r(Y_j, X_j, \theta_n)r(Y_j, X_j, \theta_n)' \quad (5.2.25)$$

$$\hat{\Omega} = \hat{\Gamma}^{-1}\hat{\Delta}(\hat{\Gamma}')^{-1} \quad (5.2.26)$$

Then

Theorem 5.2.3. Under assumption 5.2.1-5.2.6, $\text{plim}_{n \rightarrow \infty} \hat{\Omega} = \Omega$.

Proof: Exercise 7.

We now come to the null hypothesis to be tested. As said before, the null hypothesis $E(Y_j | X_j) = f(X_j, \theta_0)$ a.s. is equivalent with (5.2.6), where r is defined by (5.2.7). If H_0 is true then for $i=1, 2, \dots, m$,

$$E r_i(Y_j, X_j, \theta_0)w_i(X_j, \theta_0) = E\{E[r_i(Y_j, X_j, \theta_0) | X_j]w_i(X_j, \theta_0)\} = 0 \quad (5.2.27)$$

for all weight functions w_i for which the expectation involved is defined. If H_0 is false there exist continuous weight functions w_i for which (5.2.27) does not hold. Cf. theorem 3.1.2. Now let us specify these weight functions.

Assumption 5.2.7. The weight functions $w_i(x, \theta)$, $i=1, 2, \dots, m$, are Borel measurable real functions on $R^k \times \Theta$ such that for $i=1, 2, \dots, m$,

- (I) for each $x \in R^k$, $w_i(x, \theta)$ is continuously differentiable on Θ ;
- (II) $E \sup_{\theta \in \Theta} |r_i(Y_1, X_1, \theta)|^3 |w_i(X_1, \theta)| < \infty$;
- (III) $E \sup_{\theta \in \Theta} [r_i(Y_1, X_1, \theta) w_i(X_1, \theta)]^2 < \infty$;
- (IV) for $\ell=1, 2, \dots, m$, $E \sup_{\theta \in \Theta} |\partial/\partial \theta_\ell [r_i(Y_1, X_1, \theta) w_i(X_1, \theta)]| < \infty$;
- (V) if H_0 is false then $E r_i(Y_1, X_1, \theta_0) w_i(X_1, \theta_0) \neq 0$ for at least one i .

The conditions (I)-(IV) are regularity conditions. Condition (V), however, is the crux of the conditional M-test, because it determines the power of the test. It says that the random vector function

$$\hat{M}(\theta) = (1/n) \sum_{j=1}^n M(Y_j, X_j, \theta) \quad (5.2.28)$$

with

$$\begin{aligned} M(Y_j, X_j, \theta) &= (r_1(Y_j, X_j, \theta) w_1(X_j, \theta), \dots, r_m(Y_j, X_j, \theta) w_m(X_j, \theta))' \\ &= (M_1(Y_j, X_j, \theta), \dots, M_m(Y_j, X_j, \theta))' \end{aligned} \quad (5.2.29)$$

say, has nonzero mean at $\theta = \theta_0$ if H_0 is false. Thus, we actually test the null hypothesis

$$H_0^* : E \hat{M}(\theta_0) = 0 \quad (5.2.30)$$

against the alternative hypothesis

$$H_1^* : E \hat{M}(\theta_0) \neq 0 \quad (5.2.31)$$

However, it may occur that the choice of the weight functions is inappropriate in that H_0^* holds while H_0 is false. The choice of the weight functions is therefore more or less a matter of guesswork, as a watertight choice requires knowledge of the true model. In the next section it will be shown how the conditional M-test can be modified to a consistent test.

