

ADMISSIBLE AND NONADMISSIBLE TESTS IN UNIT-ROOT-LIKE SITUATIONS

WERNER PLOBERGER
Washington University in St. Louis

This paper investigates the asymptotic behavior of tests in situations where the likelihood is locally asymptotically quadratic. Necessary and sufficient conditions are given for a test to be admissible. Even without these restrictive parametric assumptions, it is shown that certain common procedures—such as the augmented Dickey–Fuller test in cases where no deterministic trend is present or standard tests for restrictions on cointegrating relationships—are asymptotically inadmissible. These results confirm the existence of tests that dominate these classical tests for all parameters.

1. INTRODUCTION

This paper discusses optimality properties of certain statistical tests and makes two contributions. First we analyze a restricted parametric model. These restrictions enable us to determine the class of asymptotically optimal tests. Moreover, we show that in nonclassical situations many traditional tests for testing simple hypotheses, such as, e.g., the usual generalizations of the t -test, are not in our class and are inadmissible. Second, we use these results to show the inadmissibility of some familiar procedures, such as the Dickey–Fuller and related tests, when there is no deterministic trend in the regression or standard t -type tests for restrictions on cointegrating relations.

We begin by outlining the asymptotic testing problem for which we can exactly describe the class of all admissible tests. Let us assume that our parametric family is unidimensional, and later on we move on to discuss more general cases.

Many problems in asymptotic statistics have a specific form, namely, that the logarithm of the likelihood can be approximated by a quadratic function of the (properly normalized) parameter. Classical asymptotic analysis then rests on two crucial assumptions: first, that the linear term of the quadratic function

I express my gratitude to the editors, H. Lütkepohl and especially Peter C.B. Phillips, for their help, which enormously exceeded the usual amount of support. Also I thank the referees for their helpful comments. Their contribution greatly improved the paper. All remaining errors are mine. Address correspondence to Werner Ploberger, Washington University in St. Louis, Department of Economics, Campus Box 1208, St. Louis, MO, USA 63130-4899; e-mail: wernerp@artsci.wustl.edu.

is (asymptotically) normal; and second, that the quadratic term (the Fisher information) is asymptotically constant.

In many situations relevant for econometrics—most of them connected to problems with unit roots—these assumptions are not fulfilled. Although the log-likelihood can asymptotically be approximated by a quadratic function, the linear term in the parameter is not Gaussian, and the matrix describing the quadratic term is not a constant. The most typical example arises with the simplest autoregressive (AR) model

$$\Delta y_t = \delta y_{t-1} + u_t,$$

where u_t is normal with zero expectation and known variance. Then it is straightforward to show that the likelihood is quadratic around $\delta = 0$. The properly normalized coefficients of this quadratic function are neither normal nor constant. The prominence of this example indicates that there should be a significant number of possible applications. These will be discussed in more detail in Section 3.

So let us denote by $p_n(\theta)$ the density corresponding to parameter θ . For simplicity, let us assume that the parameter θ is unidimensional. So, for a given true parameter θ_0 , we assume that asymptotically with a suitable scaling sequence $D_n \uparrow \infty$

$$\log p_n(\theta_0 + D_n^{-1}h) - \log p_n(\theta_0) \sim hW_n - \frac{1}{2}A_n h^2, \quad (1)$$

where we assume that the pair (W_n, A_n) converges in distribution to a nontrivial limit (W, A) . We need not discuss here the precise meaning of the approximation or convergence in (1), for which the reader is referred to Le Cam and Yang (1990), Ploberger (2004), and Jeganathan (1991, 1995). In Ploberger (2004), specific problems associated with testing

$$\theta = \theta_0 \quad (2)$$

were discussed.

First of all let us analyze (1). If the log-likelihood were exactly quadratic (i.e., in (1) we would replace \sim with $=$), then the likelihood only depends on (W_n, A_n) . Consequently, this pair would be a *sufficient* statistic. In that case, we would be able to find for every test φ another test ψ depending only on (W_n, A_n) so that φ and ψ have identical power. Hence, we would only need to analyze tests depending on (W_n, A_n) .

Modern statistical theory allows us to use similar reasoning in the asymptotic case. We can, however, no longer analyze the tests themselves; the theory only gives us convergence results about the power functions of the tests. Typically, the theory allows us to show that the behavior of power functions of problems like (2) is similar to a fixed “asymptotic” testing problem. The theory

shows that one can approximate the *power functions* of tests for (2) with *power functions* of tests for the “asymptotic” problem up to an arbitrary accuracy.

We now briefly sketch the reason for this kind of analysis, emphasizing that the arguments in the following paragraph are heuristic. A rigorous treatment of the subject can be found in Strasser (1985).

If a test ψ is admissible for the asymptotic problem, then we know that we can find a sequence φ_n of tests for (2) such that with increasing sample size the difference between the power functions of the φ_n and the power function of ψ converges to zero. (Although the tests are defined on different sample spaces, their power functions have the same domain, namely, the parameter space!) Then the φ_n are in a certain sense asymptotically admissible. Assume one has given some “competitors” ρ_n . Then the power functions of the ρ_n can be approximated by power functions of tests ν_n for the asymptotic problem. There are compactness results available that show that the power functions of the ν_n must contain a convergent subsequence, converging to the power function of a test ν . We did assume ψ to be admissible. Hence it cannot be dominated by ν , and so it must have better power properties for some parameters. Hence, *asymptotically*, the same must be true for the power functions of φ_n and ρ_n . The same kind of argument applies, vice versa, for tests that are not admissible for the asymptotic problem.

This type of testing problem was analyzed for a one-dimensional parameter in Ploberger (2004), and one of the main results was that power functions of tests for (2) can either be uniformly approximated or dominated by power functions of tests for the following testing problem. Assume we have given a parametrized family Q_λ , $\lambda \in \mathbf{R}$ on the space

$$\mathbf{R} \times \mathbf{R}_+ = \{(W, A) : A > 0\}$$

so that

$$\frac{dQ_\lambda}{dQ_0} = \exp\left(\lambda W - \frac{\lambda^2}{2} A\right).$$

We can think of this parametrized family as a statistical experiment, namely, a sample of size one of a random two-dimensional vector distributed according to a Q_λ with an unknown λ . Suppose we want to consider tests for the hypothesis $\lambda = 0$. In Ploberger (2004) a set of tests was constructed such that for every other test we can find one from our class with “better” power properties. Such a class is called “essentially complete” in Schervish (1995, p. 174) and “complete” in Strasser (1985, p. 41) and Ploberger (2004). Because the present paper heavily relies on Ploberger (2004), we will continue to use the term *complete* for this kind of set of tests.

Together with the limiting result, this complete class theorem characterizes all possible limits for power functions of tests for the problem (2). However, the problem is that every set covering (in our sense) a complete class is a com-

plete class again. Hence the result from Ploberger (2004) does not allow us to directly classify tests. If we know that a test is within our class we cannot directly conclude that it is admissible, because our definition allows the “complete class” to contain additional tests.

Moreover, a test outside the complete class may be admissible. The results of Ploberger (2004) do not preclude that there exists another test (within the class) with exactly the same power function. In this paper, we show that this is in fact not the case. The main theorem implies that every test within our class is admissible and characterized by the power function. Hence we can conclude that tests not in our class are asymptotically inadmissible. So essentially we show that for every test outside our complete class one can find a test from our class that is better *uniformly*—i.e., it is better for *all* values of the parameter!

Moreover, the same paper establishes the fact that the test based on the t - or F -test-type statistic, namely, rejecting if

$$\frac{W^2}{A} > U \quad (3)$$

for some critical values U , is not within this class. In Section 3 we discuss important examples where our results show the existence of tests that dominate standard procedures uniformly. Obviously, it is also important to know how much power one can win by using the admissible tests. For the case of unit root testing, this question was analyzed in Elliot, Rothenberg, and Stock (1996).

This result has also some interesting consequences when we are testing (2) for our family (1). Suppose we have a sequence of tests ψ_n for the finite-sample problem such that their asymptotic power function converges to the power function of (3). Then the standard convergence results of decision problems (cf. Strasser, 1985) allow us to conclude that there exists a sequence of tests with power functions dominating the power functions of ψ_n uniformly!

