VU Research Portal

Introduction to conditioning

Bierens, H.J.

1987

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)
Bierens, H. J. (1987). *Introduction to conditioning*. (Serie Research Memoranda; No. 1987-20). Faculty of Economics and Business Administration, Vrije Universiteit Amsterdam.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

Download date: 25. Jun. 2021

ET 20

05348

1987

SERIE RESEARCH MEMORANDA

INTRODUCTION TO CONDITIONING

Herman J. Bierens

Researchmemorandum 1987-20 April '87

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

INTRODUCTION TO CONDITIONING *

Herman J.Bierens
Department of Econometrics
Free University, Amsterdam

December 1988

---

* Chapter 3 of:
PRINCIPLES OF NONLINEAR AND NONPARAMETRIC REGRESSION ANALYSIS
To be published by Marcel Dekker Inc., New York.

BIBL. VRIJE UNIVERSITEIT
ET
05348.-
AMSTERDAM

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.

3. INTRODUCTION TO CONDITIONING

The concept of conditional expectation is basic to regression analysis, as regression models essentially represent conditional expectations. The theory of conditioning, however, is one of the most abstract and difficult parts of probability theory. In particular conditioning relative to a one-sided infinite sequence of random variables requires the concept of conditional expectation relative to a Borel field. We shall need this concept when we deal with time series regression models. However, in part 2 of this book where we deal with independent samples, we only need the concept of a conditional expectation relative to a random vector, and fortunately the latter conditional expectation concept can be defined in a much more transparant way than the former. Therefore we shall discuss the abstract theory of conditioning relative to a Borel field later on. Here we shall confine attention to the easier concept of conditional expectation relative to a random vector.

3.1 *Definition of conditional expectation*

Most intermediate textbooks on mathematical statistics define conditional expectations by using conditional densities and probabilities. For our purpose this elementary conditional expectation concept is not suitable. In particular our theory of model specification testing requires a more rigorous conditioning concept. Before we introduce this rigorous concept, however, we illustrate two basic features of the elementary conditional expectation concept. Thus let $(Y,X) \in R \times R^k$ be an absolutely continuously distributed random vector with density $f(y,x)$ and marginal density $h(x)$. Then the conditional density of Y relative to the event $X = x$ is defined as:

$$f(y|x) = f(y,x)/h(x) \quad \text{if} \quad h(x) > 0,$$
$$f(y|x) = 0 \quad \text{if} \quad h(x) = 0.$$

The conditional expectation of Y relative to the event $X = x$ is

$$E(Y|X = x) = \int_{-\infty}^{+\infty} yf(y|x)dy = g(x),$$

say. Plugging in X for x, we get the conditional expectation of Y relative to X:

$$E(Y|X) = g(X).$$

Thus $E(Y|X)$ is a *function* of X. Moreover, we also have

$$E(Y-E(Y|X))\psi(X) = E\, Y\psi(X) - E\, g(X)\psi(X) = 0$$

for every function ψ for which this expectation is defined, as easily follows from the above elementary definition. These two properties are basic to conditional expectations. In fact they form the defining properties:

Definition 3.1.1. Let Y be a random variable satisfying $E|Y| < \infty$ and let X be a random vector in R^k. The conditional expectation of Y relative to X, denoted by $E(Y|X)$, is defined as $E(Y|X) = g(X)$, where g is a Borel measurable real function on R^k such that for every bounded Borel measurable real function on ψ on R^k,

$$E(Y-g(X))\psi(X) = 0. \tag{3.1.1}$$

Example: Draw randomly a pair (Y,X) from the set

$$\{(1,1),(2,1),(3,2),(4,2)\}.$$

Since X takes only two values, any Borel measurable function φ of X is a.s. equal to a simple function of X, i.e.

