A Complete Class of Tests when the Likelihood is Locally Asymptotically Quadratic

Werner Ploberger
Dept. of Economics
University of Rochester

April 28, 1999

Abstract

This paper constructs an asymptotically complete class of tests (i.e. for a test not within this class one can …nd a test within the class that has better or equal asymptotical powerfunction) for testing one parametric restriction when the likelihoodfunction can asymptotically be approximated by a quadratic function: The coe¢cients of this quadratic form may be random variables, so that our results also apply to the problem of unit-root testing.

1 Introduction

In econometric theory it is standard practice to construct tests based on by e.g. the likelihood-principle: Then one usually analyzes their powerfunctions in order to establish some optimality properties. In this paper we will, however, work the other way around: We will analyze the possible (asymptotic) powerfunctions of a testing problem and then - as a by-product - get the corresponding tests.

We will investigate testing problems where the corresponding likelihood is locally asymptotic quadratic - i.e. for sample size converging to in…nity the likelihoodfunction can - around the true value - be approximated by a quadratic function (i.e. a second-order Taylor expansion)

If the matrix de…ning the second-order-term (i.e. the matrix of second derivatives of the likelihoodfunction) (after proper normalization) converges after proper normalization to a constant (the Fisher-information) and the the …rst-order term is asymptotically normal the optimality theory of tests is well known.

In the analysis of econometric time series, however, often situations arise where the above assumptions are not ful…lled - like in the so called unit-root cases: Assume our data are generated by an autoregressive model

$$x_t = \mu x_{t-1} + u_t \quad (1)$$

where $t = 1, \ldots, T$ and the u_t are i.i.d $N(0, \sigma^2)$ (and - for simplicity - let us assume that σ^2 is known and equal 1). Suppose one wants to test whether $\theta = 1$: then with transforming the parameter by $\theta = 1 + h/T$ it is easily seen that the likelihoodfunction is quadratic in h and the coefficients of the linear and quadratic term converge in distribution. The first-order terms $\frac{1}{T}\sum_{t \leq T} u_t x_t$ are, however, asymptotically non-normal (they converge in distribution to the Ito-integral $\int_0^1 W dW$, wheter W is a standard Wienerprocess) and neither the second order terms $\frac{1}{T^2}\sum x_t^2$ converge to a constant (they converge in distribution to $\int_0^1 W(z)^2 dz$, where W is (as above) a standard Wienerprocess).

Based on the path-breaking work of Dickey and Fuller(1979),(1981), a lot of tests for models (1) were developed: For a survey cf. Phillips-Xaio(1998) and Stock(1995)(. Although some tests have been proved to be asymptotically optimal in certain ways cf. King(1982), Dufour and King(1991) and Elliot-Graham-Stock(1996), no systematic study of the class of admissible tests has been undertaken: This paper tries to characterize all asymptotically admissible tests - i.e. all tests for which there does not exist an asymptotically better one.

We will limit ourselves to the simplest case, namely when the parameter in question is one-dimensional. Then we show a class of tests to be (asymptotically) complete: A test outside this class has a powerfunction which is either dominated by one of tests within the class or the powerfunction of the (outside)test is equal to the powerfunction of a test from our class (i.e. given an "outside" test we can find a test within our class with equivalent or better powerfunction): So from the point of view of power we can limit ourselves to this class.

Our second result - which is a bit surprising - is that the traditional LR-test (or - in case of our model (1) the DF τ-test) is not within our class: So either this test is inadmissible (i.e. there exists a strictly better one) or we can find an equivalent one within our class.

2 Statistical experiments

This paper will use some methodology developed in mathematical statistics to characterize admissible tests (A detailed exposition of this theory can be found in the books of LeCam and Yang (1990) . LeCam(1986) and Strasser(1986)). In this section we will introduce the basic concepts of this theory in a formal way: I think that most readers will be familiar with the concepts introduced here in one way or another: Hence I will not bring any examples. In the field of econometrics, however, the use of some of these ideas (e.g. the Δ-metric for testing problems) is rather novel, so I think it is necessary to give precise formal definitions of the concepts involved.

Definition 1 An experiment E consists of a sample space Ω, a σ-algebra $\mathcal{F}$, a parametrized set of probability measures P_θ, $\theta \in \Theta$ on the sample space. Θ is called the parameterspace. We will denote the expectation with respect to P_θ by E_θ.

Our testing problem is a partition of the parameterspace in two nonempty, disjoint subsets Θ_0 (the hypothesis or null) and (the alternative) $\Theta_1 = \Theta - \Theta_0$. Then a test to the level α for Θ_0 against Θ_1 is a mapping from Ω in the interval $[0,1]$ so that

$$\sup_{\theta \in \Theta_0} E_\theta \varphi \leq \alpha \tag{2}$$

The powerfunction of the test φ is the function which maps each $\theta \in \Theta_1$ to $E_\theta \varphi$.

As mentioned in the introduction, our first concern is not the construction of tests but the analysis of their powerfunctions. We assume to have given some testing problem with hypothesis Θ_0 and alternative Θ_1 and we try to find tests that satisfy (2) and have "best possible" powerfunction. In general, there will be no test with uniformly maximal powerfunction. One test will be optimal for a certain parameter value, whereas another one may have the same property for another parameter value. It is, however, important to exclude tests which are uniformly (i.e.. for all parameter values of the alternative) dominated by another tests: these tests are called inadmissible. In "classical" situations it is possible to construct (asymptotically) admissible tests. In this paper, however, we do not want to construct specific tests. We want to get an overview over all possible tests. Therefore the following concept proved to be useful:

Definition 2 Suppose one has given an experiment $E = (\Omega, F, \{P_\theta : \theta \in \Theta\})$ and one wants to test Θ_0 against $\Theta_1 = \Theta - \Theta_0$ with level α. Then a set C of level-α tests is called a complete class if for every level-α-test φ there exist a $\psi \in C$ so that for all $\theta \in \Theta_1$

$$E_\theta \varphi \leq E_\theta \psi$$

So for an arbitrary test we can find a test from the complete class which either dominates the original test or has identical power as the original test. Therefore tests outside the complete class are at most equivalent to our test. Therefore - in general we do not have to bother with tests outside the complete class.

One should note that any superset of a complete class is a complete class, too (and the set of all tests is a -trivially - a complete class). The value of complete class theorems therefore depends of the size of the class!

Since we are comparing tests by their powerfunctions it seems reasonable not to concentrate on the tests themselves but their powerfunctions. This led to the following definition:

Definition 3 Let $\varepsilon \geq 0$ and let $E_1 = (\Omega_1, F_1, \{P_\theta : \theta \in \Theta\})$ and $E_2 = (\Omega_2, F_2, \{Q_\theta : \theta \in \Theta\})$ be two experiments with identical parameterspace Θ. Let $\Theta_0 \subset \Theta$ and $\Theta_1 = \Theta - \Theta_0$. Then E_2 is Θ_0 ε-divergent with respect to E_1 if for every test φ defined on Ω_1 there exists a test ψ on Ω_2 so that

- for all $\theta \in \Theta_0$

$$\int \psi dQ_\theta \leq \int \varphi dP_\theta + \varepsilon \qquad (3)$$

- for all $\theta \in \Theta_0$

$$\int \psi dQ_\theta \geq \int \varphi dP_\theta - \varepsilon \qquad (4)$$

This definition allows us to compare experiments with different sample spaces: Heuristically speaking, the property of ε-divergence guarantees that for the second experiment. This definition allows us to compare experiments with different sample spaces: Heuristically speaking, the property of ε-divergence guarantees that for every test φ for our testing problem with the first experiment there exists a test ψ with - give or take ε - the same or better characteristics: Inequality (3) guarantees that - if the null is true (the parameter θ is from θ_0) the test ψ wrongly rejects with an additional probability of ε. On the other hand, (4) guarantees that - if the alternative is true - the test only loses ε in power.

We can now define a (pseudo)metric $\Delta(E_1; E_2)$ between two experiments:

Definition 4 Let $E_1; E_2$ be two experiments with identical parameter space Θ and let $\Theta_0 \subset \Theta$ and $\Theta_1 = \Theta - \Theta_0$. Then let $\Delta(E_1; E_2)$ be the infimum of the set of all ε so that E_1 is ε-divergent with respect to E_2 and vice versa.

One can immediately see that $\Delta(.;.)$ has the usual properties of a pseudometric: It is nonnegative, fulfills the triangle inequality and is symmetric. The logical next step is to investigate the corresponding convergence: So suppose we have a sequence $E_n = \left(\Omega_n; F_n; \left\{P_\theta^{(n)} : \theta \in \Theta\right\}\right)$ which converges to $E = (\Omega; F; \{P_\theta : \theta \in \Theta\})$ (i.e. $\Delta(E_n; E) \to 0$). Let us assume that C is a complete class for a testing problem. Then we have the following.

Proposition 5 Let us assume φ_n are tests for our testing problem from our experiment E_n. Then there exist tests

$$\psi_n \in C \qquad (5)$$

so that for all $\varepsilon > 0$ for all but a finite number of n

$$\int \varphi_n dP_\theta^{(n)} \leq \int \psi_n dP_\theta + \varepsilon$$

The proof follows immediately from the definition of ε-divergence (which is the basis of the definition of our distance Δ) and the definition of a complete.