In Section 3, we then extend these results to show the inadmissibility of certain familiar procedures.

1. The Dickey–Fuller τ -tests when no trend is present: Consider a univariate time series y_t and consider the models

$$\Delta y_t = \delta y_{t-1} + \pi_1 \Delta y_{t-1} + \cdots + \pi_p \Delta y_{t-p} + u_t \quad (4)$$

and suppose we want to test whether $\delta = 0$ against $\delta \neq 0$. Then it seems natural to compute the t -test statistic based on an estimator for δ and use the critical values established by Dickey and Fuller (1979, 1981).

2. Suppose we have an estimator—say, $\hat{\rho}_n$ —for a scalar parameter ρ and there exists a *scaling statistic* $\hat{\sigma}_n^2$. Let us assume that (with some normalizing sequence D_n)

- $\sqrt{\hat{\sigma}_n^2}^{-1}((\hat{\rho}_n - \rho))$ converges in distribution to a standard normal distribution.
- $D_n^{-2} \hat{\sigma}_n^2$ converges in distribution to some random variable σ , which is almost surely positive and not a constant.
- Asymptotically, the distributions of $\sqrt{\hat{\sigma}_n^2}^{-1}(\hat{\rho}_n - \rho)$ and $D_n^{-2} \hat{\sigma}_n^2$ are independent.

Although there are some other applications, the most important case where this kind of phenomenon occurs is the estimation of parameters describing cointegrating relationships. So when testing a restriction—say, $\rho = 0$ —one is tempted to use the analogue to the Wald test. Indeed, one can easily see that in case the null is true,

$$\frac{\hat{\rho}_n}{\sqrt{\hat{\sigma}_n^2}} \tag{5}$$

converges in distribution to a standard normal distribution. Therefore it seems natural to use tests based on (5) and use one- or two-sided critical values. Here I show that these procedures are inadmissible in the two-sided case.

Here we only deal with the simplest cases of unit root testing. Clearly, the results would be more valuable if we allowed for deterministic terms. Typically, one is interested in tests invariant with respect to these terms. The advantage of this approach would be a reduction in the parameter space, which would suit us well. However, the “reduced” problem no longer satisfies (1). So a more general analysis of the problem would be beyond the scope of the paper.

2. THE MAIN THEOREMS FOR THE UNIDIMENSIONAL CASE

Tests are functions φ from the set $\mathbf{R} \times \mathbf{R}_+$ to the interval $[0,1]$ ($\varphi = 0$ means that we accept the null; $\varphi = 1$ means that we reject the null; and in between we randomize).

The power function of a test φ is the function that attaches to each $\lambda \in \mathbf{R}$ the value $\int \varphi dQ_\lambda$. The concept of a power function enables us to compare tests. A test φ is better than a test ψ if

$$\begin{aligned} \int \varphi dQ_\lambda &\geq \int \psi dQ_\lambda \quad \text{for } \lambda \neq 0, \\ \int \varphi dQ_\lambda &\leq \int \psi dQ_\lambda \quad \text{for } \lambda = 0. \end{aligned} \tag{6}$$

So a better test dominates the worse test for all possible values of the parameter. In Ploberger (2004), it was established that the set C of tests consisting of the test functions defined subsequently is complete (or essentially complete according to Schervish, 1995).

These test functions are described by the following parameters: U, a, b, c and measures μ, ν subject to the following restrictions.

1. ν is a σ -finite measure on $[-1, 1] - \{0\}$ (i.e., no mass outside the unit interval with the number zero excluded) such that for all compact sets $K \subset [-1, 1] - \{0\}$ $\nu(K) < \infty$ and

$$c \geq \int_{[-1, 1]} \lambda^2 d\nu(\lambda) \tag{7}$$

(in particular $\int_{[-1, 1]} \lambda^2 d\nu(\lambda)$ is finite and $c \geq 0$).

2. μ is defined on $\mathbf{R} - [-1, 1]$ (i.e., the real line except the unit interval), and for all $M > 1$

$$\mu([-M, -1] \cup [1, M]) < \infty. \tag{8}$$

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

$$\begin{aligned} u(A, W) = & bW + \frac{c}{2} (W^2 - A) \\ & + \int_{[-1, 1]} \left(\exp\left(\lambda W - \frac{\lambda^2}{2} A\right) - \left(1 + \lambda W + \frac{\lambda^2}{2} (W^2 - A)\right) \right) d\nu(\lambda) \\ & + \int_{\mathbf{R} - [-1, 1]} \exp\left(\lambda W - \frac{\lambda^2}{2} A\right) d\mu(\lambda). \end{aligned} \tag{9}$$

Now let the class of tests be characterized by the following properties: (i) the test should reject the null for all (A, W) , so that $A < a$; (ii) if $A > a$, then the test should **reject** if

$$u(A, W) > U \tag{10}$$

and **accept** if

$$u(A, W) < U. \tag{11}$$

These parameters do not describe the test completely. In particular, we did not define what happens if $A = a$ or if $u(A, W) = U$. So if these equations do not describe null sets (with respect to all Q_h), our complete class defined previously may contain some additional (not optimal) tests. This, however, is not a contradiction to the definition of a complete class—any larger set of a complete class is a complete class, too!

Remark 1. Because of our restrictions 1 and 2, in particular expressions (7) and (8), we can easily see that every function $u(A, W)$ satisfying (9) can for every $\beta, \gamma > 0$ be written as

$$\begin{aligned}
 u(A, W) &= b'W + \frac{c'}{2} (W^2 - A) \\
 &+ \int_{[-\beta, \gamma]} \left(\exp\left(\lambda W - \frac{\lambda^2}{2} A\right) - \left(1 + \lambda W + \frac{\lambda^2}{2} (W^2 - A)\right) \right) d\nu'(\lambda) \\
 &+ \int_{\mathbf{R} - [-\beta, \gamma]} \exp\left(\lambda W - \frac{\lambda^2}{2} A\right) d\mu'(\lambda), \tag{12}
 \end{aligned}$$

where ν', μ' are the restrictions of the measure $\mu + \nu$ on $[-\beta, \gamma]$ and $\mathbf{R} - [-\beta, \gamma]$, respectively.

Remark 2. Because

$$\frac{\partial^3}{\partial \lambda^3} \exp\left(\lambda W - \frac{\lambda^2}{2} A\right) = \exp\left(\lambda W - \frac{\lambda^2}{2} A\right) \{(W - \lambda A)[(W - \lambda A)^2 - 3A]\},$$

we can replace the first integral in (9) by

$$\begin{aligned}
 &\int_{[-\beta, \gamma]} \left(\exp\left(\lambda W - \frac{\lambda^2}{2} A\right) - \left(1 + \lambda W + \frac{\lambda^2}{2} (W^2 - A)\right) \right) d\nu'(\lambda) \\
 &= \int_{[-\beta, \gamma]} \left\{ \int_0^\lambda \int_0^w \int_0^v \exp\left(zW - \frac{z^2}{2} A\right) \right. \\
 &\quad \left. \times (W - zA)[(W - zA)^2 - 3A] dzdvdw \right\} d\nu'(\lambda). \tag{13}
 \end{aligned}$$

Additionally, we assume that the following three conditions are fulfilled.

Condition C1. For some $\varepsilon > 0$ we have

$$\sup_{|\lambda| \leq \varepsilon} \int A^3 dQ_\lambda < \infty, \tag{14}$$

$$\sup_{|\lambda| \leq \varepsilon} \int |W|^3 dQ_\lambda < \infty. \tag{15}$$

Condition C2. The support of the measure Q_0 (and hence the support of all the measures Q_λ) is the whole $\mathbf{R} \times \mathbf{R}_+ = \{(W, A) : A > 0\}$ (i.e., every open set of $\mathbf{R} \times \mathbf{R}_+$ has positive probability under Q_0).

Condition C3. The distribution function of A is continuous (i.e., for all a $Q_0([A = a]) = 0$).

