Addendum to:

A Consistent Conditional Moment Test
of Functional Form

1. Introduction

In this addendum to Bierens (1990) [B90 hereafter] I will show that the tests in Theorems 4 and 5 have non-trivial power against $\sqrt{n}$-local alternatives. Also, I will derive a consistent Hausman-type test, based on an earlier unpublished paper of mine.

The paper under review is a follow-up on my earlier paper Bierens (1982) on consistent testing of the null hypothesis that the functional form of a (non-)linear regression model is correctly specified as a conditional expectation, against all deviations from the null hypothesis. In the latter paper I was unable to derive the asymptotic null distribution of the Integrated Conditional Moment (ICM) test involved. This problem has puzzled me until I came across Billingsley’s (1968) book on convergence of probability measures on metric spaces. This book taught me that a sequence of continuous random functions $bZ_n(t)$ on a compact subset $T$ of a Euclidean space converges weakly to a random element $Z(t)$ of the space $C(T)$ of continuous functions on $T$, endowed with the ”sup” metric, if and only if $bZ_n$ is tight and its finite distributions converge to the corresponding finite distributions of $Z$. See the addendum to Bierens (1982) in Chapter 2 for a discussion of these concepts.

In the present case the random function $\hat{Z}_n(t)$ (without standardization) takes the form

$$\hat{Z}_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \hat{U}_j \exp(t'\Phi(X_j)),$$

(1.1)

where $\hat{U}_j$ is the NLLS residual of the nonlinear regression model

$$Y_j = f(X_j, \theta_0) + U_j, \quad X_j \in \mathbb{R}^k, \quad \theta_0 \in \Theta \subset \mathbb{R}^m,$$

(1.2)

with $\Theta$ is a compact parameter space. As in Bierens (1982) and B90, $\{(Y_j, X_j)\}_{j=1}^{n}$ is a random sample from $(Y, X) \in \mathbb{R} \times \mathbb{R}^k$ with $E[Y^2] < \infty$, and $\Phi : \mathbb{R}^k \to \mathbb{R}^k$ is a bounded one-to-one mapping such that $X$ and $\Phi(X)$ generate the same $\sigma$-algebra.
A difference with Bierens (1982) is that instead of the complex \( \exp(.) \) function
the real one is used.

The null hypothesis that the functional form of this model is correct can be
formalized as

\[ H_0 : \text{There exists a } \theta_0 \in \Theta \text{ such that } E[Y|X] = f(X, \theta_0) \text{ a.s.} \]

which is equivalent to \( E[U_j|X_j] = 0 \text{ a.s. in } (1.2) \). Denoting

\[ Z_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_j \phi_j(t) \]

where

\[ \phi_j(t) = \exp(t' \Phi(X_j)) - b(t)' A^{-1} \nabla f(X_j, \theta_0) \]

with \( b(t) \) and \( A \) defined in the paper and

\[ \nabla f(x, \theta) = (\partial/\partial \theta') f(x, \theta), \]

it follows that under \( H_0 \) and standard NLLS conditions (c.f. Assumption A in
B90),

\[ \sup_{t \in T} \left| \tilde{Z}_n(t) - Z_n(t) \right| = o_p(1) \quad (1.3) \]

for any compact subset \( T \) of \( \mathbb{R}^k \). Moreover, similar to Theorem 8.3 in
Billingsley (1968) it follows that under \( H_0 \) and standard NLLS conditions, \( Z_n \)
converges weakly to a zero mean real-valued Gaussian process \( Z(t) \) on \( T \), denoted by

\[ Z_n \Rightarrow Z, \quad (1.4) \]

with covariance function

\[ \Gamma(t_1, t_2) = E[Z(t_1) Z(t_2)] = E[U_1^2 \phi_1(t_1) \phi_1(t_2)] \]

and variance function

\[ s^2(t) = \Gamma(t, t) = E[U_1^2 \phi_1(t)^2] \quad (1.5) \]

Consequently, for each \( t \in T \), \( Z_n(t) \xrightarrow{d} N[0, s^2(t)] \), hence

\[ W_n(t) = Z_n(t)^2 / s^2(t) \xrightarrow{d} \chi^2_1, \]
provided that \( s^2(t) > 0 \). As to the latter condition, it is shown in Lemma 2 in B90 that under Assumption B the set
\[
S_* = \{ t \in \mathbb{R}^k : s^2(t) = 0 \}
\]  
has Lebesgue measure zero and is nowhere dense. This result holds regardless whether \( H_0 \) is true or not, because in general, \( U = Y - f(X, \theta_0) \), with
\[
\theta_0 = \arg \min_{\theta \in \Theta} E \left[ (Y - f(X, \theta))^2 \right].
\]  
Thus, \( \Pr (E[U|X] = 0) = 1 \) under \( H_0 \), whereas \( \Pr (E[U|X] = 0) < 1 \) if \( H_0 \) is not true.

Due to (1.3) and (1.4),
\[
\hat{Z}_n \Rightarrow Z \text{ on } T
\]  
as well. Therefore, for any compact subset \( T \) of \( \mathbb{R}^k \) and any \( t \in T \setminus S_* \),
\[
\hat{W}_n(t) = \hat{Z}_n(t)^2 / \hat{s}_n^2(t) \overset{d}{\rightarrow} \chi^2_1,
\]  
where \( \hat{s}_n^2(t) \) is defined in equation (14) in B90. This result carries over for all fixed \( t \in \mathbb{R}^k \setminus S_* \), because for each \( t_0 \in \mathbb{R}^k \) we can choose \( T \) such that \( t_0 \in T \).