$$\psi(X) = a\, I(X=1) + b\, I(X=2), \quad a,b \in R.$$

Now we have

$$E(Y-g(X))\psi(X) =$$

$$1/4\{(1-g(1))a + (2-g(1))a + (3-g(2))b + (4-g(2))b\}$$

$$= (3/4 - \tfrac{1}{2}g(1))a + (7/4 - \tfrac{1}{2}g(2))b = 0$$

for every $a \in R$, $b \in R$, hence $g(1) = 3/2$, $g(2) = 7/2$. Thus

$$E(Y|X) = 1.5\, I(X=1) + 3.5\, I(X=2).$$

Two problems now arise. First, does this function g always exist? The answer is yes, but for the proof we actually need the notion of a conditional expectation relative to a Borel field [see, e.g., Chung (1974,chapter 9)], together with the Radon-Nikodym theorem [see Royden (1968, p. 238)]. We shall not pursuit this point further here. Second, is $g(X)$ unique? Also the answer to this question is yes, due to the Radon-Nikodym theorem, in the sense that if there are two Borel measurable real functions g_1 and g_2 on R^k satisfying the definition then $g_1(X) = g_2(X)$ a.s. An alternative proof of the uniqueness of $g(X)$ is given by the following theorem of Bierens (1982), which is also of intrinsic interest and moreover is basic to our theory of model specification testing, in chapter 5.

Theorem 3.1.1. Let g_1 and g_2 be Borel measurable real functions on R^k. Let X be a random vector in R^k such that $E|g_1(X)| < \infty$, $E|g_2(X)| < \infty$. Let for non-random $t \in R^k$,

$$\varphi_1(t) = E\, g_1(X)\exp(i\cdot t'X), \quad \varphi_2(t) = E\, g_2(X)\exp(i\cdot t'X).$$

Then $P(g_1(X) = g_2(X)) < 1$ if and only if $\varphi_1(t) \neq \varphi_2(t)$ for some $t \in R^k$.

Now suppose that there exist two Borel measurable real functions g_1 and g_2 satisfying (3.1.1) for every bounded Borel measurable real function $\psi(x)$ on R^k. Then also

$$E(g_1(X) - g_2(X))\psi(X) = 0$$

and consequently

$$E\,(g_1(X) - g_2(X))\cos(t'X) = 0 \quad \text{for all} \quad t \in R^k,$$

$$E\,(g_1(X) - g_2(X))\sin(t'X) = 0 \quad \text{for all} \quad t \in R^k.$$

Since $\exp(i\cdot t'x) = \cos(t'x) + i\cdot\sin(t'x)$, it follows now that

$$\varphi_1(t) = \varphi_2(t) \text{ for all } t \in R^k,$$

which by theorem 3.1.1 implies $P[g_1(X) = g_2(X)] = 1$. Thus $g(X)$ is a.s. unique.

A byproduct of this argument is:

Theorem 3.1.2. Let $Y \in R$ be a random variable satisfying $E|Y| < \infty$ and let $X \in R^k$ be a random vector. Let g be a Borel measurable real function on R^k. If $E(Y - g(X))\psi(X) = 0$ for all bounded *continuous* functions ψ on R^k then $E(Y|X) = g(X)$.

Proof of theorem 3.1.1. Let

$$r(x) = g_1(x) - g_2(x).$$

Then r is Borel measurable, and so are

$$r_1(x) = \max(r(x),0), \quad r_2(x) = \max(-r(x),0).$$

Clearly we have $r = r_1 - r_2$, where r_1 and r_2 are non-negative. Now assume for the moment

$$c_1 = E\, r_1(X) > 0, \quad c_2 = E\, r_2(X) > 0. \tag{3.1.2}$$

Then we can define probability measures ν_1 and ν_2 on the Euclidean Borel field B^k by (cf. exercise 1.a)

$$\nu_j(B) = \int_B r_j(x)\nu(dx)/c_j, \quad j = 1,2, \tag{3.1.3}$$

where ν is the probability measure induced by X (cf. section 1.1) and B is an arbitrary Borel set in R^k. We may now write (cf. exercise 1.c)

$$E\, r(X)\exp(i\cdot t'X) = \int r(x)\exp(i\cdot t'x)\nu(dx)$$

$$= \int r_1(x)\exp(i\cdot t'x)\nu(dx) - \int r_2(x)\exp(i\cdot t'x)\nu(dx)$$

$$= c_1\int \exp(i\cdot t'x)\nu_1(dx) - c_2\int \exp(i\cdot t'x)\nu_2(dx)$$

$$= c_1\eta_1(t) - c_2\eta_2(t),$$

say, where

$$\eta_j(t) = \int \exp(i\cdot t'x)\nu_j(dx), \quad j = 1,2,$$

is the characteristic function of ν_j, j=1,2. If

$$E\ r(X)\exp(i\cdot t'X) = 0$$

then $c_1\eta_1(t) - c_2\eta_2(t) = 0$ for all $t \in R^k$. Substituting $t = 0$ yields:

$$c_1\eta_1(0) - c_2\eta_2(0) = c_1 - c_2 = 0, \tag{3.1.4}$$

hence

$$\eta_1(t) = \eta_2(t) \text{ for all } t \in R^k.$$

The latter implies that the corresponding probability measures are equal, i.e.,

$$\nu_1(B) = \nu_2(B) \text{ for all Borel sets B in } R^k. \tag{3.1.5}$$

From (3.1.3), (3.1.4) and (3.1.5) it follows now:

$$\int_B r_1(x)\nu(dx) = \int_B r_2(x)\nu(dx) \text{ for all } B \in \mathcal{B}^k$$

and consequently

$$\int_B r(x)\nu(dx) = 0 \quad \text{for all } B \in \mathcal{B}^k. \tag{3.1.6}$$

Now take

$$B_1 = \{x \in R^k : r(x) > 0\}.$$

This is a Borel set, for r is Borel measurable. Hence by (3.1.6), $\int_{B_1} r(x)\nu(dx) = 0$. This implies that $\nu(B_1) = 0$. Similarly, we have for

$$B_2 = \{x \in R^k : r(x) < 0\}$$

that $\nu(B_2) = 0$. Since B_1 and B_2 are disjoint we now have $\nu(B_1 \cup B_2) = \nu(B_1) + \nu(B_2) = 0$ or equivalently:

$$P(r(X) \neq 0) = 0.$$

This proves that $r(X) = g_1(X) - g_2(X) = 0$ a.s. if (3.1.2) holds. The proof for the case that $E\ r_1(X) = 0$ and/or $E\ r_2(X) = 0$ is left to the reader as an easy exercise. Q.E.D.

Exercises:
1. Let ν be a probability measure on $(R^k, \mathcal{B}^k)$ and let f be a non-negative Borel measurable function on R^k such that $\int f(x)\nu(dx) = c$ with $0 < c < \infty$. Define for $B \in \mathcal{B}^k$,

$$\mu(B) = \int_B f(x)\nu(dx)/c$$

a) Prove that μ is a probability measure on $(R^k, \mathcal{B}^k)$.
b) Prove that for every simple function ψ on R^k,

$$\int\psi(x)f(x)\nu(dx) = c\int\psi(x)\mu(dx)$$

c) Prove the same for bounded Borel measurable real functions ψ on R^k.
2. Check the proof of theorem 3.1.1 for the cases $c_1 = 0$; $c_2 > 0$, $c_1 > 0$; $c_2 = 0$ and $c_1 = c_2 = 0$.
3. Prove theorem 3.3.2.

3.2 *Basic properties of conditional expectations*

All the basic properties of conditional expectations, as are well-known from intermediate statistical textbooks, can easily be derived from definition 3.1.1 and theorem 3.1.1. We list them in theorem 3.2.1 below. The proofs are left to the reader as exercises.

Theorem 3.2.1. Let $Y \in R$ and $V \in R$ be random variables satisfying $E|Y| < \infty$, $E|V| < \infty$, and let $X \in R^k$ and $Z \in R^m$ be random vectors. We have:

(I) $E[E(Y|X,Z)|X] = E(Y|X) = E[E(Y|X)|X,Z]$;

(II) $E[E(Y|X)] = E\ Y$; $E(Y|Y) = Y$;

(III) Let $U = Y - E(Y|X)$. Then $E(U|X) = 0$ a.s.;

(IV) $E(Y + V|X) = E(Y|X) + E(V|X)$;

(V) $Y \le V$ a.s. implies $E(Y|X) \le E(V|X)$ a.s.;

(VI) $|E(Y|X)| \le E(|Y| \; |X)$ a.s.;

(VII) $E(Y\, f(X)|X) = f(X)E(Y|X)$ a.s. for every Borel measurable real function f on R^k satisfying $E|f(X)| < \infty$;

(VIII) Let Γ be a Borel measurable mapping from R^k into a subset of R^m. Then $E[E(Y|X)|\Gamma(X)] = E[Y|\Gamma(X)]$ a.s. If Γ is a one-to-one mapping then $E(Y|\Gamma(X)) = E(Y|X)$ a.s.