THEOREM 1. *Suppose restriction 1 and Conditions C1, C2, and C3 hold and assume that ψ is a test that has the same or a better power function than the test φ , which is from our class and is determined by the parameters a, b, c, U, μ, ν . Then ψ is from our class, too, and determined by the same parameters.*

Proof. The proof is provided in the Appendix.

It is, however, a nontrivial task to show that tests are not in our complete class. For a proof of this fact, it is *not* sufficient to show that the corresponding test statistic cannot be written in the form (9). One has to show that the corresponding test (or the critical set) cannot be generated as a level set of any function (9)!

Suppose now that we have given a family $Q_\lambda, \lambda \in \mathbf{R}$ and a test φ (which is a function from $\mathbf{R} \times \mathbf{R}_+$ into $[0, 1]$) satisfying the following requirements.

1. All the assumptions of the previous theorem are fulfilled.
2. Every neighborhood of the point (0,0) has positive probability with respect to Q_0 .
3. There exists a constant $K > 1$ such that

$$\varphi(W, A) = 0 \quad \text{if} \quad \frac{W^2}{A} < K,$$

$$\varphi(W, A) = 1 \quad \text{if} \quad \frac{W^2}{A} > K.$$

Then the following theorem holds.

THEOREM 2. *The test φ is not in our class and is therefore inadmissible.*

Proof. The assumptions of this theorem are quite similar to those of Ploberger (2004), and only our second assumption here is stronger. Here we only postulate that all neighborhoods of the origin have positive mass, instead of assuming that every open set of the half plan $\mathbf{R} \times \mathbf{R}_+$ has positive mass. One can easily see that this requirement is sufficient, because only the behavior of the power function for alternatives near the origin is used in the proof. ■

In case the likelihoods are given by (1), one consequence of the preceding theorems is that the analogues to the “classical” Lagrange multiplier (LM), likelihood ratio (LR), and Wald tests for testing $\theta = \theta_0$ are asymptotically inadmissible. More precisely, let us denote by φ_n such a test. The “classical” tests reject when test statistics exceed the critical values. Moreover, it is an easy exercise—very similar to the standard proof of asymptotic equivalence in the stationary

case—to show that differences between suitably normed and transformed test statistics and W_n^2/A_n converge to zero under the null. Let us denote the asymptotic critical value for the test statistic W_n^2/A_n by K and assume that K is such that the cumulative distribution function of W^2/A is continuous at K . This is not a very strong restriction, because a cumulative distribution function can have at most countably many discontinuities. Let us define φ_n by

$$\varphi_n(W_n, A_n) = 0 \quad \text{if } \frac{W_n^2}{A_n} < K,$$

$$\varphi_n(W_n, A_n) = 1 \quad \text{if } \frac{W_n^2}{A_n} > K.$$

Then we can easily see that under the null the probability of the LM, LR, and Wald tests and the tests φ_n giving different results converges to zero. (So, with probability converging to one under the null, either all of them accept or all of them reject!) Hence the power functions of φ_n and the classical tests are asymptotically the same under the null.

We now have to show that the same holds true for the alternatives. Consider local alternatives of the form

$$\theta_n = \theta_0 + \frac{C}{D_n}.$$

In Ploberger (2004), *contiguity* of the probability measures corresponding to θ_0 and θ_n was established. This means that for any sequence E_n of events such that $P_{\theta_0}(E_n) \rightarrow 0$ $P_{\theta_n}(E_n) \rightarrow 0$, too. (Contiguity is one of the cornerstones of modern theoretical asymptotic theory. For a more detailed discussion see Le Cam and Yang, 1990; Strasser, 1985.) Because the probability of conflict between the LM, LR, and Wald tests and φ_n converges to zero under the null, the result quoted previously allows us to conclude that the probability of a conflict under the local alternatives converges to zero, too. Hence, the difference of the probabilities of rejection—which is exactly the definition of a power function—must converge to zero for $n \rightarrow \infty$. Therefore, the power functions are the same asymptotically. We assumed that $(W_n, A_n) \rightarrow (W, A)$ in distribution under the null, and so $W_n^2/A_n \rightarrow W^2/A$ in distribution, too. Because the cumulative distribution function of the limiting distribution is continuous at K , the power function of the φ_n (and, consequently, the power functions of the LM, LR, and Wald tests) must converge to the power function of our test φ under the null.

To show the convergence of the power functions of the φ_n to the power function of φ at the local alternatives requires only a bit more technical effort. Let us fix the sequence of local alternatives and consider the densities

$$\ell_n = \frac{dP_{\theta_n}}{dP_{\theta_0}}.$$

Then we can easily conclude from (1) that

$$\ell_n \rightarrow \ell = \exp\left(CW - \frac{C^2}{2} A\right) \tag{16}$$

in distribution. We did assume, however, that

$$\int \ell dQ_0 = \int \exp\left(CW - \frac{C^2}{2} A\right) dQ_0 = 1,$$

because otherwise Q_C would not be a probability measure. Now let $\varepsilon > 0$ be arbitrary. Then we can find an L so that

$$\int \inf(\ell(A, W), L) dQ_0 \geq 1 - \varepsilon.$$

Consequently, as $\ell_n \rightarrow \ell$ in distribution,

$$\lim_{n \rightarrow \infty} \int \min(\ell_n(A_n, W_n), L) dP_{\theta_n} \geq 1 - \varepsilon,$$

and as $\int \ell_n dP_{\theta_n} = 1$,

$$\lim_{n \rightarrow \infty} \int (\ell_n - \min(\ell_n, L)) dP_{\theta_n} \leq \varepsilon. \tag{17}$$

Hence we have

$$\lim \int \varphi_n dP_{\theta_n} = \lim \int \varphi_n \ell_n dP_{\theta_0}.$$

One can easily see that

$$\ell_n - \min(\ell_n(A, W), L) + \varphi_n \min(\ell_n(A, W), L) \geq \varphi_n \ell_n \geq \varphi_n \min(\ell_n(A, W), L).$$

Further, because of the convergence of (A_n, W_n) and ℓ_n ,

$$\lim \int \varphi_n \min(\ell_n(A, W), L) dP_{\theta_n} = \int \varphi \min(\ell(A, W), L) dQ_0.$$

We have that (when using (17))

$$\begin{aligned} \varepsilon + \int \varphi \min(\ell(A, W), L) dQ_0 &\geq \limsup \int \varphi_n dP_{\theta_n} \\ &\geq \liminf \int \varphi_n dP_{\theta_n} \\ &\geq \int \varphi \min(\ell(A, W), L) dQ_0. \end{aligned}$$

As $\varepsilon + \int \varphi \min(\ell, L) dQ_0 \geq \int (\ell - \min(\ell, L)) dQ_0 + \int \varphi \min(\ell, L) dQ_0 \geq \int \varphi \ell dQ_0$, we can conclude that

$$\varepsilon + \int \varphi \ell dQ_0 \geq \limsup \int \varphi_n dP_{\theta_n} \geq \liminf \int \varphi_n dP_{\theta_n} \geq \int \varphi \ell dQ_0 - \varepsilon.$$

Because ε was arbitrary, we have shown that

$$\lim \int \varphi_n dP_{\theta_n} = \int \varphi \ell dQ_0.$$

This establishes the fact that the power function of the φ_n converges to the power function of φ for all local alternatives. Because we already know that φ is not within our class, we can find a test ψ with (strictly) better power properties. In Theorem 3 in Section 3 we show that in this situation it is possible to construct tests ψ_n with strictly better asymptotic power properties than the φ_n , which establishes our result, namely, that the φ_n are asymptotically inadmissible.