So suppose we have a converging sequence of experiments and a complete class for the limiting experiment: Then - according to our proposition we can - give or take an arbitrary $\varepsilon > 0$ - bound the powerfunctions on arbitrary tests by powerfunctions of the tests from the complete class.

So the powerfunctions of the tests from the complete class (plus an arbitrary $\varepsilon > 0$) are for large n upper bounds for the powerfunctions of tests for the experiments E_n.

Can we reach these bounds ? Consider a sequence $\psi_n \in C$. The definition of ε-divergence immediately shows the existence of tests -say φ^*_n - so that

$$\int \varphi^*_n dP^{(n)}_\theta \geq \int \psi_n dP_\theta - \varepsilon$$

for all but finitely many $n \in N$. Therefore we may conclude that for sufficiently large n we can find tests for the experiments E_n which have - up an arbitrary $\varepsilon > 0$ - the same powerfunctions as tests of the complete class for the limiting experiment!

So a complete class for the limiting problem allows us to - essentially - characterize the behaviour of tests for the experiments: For the "finite" experiments E_n we cannot find (up to an arbitrary $\varepsilon > 0$) better tests, but we can also reach these bounds: the big theoretical advantage of this theory is that one does not have to explicitly construct the approximating tests φ^*_n for proving these statements. In our case, however, the construction of the tests for E_n follows immediately from the definition of our complete class: therefore it is possible to use the tests directly.

In the next sections, we therefore will derive two results: First of all we will construct the limiting experiment for a family of experiments (which includes (1)), and then we will derive a complete class for this limiting experiment.

3 Families with LAQ likelihood and their asymptotics

Usually one assumes the parameter to be an open subset of a R^k and the likelihood to be a twice continuously differentiable function of the parameter. We will not directly follow this tradition: We will assume that the likelihood can be approximated (locally) by a quadratic function: obviously (if the likelihoodfunction is twice continuously differentiable) we can use Taylor series expansion to get this approximation: The technical details of a full discussion of this subject would lead us to much away from our goal, namely discussing the optimality of tests. Lateron, we will

Furthermore we assume that $k = 1$, i.e. we assume that our parametric family is onedimensional and that our null consists only of one element, i.e. $\Theta_0 = \{\theta_0\}$. To simplify the notation we assume (without limitation of generality) that $\theta_0 = 0$ and $\Theta = R$.

Then we assume that the following conditions hold true:

1. For each $n \in N$ and $\theta \in R$ we have given probability measures $P_{n,\theta}$ on sample spaces Ω_n: these probability measures are continuous to some measures μ_n with densities $f_{n,\theta}$ and likelihoods $\log f_{n,\theta}$.

2. These likelihoodfunctions are locally asymptotic quadratic: i.e. there exists a sequence $D_n \uparrow \infty$ of positive numbers so that for all $h \in \mathbb{R}$

$$\ell_{n;h} = \log f_{n;D_n h}$$

is "locally quadratic" in the following sense: There exist statistics W_n and A_n (i.e. random variables defined on the sample space independent of h) so that for all $M \in \mathbb{R}$

$$\sup_{|h| \le M} \left| \ell_{n;h} - (\ell_{n;0} + hW_n - \frac{h^2}{2} A_n) \right| \to 0 \tag{6}$$

where the convergence is stochastic in $P_{n;0}$.

1. W_n and A_n converge in distribution to some random variables W and A and

$$A > 0 \text{ almost surely}$$

2. We can "easily recognize" alternatives far away: For each $\varepsilon > 0$ there exists a $M = M(\varepsilon)$ and a sequence of tests $\rho_n = \rho_{n;\varepsilon}$ depending only on W_n and A_n so that

$$\limsup_{n \to \infty} E_{0;n} \rho_n \le \varepsilon$$

and for $h \ge M$ (where $E_{h;n}(.)$ should denote the expectation with respect to $P_{n;D_n h}$

$$\liminf_{n \to \infty} E_{h;n} \rho_n \ge 1 - \varepsilon.$$

3. For every $h \in \mathbb{R}$ the probability measures $P_{h;n}$ and $P_{0;n}$ are contiguous (i.e. for every sequence of events A_n so that $P_{0;n}(A_n) \to 0$ $P_{1;n}(A_n) \to 0$ and vice versa, for a precise discussion cf. Le Cam and Yang (19900). Although this concept is not that widely used in econometric theory, it is indeed one of the cornerstones of modern asymptotic theory, since on the one hand it allows to derive interesting results (cf e.g. Jennagathan(1981),(1985), LeCam and Yang (1990)), and on the other hand it is relatively easy to verify: A criterion for continuity is the convergence in distribution with respect to $P_{0;n}$ of

$$\frac{dP_{h;n}}{dP_{0;n}} \to f \tag{7}$$

with

$$Ef = 1 \tag{8}$$

and the analogous property when we interchange $P_{h;n}$ with $P_{0;n}$.

4. Let us consider the mappings $\phi_n : h \to P_{h;n}$. The we want these mappings to be uniformly continuous with respect to the total variation (on every compact subset of the parameter space). We assume that for every $M \in \mathbb{R}$ the following condition holds true: For all $\varepsilon > 0$ there exists a $\delta = \delta(\varepsilon; M) > 0$ so that for $h_1; h_2$ with $|h_1|; |h_2| \leq M$ and $|h_1 - h_2| < \delta$

$$\|P_{h_1;n} - P_{h_2;n}\| < \varepsilon.$$

where $\|.\|$ should denote the totalvariation of the argument (a signed measure)). Since $\|P_{h_1;n} - P_{h_2;n}\| = E_{0;n} |f_{n;D_n h_1} - f_{n;D_n h_2}|$ this assumption can easily be checked. It is easily seen that in most of the "reasonable" models this condition is indeed fulfilled..

All of the above conditions are easily seen to hold true in the "classical" case of e.g. experiments E_n consisting of the observation of n i.i.d random variables $X_1; \ldots; X_n$ each distributed with density g_θ, we can easily see that it is possible to choose $D_n = \sqrt{n}$ and

$$g^{(1)} = \frac{\partial \log g_\theta}{\partial \theta}|_{\theta=0}$$

$$g^{(2)} = \frac{\partial^2 \log g_\theta}{\partial \theta^2}|_{\theta=0}$$

$$W_n = \frac{1}{\sqrt{n}} \sum_{i \leq n} g^{(1)}(X_i)$$

and

$$A_n = -E\left(g^{(2)}(X_i)\right) = E\left(g^{(1)}(X_i)^2\right)$$

The property (6) can be shown to be fulfilled if the modulus of continuity of $g^{(2)}$ is not to ill-behaved: a more detailed discussion and examples generalizing the above one can be found in Hall&Heyde(). From the CLT we can see that that W_n are asymptotically normal. The A_n are immediately seen to be positive.

The contiguity of the probability measures is easily checked with the help of the criterion (7),(8). one can easyly show that $\frac{dP_{h;n}}{dP_{0;n}}$ converges in distribution to $\exp(hG(0; A) - \frac{h^2}{2}A)$ (where $G(0; A)$ is a random variable following a Gaussian distribution with expectation zero and variance A).

Another example is aur model (1). here the we have $D_n = n$ and

$$W_n = \frac{1}{n} \sum_{t \leq n} u_t x_t$$

and

$$A_n = \frac{1}{n^2} \sum_{t \leq n} x_t^2.$$

Again, it is well known (cf. Philips and Xiao(1998) for a list of refrences) that W_n and A_n converge in distribution: W_n converge in distribution to he Ito-integral $\int_0^1 W dW$, (where W is a standard Wienerprocess) and the A_n converge in distribution to $\int_0^1 W(z)^2 dz$, where W is the same standard Wienerprocess.

The fourth condition can be checked relatively easy, too: In the "stationary" case (where $D_n = \frac{1}{\sqrt{n}}$, A_n converges in general (uniformly in h) to a positive constant A and W_n converges in distribution to $G(0; A)$) it is easily seen (if the normalized second derivative of the likelihoodfunction remains equicontinuous!) tests rejecting if $n\hat{\mu}_n^2$ gets "too large".

For our basic model (1) we can choose a modified Sargan-Bhagarva test (cf.Sargan-Bhagarva(1983) : Under P_0 $\frac{1}{n^2}\sum_{t\le n} x_t^2$ converges in distribution. For negative h, the alternative P_h is a stable autoregressive process with parameter $1 - \frac{h}{n}$ (heuristically, the larger h the more we are on the "stable side"). Therefore it is easy to check that for all $b > 0$

$$\lim_{M\to\infty} \sup_{h<-M} P_h\left(\left[\frac{1}{n^2}\sum_{t\le n} x_t^2 > b\right]\right) = 0 \tag{9}$$

On the other hand, if h is positive, the alternative P_h is an explosive autoregressive process with parameter $1 + \frac{h}{n}$ (or, heuristically: the larger h the more the process is on the stable side: It is therefore easy to check that for all $B \in \mathbb{R}$

$$\lim_{M\to\infty} \sup_{h>M} P_h\left(\left[\frac{1}{n^2}\sum_{t\le n} x_t^2 < B\right]\right) = 0 \tag{10}$$

Due to the fact that $\frac{1}{n^2}\sum_{t\le n} x_t^2$ converges in distribution under P_0, we can find for every $\epsilon > 0$ some $0 < b(\epsilon) < B(\epsilon)$ so that

$$\liminf P_h\left(\left[b(\epsilon) < \frac{1}{n^2}\sum_{t\le n} x_t^2 < B(\epsilon)\right]\right) > 1 - \epsilon$$

Then we can define the test φ_ϵ to accept if $b(\epsilon) < \frac{1}{n^2}\sum_{t\le n} x_t^2 < B(\epsilon)$ and reject otherwise. Then (9) and (10) guarantee the existence of an appropriate M so that our fourth condition is fulfilled.