We are now going to construct a test statistic on the basis of the statistic $\hat{M}(\theta_n)$. Consider its i -th component $\hat{M}_i(\theta)$. By the mean value theorem we have

$$\begin{aligned} \sqrt{n} \hat{M}_i(\theta_n) &= \sqrt{n} \hat{M}_i(\theta_0) \\ &+ (1/n) \sum_{j=1}^n (\partial/\partial \theta) M_i(Y_j, X_j, \theta_n^{(i)}) \sqrt{n}(\theta_n - \theta_0) \end{aligned} \quad (5.2.32)$$

where $|\theta_n^{(i)} - \theta_0| \leq |\theta_n - \theta_0|$. Denoting

$$\begin{aligned} A &= [E(\partial/\partial \theta') r_1(Y_1, X_1, \theta_0) w_1(X_1, \theta_0), \dots \\ &\dots, E(\partial/\partial \theta') r_m(Y_1, X_1, \theta_0) w_m(X_1, \theta_0)]' \end{aligned} \quad (5.2.33)$$

it is not hard to show, similarly to (5.2.21), that (5.2.32) implies

$$\text{plim}_{n \rightarrow \infty} (\sqrt{n} \hat{M}(\theta_n) - \sqrt{n} \hat{M}(\theta_0) - A \sqrt{n}(\theta_n - \theta_0)) = 0 \quad (5.2.34)$$

Cf. exercise 8. Substituting

$$-(1/\sqrt{n}) \sum_{j=1}^n \Gamma^{-1} r(Y_j, X_j, \theta_0)$$

for $\sqrt{n}(\theta_n - \theta_0)$ (cf. (5.2.21)) it follows from (5.2.34) and (5.2.28),

$$\text{plim}_{n \rightarrow \infty} (\sqrt{n} \hat{M}(\theta_n) - (1/\sqrt{n}) \sum_{j=1}^n Z_j) = 0, \quad (5.2.35)$$

where

$$Z_j = M(Y_j, X_j, \theta_0) - A \Gamma^{-1} r(Y_j, X_j, \theta_0).$$

If H_0 is true then $E Z_j = 0$, and moreover it follows from assumption 5.2.7 that $E Z_j Z_j' = \Delta_*$, where

$$\begin{aligned} \Delta_* &= E M(Y_1, X_1, \theta_0) M(Y_1, X_1, \theta_0)' \\ &\quad - E M(Y_1, X_1, \theta_0) r(Y_1, X_1, \theta_0)' (\Gamma')^{-1} A' \\ &\quad - A \Gamma^{-1} E r(Y_1, X_1, \theta_0) M(Y_1, X_1, \theta_0)' + A \Gamma^{-1} \Delta (\Gamma')^{-1} A', \end{aligned} \quad (5.2.36)$$

is well-defined. By the central limit theorem and (5.2.35) we now have

$$\sqrt{n} \hat{M}(\hat{\theta}_n) \rightarrow N_m[0, \Delta_*] \text{ in distr. under } H_0. \quad (5.2.37)$$

Moreover, under H_1^* we have

$$\text{plim}_{n \rightarrow \infty} \hat{M}(\hat{\theta}_n) = E M(Y_1, X_1, \theta_0) \neq 0. \quad (5.2.38)$$

A consistent estimator of Δ_* can be obtained as follows. Let $\hat{\Gamma}$, $\hat{\Delta}$ and $\hat{\Omega}$ be defined in (5.2.23)-(5.2.25), let

$$\begin{aligned} \hat{A} &= [(1/n) \sum_{j=1}^n (\partial/\partial \theta') \{r_1(Y_j, X_j, \theta_n) w_1(X_j, \theta_n)\}, \dots \\ &\quad \dots, (1/n) \sum_{j=1}^n (\partial/\partial \theta') \{r_m(Y_j, X_j, \theta_n) w_m(X_j, \theta_n)\}]' \\ \hat{B} &= (1/n) \sum_{j=1}^n M(Y_j, X_j, \theta_n) M(Y_j, X_j, \theta_n)' \\ \hat{C} &= (1/n) \sum_{j=1}^n M(Y_j, X_j, \theta_n) r(Y_j, X_j, \theta_n)' \end{aligned}$$

and

$$\hat{\Delta}_* = \hat{B} - \hat{C} (\hat{\Gamma}')^{-1} \hat{A}' - \hat{A} \hat{\Gamma}^{-1} \hat{C}' + \hat{A} \hat{\Omega} \hat{A}'. \quad (5.2.39)$$

Then:

Theorem 5.2.4. Under assumption 5.2.1-5.2.7, $\text{plim}_{n \rightarrow \infty} \hat{\Delta}_* = \Delta_*$.

Proof: Exercise 9.