In view of the preceding results one is tempted to look at other “generalizations” of the classical LR test to situations with stochastic information. One such generalization is the test based on the posterior information criterion (PIC) (cf. Phillips, 1996; Phillips and Ploberger, 1996; Ploberger and Phillips, 2003). Consider tests that reject if

$$\frac{W^2}{2A} - \frac{1}{2} \ln A > K \tag{18}$$

and accept if

$$\frac{W^2}{2A} - \frac{1}{2} \ln A < K, \tag{19}$$

where K is a constant determined by the desired significance level. These tests are all admissible. Simply consider our measures μ, ν to be the Lebesgue measure and choose the other constants appropriately. Then it is a simple but tedious calculation to show that (12) yields a test based on

$$\frac{1}{\sqrt{A}} \exp\left(\frac{W^2}{2A}\right),$$

which is immediately seen to be equivalent to the test determined by (18) and (19).

It should be noted that in the “classical” case of A being constant this test is equivalent to a test based on W^2/A , too. So we can view the PIC test as a generalization of the classical tests, too.

3. APPLICATIONS

The result of the previous section establishes a minimal complete class theorem for a rather large class of testing problems. Apart from the three technical conditions we only require the likelihood function to be locally asymptotically quadratic. The main point of interest seems to be that this class of admissible tests is quite different from the one for the classical problem where W is Gaussian and A constant. In this kind of situation the set of all admissible tests is well known even for multivariate parameters (Birnbaum’s theorem; cf. Strasser, 1985). Each of these tests is characterized by a convex set (in the univariate case an interval) and rejects if W lies outside the interval. Hence Theorem 1 of the previous section seems to be of considerable theoretical interest. It shows that even a small change in our assumptions makes enormous differences for the class of admissible tests.

In the previous section we only dealt with one-parameter families. This fact represents a significant hurdle for a direct application of the result. However, we can apply the result to prove the *inadmissibility* of tests. We will show that certain tests—e.g., the Dickey–Fuller τ -test and the pseudo- t -test based on the statistic (5)—can be interpreted as tests on a unidimensional parameter. Our result then shows that there must exist a test with uniformly better power function. The drawback is that the result is nonconstructive. The only information we supply about the better test is that it has a form of (9). Nevertheless, the result may motivate researchers in the field. Efforts to find better tests are now not doomed from the start. We know that a better test exists!

Let us first analyze the Dickey–Fuller test. Unfortunately, the techniques presented here cover only the cases that do not occur that often in practice. In most cases the data contain a drift or a deterministic trend. It should be noted that especially the presence of deterministic trends changes the situation completely. I conjecture that in this case the Dickey–Fuller τ -test is admissible. The results of Müller and Elliot (2001) establish an optimality property. For a detailed analysis also see Elliot et al. (1996), Phillips and Xiao (1998), and Stock (1995). Nevertheless, I think that the results of this paper are important to the theory of testing for unit roots. It shows, in particular, that the situation is radically different from the classical one.

Let us now assume that a process y_t is defined by (4) and let us furthermore assume that the

$$u_t \text{ are i.i.d. Gaussian} \tag{20}$$

with zero expectation and that all of the roots of the polynomial

$$p(z) = 1 - \sum_1^p \pi_i z^i$$

lie outside the unit circle. Then let us consider local alternatives

$$\delta_n = \frac{C}{n}. \tag{21}$$

Jeganathan (1991, 1995) performed a rigorous asymptotic analysis for these kinds of testing problems. In particular, he established limiting properties of these experiments. Using his results, one can easily show that in many interesting cases the necessary conditions of our theorems are fulfilled. The “elementary” case of the Dickey–Fuller test without a trend is discussed in Ploberger (2004).

In many applications, however, one wants to avoid relatively narrow parametric assumptions such as (20). Standard statistical theory of optimal testing (cf. Strasser, 1985) is based on Neyman–Pearson tests, which critically depend on densities. Hence we cannot show the admissibility of tests. Under certain circumstances, however, we can show that certain tests are *not* admissible in the sense that there exists a test with a uniformly better power function.

Let us now assume that P_θ is a family of probability measures parametrized by a parameter θ . We assume that the parameter can be split into two parts,

$$\theta = \begin{pmatrix} \delta \\ \beta \end{pmatrix},$$

where δ should be one-dimensional, whereas the nuisance parameter β can be of arbitrary dimension. Suppose we want to test the hypothesis $\delta = 0$.

Suppose one has an estimator—say, $\hat{\delta}_n$ —for δ ; then one will be tempted to use an analogue to the usual t -test. One will construct some other statistic—say, $\hat{\sigma}_n^2$ —so that

$$\frac{\hat{\delta}_n}{\sqrt{\hat{\sigma}_n^2}} \text{ converges in distribution.}$$

Then it is easy to construct a test. In the “classical” case, where $\hat{\delta}_n$ is asymptotically normal, $\hat{\sigma}_n^2$ will be an appropriately scaled consistent estimator for the asymptotic variance. We are more interested in the nonstandard case, where the limiting distribution of the properly scaled $\hat{\sigma}_n^2$ is nontrivial.

In some cases, we can utilize our approach analyzing the nonoptimality of tests. In the previous section, we were mainly working with the scores and the “information” (the second derivative) of the likelihood function. Here the situation is different. We start with estimators for the parameter and their accuracy.

So let us review the basic assumption of the previous section. For doing so, we must find analogues to the W_n and A_n defined earlier. We started with the model (1):

$$\log p_n(\theta_0 + D_n^{-1}h) - \log p_n(\theta_0) \sim hW_n - \frac{1}{2} A_n h^2.$$

Then it is easily seen that—in case the logarithm of the likelihood is twice continuously differentiable—the maximum likelihood estimator $\hat{\theta}_n$ can be approximated in the following way:

$$D_n(\hat{\theta}_n - \theta_0) - A_n^{-1}W_n \rightarrow 0.$$

Moreover, we can easily conclude from our assumptions that $\sqrt{A_n^{-1}}W_n$ converges in distribution, and so $\hat{\sigma}_n^2 = A_n/D_n^2$ would be a feasible normalizing estimator.

In our situation, however, we *start* with the estimators $\hat{\delta}_n, \hat{\sigma}_n^2$. So let us assume that there exists an appropriate sequence of scaling factors D_n and *define* (keeping in mind that we want to test $\delta = \delta_0$)

$$A_n = D_n^2 \hat{\sigma}_n^2,$$

$$W_n = \hat{\sigma}_n(\hat{\delta}_n - \delta_0).$$

Then we assume that the following four assumptions are fulfilled.

Assumption A1. The nuisance parameter β is fixed.

Assumption A2. (A_n, W_n) converges (for $\theta = \begin{pmatrix} \delta_0 \\ \beta \end{pmatrix}$) in distribution to some random variable (A, W) with distribution Q_0 , which fulfills the prerequisites of the previous section.

Assumption A3. For local alternatives

$$\delta_n = \delta_0 + \frac{c}{D_n} \tag{22}$$

(i.e., $\theta_n = \begin{pmatrix} \delta_n \\ \beta \end{pmatrix}$), the distributions of (A_n, W_n) converge to a measure Q_c with

$$\frac{dQ_c}{dQ_0} = \exp\left(cW - \frac{c^2}{2} A\right).$$

Assumption A4. The mapping $c \rightarrow Q_c$ is continuous (with respect to the weak topology of measures). This means that $c_n \rightarrow c$ implies $Q_{c_n} \rightarrow Q_c$ weakly.

Essentially, we assume that the statistic (A_n, W_n) behaves “the same way” as in the case of a unidimensional parameter. We will now construct some examples that show that these assumptions will be fulfilled in many cases of considerable interest. In particular, it should be noted that we do not require parametric assumptions such as, e.g., Gaussian distributions of the error terms!