One can prove contiguity relatively easy by just using (7) and (8).

In this section we want to show that the sequence of experiments defined above converges: The limiting experiment is defined on $\mathbb{R}^2$ (i.e. it consists of one realisation of a 2-dimensional random vector): This should not be very surprising: (6) shows that two random variables (namely W_n and A_n) essentially determine the likelihood-function for all h.

We did assume that the vector $(W_n; A_n)$ converges in distribution with respect to $P_{0;n}$ to some random variable $(W; A)$. Therefore we may conclude from (6) that for any fixed $h \in \mathbb{R}$

$$\ell_{n;h} - \ell_{n;0} - (hW_n - \frac{h^2}{2}A_n) \to 0$$

and therefore - as $\frac{dP_{h;n}}{dP_{0;n}} = \exp(\ell_{n;h} - \ell_{n;0})$ we conclude that $\frac{dP_{h;n}}{dP_{0;n}}$ converges (with respect to $P_{0;n}$) in distribution to $\exp(hW - \frac{h^2}{2}A)$. With the help of Theorem 1 of LeCam and Yang (1990) we can conclude that

$$E(\exp(hW - \frac{h^2}{2}A)) = 1 \tag{11}$$

We can now define the experiment which we will show to be the limiting one: The sample space is the $\mathbb{R}^2$ (i.e. we take one sample of a 2-vector) and the probability measures Q_h, $h \in \mathbb{R}$ are defined as follows:

1. Q_0 is defined to be the distribution of the vector (W, A).
2. Q_h is defined by
$$\frac{dQ_h}{dQ_0} = \exp(hW - \frac{h^2}{2}A) \tag{12}$$
(i.e. we have for all measurable sets $M \subset \mathbb{R}^2$ $Q_h(M) = E(I_M \exp(hW - \frac{h^2}{2}A))$, where I_M should denote the indicator-function of the set M. Therefore (11) shows that the Q_h are probability measures!)
3. The mapping $h \to Q_h$ is continuous with respect to the total variation.

We now are able to formulate our theorem:

Theorem 6 Assume all of the assumptions above are fulfilled. Then the sequence of experiments E_n consisting of the probability measures $P_{n;h}$, sample spaces Ω_n and parameter space $\mathbb{R}$ converge (in our Δ-metric) to the experiment E defined by the probability measures Q_h, sample space $\mathbb{R}^2$ and parameter space $\mathbb{R}$. So we have

$$\Delta(E_n, E) \to 0. \tag{13}$$

Proof: For showing (13) we have to show that for every $\varepsilon > 0$ the following holds true: There exists a $N = N(\varepsilon)$ so that for $n > N(\varepsilon)$

1. for every test sequence of test φ_n for E_n there exist tests ψ_n for E so that
$$E_{0;n}\varphi_n \geq \int \psi_n dQ_0 + \varepsilon \tag{14}$$
and for $h \neq 0$
$$E_{h;n}\varphi_n \leq \int \psi_n dQ_h - \varepsilon \tag{15}$$
and

2. for every test ψ for E there exist tests φ_n for E_n so that

$$\int \psi \, dQ_0 \ge E_{0;n}\varphi_n + \varepsilon \tag{16}$$

and for $h \neq 0$

$$\int \psi \, dQ_h \le E_{h;n}\varphi_n - \varepsilon \tag{17}$$

Let us fix $\varepsilon > 0$ and let us begin with the first showing (14), (15): First of all let us construct the existence of a test $\rho : R^2 \to [0;1]$ and a $M = M(\varepsilon)$ so that

$$\int \rho \, dQ_0 \le \frac{\varepsilon}{3} \tag{18}$$

and for all h so that $|h| \ge M$

$$\int \rho \, dQ_h \ge 1 - \frac{\varepsilon}{3} \tag{19}$$

.Define ρ in the following way: We did assume that $Q_0[A = 0] = 0$; therefore there exist constants $a_1; a_2; w$, all strictly greater than zero so that with the set $R = \{(A; W) : a_1 < A < a_2; |W| < w\}$

$$Q_0(R) \ge 1 - .\frac{\varepsilon}{3}$$

Now define ρ to be 1 (i.e. rejecting) if $(A; W) \notin R$ and 0 otherwise. (18) immediately follows from the construction of R. It is easily seen that for $(A; W) \in R$

$$\frac{dQ_h}{dQ_0} \le \exp\left(hw - \frac{h^2}{2}a_1\right).$$

Hence we can find a $M = M(\varepsilon)$ so that for $(A; W) \in R$, $|h| \ge M$ $\frac{dQ_h}{dQ_0} \le \frac{\varepsilon}{3}$ and therefore $Q_h(R) \le \int_R \frac{dQ_h}{dQ_0} dQ_0 \le \frac{\varepsilon}{3} Q_0(R) \le \frac{\varepsilon}{3}$, which proves (19).

Assumption 5 implies the existence of certain gridpoints - say - $-M = h_1 < h_2 < \dots < h_{p-1} < h_p = M$ so that for each i and for all $h; h'$ with $h_i \le h; h' \le h_{i+1}$

$$\|P_{h;n} - P_{h';n}\| < \frac{\varepsilon}{3} \tag{20}$$

and

$$\|Q_h - Q_{h'}\| < \frac{\varepsilon}{3} \tag{21}$$

We did assume in our second assumption that $(W_n; A_n)$ converge in distribution to $(W; A)$. Then (6) shows that the vector of likelihoods $\left(\ell_{h_i;n} - \ell_{0;n}\right)_{1 \le i \le p}$ converges in distribution to $\left(h_i W - \frac{h_i^2}{2} A\right)_{1 \le i \le p} = \left(\log \frac{dQ_{h_i}}{dQ_0}\right)_{1 \le i \le p}$. Let us define the experiments

F_n, F as having the same sample space as $E_n; E$, but their parameterspace should consist only of the finite sets $\{0; h_1; \ldots; h_p\}$. Then we can apply the theorem of Strasser which states that $\Delta(F_n; F) \to 0$. Hence we can find for all n large enough tests ζ_n for F (i.e. functions from R^2 to $[0; 1]$) so that

$$\int \xi_n dQ_0 \leq E_{0;n}\varphi_n + \frac{\varepsilon}{3} \tag{22}$$

and for $1 \leq i \leq p$

$$\int \zeta_n dQ_{h_i} \geq E_{h_i;n}\varphi_n - \frac{\varepsilon}{3} \tag{23}$$

So let us now define the sequence ψ_n by

$$\psi_n = \max(\zeta_n; \rho_n)$$

Then on the null we have for n large enough $\int \psi_n dQ_0 \leq \int \zeta_n dQ_0 + \int \rho_n dQ_0 \leq E_{0;n}\varphi_n + \frac{\varepsilon}{3} + \frac{\varepsilon}{3}$, where the last inequality follows from (22) and (18). We now have to prove (15): Here we have to distinguish two cases: If $|h| > M$ then $\int \psi_n dQ_0 \geq \int \rho_n dQ_h \geq 1 - \frac{\varepsilon}{3}$ (cf. (19)), and as $1 - \frac{\varepsilon}{3} > E_{h;n}\varphi_n - \varepsilon$ we have shown (15) for this case.

We now have to prove (15) for the other case: Assume $|h| \leq M$: then there exist gridpoints $h_i; h_{i+1}$ from the grid constructed above so that $h_i \leq h \leq h_{i+1}$: Then we can conclude from (20),(21) that the totalvariations $\|P_{h;n} - P_{h_i;n}\|$ and $\|Q_h - Q_{h_i}\|$ are both smaller than $\frac{\varepsilon}{3}$: Since the value of the functions representing tests lies between 0 and 1 we therefore can conclude that

$$E_{h;n}\varphi_n \leq E_{h_i;n}\varphi_n + \frac{\varepsilon}{3}$$

$$\int \psi_n dQ_h \geq \int \psi_n dQ_{h_i} - \frac{\varepsilon}{3}.$$

Directly from the definition of ψ_n we can see that

$$\int \psi_n dQ_{h_i} - \frac{\varepsilon}{3} \geq \int \zeta_n dQ_{h_i} - \frac{\varepsilon}{3}.$$

Together with (23) these three inequalities prove (15) for this case, too.

Now it remains to show (16) and (17): We will not discuss this part of the proof since it is perfectly analogous to the ideas above - with the only exception that instead of the test ρ constructed above we have to use the tests ρ_n mentioned in our fourth assumption on the properties of $P_{h;n}$.

4 The Complete Class Theorem

In the preceding paragraph (theorem 6) we established the following result: for every sequence φ_n of tests for our original experiments E_n we can find a sequence ψ_n of tests for our experiment E so that for all $\varepsilon > 0$ for all n large enough the powerfunction of φ_n lies within $\pm\varepsilon$ of the powerfunction of ψ_n: A complete class of tests therefore gives us complete information about the possible powerfunction: On the one hand, we cannot get essentially better powerfunctions by tests which are not included in the class. On the other hand we know that for every test from our complete class we can find a test for E_n which has - up to $\pm\varepsilon$ - the same powerfunction as our original test: Lateron we will discuss how to find these "tests".