Note that this result also holds if H_0 is false, although in that case Δ_* is no longer the asymptotic variance matrix of

$\sqrt{n} \hat{M}(\theta_n)$.

Finally, assume

Assumption 5.2.8. The matrix Δ_* is nonsingular,

and let

$$\hat{H} = n \hat{M}(\theta_n)' \hat{\Delta}_*^{-1} \hat{M}(\theta_n)$$

be the ultimate test statistic. Then

Theorem 5.2.5. Under assumptions 5.2.1-5.2.8,

(I) $\hat{H} \rightarrow \chi_m^2$ in distr. if H_0 is true,

(II) $\text{plim}_{n \rightarrow \infty} \hat{H}/n = E M(Y_1, X_1, \theta_0)' \Delta_*^{-1} E M(Y_1, X_1, \theta_0) > 0$ if H_0 is false.

Exercises:

1. Prove (5.2.4)
2. Prove (5.2.5)
3. Prove (5.2.12)
4. Why does (5.2.19) follow from (5.2.17)?
5. Prove (5.2.21)
6. Prove that $\Omega = \Omega_2^{-1} \Omega_1 \Omega_2^{-1}$ under the conditions in section 4.3.
7. Prove theorem 5.2.3
8. Prove (5.2.34)
9. Prove theorem 5.2.4. In particular, check which parts of assumption 5.2.7 have been used here.

5.3 A consistent conditional M-test

As mentioned before, the power of the conditional M-test heavily depends on the choice of the weight functions. Quoting Newey (1985, p.1054): "An important property of specification tests based on a finite set of moment conditions is that they may not be consistent. This inconsistency has been noted by in particular examples by Bierens (1982) and Holly (1982) and is a

pervasive phenomenon". Thus, the solution of the inconsistency problem is to use an infinite set of moment conditions. Theorem 3.3.4 suggests how to do that. Let ψ be a bounded Borel measurable one-to-one mapping from R^k into R^k , and replace $w_1(x, \theta)$ by $\exp[\xi' \psi(x)]$. Then theorem 3.3.4 says that the null hypothesis is false, i.e.,

$$H_1: P(E[r(Y_j, X_j, \theta_0) | X_j] = 0) < 1,$$

if and only if $E r(Y_j, X_j, \theta_0) \exp(\xi' \psi(X_j)) \neq 0$, except on a set S with $\mu(S) = 0$, where μ is a probability measure induced by an absolutely continuous k -variate distribution. Denoting

$$M(y, x, \theta, \xi) = r(y, x, \theta) \exp[\xi' \psi(x)] \quad (5.3.1)$$

we thus have

$$E M(Y_j, X_j, \theta_0, \xi) \neq 0 \text{ for all } \xi \notin S \text{ if } H_0 \text{ is false} \quad (5.3.2)$$

whereas clearly

$$E M(Y_j, X_j, \theta_0, \xi) = 0 \text{ for all } \xi \in R^k \text{ if } H_0 \text{ is true.} \quad (5.3.3)$$

Next, let

$$\hat{M}(\theta_n, \xi) = (1/n) \sum_{j=1}^n M(Y_j, X_j, \theta_n, \xi) \quad (5.3.4)$$

$$\hat{H}(\xi) = n \hat{M}(\theta_n, \xi)' \hat{\Delta}_*(\xi)^{-1} \hat{M}(\theta_n, \xi), \quad (5.3.5)$$

where $\hat{\Delta}_*(\xi)$ is defined in (5.2.39) with $w_1(x, \theta)$ replaced by $\exp[\xi' \psi(x)]$, and assume that in particular assumption 5.2.8 holds for every $\xi \in R^k \setminus \{0\}$. Then we have

$$\hat{H}(\xi) \rightarrow \chi_m^2 \text{ in distr. for every } \xi \in R^k \setminus \{0\} \text{ if } H_0 \text{ is true.} \quad (5.3.6)$$

and

$$\text{plim}_{n \rightarrow \infty} \hat{H}(\xi)/n = E M(Y_1, X_1, \theta_0, \xi)' \Delta_*(\xi)^{-1} E M(Y_1, X_1, \theta_0, \xi) > 0 \\ \text{for all } \xi \in R^k \setminus S \text{ if } H_0 \text{ is false,} \quad (5.3.7)$$

where $\Delta_x(\xi)$ is defined in (5.2.35).