Under the alternative hypothesis that \( H_0 \) is false, i.e.,
\[
H_1 : \text{For all } \theta \in \Theta, \Pr (E[Y_j|X_j] = f(X_j, \theta)) < 1,
\]  
it is shown in B90, similar to Bierens (1982), that, with
\[
\eta_*(t) = (E [(Y - f(X, \theta_0)) \exp(t' \Phi(X_j))] )^2,
\]  
where \( \theta_0 \) is defined by (1.7), we have
\[
\hat{Z}_n(t)^2 / n \overset{p}{\rightarrow} \eta_*(t)
\]  
uniformly on any compact subset \( T \) of \( \mathbb{R}^k \), and pointwise in any \( t \in \mathbb{R}^k \), where the set
\[
S = \{ t \in \mathbb{R}^k : \eta_*(t) = 0 \}
\]  
has Lebesgue measure zero and is nowhere dense, whereas under \( H_0, S = \mathbb{R}^k \). Since under \( H_1, \hat{s}_n^2(t) \overset{p}{\rightarrow} s^2(t) \) as well, because under standard conditions the NLLS estimator \( \hat{\theta} \) converges in probability to \( \theta_0 \) defined by (1.7) regardless whether \( H_0 \) is true or not, it follows that under \( H_1 \) we have
\[
\hat{W}_n(t)/n \overset{p}{\rightarrow} \eta_*(t)/s^2(t) = \eta(t) > 0 \text{ for all } t \in \mathbb{R}^k \setminus (S_* \cup S).
\]  

2. Randomization

Once I had figured out the weak convergence result (1.8), I wrote the first draft of the paper in a weekend, based on the following randomization trick, and submitted it to *Econometrica*.

**Lemma 2.1.** Let \( \tilde{t} \) be a random drawing from an absolutely continuous distribution on \( \mathbb{R}^k \). Then under the conditions in B90,

\[
\hat{W}_n(\tilde{t}) \overset{d}{\to} \chi_1^2 \text{ under } H_0 \text{ and } \hat{W}_n(\tilde{t}) \overset{p}{\to} \infty \text{ under } H_1.
\]

**Proof.** Let \( \mu \) be the probability measure induced by the absolutely continuous distribution involved, and let \( \varphi_{\chi_1^2}(\xi) \) be the characteristic function of the \( \chi_1^2 \) distribution.\(^1\) Then by (1.9) and the bounded convergence theorem,

\[
E \left[ \exp \left( i \xi \hat{W}_n(t) \right) \right] = \int_{\mathbb{R}^k} E \left[ \exp \left( i \xi \hat{W}_n(t) \right) \right] d\mu(t) = \int_{\mathbb{R}^k \setminus S_*} E \left[ \exp \left( i \xi \hat{W}_n(t) \right) \right] d\mu(t) \to \int_{\mathbb{R}^k \setminus S_*} \lim_{n \to \infty} E \left[ \exp \left( i \xi \hat{W}_n(t) \right) \right] d\mu(t) = \int_{\mathbb{R}^k \setminus S_*} \varphi_{\chi_1^2}(\xi) d\mu(t) = \varphi_{\chi_1^2}(\xi)
\]

pointwise in \( \xi \in \mathbb{R} \), hence \( \hat{W}_n(\tilde{t}) \overset{d}{\to} \chi_1^2 \) under \( H_0 \).

Under \( H_1 \) it follows from (1.11) that for arbitrary \( K > 0, M > 0 \) and \( j \in \mathbb{N} \),

\[
\lim_{n \to \infty} \inf \Pr \left[ \hat{W}_n(\tilde{t}) > M \right] = \lim_{n \to \infty} \inf \Pr \left[ \hat{W}_n(\tilde{t})/n > M/n \right] \geq \lim_{n \to \infty} \inf \Pr \left[ \hat{W}_n(\tilde{t})/n > 1/j \right] = \lim_{n \to \infty} \inf \int_{\mathbb{R}^k} \Pr \left[ \hat{W}_n(t)/n > 1/j \right] d\mu(t) = \lim_{n \to \infty} \inf \int_{\mathbb{R}^k \setminus (S_1 \cup S)} \Pr \left[ \hat{W}_n(t)/n > 1/j \right] d\mu(t)
\]