(IX) If X and Y are independent then $E(Y|X) = E\,Y$ a.s.

Hint. For proving (V) apply (3.1.1) with

$$\psi(X) = I[E(Y|X) > E(V|X)],$$

where I(.) is the indicator function.

Also Chebishev's, Holder's, Minkowski's, Liapounov's and Jensen's inequalities easily carry over to conditional expectations:

Theorem 3.2.2. (Chebishev's inequality) Let $Y \in R$, $X \in R^k$ and let φ be a positive monotonic increasing real function on $(0,\infty)$ such that $\varphi(y) = \varphi(-y)$ and $E\,\varphi(Y) < \infty$. Then for every $\delta > 0$,

$$E[I(|Y| > \delta)|\; X] \le E[\varphi(Y)|X]/\delta \text{ a.s.}$$

Proof: Let $\psi(X) = I(E[I(|Y| > \delta)|\; X] > E[\varphi(Y)|X]/\delta)$. Applying definition 3.1.1 we find $\psi(X) = 0$ a.s. Q.E.D.

Theorem 3.2.3. (Holder's inequality) Let $Y \in R$, $V \in R$, $E|Y|^p < \infty$, $E|V|^q < \infty$, $E|Y\cdot V| < \infty$, and $X \in R^k$, where $p > 1$ and $1/p + 1/q = 1$. Then

$$|E(Y\cdot V|X)| \le (E(|Y|^p|X))^{1/p}(E(|V|^q|\; X))^{1/q} \text{ a.s.}$$

Proof: Similarly to the unconditional case.

Theorem 3.2.4. (Minkowski's inequality) Let $Y \in R$, $V \in R$, $X \in R^k$ and $E|Y|^p < \infty$, $E|V|^p < \infty$ for some $p \geq 1$. Then

$$\{E[|Y+V|^p| X]\}^{1/p} \leq \{E[|Y|^p| X]\}^{1/p} + \{E[|V|^p| X]\}^{1/p} \text{ a.s.}$$

Proof: Similarly to the unconditional case.

Theorem 3.2.5. (Liapounov's inequality) Let $Y \in R$, $E|Y|^q < \infty$ for some $q > 1$, $X \in R^k$ and $1 \leq p < q$. Then

$$\{E(|Y|^p| X)\}^{1/p} \leq \{E(|Y|^q | X)\}^{1/q} \text{ a.s.}$$

Proof: Let $V = 1$ in theorem 3.2.3. Q.E.D.

Theorem 3.2.6. (Jensen's inequality) Let φ be a convex real function on R and let $Y \in R$, $X \in R^k$, $E|Y| < \infty$, $E|\varphi(Y)| < \infty$. Then

$$\varphi[E(Y|X)] \leq E[\varphi(Y)|X] \quad \text{a.s.}$$

Also the results in section 2.2 go through for conditional expectations. Although we do not need these generalizations we shall state and prove them here for completeness.

Theorem 3.2.7. Let Y_n, Y and Z be random variables and let X be a random vector in R^k. If $\sup_n|Y_n| \leq Z$; $E|Z|^p < \infty$ for some $p > 0$ and $Y_n \to Y$ in prob., then $E[|Y_n - Y|^p | X] \to 0$ in pr.

Proof: The theorem follows easily from theorem 2.2.1, Chebishev's inequality and theorem 3.2.1 (II). Q.E.D.