The latter result indicates that this version of the conditional M-test is 'almost surely' consistent (i.e., has asymptotic power 1 against all deviations from the null), as consistency only fails for ξ in a null set S of the measure μ . Also, note that we actually have imposed an infinite number of moment restrictions, namely the restrictions (5.3.3). Furthermore, observe that the exclusion of $\xi = 0$ is essential for (5.3.6) and (5.3.7), because by assumption 5.2.3,

$$P(\hat{M}(\theta_n, 0) = 0) \rightarrow 1,$$

hence

$$\hat{H}(0) \rightarrow 0 \text{ in prob.}$$

regardless whether or not H_0 is true. Thus, the set S contains at least the origin of R^k .

One might argue now that the problem of how to choose the weight function $w(x, \theta)$ has not been solved but merely been shifted to the problem of how to choose the vector ξ in the weight function $\exp[\xi' \psi(x)]$. Admittedly, in the present approach one has still to make a choice, but the point is that our choice will now be far less crucial for the asymptotic power of the test, for the asymptotic power will be equal to 1 for 'almost' any $\xi \in R^k$, namely all ξ outside a null set with respect to an absolutely continuous k -variate distribution. If one would pick ξ randomly from such a distribution then ξ will be an admissible choice with probability 1. In fact, the asymptotic properties of the test under H_0 will not be affected by choosing ξ randomly, whereas the asymptotic power will be 1 without worrying about the null set S :

Theorem 5.3.1. Let $\hat{H}(\xi)$ be the test statistic of the conditional M-test with weight functions $w_i(x, \theta) = \exp[\xi' \psi(x)]$, where ψ is a bounded Borel measurable one-to-one mapping from R^k into R^k . Suppose that the conditions of theorem 5.2.5 hold for $\xi \in R^k \setminus \{0\}$, possibly except assumption 5.2.7(V). Let ξ be a random drawing from an arbitrary absolutely continuous k -variate distribution. Then

$$\hat{H}(\xi) \rightarrow \chi_m^2 \text{ if } H_0 \text{ is true} \quad (5.3.8)$$

and

$$\text{plim}_{n \rightarrow \infty} \hat{H}(\xi) = \infty \text{ if } H_0 \text{ is false} \quad (5.3.9)$$

Proof: First, assume that H_0 is true, so that for every $\xi \in \mathbb{R}^k \setminus \{0\}$, $\hat{H}(\xi) \rightarrow \chi_m^2$ in distr. Then by theorem 2.3.6

$$E \exp[i \cdot t \hat{H}(\xi)] \rightarrow (1 - 2it)^{-m/2} = \varphi_m(t) \quad (5.3.10)$$

for every $t \in \mathbb{R}$ and every fixed $\xi \in \mathbb{R}^k \setminus \{0\}$, where $\varphi_m(t)$ is the characteristic function of the χ_m^2 distribution. (Cf. section 2.3, exercise 3). Now let ξ be a random drawing from an absolutely continuous k -variate distribution with density $h(\xi)$. Then for every $t \in \mathbb{R}$,

$$\begin{aligned} E \exp[i \cdot t \hat{H}(\xi)] &= \int E \exp[i \cdot t \hat{H}(\xi)] h(\xi) d\xi \\ &\rightarrow \int \varphi_m(t) h(\xi) d\xi = \varphi_m(t) \end{aligned} \quad (5.3.11)$$

by bounded convergence (cf. theorem 2.2.2). Theorem 2.3.6 says that this result implies that $\hat{H}(\xi) \rightarrow \chi_m^2$ in distr.