\(^1\)Note that \( \varphi_{\chi_1^2}(\xi) = (1 - 2i\xi)^{-1/2} \), where here and in the sequel \( i = \sqrt{-1} \).
\[
\geq \int_{\mathbb{R}^k \setminus (S_* \cup S)} \liminf_{n \to \infty} \Pr \left[ \frac{\widehat{W}_n(t)}{n} > 1/j \right] d\mu(t) \\
= \int_{\mathbb{R}^k \setminus (S_* \cup S)} I(\eta(t) > 1/j) d\mu(t) \\
= \int_{\mathbb{R}^k \setminus (S_* \cup S)} I(\eta_*(t) > (1/j)s^2(t)) d\mu(t) \\
\geq \int_{\mathbb{R}^k \setminus (S_* \cup S)} I(\eta_*(t) > (1/j)s^2(t)) I(s^2(t) < K) d\mu(t) \\
\geq \int_{\mathbb{R}^k \setminus (S_* \cup S)} I(\eta_*(t) > K/j) I(s^2(t) < K) d\mu(t) \\
= \int_{\mathbb{R}^k \setminus (S_* \cup S)} (1 - I(\eta_*(t) \leq K/j)) I(s^2(t) < K) d\mu(t) \\
= \int_{\mathbb{R}^k \setminus (S_* \cup S)} I(s^2(t) < K) d\mu(t) \\
- \int_{\mathbb{R}^k \setminus (S_* \cup S)} I(\eta_*(t) \leq K/j) I(s^2(t) < K) d\mu(t) \\
\geq \int_{\mathbb{R}^k \setminus (S_* \cup S)} I(s^2(t) < K) d\mu(t) \\
- \int_{\mathbb{R}^k \setminus (S_* \cup S)} I(\eta_*(t) \leq K/j) d\mu(t) \\
= \mu(\{t \in \mathbb{R}^k : s^2(t) < K\}) \\
- \mu(\{t \in \mathbb{R}^k : \eta_*(t) \leq K/j\}) \\
= \mu(\{t \in \mathbb{R}^k : s^2(t) < K\}) \\
- \mu(\{t \in \mathbb{R}^k : \eta_*(t) \leq K/j\}),
\]

where \(I(\cdot)\) is the indicator function. For fixed \(K > 0\) and \(j \to \infty\),

\[
\lim_{j \to \infty} \mu(\{t \in \mathbb{R}^k : \eta_*(t) \leq K/j\}) \\
= \mu\left(\bigcap_{j=1}^{\infty} \{t \in \mathbb{R}^k : \eta_*(t) \leq K/j\}\right) \\
= \mu(\{t \in \mathbb{R}^k : \eta_*(t) = 0\}) = \mu(S) = 0,
\]

hence for each \(K > 0\),

\[
\liminf_{n \to \infty} \Pr \left[ \widehat{W}_n(t) > M \right] \geq \mu(\{t \in \mathbb{R}^k : s^2(t) < K\}).
\]
Letting now $K \to \infty$ along the natural numbers it follows that
\[
\lim_{K \to \infty} \mu \left( \{ t \in \mathbb{R}^k : s^2(t) < K \} \right) = \mu \left( \bigcup_{K=1}^{\infty} \{ t \in \mathbb{R}^k : s^2(t) < K \} \right)
= \mu \left( \{ t \in \mathbb{R}^k : s^2(t) < \infty \} \right) = 1.
\]
Consequently, for all $M > 0$, \( \lim_{n \to \infty} \text{Pr} \left[ \hat{W}_n(t) > M \right] = 1 \), which implies that
\[ p \lim_{n \to \infty} \hat{W}_n(t) = \infty. \]

I used this randomization trick earlier in the unpublished paper Bierens (1987). A similar randomization approach as been used later by Domowitz and El-Gamal (1993, 2001).

Thus, the randomized test $\hat{W}_n(t)$ has an asymptotic $\chi_1^2$ null distribution and is consistent against all deviation from the null hypothesis.

The two referees liked this idea and recommended that the paper be revised for *Econometrica*. However, the editor did not like it. His objections were:

- By randomization you change the probability space on which the sequence of test statistics $\hat{W}_n(t)$ is defined.
- Due to randomization, the test cannot be replicated.
- Although the test is consistent, $\text{Pr}[\hat{W}_n(t) \leq M]$ depends on $\hat{t}$, so that for fixed $n$, $\text{Pr}[\hat{W}_n(t) \leq M]$ may not be close to zero. In other words, the small sample power of the test may be poor.

Nevertheless, the editor allowed me to resubmit a revision, but stated that he would treat this revision as a new submission, with new referees.

The main differences between the first version and the accepted revision are Theorems 4 and 5 in B90, which I will discuss below.

### 3. The Variance Function

Before discussing Theorems 4 and 5, let us have a closer look at the variance function $s^2(t)$ in (1.5) and the set $S_*$ in (1.6). As shown in the proof of Lemma 2 in B90, $s^2(t) = 0$ is equivalent to $\exp(t^t \Phi(X)) - (\nabla f(X, \theta_0))^t A^{-1} b(t) = 0$ a.s., which in its turn is equivalent to
\[
0 = E \left[ \left( \exp(t^t \Phi(X)) - (\nabla f(X, \theta_0))^t A^{-1} b(t) \right)^2 \right]
\]
As is well-known, in general \((\nabla f(X, \theta_0))' A^{-1} b(t)\) is the projection of \(\exp(t' \Phi(X))\) on \(\nabla f(X, \theta_0)\), i.e.,

\[
A^{-1} b(t) = \arg \min_{\beta \in \mathbb{R}^m} E \left[ (\exp(t' \Phi(X)) - \nabla f(X, \theta_0)' \beta)^2 \right], \text{ hence }
E \left[ (\exp(t' \Phi(X)) - \nabla f(X, \theta_0)' A^{-1} b(t))^2 \right]
= \min_{\beta \in \mathbb{R}^m} E \left[ (\exp(t' \Phi(X)) - \nabla f(X, \theta_0)' \beta)^2 \right].
\]

Consequently,

\[
S_* = \{ t \in \mathbb{R}^k : s^2(t) = 0 \}
= \left\{ t \in \mathbb{R}^k : \min_{\beta \in \mathbb{R}^m} E \left[ (\exp(t' \Phi(X)) - \nabla f(X, \theta_0)' \beta)^2 \right] = 0 \right\}.
\]

Therefore, \(s^2(t) = 0\) if and only if \(\exp(t' \Phi(X))\) is a.s. an exact linear combination of the components of \(\nabla f(X, \theta_0)\), which seems highly unlikely if \(t \neq 0\).