1. The generalized augmented Dickey–Fuller (ADF) test with no deterministic trend. Said and Dickey (1984) investigated the ADF test for unit roots for autoregressive moving average (ARMA) processes with independent and identically distributed (i.i.d.) innovations. Lately, in Chang and Park (2002) further generalizations of the model (4) are considered (e.g., u_t are martingale differences with a very general structure, and the index p may depend on the sample size and increase to infinity). So let us consider the model (4) and define

$$\beta = (\pi_1, \dots, \pi_p)'$$

They show that, even under the general circumstances considered in their paper (using our notation), letting $\hat{\delta}_n$ be the OLS estimator for δ and using an estimator \hat{A}_n , where \hat{A}_n^{-1} equals the denominator in the t -test (i.e., “variance estimator”), the properly normalized distributions of \hat{A}_n and the estimation error jointly converge in distribution to the corresponding distributions for the (simple) Dickey–Fuller test. This convergence holds for $\delta = 0$ and also for the local alternatives (22) (cf. Chang and Park, 2002). So let us define the “scores” by

$$\hat{W}_n = n\hat{\delta}_n \hat{A}_n$$

and denote by Q_c the limiting distribution of

$$(\hat{W}_n, \hat{A}_n).$$

The measures Q_c are defined on the set $\{(A, W) : A \geq 0\} \subset \mathbf{R}^2$. As mentioned before, the limiting distributions of our statistics are the same as for the Dickey–Fuller test. Hence we can conclude that the measures Q_c are the same as the asymptotic distributions for the simple random walk model, which was discussed in Ploberger (2004). So we may conclude that

$$\frac{dQ_c}{dQ_0} = \exp\left(cW - \frac{c^2}{2}A\right).$$

Moreover, it can easily be seen that other assumptions of our main theorem are fulfilled, too. Elementary calculations show that the two-sided ADF test rejects when

$$\frac{\hat{W}_n^2}{\hat{A}_n} = n^2 \frac{\hat{\delta}_n^2}{1/\hat{A}_n}$$

exceeds the critical value—say, U . Hence the power of the ADF test against a local alternative (22) will converge to

$$Q_c\left(\left\{\frac{W^2}{A} > U\right\}\right).$$

This only holds for all those U for which $Q_c(\{W^2/A = U\}) = 0$. Because, however, the Q_c have densities with respect to Q_0 , $Q_c(\{W^2/A = U\}) > 0$ implies that $Q_0(\{W^2/A = U\}) > 0$. But this inequality can only be true for countably many U . Hence we have convergence for all but at most countably many U .

2. Cointegrated system testing. Another possible application is the testing of restrictions on certain parameters of cointegrated systems. Let us first consider the simplest possible case, namely, a bivariate system in Phillips’ triangular form:

$$y_{1,t} = \gamma y_{2,t} + u_t, \tag{23}$$

$$\Delta y_{2,t} = v_t.$$

For simplicity, let us assume that u_t, v_t are i.i.d. standard normal with zero expectation and the covariance being the identity matrix (which implies that u_t, v_t are uncorrelated). Then we can estimate γ with OLS, and our estimation error will be

$$\hat{\gamma}_n - \gamma = \frac{\sum y_{2,t} u_t}{\sum y_{2,t}^2}.$$

Let us now suppose that we want to test $\gamma = \gamma_0$. Then some tedious but elementary calculations show that the asymptotic distribution of $n(\hat{\gamma}_n - \gamma_0)$ converges under the null to

$$\frac{\int_0^1 W_2 dW_1}{\int_0^1 W_1^2},$$

where W_1, W_2 are independent Wiener processes. Moreover, we can also consider local alternatives of the form

$$\gamma_n = \gamma_0 + \frac{c}{n}.$$

Then $n(\hat{\gamma}_n - \gamma_0)$ converges in distribution to

$$\frac{\int_0^1 W_2 dW_1}{\int_0^1 W_1^2} + c.$$

3. Some generalizations. We can generalize the ideas of the preceding example for use in a more general context. Suppose we want to test whether a parameter

$$\delta = \delta_0$$

and we have an estimator $\hat{\delta}_n$, an estimator for the “information” \hat{A}_n , and a sequence of scaling factors D_n , so that the following properties hold true.

For local alternatives

$$\delta_n = \delta_0 + \frac{c}{D_n} \tag{24}$$

the distributions of (\hat{W}_n, \hat{A}_n) , where $\hat{W}_T = D_n(\hat{\delta}_n - \delta_0)$, converge in distribution to a (W, A) with the following properties.

- The distribution of A is independent of c .
- The conditional distribution of W given A is normal with expectation cA and variance A .
- The distribution of A satisfies all the assumptions of our main theorem.

Many popular estimators for the cointegrating relationships have these asymptotically mixed normal properties: It is well established that the standard estimators for the cointegrating relationships are normal only after conditioning: cf. Ahn and Reinsel (1990), Phillips and Hansen (1990), Phillips and Ouliaris (1990), Saikkonen (1991), Johansen (1988, 1991), Stock and Watson (1993), and also Hamilton (1994, Sect. 19.3). Testing linear restrictions on these parameters with Wald-type tests (cf. Davidson, 1998) therefore will be covered by our theory, at least for one-dimensional restrictions.

A testing problem satisfying the preceding conditions is called locally asymptotically mixed normal (LAMN) (cf. Le Cam and Yang, 1990; Jeganathan, 1991, 1995). Apart from the economic applications quoted later, this type of model has also received much attention in the statistical literature (cf. the references cited previously).

4. Other cases. Asymptotically mixed normal families do not only occur in connection with unit root tests or cointegration. Other cases of interest in economics are given in Park and Phillips (2001) and Ait-Sahalia (1999).

So let us now assume that we have given a testing problem satisfying the preceding assumptions. We might be tempted to use the generalized t -test φ_T

for our testing problem. This means we reject when the absolute value or, *equivalently*, the *square* of

$$\frac{\hat{W}_n}{\sqrt{\hat{A}_n}} = D_n A_n (\delta_n - \delta_0)$$

is larger than the critical value. Furthermore, let us assume that this critical value has an absolute value larger than one, i.e.,

$$U > 1. \tag{25}$$

Let us additionally assume that the cumulative distribution function of the limiting distribution of our test statistic is continuous at U . This is equivalent to

$$Q_0 \left[\left(\frac{W}{\sqrt{A}} \right)^2 = U \right] = 0. \tag{26}$$

Hence the cumulative distribution function of the limiting distribution of our test statistic is continuous in U .

THEOREM 3. *Suppose Assumptions A1–A4 and also (25) and (26) are satisfied. Then the tests*

$$\varphi_n = I \left[\frac{\hat{W}_n^2}{\hat{A}_n} > U \right]$$

defined earlier are inadmissible. There exist tests

$$\psi_n = \psi_n(\hat{W}_n, \hat{A}_n)$$

depending on W_n and A_n alone that have a strictly better asymptotic power function than φ_n . For alternatives (24) we have

$$\limsup \int \varphi_n dP(\delta_n) \leq \liminf \int \psi_n dP(\delta_n) \quad \text{if } c \neq 0, \tag{27}$$

where for at least one $c \neq 0$ the left-hand side is smaller than (and not equal to) the right-hand side. Moreover, we also have

$$\liminf \int \varphi_n dP(\delta_n) \geq \limsup \int \psi_n dP(\delta_n) \quad \text{if } c = 0. \tag{28}$$

Remark 3. In many cases, it is not that surprising that the tests φ_n are inadmissible. As an example take, e.g., one of the preceding models with non-Gaussian error terms u_t, v_t , when their distribution is known up to a finite-dimensional parameter. Then estimators based on parametric models for u_t, v_t should be more accurate, and hence the corresponding tests would yield more power. How-

ever, it is interesting to observe that there exists a test better than φ_n depending on (W_n, A_n) alone! Therefore there exists some way of computing a test statistic with (W_n, A_n) other than the t -test statistic, which defines a test with a uniformly better asymptotic power function!

Remark 4. Unfortunately, the proof is nonconstructive. I only assert the existence of such a test. The proof does not provide a way to construct this dominating test.

Remark 5. Here we will prove pointwise convergence in (27) and (28). With some technical refinements, the proof can be modified to show that the convergence in (27) is uniform on bounded sets of C .

Proof of Theorem 3. The proof can be found in the Appendix.