Theorem 3.2.8. (Dominated convergence theorem). Let the conditions of theorem 3.2.5 with $p = 1$ be satisfied. Then

$$E(Y_n|X) \to E(Y|X) \text{ in pr.}$$

Proof: By theorems 3.2.1 (IV and VI) and 3.2.7 it follows

$$|E(Y_n|X) - E(Y|X)| = |E[(Y_n - Y)|X]|$$

$$\leq E(|Y_n - Y||X) \to 0 \text{ in pr.} \qquad \text{Q.E.D.}$$

Theorem 3.2.9. (Fatou's lemma). Let Y_n be a random variable satisfying $Y_n \geq 0$ a.s. and let X be a random vector in R^k. Then

$$E(\liminf_{n\to\infty} Y_n \,|X) \leq \liminf_{n\to\infty} E(Y_n|X) \text{ a.s.}$$

Proof: Put $Y = \liminf_{n\to\infty} Y_n$ and let $\varphi(y)$ be any simple function satisfying $0 \leq \varphi(y) \leq y$ and put $Z_n = \min(\varphi(Y), Y_n)$. Then $Z_n \to \varphi(Y)$ in pr. and $E\,\varphi(Y) < \infty$ (see the proof of theorem 2.2.3). From theorem 3.2.7 it now follows:

$$E(Z_n|X) \to E[\varphi(Y)|X] \quad \text{in prob.}$$

Moreover, $Z_n \geq \varphi(Y)$, hence $E(Z_n|X) \geq E[\varphi(Y)|X]$. Thus for every $\varepsilon > 0$ we have:

$$P[E[\varphi(Y)|X] \leq \liminf_{n\to\infty} E(Z_n|X) \leq E[\varphi(Y)|X] + \varepsilon]$$

$$\leq P[E(\varphi(Y)|X) \leq E(Z_n|X) \leq E(\varphi(Y)|X)] \to 1,$$

hence

$$\liminf_{n\to\infty} E(Z_n|X) = E(\varphi(Y)|X) \quad \text{a.s.}$$

The rest of the proof is now similar to the proof of theorem 2.2.3. Q.E.D.

Theorem 3.2.10. (Monotone convergence theorem). Let (Y_n) be a non-decreasing sequence of random variables satisfying $E|Y_n| < \infty$ and let X be a random vector in R^k. Then

$$E(\lim_{n\to\infty} Y_n|\, X) = \lim_{n\to\infty} E(Y_n|X) \leq \infty \text{ a.s.}$$

Proof: Similarly to the proof of theorem 2.2.4, using theorem 3.2.9 instead of theorem 2.2.3. Q.E.D.

Exercises:
1. Prove theorem 3.2.1.
2. Complete the proof of theorem 3.2.2.
3. Prove theorem 3.2.7.

3.3 *Identification of conditional expectations*

In parametric regression analysis the conditional expectation $E(Y|X)$ is specified as a member of a parametric family of functions of X. In particular the family of linear functions is often used in empirical research. The question we now ask is how a given specification can be identified as the conditional expectation involved. In other words: given a dependent variable Y, a k-vector X of explanatory variables and a Borel measurable functional specification $f(X)$ of $E(Y|X)$, how can we distinguish between

$$P(E(Y|X) = f(X)) = 1 \qquad (3.3.1)$$

and

$$P(E(Y|X) = f(X)) < 1 \; ? \qquad (3.3.2)$$

An answer is given by theorem 3.1.1, i.e. (3.3.1) is true if

$$E(Y - f(X))\exp(i \cdot t'X) = 0$$

and (3.3.2) is true if

$$E(Y - f(X))\exp(i \cdot t'X) \neq 0 \quad \text{for some} \quad t \in R^k .$$

Verifying this, however, requires searching over the entire space R^k for such a points t. So where should we look ? For the case that X is bounded the answer is this:

Theorem 3.3.1. Let X be bounded. Then (3.3.2) is true if and only if

$$E(Y - f(X))\exp(i \cdot t_0 X) \neq 0$$

for some $t_0 \neq 0$ in an arbitrary small neighborhood of the origin of R^k.

Thus in this case we may confine our search to an arbitrary neighborhood of $t = 0$. If we do not find such a t_0 in this neighborhood then

$$E(Y - g(X))\exp(i\cdot :'X) = 0$$

and thus (3.3.1) is true.