So we will construct a complete class of tests for our testing problem E: Its sample space is the $\{(A;W)\} = \mathbb{R}^2$ (i.e. our sample is a pair of real numbers: The first component we will call A and the second one W), the parametrized family $(Q_h)_{h\in\mathbb{R}}$ is defined by the following properties:

1. Q_0 is the distribution of $(A;W)$
2. Q_h is defined by $\frac{dQ_h}{dQ_0} = \exp(hW - \frac{h^2}{2}A)$

Then we will assume that all of the assumptions of the previous section are fulfilled..

Theorem 7 : The following class of tests is **complete** for our experiment E: Choose a level of significance α.

The class of tests are characterized by constants $a;b;c;U$ and two σ-finite measures μ and ν so that the following conditions are fulfilled:

1. ν is a σ-finite measure on $[-1;1] - \{0\}$ (i.e. no mass outside the unit interval with the number zero taken out) so that for all compact sets $K \subset [-1;1] - \{0\}$ $\nu(K)$ and
$$c \geq \int_{[-1;1]} \lambda^2 d\nu(\lambda) \tag{24}$$
(in particular $\int_{[-1;1]} \lambda^2 d\nu(\lambda)$ is finite and $c \geq 0$)
2. μ is defined on $\mathbb{R} - (-1;1)$ (i.e. the real line except the unit interval:) and for all $M > 1$
$$\mu([-M;-1] \cup [1;M]) < \infty$$
3. Not all the numbers $a;b;c;u$ and measures μ and ν are trivial
4. The test has the correct size - say $\alpha < 1$ - under the null.

Then define the function

$$\begin{aligned} u(A;W) = {} & bW + \frac{c}{2}\left(W^2 - A\right) + \\ & \int_{[-1;1]} \left(\exp(\lambda W - \frac{\lambda^2}{2}A) - \left(1 + \lambda W + \frac{\lambda^2}{2}(W^2 - A)\right)\right) d\nu(\lambda) + \\ & \int_{\mathbb{R}-(-1;1)} \exp(\lambda W - \frac{\lambda^2}{2}A) d\mu(\lambda) \end{aligned} \tag{25}$$

Now let the class of tests be characterized by the following properties:

1. The test should reject for all $(A; W)$ so that

$$A < a$$

2. If $A > a$ then the test should reject if

$$u(A; W) > U$$

and accept if

$$u(A; W) < U$$

Remark 1 The parameters describe the test not completely: We did not define what happens if $A = a$ or if $u(A; W) = U$: so if these equations do not describe nullsets (with respect to all Q_h), our complete class defined above may contain some additional (nonoptimal) tests: This is not a contradiction to the definition of a complete class: Any superset of a complete class is a complete class, too! It may be the case that the requirement 4 above will not (unlike e.g. in the analogous situation with the Neyman-Pearson test) determine the value of the test for these exceptional parameters!

Remark 2 This ambiguity highlights the difference of our situation to the "classical" case where A is almost surely constant - say A_0: Then any test $\varphi = \varphi(W)$ depending on W alone falls within our class: It is almost surely equal to a test $\psi(A; W)$ within ou class: reject when $A < A_0$, accept when $A > A_0$ (since both events occur with probability zero) and if $A = A_0$ decide according to φ; so the theorem is not very helpful in cases where A is deterministic!

Remark 3 It may be the case that the parameter a is not uniquely determined: Suppose that our measure μ is such that $\int \exp\left(-\frac{\lambda^2}{2} K\right) d\mu(\lambda) = \infty$ if $K < M$: Then the value of the teststatistic for $A < M$ is ∞ and therefore we reject for all $(A; W)$ with $A < M$: Hence any choice of a between 0 and M does not influence the test.

The proof of the theorem is relatively complicated: Hence we present it in Appendix A.

The theorem proves that we can dominate any test for the limiting problem with a test from our class: According to the discussion of the previous section, we know that we can for $n \to \infty$ either dominate or approximate any test for our experiments E_n up to $\pm\varepsilon$ with a test that has (up to $\pm\varepsilon$) the same power functions as the tests constructed above. We can, however, can in most cases construct tests for E_n having these characteristics: simply consider tests rejecting when $u(A_n; W_n) > U$, where $u; U$ are defined above: given our assumptions it is an elementary exercise to show that the powerfunctions of these tests converge to the powerfunction of the test from our complete class if the distribution function of $u(A; W)$ is continuous in U, which will be the fact in almost all cases of practical interest.

5 The Likelihood-Ratio Test

As mentioned already, a superset of a complete class is a complete class again: Therefore our theorem is "stronger" the "smaller" the complete class: It is therefore interesting what tests are not within the class: We will analyze the case where Θ is onedimensional: Then our sample space $(A; W)$ is essentially the R^2 and our teststatistics u defined in (25) are essentially functions from the R^2 into R: In the classical case (A_n converging to a constant) the distribution of A is degenerate: therefore we only have to evaluate u for one certain parameter. We, however, do not want to discuss these type of cases: We therefore assume that

$$\text{The support of } Q_0 \text{ is } R^+ \times R \tag{26}$$

One can easily see that the limit of our basic model (1) satisfies this requirement.

An easy calculation shows that the function u satisfy the partial differential equation

$$\frac{\partial u}{\partial A} + \frac{1}{2}\frac{\partial^2 u}{\partial W^2} = 0 \tag{27}$$

the so-called "reverse heat equation": so our result in theorem 7 imposes considerable restrictions on possible teststatistics: Therefore we have an interesting necessary criterion for tests being in our class:

Theorem 8 Suppose we have given a set $C \subset R^+ \times R$ with boundary ∂C: Then a necessary conditon for C to be the complement of a critical region for a test from the class described in the previous section is that there exists a solution u of equation (27) so that

$$u(x) = 1 \text{ for } x \in \partial C$$

and

$$u(x) < 1 \text{ for } x \in C$$

The proof is trivial: What makes the theorem interesting is the fact that the constant functions are always solutions of (27): Therefore the region C only can result from one of our tests if the corresponding boundary-value problem for (27) is not unique: since (27) plays a prominent role in mathematical physics this problem was - and is - analyzed extensively by mathematicians and theoretical physicists.

We will, however, analyze another problem,too: We show that - in general - the LR-test (-which - in the case of the model (1) is the Dickey-Fuller τ-test) is not within our class: This, however, does not imply that the test is not admissible: It might be possible that there exists a test within our class with identical powerfunction. To show the existence - or nonexistence - of such a test is beyond the scope of this paper: One either has to "heal" the LR-test (i.e. show its admissibility by showing that it is equivalent to one of our test) or "kill" it (i.e. find a test in our class which is uniformly better than the LR-test).

In any case, however, it is impossible that the LR-test "beats" all of the tests from the class: The first option (namely asymptotic equivalence) is the "best" possible outcome for the LR-test.

As usual in the theory of maximum-likelihood approximations let us assume that the approximation (6) is accurate enough to yield an approximation to the maximum of the likelihood, i.e. we assume that

$$\max_h(\ell_{n;h} - \ell_{n;0}) - \frac{W_n^2}{2A_n} \to 0.$$

We will not discuss this condition: It can easily be verified for practically all models of practical interest: Therefore the LR-tests essentially reject when $\frac{W_n^2}{2A_n}$ is "large enough": Since we did assume that $(A_n; W_n)$ converges in distribution to $(A; W)$ we can easily see that the powerfunctions of the LR-tests for E_n converge to the powerfunction of the following test (for experiment E): reject when $\frac{W^2}{2A} > z$, where z is the appropriate critical level (to be precise: the convergence only holds for all z beeing continuity points for the distribution function of $\frac{W^2}{2A}$). Hence it is interesting to decide whether this test is in our class: The answer is, in general negative:

Theorem 9 : Assume condition (26) holds true. Then the test rejecting when $\frac{W^2}{2A} > z$, accepting when $\frac{W^2}{2A} < z$ is - for $z > 1/2$ not an element of our complete class.

Proof: Assune the theorem is not true: Then - as we maintained the assumption of (26) being true -we can conclude that $a = 0$ and there exists a fuction $u(.;.)$ defined by (25) so that

$$u\left(\frac{W^2}{2z}; W\right) = \text{const} \tag{28}$$

for all W (here we need our assumption (26): if it would not hold then we could not postulate (28) to be true for all values of W!): We will now analyze $u(\frac{W^e}{2z}; W)$ for $W \to 0$. In particular we want to analyze the behaviour of

$$I(\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp\left(-\frac{W^2}{2\sigma^2}\right) u\left(\frac{W^2}{2z}; W\right) dW \tag{29}$$

for $\sigma^2 \to 0$. A consequence of (28) would imply that

$$I(\sigma^2) = \text{const} = u\left(\frac{W^2}{2z}; W\right). \tag{30}$$

We are now going and show that (25) implies and (30) imply $c = 0$ (and consequentlly $\nu = 0$ (because of (24)) as well as $\mu = 0$. Consequently our test either would - as we assumed it to be nontrivial - be determined by $u = bW$. But then $[u > U]$ is a half-plane and not the complement of a parabola as required by the assumptions of this theorem!