Second, assume that H_0 is false. Then there exists a null set S of the distribution of ξ such that for every $\xi \in \mathbb{R}^k \setminus S$,

$$E M(Y_1, X_1, \theta_0, \xi) \neq 0.$$

Hence

$$\text{plim}_{n \rightarrow \infty} \hat{H}(\xi)/n = E M(Y_1, X_1, \theta_0, \xi)' \Delta_{**}(\xi)^{-1} E M(Y_1, X_1, \theta_0, \xi) = T(\xi), \quad (5.3.12)$$

say, where

$$T(\xi) > 0 \text{ if } \xi \in \mathbb{R}^k \setminus S, \quad T(\xi) = 0 \text{ if } \xi \in S. \quad (5.3.13)$$

Again using theorem 2.3.6 we see that

$$\hat{H}(\xi)/n \rightarrow T(\xi) \text{ in distr.} \quad (5.3.14)$$

and since S is a null set of the distribution of ξ we have

$$P(T(\xi) > 0) = 1 \quad (5.3.15)$$

It is not hard to show now that (5.3.14) and (5.3.15) imply (5.3.9). Q.E.D.

Next, we have to deal with a practical problem regarding the choice of the bounded Borel measurable mapping ψ . Suppose for example that we would have chosen

$$\psi(x^{(1)}, \dots, x^{(k)}) = (\tan^{-1}(x^{(1)}), \dots, \tan^{-1}(x^{(k)}))'.$$

This mapping is clearly admissible. However, if the components X_{ij} of X_j are large then $\tan^{-1}(X_{ij})$ will be close to the upper-bound $\frac{1}{2}\pi$, hence $\exp(\xi' \psi(X_j))$ will be almost constant, i.e.,

$$\exp(\xi' \psi(X_j)) \approx \exp(\frac{1}{2}\pi \sum_{i=1}^k \xi_i)$$

and consequently

$$\hat{M}(\theta_n, \xi) \approx \{(1/n) \sum_{j=1}^n r(Y_j, X_j, \theta_n)\} \exp(\frac{1}{2}\pi \sum_{i=1}^k \xi_i).$$

Since the mean between the curled brackets equals the zero vector with probability converging to 1 [cf. assumption 5.2.3], $\hat{M}(\theta_n, \xi)$ will be close to the zero vector and consequently $H(\xi)$ will be close to zero. This will obviously destroy the power of the test. A cure for this problem is to standardize the X_j 's in $\psi(X_j)$. Thus let \bar{X}_i be the sample mean of the X_{ij} 's, let S_i be the sample standard deviation of the X_{ij} 's ($i=1, 2, \dots, k$), and let

$$\hat{Z}_j = (\tan^{-1}\{(X_{1j} - \bar{X}_1)/S_1\}, \dots, \tan^{-1}\{(X_{kj} - \bar{X}_k)/S_k\})'. \quad (5.3.16)$$

Then the proposed weight function is:

$$\exp(\xi' \hat{Z}_j). \quad (5.3.17)$$

It can be shown [cf. exercise 3] that using this weight function is asymptotically equivalent with using the weight function

$$\exp(\xi'Z_j) \quad (5.3.18)$$

where

$$Z_j = (\tan^{-1}[(X_{1j} - E X_{1j})/\sqrt{\text{var}(X_{1j})}], \dots, \dots, \tan^{-1}[(X_{kj} - E X_{kj})/\sqrt{\text{var}(X_{kj})}])' \quad (5.3.19)$$

Exercises:

1. Check the conditions in section 5.2.2 for the weight function $\exp[\xi'\psi(x)]$, and in particular verify that only assumption 5.2.8 is of concern.
2. Show that (5.3.12) holds uniformly on any compact subset of \mathbb{R}^k .
3. Verify that using the weight function (5.3.17) is asymptotically equivalent with using the weight function (5.3.18), provided that for $i=1,2,\dots,k$,

$$\text{plim}_{n \rightarrow \infty} \bar{X}_i = E X_{i1} \text{ and } \text{plim}_{n \rightarrow \infty} S_i = \sqrt{\text{var}(X_{i1})} > 0.$$

5.4 The integrated M-test

An alternative to plugging a random ξ in the test statistic $\hat{H}(\xi)$ defined in (5.3.5) is to take a weighted integral of $H(\xi)$, say

$$\hat{b} = \int \hat{H}(\xi) h(\xi) d\xi, \quad (5.4.1)$$

where $h(\xi)$ is a k -variate density function. This is the approach in Bierens (1982). This idea seems attractive because under H_1 it is possible to draw a ξ for which the function

$$T(\xi) = \text{plim}_{n \rightarrow \infty} \hat{H}(\xi)/n$$

[cf. (5.3.12)] is close to zero, despite the fact that

$$P(T(\xi) > 0) = 1.$$

In that case the small sample power of the test may be rather

poor. By integrating over a sufficient large domain of $T(\xi)$ we will likely cover the areas for which $T(\xi)$ is high, hence we may expect that a test based on b will in general have higher small sample power than the test in section 5.3. A disadvantage of this approach, however, is firstly that the limiting distribution of \hat{b} under H_0 is of an unknown type, and secondly that calculating b can be quite laborious.