On the other hand, if one of the components of \(\theta_0\) is a constant term, which is usually the case, then one of the components of \(\nabla f(X, \theta_0)\) is equal to 1, so that trivially,

\[
E \left[ (1 - (\nabla f(X, \theta_0))' A^{-1} b(0))^2 \right] = \min_{\beta \in \mathbb{R}^m} E \left[ (1 - (\nabla f(X, \theta_0))' \beta)^2 \right] = 0.
\]

Thus, in this case \(s^2(0) = 0\).

In view of this discussions it is hardly a restriction to assume that

**Assumption 3.1.** \(s^2(t) > 0\) for all \(t \in \mathbb{R}^k \setminus \{0\}\).

Thus, the compact set \(T\) in Theorems 3-5 should not contain the zero vector.

**4. Theorems 4 and 5**

**4.1. The power under \(\sqrt{n}\)-local alternatives**

In Bierens and Ploberger (1997) it has been shown that the ICM test involved has nontrivial power against \(\sqrt{n}\)-local alternatives. The latter take the form

\[
Y = f(X, \theta_0) + g(X)/\sqrt{n} + U, \text{ where } E[U|X] = 0 \text{ a.s.} \quad (4.1)
\]
and
\[
\min_{\beta \in \mathbb{R}^m} E \left[ \left( g(X) - (\nabla f(X, \theta_0))' \beta \right)^2 \right] > 0. \tag{4.2}
\]

The latter condition implies that we can write \( g(X) \) as
\[
g(X) = (\nabla f(X, \theta_0))' \beta_0 + U_g,
\]
where
\[
\beta_0 = A^{-1} E[g(X)\nabla f(X, \theta_0)],
\]
\[
U_g = g(X) - (\nabla f(X, \theta_0))' \beta_0,
\]
\[
E[U_g \nabla f(X, \theta_0)] = 0, \tag{4.3}
\]
\[
\text{Pr}[E(U_g | X) = 0] < 1. \tag{4.4}
\]

As shown by Bierens and Ploberger (1997), under this local alternative the empirical process \( \hat{Z}_n(t) \) defined by (1.1) converges weakly on any compact subset \( T \) of \( \mathbb{R}^k \) to a Gaussian random process \( Z(t) \), but the latter is no longer a zero-mean process. In this case \( Z \) is characterized by the mean function
\[
\lambda(t) = E[g(X_1)\phi_1(t)] = E[g(X) \exp(t' \Phi(X))] - b(t)' A^{-1} E[g(X) \nabla f(X, \theta_0)]
\]
and the same covariance function \( \Gamma(t_1, t_2) \) as under the null hypothesis. Note that by (4.5) and Lemma 1 in B90, the set
\[
\Lambda_0 = \{ t \in T : \lambda(t) = 0 \}
\]
has Lebesgue measure zero and is nowhere dense in \( T \).

Thus,
\[
\hat{Z}_n(t) \Rightarrow Z(t) = \lambda(t) + Z_0(t),
\]
where \( Z_0(t) \) is a zero-mean Gaussian process with covariance function \( \Gamma(t_1, t_2) = E[Z_0(t_1)Z_0(t_2)] \). Then, with \( T \) chosen such that \( s^2(t) > 0 \) for all \( t \in T \),
\[
\widehat{W}_n(t) = \hat{Z}_n(t)^2/\hat{s}_n(t) \Rightarrow W(t) = (\lambda(t) + Z_0(t))^2/s^2(t)
\]
on \( T \), and thus by the continuous mapping theorem,
\[
\sup_{t \in T} \widehat{W}_n(t) \overset{d}{\to} \sup_{t \in T} W(t)
\]

Now consider the trick in Theorem 4: Choose real numbers \( \gamma > 0, \rho \in (0, 1) \) and a point \( t_0 \in T \). Let \( \hat{t} = \arg \max_{t \in T} \widehat{W}_n(t) \) and
\[
\bar{t} = t_0 \text{ if } \widehat{W}_n(\hat{t}) - \widehat{W}_n(t_0) \leq \gamma n^\rho, \quad \bar{t} = \hat{t} \text{ if } \widehat{W}_n(\hat{t}) - \widehat{W}_n(t_0) > \gamma n^\rho,
\]
8
and use $\hat{W}_n(\hat{t})$ as the test statistic. Then under the local alternative (4.1),

$$\Pr[\hat{t} = t_0] = \Pr[\hat{W}_n(\hat{t}) - \hat{W}_n(t_0) \leq \gamma n^\rho] \to 1$$

because $\hat{W}_n(\hat{t}) - \hat{W}_n(t_0) \overset{d}{\to} \sup_{t \in T} W(t) - W(t_0)$, hence $\hat{W}_n(\hat{t}) - \hat{W}_n(t_0) = O_p(1)$, and thus