REFERENCES

- Aït-Sahalia, Y. (1999) Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approach. Discussion paper, Princeton University.
- Ahn, S.K. & G.C. Reinsel (1990) Estimation for partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association* 85, 813–823.
- Chang, Y. & J. Park (2002) On the asymptotics of ADF tests for unit roots. *Econometric Reviews* 21, 432–477.
- Davidson, J. (1998) A Wald test of restrictions on the cointegrating space based on Johansen's estimator. *Economics Letters* 59, 183–187.
- Dickey, D.A. & W.A. Fuller (1979) Distribution of estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.
- Dickey, D.A. & W.A. Fuller (1981) Likelihood ratio tests for autoregressive series with a unit root. *Econometrica* 49, 1057–1072.
- Elliot, G., T.J. Rothenberg, & J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.
- Hamilton, J.D. (1994) *Time Series Analysis*. Princeton University Press.
- Jeganathan, P. (1991) Some aspects of asymptotic theory with application to time series models. *Econometric Theory* 7, 269–306.
- Jeganathan, P. (1995) Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* 11, 818–867.
- Johansen, S. (1988) Statistical analysis of cointegrating vectors. *Journal of Economics Dynamics and Control* 12, 231–254.
- Johansen, S. (1991) Cointegration and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.
- Le Cam, L. & G.L. Yang (1990) *Asymptotics in Statistics: Some Basic Concepts*. Springer-Verlag.
- Müller, U. & G. Elliot (2001) Tests for Unit Root and the Initial Observation. Discussion paper, University of St. Gallen.
- Park, J. & P.C.B. Phillips (2001) Nonlinear regressions with integrated time series. *Econometrics* 69, 117–161.
- Phillips, P.C.B. (1996) Econometric model determination. *Econometrica* 64, 763–812.
- Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99–125.
- Phillips, P.C.B. & S. Ouliaris (1990) Asymptotic properties of residual based tests for cointegration. *Econometrica* 58, 165–193.

- Phillips, P.C.B. & W. Ploberger (1996) An asymptotic theory of Bayesian inference for time series. *Econometrica* 64, 381–412.
- Phillips, P.C.B. & Z. Xiao (1998) A Primer on Unit Root Testing. Cowles Foundation Discussion paper, Yale University.
- Ploberger, W. (2004) A complete class of tests when the likelihood is locally asymptotically quadratic. *Journal of Econometrics* 118, 67–94.
- Ploberger, W. & P.C.B. Phillips (2003) Empirical limits for time series models. *Econometrica* 71, 627–674.
- Rudin, W. (1974) *Real and Complex Analysis*, 2nd ed. Reprinted by Tata-McGraw-Hill.
- Said, S.E. & D.A. Dickey (1984) Testing for unit roots in ARMA models of unknown order. *Biometrika* 71, 599–607.
- Saikkonen, P. (1991) Asymptotically efficient estimation of cointegrating regressions. *Econometric Theory* 7, 1–21.
- Schervish, M.J. (1995) *Theory of Statistics*. Springer-Verlag.
- Stock, J.H. (1995) Unit roots, structural breaks and trends. In R.F. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2739–2841. North-Holland.
- Stock, J.H. & M. Watson (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 51, 783–820.
- Strasser, H. (1985) *Mathematical Theory of Statistics*. De Gruyter.

APPENDIX

Proof of Theorem 1. We give an indirect proof. Let us assume that the test φ is from our class and has the same (or worse) power function as ψ .

Let us first deal with the extreme cases. So let us assume that the parameters describing the test φ are zero except for U . Then U must be nonzero (because we did assume that not all of the parameters vanish), and we can conclude that φ must be a trivial test, either accepting for all (A, W) if U is positive or rejecting in every case if U is negative. Hence we can easily see that—as ψ must have the same or better power function as φ — ψ must be trivial, too.

PROPOSITION 1. *Let us assume that the parameter a for our test φ is nonzero. Then the test $\psi = 1$ on the set $\{(A, W) : A < a\}$ Q_0 almost surely.*

For the proof of our proposition the following lemma is helpful.

LEMMA 1. *There exist $\kappa_1, C_1 > 0$, so that the test φ rejects on the set*

$$\{(A, W) : W > \kappa_1 A + C_1, A > a\},$$

or there exist $\kappa_2, C_2 < 0$, so that the test φ rejects on the set

$$\{(A, W) : W < \kappa_2 A + C_2, A > a\}.$$

Let us now prove the lemma. First, let us look at the trivial cases. If $\mu + \nu = 0$, the lemma is trivial. Let us now assume that $\mu + \nu$ is nontrivial. Then, at least one of the

sets $\{\lambda : \lambda > 0\}$ and $\{\lambda : \lambda < 0\}$ is not a $\mu + \nu$ null set. Let us assume without limitation in generality that

$$(\mu + \nu)(\{\lambda : \lambda > 0\}) > 0.$$

If this is true, then there exist $\alpha, \beta > 0$ such that

$$(\mu + \nu)(\{\lambda : \beta > \lambda > 2\alpha\}) > 0. \tag{A.1}$$

Consequently using (13), we may write our criterion function u as

$$\begin{aligned} u(A, W) = & b'W + \frac{c'}{2} (W^2 - A) \\ & + \int_{[-\alpha, \alpha]} \frac{1}{\lambda^2} \left\{ \int_0^\lambda \int_0^w \int_0^v \exp\left(zW - \frac{z^2}{2} A\right) \right. \\ & \quad \left. \times (W - zA)[(W - zA)^2 - 3A] dz dv dw \right\} (\lambda^2 d\nu'(\lambda)) \\ & + \int_{\mathbf{R}_{-[-\alpha, \alpha]}} \exp\left(\lambda\left(W - \frac{\lambda}{2} A\right)\right) d\mu'(\lambda). \tag{A.2} \end{aligned}$$

With

$$(\mu + \nu)(\{\lambda : \beta > \lambda > 2\alpha\}) = \mu'(\{\lambda : \beta > \lambda > 2\alpha\}) > 0. \tag{A.3}$$

So let us define

$$\kappa_1 = \max(\beta, \alpha + 3)$$

and let us define C_1 later on. We will, however, without limitation of generality, assume that

$$C_1 > 1.$$

We now want to show that our test rejects if

$$W > \kappa_1 A + C_1. \tag{A.4}$$

To do so, we have to show that for all values of (W, A) satisfying (A.4) $u(A, W)$ becomes bigger than the critical value. Hence, we have to find lower bounds for all the terms (A.2). Subsequently we will show that all terms—except the last one—are bounded from below by polynomials in $W + A$, at least for the values of (W, A) that are of interest to us.

Because (A.4) implies that $W > 0$ and $W > A$ we have

$$bW > -|b'| (W + A),$$

and, as $c' \geq 0$,

$$\frac{c'}{2} (W^2 - A) \geq 0.$$

Now observe that for $W - \kappa_1 A > 1$ for z with $z < \alpha$

$$W - zA > 1 > 0,$$

and, because of our definition of κ_1 ,

$$(W - zA)^2 - 3A > 0.$$

To analyze the first integral in (A.2) let us distinguish two cases: $\lambda < 0$ and $\lambda \geq 0$. For the first case, observe that for $z < 0, W > 0$

$$zW - \frac{z^2}{2} A < 0.$$

Hence, for $-\alpha \leq \lambda < 0$

$$\begin{aligned} & \frac{1}{\lambda^2} \left\{ \int_0^\lambda \int_0^w \int_0^v \exp\left(zW - \frac{z^2}{2} A\right) (W - zA) [(W - zA)^2 - 3A] dz dv dw \right\} \\ & \geq -(W + \alpha A) [(W + \alpha A)^2 + 3A] \alpha / 6 \\ & \geq \max(\alpha, 1)^4 \{ (W + A)^3 + (W + A)^2 \}. \end{aligned}$$

On the other hand, for $W > 0$ and $\lambda > 0$

$$\frac{1}{\lambda^2} \left\{ \int_0^\lambda \int_0^w \int_0^v \exp\left(zW - \frac{z^2}{2} A\right) (W - zA) [(W - zA)^2 - 3A] dz dv dw \right\} > 0,$$

because the expression inside the integral is nonnegative and all the bounds of the integrals are nonnegative. Because the integrand in the last term on the right-hand side of (A.2) is nonnegative, we have for $W > 0$

$$u(A, W) \geq p(W + A) + \int_{(2\alpha, \beta)} \exp\left(\lambda \left(W - \frac{\lambda}{2} A\right)\right) d\mu'(\lambda),$$

where $p(\cdot)$ is a polynomial (of third order) representing our estimates for the first three terms of (A.2). Let us fix the constant $C_1 > 0$ later on. For any nonnegative C_1 , if