So let us define

$$K = \frac{1}{2z}.$$

Then (25) allows us to write our function $u(.,.)$ as

$$\begin{aligned} u(KW^2, W) &= bW + \frac{c}{2}W^2(1 - K) + \\ &\int_{[-1,1]} \left(\exp(\lambda W - \frac{(\lambda W)^2}{2}K) - \left(1 + \lambda W + \frac{(\lambda W)^2}{2}(1 - K)\right)\right) d\nu(\lambda) \\ &\int_{R-(-1,1)} \exp(\lambda W - \frac{(\lambda W)^2}{2}K) d\mu(\lambda) \end{aligned} \quad (31)$$

We will now evaluate $I(\sigma^2)$ by evaluating the integral for each of the terms on the right hand side. It is easily seen that

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int \exp(-\frac{W^2}{2\sigma^2}) W dW = 0 \quad (32)$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int \exp(-\frac{W^2}{2\sigma^2}) W^2 dW = \sigma^2 \quad (33)$$

For the evaluation of the third term let us define for real $u \neq 0$

$$f(u) = \frac{\exp(u - Ku^2/2) - (1 + u + (1 - K)u^2/2)}{u^2}$$

and $f(0) = 0$. Then it is easily seen that f is a continuous function and therefore the function $\theta(W) = \sup_{|u| \leq |W|} |f(u)|$ converges to zero for $W \to 0$:

$$\lim_{W \to 0} \theta(W) = 0 \quad (34)$$

Furthermore $\lim_{u \to \pm\infty} f(u) \exp(-|u|) = 0$. Hence we can find a M so that for all $u \in R$

$$|f(u)| \leq M \exp(|u|) \quad (35)$$

The third term on the righthand-side of (31) equals

$$W^2 \int_{[-1,1]} f(W\lambda) \left[\lambda^2 d\nu(\lambda)\right].$$

Let $\epsilon > 0$ be arbitrary. Then (34) implies that there exists a $\delta > 0$ so that $\theta(\delta) \leq \epsilon$. Then we can conclude that with

$$K(\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp(-\frac{W^2}{2\sigma^2}) W^2 \int_{[-1,1]} f(W\lambda) \left[\lambda^2 d\nu(\lambda)\right] dW$$

$$\left|K(\sigma^2)\right| \leq K_1(\sigma^2) + K_2(\sigma^2);$$

where

$$K_1(\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{|W|<\delta} \exp(-\frac{W^2}{2\sigma^2}) W^2 \left\{ \int_{[-1,1]} \varepsilon \left(\lambda^2 d\nu(\lambda) \right) \right\} dW$$

and (using (35))

$$K_2(\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{|W|\geq\delta} \exp(-\frac{W^2}{2\sigma^2}) W^2 \left\{ \int_{[-1,1]} M.\exp(W) \left(\lambda^2 d\nu(\lambda) \right) \right\} dW$$

It is now easily seen that

$$\lim_{\sigma^2 \to 0} \frac{K_2(\sigma^2)}{\sigma^2} = 0$$

and

$$\limsup_{\sigma^2 \to 0} \frac{K_1(\sigma^2)}{\sigma^2} \leq \varepsilon \int_{[-1,1]} \lambda^2 d\nu(\lambda)$$

so that we can conclude that for σ^2 small enough (i.e. for every ε there exists a $\sigma_0^2(\varepsilon)$ so that for all $\sigma^2 < \sigma_0^2(\varepsilon)$)

$$\left| K\left(\sigma^2\right) \right| \leq 2\varepsilon \int_{[-1,1]} \lambda^2 d\nu(\lambda) \tag{36}$$

Now we have to evaluate the fourth term on the right hand side of (31): First of all we can easily see from (31) that

$$\lim_{W \to 0} u(W; KW^2) = \int_{R-(-1,1)} d\mu(\lambda) \tag{37}$$

and therefore

$$u(W; KW^2) = \int_{R-(-1,1)} d\mu(\lambda) \tag{38a}$$

since $u(W; KW^2)$ was assumed to be constant. Hence we may conclude that μ is a finite measure:

$$\mu(R - (-1,1)) < \infty$$

Let us now define

$$J(\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int \left\{ \int_{R-(-1,1)} \exp(\lambda W - K\frac{(\lambda W)^2}{2}) d\mu(\lambda) \right\} \exp(-\frac{W^2}{2\sigma^2}) dW.$$

Then we have

$$J(\sigma^2) = \int_{R-(-1,1)} \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp(\lambda W - (K\lambda^2 + \frac{1}{\sigma^2})\frac{W^2}{2}) dW \right\} d\mu(\lambda) =$$

$$= \int_{R-(-1,1)} \frac{1}{\sqrt{\sigma^2}} \left(\sqrt{(K\lambda^2 + \frac{1}{\sigma^2})}\right)^{-1} \exp\left(\frac{\lambda^2}{2(K\lambda^2 + \frac{1}{\sigma^2})}\right) d\nu(\lambda) =$$

$$= \int_{R-(-1,1)} \frac{1}{\sqrt{(K\lambda^2\sigma^2+1)}} \exp\left(\frac{\sigma^2\lambda^2}{K\lambda^2\sigma^2+1}\frac{1}{2}\right) d\nu(\lambda)$$

and therefore

$$J(\sigma^2) - \nu(R-(-1,1)) = \int_{R-(-1,1)} \left\{ \frac{1}{\sqrt{(K\lambda^2\sigma^2+1)}} \exp\left(\frac{\sigma^2\lambda^2}{K\lambda^2\sigma^2+1}\frac{1}{2}\right) - 1 \right\} d\nu(\lambda)$$

Let now C be an arbitrary real number. Then let us define the sets $D_1(C)$ and $D_2(C)$ as $D_1(C) = [-M, -1] \cup [1, M]$ and $D_2(C) = (R-(-1,1)) - D_1(C)$. Define $J_1(\sigma^2)$, $J_2(\sigma^2)$ by

$$J_1(\sigma^2) = \int_{D_1(C)} \left\{ \frac{1}{\sqrt{(K\lambda^2\sigma^2+1)}} \exp\left(\frac{\sigma^2\lambda^2}{K\lambda^2\sigma^2+1}\frac{1}{2}\right) - 1 \right\} d\nu(\lambda)$$

$$J_2(\sigma^2) = \int_{D_2(C)} \left\{ \frac{1}{\sqrt{(K\lambda^2\sigma^2+1)}} \exp\left(\frac{\sigma^2\lambda^2}{K\lambda^2\sigma^2+1}\frac{1}{2}\right) - 1 \right\} d\nu(\lambda).$$

Then define

$$J(\sigma^2) - \nu(R-(-1,1)) = J_1(\sigma^2) + J_2(\sigma^2).$$

It now can easily be seen that

$$\lim_{\sigma^2 \to 0} \frac{J_1(\sigma^2)}{\sigma^2} = \frac{1-K}{2} \int_{D_1(C)} \lambda^2 d\nu(\lambda). \qquad (39)$$

Elementary calculus shows that the function $h(z) = \frac{1}{\sqrt{Kz+1}} \exp(\frac{z}{Kz+1}\frac{1}{2}) - 1$ is - for $K > 1$ -monotonously decreasing for nonegative z. Hence we have

$$J_2(\sigma^2) \le \left\{ \frac{1}{\sqrt{(KC^2\sigma^2+1)}} \exp\left(\frac{\sigma^2 C^2}{KC^2\sigma^2+1}\frac{1}{2}\right) - 1 \right\} \nu(D_2(C))$$

and therefore $\limsup_{\sigma^2 \to 0} \frac{J_2(\sigma^2)}{\sigma^2} \le \frac{1-K}{2} C^2 \nu(D_2(C))$
and therefore

$$\limsup_{\sigma^2 \to 0} \frac{J(\sigma^2) - \nu(R-(-1,1))}{\sigma^2} \le \frac{1-K}{2} \int_{R-(-1,1)} \min(\lambda^2, C^2) d\nu(\lambda) \qquad (40)$$

Now observe that (31) allows us to express

$$\frac{I(\sigma^2) - \nu(R-(-1,1))}{\sigma^2}$$

as a sum of the terms on the left hand side of (32), (33), (36), (40): Hence we may conclude that on the one hand

$$\limsup_{\sigma^2 \to 0} \frac{J(\sigma^2) - \mu(R - (-1;1))}{\sigma^2} \leq \frac{1-K}{2} \left\{ \left(\int_{R-(-1;1)} \min(\lambda^2; C^2) d\mu(\lambda) \right) + \frac{c}{2} - 2\varepsilon \int_{[-1;1]} \lambda^2 d\nu(\lambda) \right\}$$

If either μ or c are nontrivial (i.e. positive) we can choose C and ε in such a way that the righthand side of the above inequality is strictly negative, so that

$$\limsup_{\sigma^2 \to 0} \frac{J(\sigma^2) - \mu(R - (-1;1))}{\sigma^2} < 0 \tag{41}$$

On the other hand, we know from (37) and (30) that $\frac{J(\sigma^2) - \mu(R-(-1;1))}{\sigma^2} = 0$ for all σ^2, which contradicts (41): Hence our assumption - that either μ or c are nontrivial - must be wrong, which was exactly what we wanted to show.