$$\hat{W}_n(\hat{t}) \overset{d}{\to} W(t_0) = \left(\lambda(t_0) / \sqrt{s^2(t_0)} + V\right)^2,$$

where

$$V = Z_0(t_0) / \sqrt{s^2(t_0)} \sim N(0,1)$$

Next, note that for arbitrary $c > 0$,

$$\Pr \left[ \left(\lambda(t_0) / \sqrt{s^2(t_0)} + V\right)^2 > c \right]$$

$$= 1 - \Pr \left[ -\sqrt{c} - \lambda(t_0) / \sqrt{s^2(t_0)} \leq V \leq \sqrt{c} - \lambda(t_0) / \sqrt{s^2(t_0)} \right]$$

$$= 1 - \int_{-\sqrt{c} - \lambda(t_0) / \sqrt{s^2(t_0)}}^{\sqrt{c} - \lambda(t_0) / \sqrt{s^2(t_0)}} \frac{\exp(-v^2/2)}{\sqrt{2\pi}} \, dv$$

$$\geq 1 - \int_{-\sqrt{c}}^{\sqrt{c}} \frac{\exp(-v^2/2)}{\sqrt{2\pi}} \, dv = \Pr[V^2 > c], \quad (4.6)$$

where the inequality is strict if $\lambda(t_0) \neq 0$. The inequality in (4.6) follows from

$$\frac{d}{dx} \int_{-\sqrt{c} - x}^{\sqrt{c} - x} \frac{\exp(-v^2/2)}{\sqrt{2\pi}} \, dv = \frac{\exp(-(\sqrt{c} + x)^2/2)}{\sqrt{2\pi}} - \frac{\exp(-(\sqrt{c} - x)^2/2)}{\sqrt{2\pi}}$$

$$> 0 \text{ if } (\sqrt{c} + x)^2 < (\sqrt{c} - x)^2$$

$$< 0 \text{ if } (\sqrt{c} + x)^2 > (\sqrt{c} - x)^2$$

and the trivial facts that $(\sqrt{c} + x)^2 < (\sqrt{c} - x)^2$ is equivalent to $x < 0$ and $(\sqrt{c} + x)^2 > (\sqrt{c} - x)^2$ is equivalent to $x > 0$.

Therefore, the test in Theorem 4 has nontrivial $\sqrt{n}$-local power as long as $t_0$ is chosen such that $\lambda(t_0) \neq 0$. The only way to guarantee the latter is to draw $t_0$ randomly from an absolutely continuous distribution on $T$. However, with $\tilde{t}_0$ such a random drawing, under the local alternative (4.1) the test statistic $\hat{W}_n(\tilde{t})$ is asymptotically equivalent to $\hat{W}_n(\tilde{t}_0)$, so that effectively we are then back to the original randomization procedure. The same applies to the test in Theorem 5.
4.2. Theorem 5 with random search

The difference between Theorems 4 and 5 is that in the latter case \( \hat{W}_n(t) \) is maximized over \( \{t_1, t_2, \ldots, t_{K_n}\} \), where the sequence \( \{t_i\}_{i=1}^\infty \) is dense in \( T \) and \( K_n \) is an arbitrary subsequence of \( n \). As mentioned in B90, Theorem 5 carries over if \( t_0 \) and the \( t_i \)'s for \( i \geq 1 \) are drawn randomly from an absolutely continuous distribution on \( T \). In this subsection I will give a formal proof of this conjecture:

**Theorem 4.1.** Let the conditions in Theorem 5 in B90 be satisfied, with \( T \) a compact subset of \( \mathbb{R}^k \) such that \( s^2(t) > 0 \) for all \( t \in T \). Let \( \tilde{t}_i, i = 0, 1, \ldots, K_n \) be random drawings from an absolutely continuous distribution \( \mu \) with support \( T \), and let \( \hat{W}_n(t) \) be defined by (1.9). Denote \( \tilde{t} = \max_{t \in [\tilde{t}_1, \tilde{t}_2, \ldots, \tilde{t}_{K_n}]} \hat{W}_n(t) \) and

\[
\tilde{t} = \tilde{t}_0 \text{ if } \hat{W}_n(\tilde{t}_0) - \hat{W}_n(\tilde{t}) \leq \gamma n^\alpha, \quad \tilde{t} = \hat{t} \text{ if } \hat{W}_n(\hat{t}_n) - \hat{W}_n(\tilde{t}_0) > \gamma n^\alpha.
\]

Then under \( H_0 \), \( \hat{W}_n(\hat{t}) \xrightarrow{d} \chi_1^2 \), whereas under \( H_1 \), \( \lim_{n \to \infty} \hat{W}_n(\hat{t})/n = \sup_{t \in T} \eta(t) > 0 \), where \( \eta(t) \) is defined by (1.11).

**Proof.** Recall that under \( H_0 \), \( \hat{W}_n(t) \Rightarrow W(t) = Z(t)^2/s^2(t) \), hence \( \sup_{t \in T} \hat{W}_n(t) \xrightarrow{d} \sup_{t \in T} W(t) \). Then

\[
\max_{i=1, \ldots, K_n} \hat{W}_n(t_i) - \hat{W}_n(\tilde{t}_0) \leq \sup_{t \in T} \hat{W}_n(t) - \hat{W}_n(\tilde{t}_0) = O_p(1),
\]

which implies that under \( H_0 \), \( \Pr[\hat{t} = \tilde{t}_0] \to 1 \) and thus \( \hat{W}_n(\hat{t}) \xrightarrow{d} \chi_1^2 \).