$W > \kappa_1 A + C_1$, then for $\lambda \in (2\alpha, \beta)$, $W - (\lambda/2)A \geq (\beta/2)A$, and hence $W + A = W - (\lambda/2)A + (1 + \lambda/2)A \leq (W - (\lambda/2)A)(1 + (2/\beta)(1 + \lambda/2))$. Hence, for all $W > \kappa_1 A + C_1$

$$\left(W - \frac{\lambda}{2}A\right) \geq (W + A)\gamma,$$

with $\gamma = 1/(1 + (2/\beta)(1 + \kappa_1/2))$ and

$$u(A, W) \geq p(W + A) + \mu'(\{\lambda : \beta > \lambda > \alpha\})\exp(\alpha\gamma(W + A)).$$

If $W > \kappa_1 A + C_1$, then obviously $W + A > C_1$. Because an exponential function grows faster than any polynomial, we can find a C_1 such that for all $W + A > C_1$ the right-hand side of the preceding inequality is strictly bigger than U —our critical value. Hence, our test φ will reject.

Now we can proceed to prove Proposition 1. Let us assume that the proposition does not hold true and $\psi = 0$ on the set $B \subset \{(A, W) : A < a\}$ with $Q_0(B) > 0$. Because with

$$B_n = B \cap \left\{ (A, W) : A < a - \frac{1}{n}, |W| < n \right\}$$

$B_n \uparrow B$ and consequently $Q_0(B_n) \uparrow Q_0(B) > 0$, we can find an n so that

$$Q_0(B_n) > 0.$$

Now let us consider the power functions of our tests φ, ψ for probability measures Q_λ , where we choose λ depending on the behavior of the test φ . Lemma 1 guarantees that the test $\varphi = 1$ on the set $\{(A, W) : W > \kappa_1 A + C_1\}$ (in this case choose $\lambda \rightarrow \infty$) or $\varphi = 1$ on the set $\{(A, W) : W < \kappa_2 A + C_2\}$ (in this case choose $\lambda \rightarrow -\infty$). We will discuss the first case only, because the second one is perfectly analogous. As the power function of ψ is not worse than the one of φ for all values of the alternative, we have

$$\int (\varphi - \psi) dQ_\lambda \leq 0. \tag{A.5}$$

The fact that $\varphi = 1$ on B_n implies that

$$\int (\varphi - \psi) dQ_\lambda \geq \int_{B_n} dQ_\lambda - \int_{\{(A, W) : A \geq a, W \leq \kappa_1 A + C_1\}} dQ_\lambda. \tag{A.6}$$

We can easily see that for $\lambda \rightarrow \infty$

$$\inf_{(A, W) \in B_n} \frac{\exp\left(\lambda W - \frac{\lambda^2}{2}A\right)}{\exp\left(-\frac{\lambda^2}{2}\left(a - \frac{1}{2n}\right)\right)} \rightarrow \infty$$

and hence

$$\exp\left(\frac{\lambda^2}{2}\left(a - \frac{1}{2n}\right)\right) \int_{B_n} dQ_\lambda \rightarrow \infty.$$

On the other hand

$$\begin{aligned} & \exp\left(\frac{\lambda^2}{2}\left(a - \frac{1}{2n}\right)\right) \int_{\{(A,W):A \geq a, W \leq \kappa_1 A + C_1\}} dQ_\lambda \\ &= \exp\left(\frac{\lambda^2}{2}\left(a - \frac{1}{2n}\right)\right) \int_{\{(A,W):A \geq a, W \leq \kappa_1 A + C_1\}} \exp\left(\lambda W - \frac{\lambda^2}{2} A\right) dQ_0 \\ &\leq \exp\left(\frac{\lambda^2}{2}\left(a - \frac{1}{2n}\right)\right) \int_{\{(A,W):A \geq a, W \leq \kappa_1 A + C_1\}} \exp\left(\lambda(\kappa_1 A + C_1) - \frac{\lambda^2}{2} A\right) dQ_0 \\ &= \exp\left(-\frac{\lambda^2}{2}\left(\frac{1}{2n}\right) + C_1 \lambda + \lambda \kappa_1 a + \frac{\kappa_1^2}{2}\right) \\ &\quad \times \int_{\{(A,W):A \geq a, W \leq \kappa_1 A + C_1\}} \exp\left(-\frac{\kappa_1^2}{2} + \lambda\left(\kappa_1(A - a) - \frac{\lambda^2}{2}(A - a)\right)\right) dQ_0. \end{aligned}$$

Now we can easily see that $\exp(-(\lambda^2/2)(1/2n) + C_1 \lambda + \lambda \kappa_1 a + (\kappa_1^2/2)) \rightarrow 0$ for $\lambda \rightarrow \infty$. Because $\exp(-(\kappa_1^2/2) + \lambda(\kappa_1(A - a) - (\lambda^2/2)(A - a))) \leq 1$, one can conclude that the preceding estimates guarantee that

$$\exp\left(\frac{\lambda^2}{2}\left(a - \frac{1}{2n}\right)\right) \int_{\{(A,W):A \geq a, W \leq \kappa_1 A + C_1\}} dQ_\lambda \rightarrow 0.$$

Therefore (A.6) implies that

$$\exp\left(\frac{\lambda^2}{2}\left(a - \frac{1}{2n}\right)\right) \int (\varphi - \psi) dQ_\lambda \rightarrow \infty,$$

which contradicts (A.5). Hence our assumption that there exists a nontrivial set B must be wrong, which is just the proposition we wanted to prove.

Now that our main tools are established, we can proceed with the proof of the theorem. We will show that $\varphi = \psi$. So let us start with an extreme case, namely, that

$$\varphi = I(A < a).$$

Then the proposition we just proved shows that ψ rejects on $(A < a)$, too. Hence, we may conclude that $\varphi \leq \psi$. Moreover, the power function of ψ is better than the power function of φ on the null; hence $\int(\psi - \varphi) dQ_0 \leq 0$. Therefore we can conclude that $\psi = \varphi Q_0$ almost surely.

Next we investigate the general case, namely, where at least one of the parameters b, c, μ, ν is nontrivial. For this purpose we introduce the following notation. For each test ρ let

$$P(\rho)(\lambda) = \int \rho \exp\left(\lambda W - \frac{\lambda^2}{2} A\right) dQ_0$$

and define the functional L_n by

$$\begin{aligned} L_n(\rho) &= bP(\rho)'(0) + \frac{c}{2} P(\rho)''(0) \\ &+ \int_{[-1,1]} \left(\int_0^\lambda \int_0^w \int_0^v (P(\rho)'''(z)) dz dv dw \right) d\nu(\lambda) \\ &+ \int_{[-n,-1) \cup (-1,n)} P(\rho)(\lambda) d\mu(\lambda). \end{aligned}$$

It is easily seen that our assumptions (14) and (15) guarantee that all power functions of tests are three times differentiable. Hence our functionals are well defined. The first integral on the right-hand side of the preceding definition is finite because of the definition of the integrand ($o(\lambda^2)$) and (7). Moreover, because our functionals only depend on the power functions, we have

$$L_n(\varphi) = L_n(\psi).$$

Moreover, we can easily see that with

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

and

$$\begin{aligned} u_n(A, W) &= bW + \frac{c}{2} (W^2 - A) \\ &+ \int_{[-1,1]} \left(\exp\left(\lambda W - \frac{\lambda^2}{2} A\right) - \left(1 + \lambda W + \frac{\lambda^2}{2} (W^2 - A)\right) \right) d\nu(\lambda) \\ &+ \int_{[-n,n] - [-1,1]} \exp\left(\lambda W - \frac{\lambda^2}{2} A\right) d\mu(\lambda) \end{aligned}$$

then

$$L_n(\rho) = \int \rho u_n(A, W) dQ_0$$

and

$$u_n \uparrow u.$$

Hence,

$$\begin{aligned} 0 &= L_n(\varphi) - L_n(\psi) - U \int (\varphi - \psi) dQ_0 \\ &= \iint (\varphi - \psi)(u_n - U) dQ_0 \\ &= \int_{[u < U]} (\varphi - \psi)(u_n - U) dQ_0 + \int_{[u \geq U]} (\varphi - \psi)(u_n - U) dQ_0. \end{aligned} \tag{A.7}$$

We will now analyze the behavior for $n \rightarrow \infty$.