6 References

1. Bauer, H. (1972): Probability Theory and Elements of Measure Theory, Holt, Rinehart and Winston, New York

2. Dickey, D.A. and W.A. Fuller(1979):Distribution of estimators for autoregressive time series with a unit root, JASA, 74, 427-431

3. Dickey, D.A. and W.A. Fuller(1981): Likelihood ratio tests for autoregressive series with a unit root, Econometrica 49, 1057-1072

4. Dufour, J. and M.L. King (1991): Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary and nonstationary AR(1) errors

5. Elliot, G., T.J. Rothenberg and J.H. Stock(1996): efficient tests for and autoregresive unit root, Econometrica 64, 813-836.

6. Hall, P. and C.C. Heyde(1980): Martingale Limit theory and its Applications, Academic Press, New York

7. Jeganathan, P.(1991): On the asymptotic behavior of least squares estimators in AR time series with roots near the unit circle, Econometric Theory, 7, 269-306

8. Jeganathan, P.(1995): Some aspects of asymptotic theory with applications to time series models" Econometric Theory 11, 818-867

9. King, M.L. (1988): Towards a theory of point-optimal testing, Econometric Reviews, 6, 169-218

10. LeCam, L.(1986): Asymptotic Methods in Statistical Decision theory, Springer, New York

11. LeCam, L. and G.L. Yang(1990): Asymptotics in Statistics: some basic concepts, Springer, New York

12. Phillips, P.C.B and Z. Xiao(1998): A Primer on Unit Root Testing, Cowles Foundation Discussion Paper, Yale University, New Haven.

13. Sargan, J.D. and A. Bhargava(1983): Testing residuals from least squares regression for being generated by a Gaussian random walk, Econometrica 51, 153-174

14. Stock,J.H.(1995): Unit roots, structural breaks and trends: In: R.F. Engle and D. McFadden (eds), Handbook of Econometrics, Vol. 4, 2739-2841: North-holland, Amsterdam

15. Strasser, H. (1985): Mathematical Theory of Statistics, de Gruyter, Amsterdam

Appendix A: Proof of theorem 7

First of all let us introduce the following notation (here we follow Strasser(1985)): Let us fix $\alpha > 0$ and assume that P and Q are probability measures defined on the same space. Additionally let P and Q be dominated by some measure μ. Then let

$$NP(P;Q)$$

be defined as the Neyman-Pearson test for P against Q: Then there exists at least one $k_\alpha \in \mathbb{R}$ and one p with $0 \leq p \leq 1$ so that the test

$$\varphi = \begin{cases} 1 & \text{if } \frac{dQ/d\mu}{dP/d\mu} > k_\alpha \\ p & \text{if } \frac{dQ/d\mu}{dP/d\mu} = k_\alpha \\ 0 & \text{if } \frac{dQ/d\mu}{dP/d\mu} < k_\alpha \end{cases}$$

We will use the version from Strasser(1985) going back to a result of Wald(1949): This theorem states (if we take into account that our null is simple) that the hull (in the topological sense) (with respect to the so-called weak topology of the M-space, to be explained below) of the set

$$\left\{ NP(Q_0; \sum_{i=1} p_i Q_{h_i}) : p_i \geq 0; \sum p_i = 1; h_i \neq 0 \right\} \quad (42)$$

is a complete class: So we for proving our theorem we have to show that the class of our tests

1. contains all the tests $NP(Q_0; \sum_{i=1}^k p_i Q_{h_i})$ (i.e. the Neyman-Pearson tests of the null against Bayesian mixtures of alternatives and

2. is closed with respect to the weak topology of the M-space.

The first assertion is trivial: simply choose μ to be the measure having mass p_i in h_i if $|h_i| \geq 1$, ν to have mass p_i in h_i if $|h_i| < 1$, $a = 0$, and

$$b = \sum_{\{i:|h_i|<1\}} h_i p_i$$

as well as

$$c = \sum_{\{i:|h_i|<1\}} h_i^2 p_i.$$

Then choose U so that the size of the test under the null equals α.

The proof of the second assertion is a bit more complicated, and part of it may be due to the unfamiliar notation: First of all I would like to give some background information on the concepts of "M-space" and "L-space": We will not use many properties of these mathematical objects, so I will not prove any of these properties. A more detailed discussion can be found in Strasser(1985), LeCam(1986), LeCam and Yang(1990) . We have given an experiment defined by some probability measures Q_h, $h \in \mathbb{R}$. (12) shows that these measures all are equivalent to Q_0 (i.e. $Q_h \ll Q_0$ and $Q_0 \ll Q_h$). Then the L-space of the experiment is (in our case) defined as the space $L_1(\mathbb{R}^2, Q_0)$ of all integrable functions from $\mathbb{R}^2$ to $\mathbb{R}$, with norm $\|f\| = \int |f| dQ_0$. Then the M-space of the experiment is defined as the dual of the L-space. One can show that this space is canonically isomorphic to the space $L_\infty(\mathbb{R}^2, Q_0)$ of all essentially bounded functions. It can easily be shown that even the the norms (the usual L_∞-norm and the norm defined by the duality) coincide.

Tests were defined as functions from the sample space into $[0, 1]$. Hence they are elements of the unit sphere of the M-space. It can easily be seen that the L-space is separable. Hence the weak topology on the unit sphere of the M-space is metrizable. As a consequence, we have the following way to prove that a set B of tests is closed: We only have to show that for every sequence $\varphi_n \in B$ which converges to some $\varphi \in L_\infty(\mathbb{R}^2, Q_0)$ $\varphi \in B$.

We can describe this isorphism as follows: Suppose $g \in L_\infty(\mathbb{R}^2, Q_0)$, then the element of the M-space (which is a functional on the L-space $L_1(\mathbb{R}^2, Q_0)$) attaches to each $f \in L_1(\mathbb{R}^2, Q_0)$ the value $\int g f \, dQ_0$. So if we have $\varphi_n \to \varphi$ in the weak topology, then for every measurable set $A \subset \mathbb{R}^2$ we have

$$\int_A \varphi_n dQ_0 \to \int_A \varphi \, dQ_0 \tag{43}$$

So let us assume we have given a sequence φ_n of our tests converging to some φ in the weak topology. So let us assume that there exist constants a_n, b_n, c_n, U_n and measures μ_n and ν_n so that φ_n rejects when $A < a_n$ or when $u_n(A;W) > U_n$, (and accepts if $A > a_n$ and $u_n(A;W) < U_n$), where u_n is defined according to (25). We tacitly assume that the constants and the measures fulfill all the prerequisites of our theorem. Then we have to show that there exist $a;b;c;U$ and measures μ and ν satisfying all the assumptions of our theorem so that the limiting test φ equals the test corresponding to these parameters.

First of all we can assume that for each n $b_n;c_n;U_n$ are not all equal zero: Would this be the case then according to (24) ν would be the trivial measure: Consequently (according to assumption 3) μ would be nontrivial: Hence $u_n(.;.)$ would be strictly positive - and since $U_n = 0$ - the test would reject for all values of $(A;W)$ - which contradicts our assumption 4.

It can easily be seen that the test corresponding to $a_n;b_n;c_n;U_n$ and measures μ_n and ν_n is the same as the test corresponding to $a_n;C_nb_n;C_nc_n;C_nU_n$ and measures $C_n\mu_n$ and $C_n\nu_n$ for any sequence of positve numbers C_n: Hence we can assume without limitation of generality that

$$|b_n| + |c_n| + |U_n| = 1: \tag{44}$$

We first of all want to prove the following proposition:

Proposition 10 Suppose our parameters satisfy (44). Then there exists an infinite subsequence $I \subseteq N$ so that for $n \in I$

1. $a_n;b_n;c_n;U_n$ converge to $a';b;c;U$, where these parameters fulfill (44) and
2. the measures ρ_n defined by
$$\frac{d\rho_n}{d\nu_n}(\lambda) = \lambda^2$$
(and $\rho_n(\{0\}) = 0$) converge vaguely (cf. appendix B!) to some measure ρ (and this implies (theorem 15) that for all bounded, continuous f
$$\int_{-\infty}^{\infty} f\, d\rho_n \to \int_{-\infty}^{\infty} f\, d\rho \tag{45}$$
and
$$\rho([-1;1]) \leq b \tag{46}$$

Let us now prove the proposition: First of all let us show that the sequence a_n is bounded: Let us assume the contrary would be true, namely that there exists an infinite subsequence J so that for $n \in J$ $a_n \to \infty$: Then $\liminf_{n\in J} \int \varphi_n dQ_0 \geq \liminf_{n\in J} Q_0([A < a_n]) = 1$, which contradicts (4).