It will be shown that under H_0 the test statistic \hat{b} is asymptotically equivalent to an integral of the form

$$\hat{b}^* = \int \{[(1/\sqrt{n})\sum_{j=1}^n Z_j(\xi)]'\} [(1/\sqrt{n})\sum_{j=1}^n Z_j(\xi)] h(\xi) d\xi, \quad (5.4.2)$$

provided $h(\xi)$ vanishes outside a compact set, where the $Z_j(\xi)$'s are for each $\xi \in R^k \setminus \{0\}$ independent random vectors in R^m with zero mean vector and unit variance matrix:

$$E Z_j(\xi) = 0, \quad E Z_j(\xi) Z_j(\xi)' = I_m. \quad (5.4.3)$$

Although

$$[(1/\sqrt{n})\sum_{j=1}^n Z_j(\xi)]' [(1/\sqrt{n})\sum_{j=1}^n Z_j(\xi)] \rightarrow \chi_m^2 \text{ in distr.}$$

for each $\xi \in R^k \setminus \{0\}$, this result does not imply that

$$\hat{b}^* \rightarrow \chi_m^2 \text{ in distr.}$$

On the other hand, the first moment of \hat{b}^* equals m , hence by Chebishev's inequality

$$P(\hat{b}^* \geq m/\varepsilon) \leq E \hat{b}^* / (m/\varepsilon) = \varepsilon \quad (5.4.4)$$

for every $\varepsilon > 0$. Since under H_0 , $\text{plim}_{n \rightarrow \infty} (\hat{b} - \hat{b}^*) = 0$, we may conclude that for every $\varepsilon > 0$,

$$\limsup_{n \rightarrow \infty} P(\hat{b} \geq m/\varepsilon) \leq \varepsilon \text{ under } H_0. \quad (5.4.5)$$

Moreover, if H_0 is false then $\text{plim}_{n \rightarrow \infty} \hat{b} = \infty$, hence

$$\lim_{n \rightarrow \infty} P(\hat{b} \geq m/\varepsilon) = 1 \text{ under } H_1. \quad (5.4.6)$$

These results suggest to use m/ε as a critical value for tes-

ting H_0 at the $\epsilon \times 100\%$ significance level, i.e.,

reject H_0 if $\hat{b} \geq m/\epsilon$ and accept H_0 if $\hat{b} < m/\epsilon$.

Admittedly, the actual type I error will be (much) smaller than ϵ , because Chebishev's inequality is not very sharp an inequality, but this is the price we have to pay for possible gains of small sample power.

The problem regarding the calculation of the integral (5.4.1) can be solved by drawing a sample $(\xi_1, \dots, \xi_{N_n})$ of size N_n ($N_n \rightarrow \infty$ as $n \rightarrow \infty$) from $h(\xi)$ and to use

$$\bar{b} = (1/N_n) \sum_{\ell=1}^{N_n} \hat{H}(\xi_\ell) \quad (5.4.7)$$

instead of \hat{b} . This will be asymptotically equivalent, i.e.,

$$\text{plim}_{n \rightarrow \infty} (\hat{b} - \bar{b}) = 0 \quad \text{under } H_0 \quad (5.4.8)$$

and

$$\text{plim}_{n \rightarrow \infty} \bar{b} = \int T(\xi) h(\xi) d\xi > 0 \quad \text{under } H_1. \quad (5.4.9)$$