Let us first analyze $\int_{[u < U]} (\varphi - \psi)(u_n - U) dQ_0 = \int_{[u < U]} \psi(U - u_n) dQ_0$. We can apply the dominated convergence theorem ($U - u_n$ is monotonically *falling* and on the event $[u < U]$ between 0 and U). We therefore have

$$\int_{[u < U]} (\varphi - \psi)(u_n - U) dQ_0 \rightarrow \int_{[u < U]} \psi(U - u) dQ_0 \geq 0. \tag{A.8}$$

For the second term, observe that

$$\begin{aligned} &\int_{[u \geq U]} (\varphi - \psi)(u_n - U) dQ_0 \\ &= \int_{[u \geq U]} c(A, W)(1 - \psi) dQ_0 \\ &\quad + \int_{[u \geq U]} (1 - \psi) \int_{[-n, n] - [-1, 1]} \exp\left(\lambda W - \frac{\lambda^2}{2} A\right) d\mu(\lambda) dQ_0, \end{aligned} \tag{A.9}$$

with

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

and $c(A, W)$ does not depend on n . Moreover, assumptions (14) and (15) for the integral in conjunction with Remark 2 guarantee the absolute integrability of c with respect to Q_0 . For the second term on the right-hand side of (A.9) we may use the monotone convergence theorem and conclude that

$$\int_{[u \geq U]} (1 - \psi) \int_{[-n, n] - [-1, 1]} \exp\left(\lambda W - \frac{\lambda^2}{2} A\right) d\mu(\lambda) dQ_0$$

converges to

$$\int_{[u \geq U]} (1 - \psi) \int_{[-\infty, \infty] - [-1, 1]} \exp\left(\lambda W - \frac{\lambda^2}{2} A\right) d\mu(\lambda) dQ_0$$

(or ∞ , if the integral is not finite). Hence, we may conclude that

$$\int_{[u \geq U]} (\varphi - \psi)(u_n - U) dQ_0 \rightarrow \int_{[u \geq U]} (1 - \psi)(u - U) dQ_0 \geq 0,$$

where again the integral on the right-hand side may be infinite (as the integrand is non-negative, this is not a serious problem). Because in (A.8) the limit is finite, we may add (A.8) to the preceding equation and conclude that

$$\begin{aligned} & \int_{[u < U]} (\varphi - \psi)(u_n - U) dQ_0 + \int_{[u \geq U]} (\varphi - \psi)(u_n - U) dQ_0 \\ & \rightarrow \int_{[u < U]} \psi(U - u) dQ_0 + \int_{[u \geq U]} (1 - \psi)(u - U) dQ_0. \end{aligned}$$

However equation (A.7) guarantees that all the terms of the preceding sequence are zero; hence the limit has to be zero, too. Therefore

$$\int_{[u < U]} \psi(U - u) dQ_0 + \int_{[u \geq U]} (1 - \psi)(u - U) dQ_0 = 0,$$

which implies that $\psi = 0$ on $[u < U]$ and $\psi = 1$ on $[u > U]$. So ψ and φ can differ only on the set $[u = U]$, which was just the result we wanted to establish.

Proof of Theorem 3. Assumption A3 states that the distributions of (W_T, A_T) under P_{δ_T} converge weakly to Q_c . Hence,

$$\begin{aligned} \lim \int \varphi_T dP(\delta_T) &= \lim P_{\delta_T}\{(W, A) : W^2/A > U\} \\ &= Q_c\{(W, A) : W^2/A > U\}, \end{aligned}$$

because we did assume that the measure of the boundary of the set $\{(W, A) : W^2/A > U\}$ has Q_0 -measure, and hence Q_c -measure 0. Hence the limiting power function of the test φ_T is the power function of the test $\varphi = I\{(W, A) : W^2/A > U\}$ with respect to the measures Q_c . Ploberger (2004) establishes that this test is not within the complete class described in the previous section, and hence our main theorem establishes that this test cannot be admissible: Hence there must exist a test $\psi = \psi(A, W)$, such that

$$\begin{aligned} \int \varphi dQ_c &\leq \int \psi dQ_c \quad \text{for all } c \neq 0, \\ \int \varphi dQ_0 &= \int \psi dQ_0, \end{aligned} \tag{A.10}$$

and for at least one c the left-hand side in (A.10) is strictly smaller than the right-hand side. To prove our theorem, we have to establish the existence of tests $\psi_n = \psi_n(A_n, W_n)$, such that

$$\lim \int \psi_n dP(\delta_n) = \int \psi dQ_c$$

for all c . The test ψ is a measurable function of \mathbf{R}^2 to the interval $[0, 1]$. Hence, Lusin's theorem (cf. Rudin, 1974, p. 56) guarantees the existence of continuous functions ϕ_n , such that

$$Q_0([\psi \neq \phi_n]) < 1/n.$$

Because $0 \leq \psi \leq 1$, without limitation of generality we can assume that $0 \leq \phi_n \leq 1$, too (otherwise replace ϕ_n by $\max(0, \min(\phi_n, 1))$). Hence, we can interpret $\phi_n(A_n, W_n)$ as tests. By applying Chebyshev's inequality, we can easily see that for every $\epsilon > 0$ and all M there exists a $K = K(\epsilon)$, such that $\sup_{|c| \leq M} Q_c([|W| > K]) < \epsilon$. Because for $|W| \leq K, |c| \leq M \exp(cW - c^2A/2) \leq \exp(KM)$, we have for arbitrary $\epsilon > 0$ and arbitrary M

$$\begin{aligned} Q_c([\psi \neq \phi_n]) &\leq Q_c([|W| > K(\epsilon)]) + \int_{[|W| \leq K(\epsilon)] \cap [\psi \neq \phi_n]} \exp(cW - c^2A/2) dQ_0 \\ &\leq \epsilon + \exp(KM)Q_0([\psi \neq \phi_n]) < \epsilon + \exp(KM)/n, \end{aligned}$$

and therefore (as $|\int \psi dQ_c - \int \phi_n dQ_c| \leq Q_c([\psi \neq \phi_n])$)

$$\lim_{n \rightarrow \infty} \sup_{|c| \leq M} \left| \int \psi dQ_c - \int \phi_n dQ_c \right| = 0. \tag{A.11}$$

On the other hand, ϕ_n are continuous functions. So our fourth assumption and Prohorov's theorem guarantee the relative compactness of the set of distributions of the (W_n, A_n) under $P_n(\delta_n)$, where $\delta_n = D_n^{-1}c + \delta_0$ and $|c| \leq M$. Moreover, we did assume that the distribution of (W_n, A_n) under $P_n(\delta_n)$ converges (for all c) to the distributions to the Q_c ; hence it is easy to see that the set containing the

$$P_n(\delta_n)(W_n, A_n)^{-1}$$

and the Q_c , both for $|c| \leq M$, is compact. Therefore, it can easily be seen that

$$\lim_{T \rightarrow \infty} \sup_{|c| \leq M} \left| \int \phi_n(W_n, A_n) dP_n(\delta_n) - \int \phi_n dQ_c \right| = 0.$$

This limiting relation and (A.11) together imply that there exist sequences $M_n \uparrow \infty, m_n \rightarrow \infty$ such that

$$\lim_{T \rightarrow \infty} \sup_{|c| \leq M_n} \left| \int \phi_{m_n}(W_n, A_n) dP_n(\delta_n) - \int \psi dQ_c \right| = 0,$$

which is just the result we wanted to show.