Under \( H_1 \) and the conditions in Theorem 5 in B90 it can be shown that

\[
\sup_{t \in T} \left| \hat{W}_n(t)/n - \eta(t) \right| = o_p(1), \text{ where } \eta(t) > 0 \text{ a.e. on } T,
\]

and thus

\[
\max_{i=1, \ldots, K_n} \hat{W}_n(t_i)/n = \max_{i=1, \ldots, K_n} \left\{ \hat{W}_n(\tilde{t}_i)/n - \eta(\tilde{t}_i) + \eta(\tilde{t}_i) \right\}
\]

\[
\leq \max_{i=1, \ldots, K_n} \left\{ \sup_{t \in T} \left| \hat{W}_n(t)/n - \eta(t) \right| + \eta(\tilde{t}_i) \right\}
\]

\[
= \sup_{t \in T} \left| \hat{W}_n(t)/n - \eta(t) \right| + \max_{i=1, \ldots, K_n} \eta(\tilde{t}_i)
\]

\footnote{Recall that under Assumption 3.1 we may choose for \( T \) any compact subset of \( \mathbb{R}^k \setminus \{0\} \).}
≥ \max_{i=1,...,K_n} \left\{ -\sup_{t \in T} \left| \widehat{W}_n(t)/n - \eta(t) \right| + \eta(\tilde{t}_i) \right\} \\
= -\sup_{t \in T} \left| \widehat{W}_n(t)/n - \eta(t) \right| + \max_{i=1,...,K_n} \eta(\tilde{t}_i)

Hence

\left| \max_{i=1,...,K_n} \widehat{W}_n(\tilde{t}_i)/n - \max_{i=1,...,K_n} \eta(\tilde{t}_i) \right| \leq 2\sup_{t \in T} \left| \widehat{W}_n(t)/n - \eta(t) \right| = o_p(1)

Moreover, for any constant \( c > 0 \),

\begin{align*}
\Pr \left[ \max_{i=1,...,K_n} \eta(\tilde{t}_i) \leq c \right] &= \prod_{i=1}^{K_n} E \left[ I (\eta(\tilde{t}_i) \leq c) \right] \\
&= \left( \int_{T} I (\eta(t) \leq c) \, d\mu(t) \right)^{K_n} \\
&= \mu (\{ t \in T : \eta(t) \leq c \})^{K_n} \\
&= 1 \text{ if } c \geq \sup_{t \in T} \eta(t) \\
&= 0 \text{ if } c < \sup_{t \in T} \eta(t)
\end{align*}

hence for arbitrary \( \delta > 0 \),

\begin{align*}
\lim_{n \to \infty} \Pr \left[ \left| \max_{i=1,...,K_n} \eta(\tilde{t}_i) - \sup_{t \in T} \eta(t) \right| \leq \delta \right] \\
= \lim_{n \to \infty} \Pr \left[ \sup_{t \in T} \eta(t) - \delta < \max_{i=1,...,K_n} \eta(\tilde{t}_i) \leq \sup_{t \in T} \eta(t) + \delta \right] \\
= \lim_{n \to \infty} \Pr \left[ \sup_{t \in T} \eta(t) - \delta < \max_{i=1,...,K_n} \eta(\tilde{t}_i) \right] \\
= 1 - \lim_{n \to \infty} \Pr \left[ \max_{i=1,...,K_n} \eta(\tilde{t}_i) \leq \sup_{t \in T} \eta(t) - \delta \right] \\
= 1
\end{align*}

Thus, under \( H_1 \), \( p \lim_{n \to \infty} \max_{i=1,...,K_n} \widehat{W}_n(\tilde{t}_i)/n = p \lim_{n \to \infty} \max_{i=1,...,K_n} \eta(\tilde{t}_i) = \sup_{t \in T} \eta(t) > 0 \). □

Moreover, in view of the argument in the previous subsection the following result also holds.
**Theorem 4.2.** Under the conditions in Theorem 4.1 the test in the latter theorem has nontrivial asymptotic power against the local alternative (4.1).

5. A Consistent Hausman-Type Model Specification Test

5.1. The Hausman-White test

Hausman (1978) has shown that, with \( \hat{\theta}_n \) an efficient asymptotically normal estimator of a parameter vector \( \theta_0 \in \mathbb{R}^m \) and \( \tilde{\theta}_n \) an inefficient but also asymptotically normal estimator of \( \theta_0 \),
\[
\sqrt{n}(\tilde{\theta}_n - \hat{\theta}_n) \xrightarrow{d} N_m[0, V_1 - V_0],
\]
where \( V_1 \) is the asymptotic variance matrix of \( \sqrt{n}(\tilde{\theta}_n - \theta_0) \) and \( V_0 \) is the asymptotic variance matrix of \( \sqrt{n}(\hat{\theta}_n - \theta_0) \). Then, with \( \hat{V}_{1,n} \) and \( \hat{V}_{0,n} \) consistent estimators of \( V_1 \) and \( V_0 \), respectively, and assuming that \( V_1 - V_0 \) is nonsingular,
\[
\hat{H}_n = n \left( \tilde{\theta}_n - \hat{\theta}_n \right) \left( \hat{V}_{1,n} - \hat{V}_{0,n} \right)^{-1} \left( \tilde{\theta}_n - \hat{\theta}_n \right) \xrightarrow{d} \chi^2_m
\]
Hausman proposed to use \( \hat{H}_n \) as test statistic of the null hypothesis that the model involved is correctly specified, against the alternative hypothesis that the model is so severely misspecified that
\[
p \lim_{n \to \infty} (\tilde{\theta}_n - \hat{\theta}_n) \neq 0,
\]
so that in the latter case \( p \lim_{n \to \infty} \hat{H}_n = \infty \).