Now let us turn to the proof of the asymptotic equivalence of \hat{b} and \hat{b}^* under H_0 . Observe that similarly to (5.2.32)

$$\begin{aligned} \sqrt{n} \hat{M}_i(\theta_n, \xi) &= \sqrt{n} \hat{M}_i(\theta_0, \xi) \\ &+ [(1/n) \sum_{j=1}^n (\partial/\partial \theta) M_i(Y_j, X_j, \theta_n^{(i)}(\xi), \xi)] / \sqrt{n}(\theta_n - \theta_0), \end{aligned} \quad (5.4.10)$$

where $\hat{M}_i(\theta, \xi)$ is the i -th component of $\hat{M}(\theta, \xi)$ defined in (5.3.4) and $\theta_n^{(i)}(\xi)$ is a mean value satisfying

$$|\theta_n^{(i)}(\xi) - \theta_0| \leq |\theta_n - \theta_0| \quad \text{a.s., for all } \xi \in R^k. \quad (5.4.11)$$

Let Ξ be a compact subset of R^k . Under the conditions of theorem 5.3.1 we have

$$(1/n) \sum_{j=1}^n (\partial/\partial \theta') M_i(Y_j, X_j, \theta, \xi) \rightarrow E(\partial/\partial \theta') M_i(Y_1, X_1, \theta, \xi) \quad (5.4.12)$$

a.s., uniformly on $\Theta \times \Xi$. Cf. theorem 2.7.5. Denoting

$$a_i(\theta, \xi) = E(\partial/\partial\theta') M_i(Y_1, X_1, \theta, \xi) \quad (5.4.13)$$

we thus have

$$\begin{aligned} & \text{plim}_{n \rightarrow \infty} \sup_{\xi \in \Xi} |(1/n) \sum_{j=1}^n (\partial/\partial\theta') M_i(Y_j, X_j, \theta_n^{(i)}(\xi), \xi) - a_i(\theta_0, \xi)| \\ &= \text{plim}_{n \rightarrow \infty} \sup_{\xi \in \Xi} |a_i(\theta_n^{(i)}(\xi), \xi) - a_i(\theta_0, \xi)| \\ &\leq \text{plim}_{n \rightarrow \infty} \sup_{\xi \in \Xi} \sup_{|\theta_* - \theta_0| \leq |\theta_n - \theta_0|} |a_i(\theta_*, \xi) - a_i(\theta_0, \xi)| = 0, \end{aligned} \quad (5.4.14)$$

where the last step follows from the continuity of $a_i(\theta, \xi)$ on the compact set $\Theta \times \Xi$ (hence $a_i(\theta, \xi)$ is uniformly continuous on $\Theta \times \Xi$), and the consistency of θ_n . Consequently, denoting

$$A(\xi) = [a_1(\theta_0, \xi), \dots, a_m(\theta_0, \xi)]', \quad (5.4.15)$$

cf. (5.2.33), we have

$$\text{plim}_{n \rightarrow \infty} \sup_{\xi \in \Xi} |\sqrt{n} \hat{M}(\theta_n, \xi) - \sqrt{n} \hat{M}(\theta_0, \xi) + A(\xi)/n(\theta_n - \theta_0)| = 0. \quad (5.4.16)$$

Next, let

$$c_j(\xi) = M(Y_j, X_j, \theta_0, \xi) + A(\xi) \Gamma^{-1} r(Y_j, X_j, \theta_0). \quad (5.4.17)$$

Then it follows from (5.2.21) and (5.4.16) that

$$\text{plim}_{n \rightarrow \infty} \sup_{\xi \in \Xi} |\sqrt{n} \hat{M}(\theta_n, \xi) - (1/\sqrt{n}) \sum_{j=1}^n c_j(\xi)| = 0. \quad (5.4.18)$$

Moreover, we have shown in section 5.2 that under H_0 ,

$$E c_j(\xi) = 0, \quad (5.4.19)$$

$$E c_j(\xi) c_j(\xi)' = \Delta_*(\xi).$$

Furthermore, it is not hard to show that the consistent estimator $\hat{\Delta}_*(\xi)$ of $\Delta_*(\xi)$ is also uniformly consistent on Ξ , and thus

(cf. exercise 1)

$$\hat{\Delta}_*(\xi)^{-1} \rightarrow \Delta_*(\xi)^{-1} \text{ in prob., uniformly on } \Xi. \quad (5.4.20)$$