As a_n remains bounded we can find a subsequence - say I_1- so that

$$a' = \lim_{n\in I_1} a_n \tag{47}$$

exists. (44) guarantees that $b_n; c_n; U_n$ remain bounded, too. Therefore we can find an infinite subsequence $I_2 \subseteq I_1$ so that

$$\begin{aligned} b &= \lim_{n \in I_2} b_n \\ c &= \lim_{n \in I_2} c_n \\ U &= \lim_{n \in I_2} u_n \end{aligned} \tag{48}$$

As the modulus is a continuous function of its argument we can immediately see from (44) that

$$|b| + |c| + |U| = 1$$

which proves the first assertion of our theorem. For the proof of the second part of proposition 10 observe that (24) can also be read as

$$\rho_n([-1;1]) \leq b_n$$

Since b_n, $n \in I_2$ remains bounded, the total mass of the ρ_n remains bounded by b_n. We therefore can apply theorems 13 and 14 and conclude that there exists an infinite sub-sequence $I \subseteq I_2$ so that there exists a measure ρ defined on $[-1;1]$ so that ρ_n converge for $n \in I$ to ρ vaguely. Then (45), (46) are consequences of theorem14 and corollary 16. Since $\rho_n(\mathbb{R} - [-1;1]) = 0$ and $\rho(\mathbb{R} - [-1;1]) = 0$, too we have $\rho_n(\mathbb{R}) = \int h d\rho_n$; $\rho(\mathbb{R}) = \int h d\rho$ with $h(x)$ defined to be 1 if $x \in [-1;1]$, 0 if $x \in (-\infty;-2] \cup [2;\infty)$ and linear (and continuous) on each of the intervals $[-2;-1]$ and $[1;2]$. Vague convergence implies $\int h d\rho_n \to \int h d\rho$, so we have $\rho_n(\mathbb{R}) \to \rho(\mathbb{R})$, too. Hence we can apply corollary 16. Since $I \subseteq I_2 \subseteq I_1$ the limiting relations (47),(48) remain valid, too, which finishes our proof of proposition 10

Proposition 11 Let I be the infinite subsequence defined in proposition 10. Then there exists an infinite sub-subsequence $J \subseteq I$, a constant a'' and a measure μ on $\mathbb{R} - (-1;1)$ so that the following holds true: for an arbitrary W

1. for all $A < a''$

$$\lim_{n \in J} \int_{\mathbb{R} - (-1;1)} \exp(\lambda W - \frac{\lambda^2 A}{2}) d\mu_n \to \infty \tag{49}$$

2. for all $A > a''$

$$\lim_{n \in J} \int_{\mathbb{R} - (-1;1)} \exp(\lambda W - \frac{\lambda^2 A}{2}) d\mu_n = \int_{\mathbb{R} - (-1;1)} \exp(\lambda W - \frac{\lambda^2 A}{2}) d\mu \tag{50}$$

Let us now prove proposition 2: Let us define

$$v_n(A) = \int_{\mathbb{R}-(-1;1)} \exp\left(-\frac{\lambda^2 A}{2}\right) d\lambda$$

$$w_n(A;W) = \int_{\mathbb{R}-(-1;1)} \exp\left(\lambda W - \frac{\lambda^2 A}{2}\right) d\mu_n$$

and

$$S = \left\{ A > 0 : \limsup_{n \in I} v_n(A) < \infty \right\}$$

Observe that for all real $M > 0$ we have

$$-\frac{M}{2} - \frac{\lambda^2 W^2}{2M^2} \leq \lambda W \leq \frac{M}{2} + \frac{\lambda^2 W^2}{2M^2} \tag{51}$$

and therefore

$$\exp\left(-\frac{M}{2}\right) v_n\left(A - \frac{W^2}{M^2}\right) \leq w_n(A;W) \leq \exp\left(\frac{M}{2}\right) v_n\left(A + \frac{W^2}{M^2}\right) \tag{52}$$

Let us first show that $S \neq \emptyset$: Assume the contrary to be true, namely that $S = \emptyset$: Then for all A $\limsup_{n \in I} v_n(A) = \infty$ and due to (52) $\limsup_{n \in I} w_n(A;W) = \infty$.(44) and the construction of the sequence I guarantee that the other parameters defining our tests remain bounded: Therefore for all $A > 0; W$ $\limsup_{n \in I} u_n(A;W) = \infty$: Since our parameters U_n remain bounded, too, for arbitrary $A > 0; W$ for infinitely many $n \in I$ $u_n(A;W) > U_n$: Hence $\lim_{n \in I} \varphi_n = 1$ Q_0-almost surely and consequently (due to fact that $0 \leq \varphi_n \leq 1$ we may apply Fatou's Lemma to $1 - \varphi_n$) $\limsup_{n \in I} \int \varphi_n dQ_0 = 1$. This means, however, that the size of the tests converges to 1, which contradicts the assumption 4 of the theorem.

We now have shown that $S \neq \emptyset$: Let $A < B$: Then $v_n(A) > v_n(B)$: Therefore we can conclude that if $A \in S$ $B \in S$, too: So define

$$a'' = \inf S$$

So let us first prove (49): It is sufficient show that for $A < a''$

$$\lim_{n \in I} w_n(A;W) = \infty\ . \tag{53}$$

Then the same limiting relation is true for all infinite subsequences $J \subset I$. For a proof of (53) distinguish two cases: If $W = 0$ then $w_n(A;W) = v_n(A)$ and (53) follows directly from the fact that $A \notin S$ (since $A < \inf S$). Otherwis apply the inequalty (52) with

$$M = \frac{|W|}{\sqrt{(A - a'')/2}}.$$

Then $\exp(-M/2)v_n(a'' - \frac{a''-A}{2}) \leq w_n(A;W)$. The lefthand side of this inequality converges to ∞, since the argument of v_n is less than a''. Consequently, the right hand side must converge to ∞, too, which proves (53).

Let us now prove (50): Let us now define the family of measures $\{\kappa_n(a) : n \in N; a > a''\}$ on $R - (-\infty;\infty)$ by

$$\frac{d\kappa_n(a)}{d\mu_n} = \exp(-\frac{\lambda^2}{2a}) \tag{54}$$

Each $\kappa_n(a)$ is a measure on $R - (-\infty;\infty)$. When we integrate we will encounter formulae like $\int f\, d\kappa_n(a)$. These expression should not be interpreted as integrals with respect to a. They are integrals with respect to the measure $\kappa_n(a)$ (in the sequel, we will mainly iintegrate with respect to λ!).

Then

$$\kappa_n(a)(R - (-\infty;\infty)) = v_n(a) \tag{55}$$

. As $\limsup_{n\in I} v_n(a''+1) < \infty$ we also have

$$\sup_{n\in I} \kappa_n(a''+1)(R - (-\infty;\infty)) < \infty$$

Then we can again apply theorem 13 and conclude that there exists a measure $\psi(a''+1)$ on $R - (-\infty;\infty)$ and an infinite sequence $J \subseteq I$ so that the distribution function of $\kappa_n(a''+1)$ converges for $n \in J$ to the distribution function of $\psi(a''+1)$ in all points for which the latter one is continuous. Let us now define for $\gamma \geq a''$

$$\frac{d\psi(\gamma)}{d\psi(a''+1)} = \frac{\exp(-\frac{\lambda^2}{2\gamma})}{\exp(-\frac{\lambda^2}{2(a''+1)})} \tag{56}$$

and

$$\mu = \psi(a'')$$

Evidently the $\psi(\gamma)$ are (at most) σ-finite measures. The density with respect to the finite measure $\psi(a''+1)$ is bounded on all compact sets. Let us now establish that for all $\varepsilon > 0$ $\kappa_n(a''+\varepsilon)$ converges for $n \in J$ vaguely to $\psi(a''+\varepsilon)$.

Let $\varepsilon > 0$. Then - as $a''+\varepsilon \in S$ - $\sup_{n\in J} v_n(a''+\varepsilon)(R) < \infty$ and consequently due to (55) $\sup_{n\in J} \kappa_n(a''+\varepsilon)(R) < \infty$. Hence every infinite subsequence of $\kappa_n(a''+\varepsilon)$ must contain a (vaguely) convergent subsequence. Hence if we assume that $\kappa_n(a''+\varepsilon) \nrightarrow \psi(a''+\varepsilon)$ then existed a subsequence $K \subseteq J$ so that $\kappa_n(a''+\varepsilon); n \in K$ converged (vaguely) to some measure $\nu \neq \psi(a''+\varepsilon)$. Then theorem 14 shows that for all bounded, continuous f so that $\lim_{x\to\pm\infty} f(x) = 0$ for $n \in K$ $\int f\, d\kappa_n(a''+\varepsilon) \to \int f\, d\nu$.

So if g is an arbitrary bounded continuous function with $\lim_{x\to\pm\infty} g(x) = 0$, we have $\int g d\mu(a'' + 1)(\lambda) = \lim_{n\in K} \int g d\mu_n(a'' + 1)(\lambda)$. Using (54) the limit on the righthand side equals $\lim_{n\in K} \int g r d\mu_n(a'' + \varepsilon)(\lambda)$ with $r(\lambda) = \exp(-\frac{\lambda^2}{2}\left[\frac{1}{a''+\varepsilon} - \frac{1}{a''+1}\right])$, and this limit equals - since gr is continuous, bounded and satisfies $\lim_{x\to\pm\infty}(gr)(x) = 0$ - $\int g r d\xi$. Hence for all g $\int g d\mu(a''+1) = \int g r d\xi$ and - again using (54) $\int g r d\nu(a''+\varepsilon) = \int g r d\xi$. Sine r is strictly positive and uniformly bounded away from zero for all compact sets we may conclude that $\nu(a'' + \varepsilon) = \xi$, which contradicts our assumption that $\xi \neq \nu(a'' + \varepsilon)$. Hence we have established that $\mu_n(a'' + \varepsilon)$ converges for $n \in J$ weakly to $\nu(a'' + \varepsilon)$.