Proof: Let (3.3.2) be true. According to theorem 3.3.1 there exists a $t_* \in R^k$ for which

$$E(Y - f(X))\exp(i.t_*'X) \neq 0 \tag{3.3.3}$$

Since X is bounded we can write (3.3.3) as

$$E(Y - f(X))\exp(i\cdot t_*'X) = E(Y - f(X))\Sigma_{j=0}^{\infty}(i^j/j!)(t_*'X)^j$$

$$= \Sigma_{j=0}^{\infty}(i^j/j!)E(Y - f(X))(t_*'X)^j \neq 0.$$

Consequently, there exists at least one j_* for which

$$E\ (Y - f(X))(t_*'X)^{j*} \neq 0.$$

Then

$$(d/d\lambda)^{j_*}E(Y - f(X))\exp[i\cdot\lambda t_*'X]$$

$$= \Sigma_{j=j_*}^{\infty}\{i^{j-j_*}\lambda^{j-j_*}/[(j-j_*)!]\}E(Y - f(X))^j(t_*'X)$$

$$\to E(Y - f(X))^{j_*}(t_*'X) \neq 0 \text{ as } \lambda \to 0.$$

This result implies that there exists an arbitrarily small λ_* such that

$$E(Y - f(X))\exp(i\lambda_* t_*'X) \neq 0.$$

Taking $t_0 = \lambda_* t_*$, the theorem follows. Q.E.D.

Now observe from the proof of theorem 3.3.1 that (3.3.2) is true if and only if for a point t_0 in an arbitrarily small neighborhood of the origin of R^k and some non-negative integer j_0,

$$E(Y - f(X))(t_0'X)^{j_0} \neq 0.$$

Applying a similar argument as in the proof of theorem 3.3.1

(with i replaced by 1) it is easy to verify:

Theorem 3.3.2. Let X be bounded. Then (3.3.2) is true if and only if the function $E(Y - f(X))\exp(t'X)$ is nonzero for a t in an arbitrarily small neighborhood of the origin of R^k.

Clearly this theorem is more convenient than theorem 3.3.1, as we no longer have to deal with complex valued functions.

Next, let $t_* \in R^k$ be arbitrary, let

$$Y_* = Y \cdot \exp(t_*'X)$$

and let

$$f_*(X) = f(X)\exp(t_*'X).$$

Then (3.3.2) is true if and only if

$$P[E(Y_*|X) = f_*(X)] < 1.$$

Applying theorem 3.3.2 we see that then

$$E(Y_* - f_*(X))\exp(t_0'X) = E(Y - f(X))\exp[(t_*+t_0)'X] \neq 0$$

for some t_0 in an arbitrary neigborhood of the origin of R^k. Consequently we have:

Theorem 3.3.3. Let X be bounded and let $t_* \in R^k$ be arbitrary. Then (3.3.2) is true if and only if $E(Y - f(X))\exp(t_0'X) \neq 0$ for a t_0 in an arbitrarily small neighborhood of t_*.

Thus actually we may pick an arbitrary neighborhood and check whether there exists a t_0 in this neighborhood for which

$$E(Y - f(X))\exp(t_0'X) \neq 0.$$

If so, then (3.3.2) is true, else (3.3.1) is true. This result now leads to our main theorem.

Theorem 3.3.4. Let $X \in R^k$ be bounded, and let S be the set of all $t \in R^k$ for which $E(Y - f(X))\exp(t'X) = 0$. For any probability measure μ on B^k corresponding to an absolutely continuous distribution we have: $\mu(S) = 1$ if (3.3.1) is true and $\mu(S) = 0$ if (3.3.2) is true.

Proof: Let $V = Y - f(X)$. Suppose for the moment that $X \in R$ (thus $k = 1$). Then theorem 3.3.3 implies that if

$$P[E(V|X) = 0] < 1$$

then for every $t_0 \in R$ there exists a $\delta > 0$ such that

$$E\ V \cdot \exp(tX) \neq 0 \text{ on } (-\delta, 0) \cup (0, \delta).$$

Consequently we have:

Lemma 3.3.1. Let $V \in R$ be a random variable satisfying $E|V| < \infty$ and let $X \in R$ be a bounded random variable. If $P[E(V|X) = 0] < 1$ then the set

$$S = \{t \in R : E\ V \cdot \exp(t \cdot X) = 0\}$$

is countable.