Denoting

$$Z_j(\xi) = \Delta_*(\xi)^{-1/2} c_j(\xi) \quad (5.4.21)$$

it is now not too hard to show (cf. exercise 2) that

$$\text{plim}_{n \rightarrow \infty} \int_{\Xi} \left| \hat{H}(\xi) - \left[(1/\sqrt{n}) \sum_{j=1}^n Z_j(\xi) \right]' \left[(1/\sqrt{n}) \sum_{j=1}^n Z_j(\xi) \right] h(\xi) \right| d\xi = 0. \quad (5.4.22)$$

Moreover, we leave it as an exercise (cf. exercise 3) to show that under H_1 ,

$$\text{plim}_{n \rightarrow \infty} \sup_{\xi \in \Xi} \left| \hat{H}(\xi)/n - T(\xi) \right| = 0. \quad (5.4.23)$$

Cf. (5.3.12).

Summarizing, we now have shown:

Theorem 5.4.1. Let the conditions of theorem 5.3.1 hold and let $h(\xi)$ be a k -variate density vanishing outside a compact subset Ξ of R^k .

(I) Under H_0 we have $\text{plim}_{n \rightarrow \infty} (\hat{b} - \hat{b}^*) = 0$, where $E \hat{b}^* = m$.

(II) Under H_1 we have $\text{plim}_{n \rightarrow \infty} \hat{b}/n = \int T(\xi) h(\xi) d\xi > 0$.

(III) Replacing \hat{b} by \tilde{b} defined in (5.4.7), the above results go through.

Exercises

1. Prove (5.4.20)
2. Prove (5.4.22)
3. Prove (5.4.23)
4. Prove part III of theorem 5.4.1.

References:

- Atkinson, A.C. (1969), "A Test for Discriminating Between Models", *Biometrika* 56, 337-347.
- Atkinson, A.C. (1970), "A Method for Discriminating Between Models", *Journal of the Royal Statistical Society, Series B*, 32, 323-353.
- Bierens, H.J. (1982), "Consistent Model Specification Tests", *Journal of Econometrics* 20, 105-134.
- Cox, D.R. (1961), "Tests for Separate Families of Hypotheses", *Proceedings of the 4th Berkeley Symposium* 1, 105-123.
- Cox, D.R. (1962), "Further Results on Tests of Separate Families of Hypotheses", *Journal of the Royal Statistical Society, Series B*, 24, 406-424.
- Davidson, R. and J.G. MacKinnon (1981), "Several Tests for Model Specification in the Presence of Alternative Hypotheses", *Econometrica* 49, 781-793.
- Hausman, J.A. (1978), "Specification Tests in Econometrics", *Econometrica* 46, 1251-1271.
- Holly, A. (1982), "A Remark on Hausman's Specification Test", *Econometrica* 50, 749-760.
- Newey, W.K. (1985), "Maximum Likelihood Specification Testing and Conditional Moment Tests", *Econometrica* 53, 1047-1070.
- Pereira, B. de B. (1977), "A Note on the Consistency and on the Finite Sample Comparisons of Some Tests of Separate Families of Hypotheses", *Biometrika* 64, 109-113.
- Pereira, B. de B. (1978), "Tests and Efficiencies of Separate Regression Models", *Biometrika* 65, 319-327.
- Pesaran, M.H. and A.S. Deaton (1978), "Testing Non-nested Non-linear Regression Models", *Econometrica* 46, 677-694.
- Quandt, R.E. (1974), "A Comparison of Methods for Testing Non-nested Hypotheses", *Review of Economics and Statistics* 56, 92-99.
- Ramsey, J.B. (1969), "Tests for Specification Errors in Classical Linear Least-squares Regression Analysis", *Journal of the Royal Statistical Society, Series B*, 31, 350-371.

Ramsey, J.B. (1970), "Models, Specification Error, and Inference: A Discussion of Some Problems in Econometric Methodology", *Bulletin of the Oxford Institute of Economics and Statistics* 32, 301-318.

White, H. (1981), "Consequences and Detection of Misspecified Nonlinear Regression Models", *Journal of the American Statistical Association* 76, 419-433.

P 56979-W