The last step in our proof of proposition 11 is to establish (50), namely that for all (A, W) with $A > a''$

$$\lim_{n\in J} \int_{R-(-\infty,\infty)} \exp(\lambda W - \frac{\lambda^2 A}{2}) d\mu_n = \int_{R-(-\infty,\infty)} \exp(\lambda W - \frac{\lambda^2 A}{2}) d\mu$$

Fix A and W. If $W = 0$, then (50) simply states that $\lim_{n\in J} \int d\mu_n(a''+\varepsilon) = \int d\nu(a''+\varepsilon)$, which follows directly from the weak convergence we have established above; then define $M = \frac{|W|}{\sqrt{(A-a'')/2}}$ and $A' = \frac{A+a}{2}$, then we can apply (52) and conclude that

$$\begin{aligned}\exp(\lambda W - \frac{\lambda^2 A}{2}) &\leq \exp(\frac{M}{2}) \exp(\frac{\lambda^2}{2}(\frac{W^2}{M^2} - A)) \\ &= \exp(\frac{M}{2}) \exp(\frac{\lambda^2}{2}(-\frac{A + a''}{2}))\end{aligned}$$

Therefore $\exp(\lambda W - \frac{\lambda^2 A}{2}) \exp(\frac{\lambda^2}{2}(\frac{A+a''}{2}))$ remains bounded - and as $(\frac{A+a''}{2}) < A$ - converges to 0 for $\lambda \to \pm\infty$. As $(\frac{A+a''}{2}) > a''$, we can conclude from the weak convergence that

$$\begin{aligned}&\lim_{n\in J} \int_{R-(-\infty,\infty)} \left\{\exp(\lambda W - \frac{\lambda^2 A}{2}) \exp(\frac{\lambda^2}{2}(\frac{A + a''}{2}))\right\} \exp(-\frac{\lambda^2}{2}(\frac{A + a''}{2})) d\mu_n \\ = &\int_{R-(-\infty,\infty)} \left\{\exp(\lambda W - \frac{\lambda^2 A}{2}) \exp(\frac{\lambda^2}{2}(\frac{A + a''}{2}))\right\} \exp(-\frac{\lambda^2}{2}(\frac{A + a''}{2})) d\mu\end{aligned}$$

which is just the desired result.

Now having established proposition 1 and 2, we can show the final result, namely that the limiting test φ has the desired form with parameters $a = \max(a', a''), b, c, U$ and measures μ and ν. We have to show that the test $\varphi(A, W) = 1$ if $A < a$ or $u(A, W) > U$ and $\varphi(A, W) = 0$ if $A > a$ and $u(A, W) < U$. This is equivalent to show that for all $\varepsilon > 0$ the following holds true: Define $B(\varepsilon) = \{(A, W) : A < a - \varepsilon$ or $u(A, W) > U + \varepsilon\}$ and $C(\varepsilon) = \{(A, W) : A > a + \varepsilon$ or $u(A, W) < U - \varepsilon\}$. then we have to show that

$$\int_{B(\varepsilon)} (1 - \varphi) dQ_0 = 0 \qquad (57)$$

and
$$\int_{C(\varepsilon)} \varphi \, dQ_0 = 0. \tag{58}$$
>From (43) we can conclude that
$$\lim_{n \in J} \int_{B(\varepsilon)} (1 - \varphi_n) dQ_0 = \int_{B(\varepsilon)} (1 - \varphi) dQ_0 \tag{59}$$
and
$$\lim_{n \in J} \int_{C(\varepsilon)} \varphi_n dQ_0 = \int_{C(\varepsilon)} \varphi \, dQ_0 \tag{60}$$
Let $(A; W) \in B(\varepsilon)$. Then either $A < a - \varepsilon$ or $u(A; W) > U + \varepsilon$. In the second case 11 and (48) immediately imply that $u_n(A; W) > U_n$ for all but finitely many $n \in J$ and consequently φ_n rejects for these $n \in J$. In the first case we have to distinguish two cases again: Either $A < a' - \varepsilon$ or $A < a'' - \varepsilon$. In the first case (48) shows that $A < a_n$ for all but finitely many $n \in J$, so φ_n rejects for n large enough.: In the second case 10 and (49) show that $u_n(A; W) \to \infty$, hence $u_n(A; W) > U_n$ for all but finitely many $n \in J$ (as (48) shows that U_n remain bounded). So we have shown that φ_n rejects for n large enough.

Hence we have shown that for $(A; W) \in B(\varepsilon)$ $\varphi_n(A; W) \to 1$ and therefore we may conclude by the theorem of dominated convergence that
$$\lim_{n \in J} \int_{B(\varepsilon)} (1 - \varphi_n) dQ_0 = \int_{B(\varepsilon)} \lim_{n \in J} (1 - \varphi_n) dQ_0 = 0,$$
which (together with (59))proves (57).

With a similar argumentation we can show that for $(A; W) \in C(\varepsilon)$ $\varphi_n(A; W) \to 0$. Indeed, for $(A; W) \in C(\varepsilon)$ $A \geq a + \varepsilon \geq a' + \varepsilon$. Hence for all but finitely many $n \in J$
$$A > a_n.$$
Furthermore we have $A \geq a + \varepsilon \geq a'' + \varepsilon$. therefore (48) and (50) imply that $\lim_{n \in J} u_n(A; W) = u(A; W) > U + \varepsilon = \lim_{n \in J} U_n + \varepsilon$. hence for all but finitely many $n \in J$
$$u_n(A; W) > U_n.$$
So for all but finitely many $n \in J$ both conditions for acceptance are satisfied, and so for $n \in J$ large enough $\varphi_n(A; W) = 0$. We therefore have shown that for all $(A; W) \in C(\varepsilon)$ $\lim_{n \in J} \varphi_n(A; W) = 0$ and hence we may apply again the theorem of dominated convergence and conclude that
$$\lim_{n \in J} \int_{C(\varepsilon)} \varphi_n dQ_0 = \int_{C(\varepsilon)} \lim_{n \in J} \varphi_n dQ_0 = 0,$$
which, together with (60) proves (58). we thus have completed the proof.

Appendix B: The Vague Topology and Radon Measures

Since we are working with measures of (possibly) infinite mass we cannot use the usual weak topology of measures: Instead we have to use an analogue - the vague topology: Since all of our measures are defined on the $\mathbb{R}^k$ we do not need to introduce this concept in full generality: The interested reader is referred to Bauer(1972).

Definition 12 A Radon measure on $\mathbb{R}$ is a σ-finite, (nonnegative) measure μ on the Borel-sets of $\mathbb{R}$ so that for every compact set $K \subset \mathbb{R}$

$$\mu(K) < \infty.$$

The vague topology on the set of Radon measures is the weakest topology so that the mappings

$$\mu \mapsto \int f d\mu; \quad f \in C^{(c)}(\mathbb{R}) \tag{61}$$

are continuous, where $C^{(c)}(\mathbb{R})$ is the set of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$ which vanish outside a compact set (also called functions with compact support).

So if μ_n converges to μ in the vague topology we have for all f vanishing outside a compact set

$$\lim \int f d\mu_n = \int f d\mu$$

This topology has the following advantages:

Theorem 13 Let μ_n be a sequence of (nonegative) Radon measures and assume that

$$\sup_{n \in \mathbb{N}} \mu_n(\mathbb{R}) < \infty \tag{62a}$$

Then there exists an infinite subsequence $I \subset \mathbb{N}$ and a Radonmeasure μ so that

$$\lim_{n \in I} \mu_n = \mu \text{ in the vague topology}$$

and

$$\mu(\mathbb{R}) \leq \sup_{n \in \mathbb{N}} \mu_n(\mathbb{R})$$

(i.e. if the total masses of the measures remain bounded we can choose a convegent subsequence).

We do not give a proof here since the theorem follows immediately from corollary 7.8.3 of Bauer(1972). So the vague topology is a very attractive one: We only have to show that a sequence of measures has uniformly bounded mass: Then we can pick a vaguely convergent sequence. On the other hand, vague convergence does only guarantee the convergence of integrals of functions which vanish outside a compact set: With the additional assumption of boundedness we are, however, able to derive a sharper result:

Theorem 14 Suppose we have given a sequence of Radon measures μ_n converging to μ. Suppose that (62a) is fulfilled. Then for every continuous f so that

$$\lim_{x \to \pm\infty} f(x) = 0$$

$$\int f \, d\mu_n \to \int f \, d\mu$$

Again we do not give a proof: It can be found in Bauer(1972), theorem 7.7.5 on p.231.

The vague convergence is related to the "usual" weak convergence via the following result:

Theorem 15 Let μ_n be a vaguely convergent sequence of Radon measures so that (62a) is fulfilled. Let $\mu = \lim \mu_n$ and let F_n, F be the distribution functions of μ_n, μ. Then for all $x \in \mathbb{R}$

$$\lim F_n(x) = F(x) \tag{63}$$

for all x so that $F(\cdot)$ is continuous in x.

This result follows immediately from theorem 7.7.11 in Bauer(1972) on p.235.

Remark 4 : (63) is - in general - not true for $x = \infty$: In general, the vague limit of e.g. probabilitymeasures will not be a probabilitymeasure again!

If, however, (63) also holds for $x = \infty$ we have also weak convergence and we can apply the Helly-Bray lemma:

Corollary 16 If additionally $\lim F_n(\infty) = F(\infty)$ (i.e. $\lim \mu_n(\mathbb{R}) = \mu(\mathbb{R})$) then for all continuous, bounded $f(\cdot)$ $\lim \int f \, d\mu_n = \int f \, d\mu$.