Using the lemma it is very easy to prove theorem 3.3.4 for the case $k = 1$. So let us turn to the case $k = 2$. Let $P[E(V|X) = 0] < 1$. According to theorem 3.3.3 there exists a $t_* \in R^2$ such that

$$E\ V \cdot \exp[t_*'X] \neq 0.$$

Denote

$$V_* = V \cdot \exp[t_*'X],$$

$$\psi_*(t_1, t_2) = E\ V \cdot \exp(t_1 X_1 + t_2 X_2)$$

where X_1 and X_2 are the components of X. Moreover, let

$$S_1 = \{t_1 \in R : \psi_*(t_1, 0) = 0\},$$

$$S_2(t_1) = \{t_2 \in R : \psi_*(t_1, t_2) = 0\} \quad .$$

Since $E\, V_* \neq 0$, we have $P[E(V_*|X_1) = 0] < 1$, hence by lemma 3.3.1 the set S_1 is countable. By the same argument it follows that the set $S_2(t_1)$ is countable if $t_1 \notin S_1$. Now let (t_1, t_2) be a random drawing from an absolutely continuous distribution. We have:

$$E\, I[\psi_*(t_1, t_2) = 0] = E\, I[\psi_*(t_1, t_2) = 0] \cdot I(t_1 \in S_1)$$

$$+ E\, I[\psi_*(t_1, t_2) = 0] \cdot I[t_1 \notin S_1]$$

$$\leq E\, I(t_1 \in S_1) + E\, I(t_1 \notin S_1) \cdot I(t_2 \in S_2(t_1)).$$

Since the set S_1 is countable and t_1 is continuously distributed we have $E\, I(t_1 \in S_1) = 0$. Moreover, since the distribution of t_2 conditional on t_1 is continuous we have:

$$E\, I[t_2 \in S_2(t_1)] \cdot I(t_1 \notin S_1) = 0 \ ,$$

for $S_2(t_1)$ is countable if $t_1 \notin S_1$. Thus:

$$P[\psi_*(t_1, t_2) = 0] = 0.$$

Replacing (t_1, t_2) by $(t_1 - t_1{*},\ t_2 - t_2{*})$, where $t_1{*}$ and $t_2{*}$ are the compoments of t_*, we see now that theorem 3.3.3 holds too for the case $k = 2$. The proof of the cases $k = 3, 4, \ldots$ is similar to the case $k = 2$ and therefore left to the reader.

Q.E.D.

Finally we consider the case that X is not bounded. By theorem 3.2.1 (VIII) we have $E(Y - f(X)|X) = E(Y - f(X)|\Gamma(X))$ a.s. for every Borel measurable one-to-one mapping Γ from R^k into R^k. From this result and theorem 3.3.3 it follows now:

Theorem 3.3.4. Let the conditions of theorem 3.3.3 be satisfied, except that X is bounded. Let Γ be an arbitrary bounded Borel measurable one-to-one mapping from R^k into R^k, and let

$$S = \{t \in R^k : E(Y - f(X))\exp[t'\Gamma(X)] = 0\}.$$

For any probability measure μ on B^k corresponding to an absolutely continuous distribution we have: $\mu(S) = 1$ if (3.3.1) is true and $\mu(S) = 0$ if (3.3.2) is true.

Exercises:

1. Use the argument in the proof of theorem 3.3.1 to prove the following corollary: Let $X = (X_1, \ldots, X_k)'$. Under the conditions of theorem 3.3.1 it follows that (3.3.2) is true if and only if there exists non-negative integers $m_1, \ldots, m_k$ such that

$$E[(Y-f(X))\Pi_{j=1}^k X_j^{m_j}] \neq 0.$$

Cf. Bierens (1982, theorem 2).

2. Let Θ be a subset of R^k. A point y in R^k is called a point of closure of Θ if for every $\varepsilon > 0$ we can find an $x \in \Theta$ such that $|x-y| < \varepsilon$. The set of all points of closure of Θ is called the closure of Θ. A subset S of Θ is called *dense* in Θ if the closure of S equals Θ. Prove that the set S in theorem 3.3.3 and 3.3.4 is *not* dense in R^k.

References:

Bierens,H.J. (1982), "Consistent Model Specification Tests", *Journal of Econometrics* 20, 105-134.

Chung,K.L. (1974), *A Course in Probability Theory*, New York: Academic Press

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

P 56973-4