White (1981) showed that (5.1) carries over for homoskedastic nonlinear regression models, where \( \hat{\theta}_n \) is the NLLS estimator and \( \tilde{\theta}_n \) a weighted NLLS estimator. Also White (1981) assumes that \( V_1 - V_0 \) is nonsingular, so that (5.2) carries over.

5.2. Inconsistency of the Hausman-White test

As pointed out in Bierens (1982), Hausman’s (1978) test and White’s (1981) version of this test are not consistent. A counter example is the data generating process
\[
Y_j = X_{1,j} + X_{2,j} + X_{1,j}X_{2,j} + U_j
\]

\[\text{3Provided that } p \lim_{n \to \infty} (\hat{V}_{1,n} - \hat{V}_{0,n}) \text{ remains positive semidefinite.}\]
where the $X_{1,j}$, $X_{2,j}$ and $U_j$'s are i.i.d. standard normal, whereas the model is incorrectly specified as

$$Y_j = \theta_{0,1}X_{1,j} + \theta_{0,2}X_{2,j} + U_j.$$ 

Let $\tilde{\theta}_n$ be the OLS estimator of $\theta_0 = (\theta_{0,1}, \theta_{0,2})'$ and let

$$\tilde{\theta}_n = \arg\min_{\theta=(\theta_1, \theta_2)} \frac{1}{n} \sum_{j=1}^n (Y_j - \theta_1X_{1,j} - \theta_2X_{2,j})^2 w(X_{1,j}, X_{2,j})$$

be a weighted least squares estimator, where $w(.,.)$ is a nonnegative weight function. Suppose we have chosen

$$w(X_{1,j}, X_{2,j}) = X_{1,j}^2 + X_{2,j}^2.$$ 

Then $p \lim_{n \to \infty} \tilde{\theta}_n = p \lim_{n \to \infty} \tilde{\theta}_n = (1, 1)'$ and $\sqrt{n}(\tilde{\theta}_n - \hat{\theta}_n)$ is asymptotically bivariate zero-mean normal, as is not hard to verify.

### 5.3. A consistent Hausman-White type test

In this section I will derive a consistent Hausman-type test, by using White’s (1981) version with weight function exp($t'\Phi(X_j)$), based on the results in Bierens (1987), as follows. Consider the same nonlinear regression model (1.2) as before, with NLLS estimator $\tilde{\theta}_n$ and pseudo-true parameter vector

$$\theta_0 = \arg\min_{\theta \in \Theta} E[(Y - f(X, \theta))^2].$$

Similarly, let for $t \in \mathbb{R}^k$,

$$\tilde{\theta}_n(t) = \arg\min_{\theta \in \Theta} \frac{1}{n} \sum_{j=1}^n (Y_j - f(X_j, \theta))^2 \exp(t'\Phi(X_j)),$$

which is the weighted NLLS estimator with weight function exp($t'\Phi(X_j)$), with pseudo-true parameter vector

$$\theta_0(t) = \arg\min_{\theta \in \Theta} E[(Y - f(X, \theta))^2 \exp(t'\Phi(X))].$$

As before, $\{(Y_j, X_j)\}_{j=1}^n$ is a random sample from $(Y, X) \in \mathbb{R} \times \mathbb{R}^k$ with $E[Y^2] < \infty$, and $\Phi : \mathbb{R}^k \to \mathbb{R}^k$ is a bounded one-to-one mapping such that $X$ and $\Phi(X)$
generate the same $\sigma$-algebra. Note that under the same conditions as in B90, 
$p\lim_{n\to\infty} \tilde{\theta}_n = \theta_0$ and $p\lim_{n\to\infty} \tilde{\theta}_n(t) = \theta_0(t)$, regardless whether the model is 
correctly specified or not.

Let us first assume that the alternative hypothesis (1.10) is true, and that

**Assumption 5.1.** For each $t \in \mathbb{R}^k$, $\theta_0(t)$ is a unique interior point of $\Theta$;

**Assumption 5.2.** Under $H_1$, $\Pr[\nabla f(X, \theta_0) = 0 \mid E(U|X) \neq 0] < 1$.

Then

**Lemma 5.1.** Under $H_1$ and Assumptions 5.1 and 5.2, the set

$$S = \{t \in \mathbb{R}^k : \theta_0(t) = \theta_0\}$$

has Lebesgue measure zero and is nowhere dense in $\mathbb{R}^k$.

**Proof.** By the first-order condition for $\theta_0(t)$ and Assumption 5.1,

$$E[(Y - f(X, \theta_0))\nabla f(X, \theta_0(t)) \exp(t'\Phi(X))] = 0.$$ 

Suppose that $\theta_0(t) = \theta_0$ for all $t$ in an open set $O \subset \mathbb{R}^k$. Then for all $t \in O$,

$$E[U, \nabla f(X, \theta_0) \exp(t'\Phi(X))] = 0.$$ 

This implies, by Lemma 1 in B90, that

$$E[U|X].\nabla f(X, \theta_0) = 0 \text{ a.s.},$$

and thus also

$$|E[U|X]|.||\nabla f(X, \theta_0)|| = 0 \text{ a.s.}$$

Now

$$1 = \Pr[|E[U|X]|.||\nabla f(X, \theta_0)|| = 0]$$

$$= \Pr[(|E[U|X]| > 0 \text{ and } ||\nabla f(X, \theta_0)|| = 0)$$

or $(|E[U|X]| = 0 \text{ and } ||\nabla f(X, \theta_0)|| > 0)$

or $(|E[U|X]| = 0 \text{ and } ||\nabla f(X, \theta_0)|| = 0)]$

$$= \Pr[(|E[U|X]| > 0 \text{ and } ||\nabla f(X, \theta_0)|| = 0)]$$
\[ + \Pr [|E[U|X]| = 0 \text{ and } ||\nabla f(X, \theta_0)|| > 0] \\
+ \Pr [|E[U|X]| = 0 \text{ and } ||\nabla f(X, \theta_0)|| = 0] \\
= \Pr [(|E[U|X]| > 0 \text{ and } ||\nabla f(X, \theta_0)|| = 0)] \\
+ \Pr [|E[U|X]| = 0] \\
= \Pr [||\nabla f(X, \theta_0)|| = 0 \mid |E[U|X]| > 0] \Pr [|E[U|X]| > 0] \\
+ 1 - \Pr [|E[U|X]| > 0] \]

and therefore, since under \( H_1 \), \( \Pr [|E[U|X]| > 0] > 0 \), we have
\[ \Pr [||\nabla f(X, \theta_0)|| = 0 \mid |E[U|X]| > 0] = 1. \]

However, the latter contradicts Assumption 5.2. Consequently, there does not exists an open set \( O \) such that for all \( t \in O \), \( \theta_0(t) = \theta_0 \), which proves the lemma.

Recall that under the null hypothesis that the nonlinear regression model (1.2) is correctly specified, \( E[U_j|X_j] = 0 \) a.s. In that case, \( \theta_0(t) = \theta_0 \) for all \( t \in \mathbb{R}^k \). Then similar to White (1981) it can be shown that pointwise in \( t \in \mathbb{R}^k \),
\[
\sqrt{n} \left( \hat{\theta}_n(t) - \theta_0 \right) \\
= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_j \left( A(t)^{-1} \nabla f(X_j, \theta_0) \exp(t'\Phi(X_j)) - A(0)^{-1} \nabla f(X_j, \theta_0) \right) \\
+ o_p(1) \xrightarrow{d} N_m[0, \Omega(t)], \tag{5.3}
\]

where
\[
\Omega(t) = A(t)^{-1} B(2t) A(t)^{-1} - A(0)^{-1} B(t) A(t)^{-1} \\
- A(t)^{-1} B(t) A(0)^{-1} + A(0)^{-1} B(0) A(0)^{-1}
\]

with
\[
A(t) = E [ \langle \nabla f(X, \theta_0) \rangle (\nabla f(X, \theta_0))' \exp(t'\Phi(X))] , \\
B(t) = E \left[ U^2 (\nabla f(X, \theta_0) \rangle (\nabla f(X, \theta_0))' \exp(t'\Phi(X))] \right].
\]

Note that if we assume homoskedasticity, as in White (1981), then \( B(t) = \sigma^2 A(t) \), where \( \sigma^2 = E[U^2] \), hence \( \Omega(t) = \sigma^2 A(t)^{-1} A(2t) A(t)^{-1} - \sigma^2 A(0)^{-1} \). The latter expression is in essence the same as in White (1981).
Under Assumption A in B90 the variance matrix $\Omega(t)$ can be estimated consistently, pointwise in $t$, by

$$
\hat{\Omega}_n(t) = \hat{A}(t)^{-1} \hat{B}_n(2t) \hat{A}(t)^{-1} - \hat{A}_n(0)^{-1} \hat{B}_n(t) \hat{A}_n(t)^{-1} - \hat{A}_n(t)^{-1} \hat{B}_n(t) \hat{A}_n(0)^{-1} + \hat{A}_n(0)^{-1} \hat{B}_n(0) \hat{A}_n(0)^{-1}
$$

where

$$
\hat{A}_n(t) = \frac{1}{n} \sum_{j=1}^{n} \left( \nabla f(X_j, \hat{\theta}_n) (\nabla f(X_j, \hat{\theta}_n))' \exp(t' \Phi(X_j)) \right),
$$

$$
\hat{B}_n(t) = \frac{1}{n} \sum_{j=1}^{n} \hat{U}_j^2 (\nabla f(X_j, \hat{\theta}_n) (\nabla f(X_j, \hat{\theta}_n))' \exp(t' \Phi(X_j)).
$$

with the $\hat{U}_j$'s the NLLS residuals. Thus,

$$
p \lim_{n \to \infty} \hat{\Omega}_n(t) = \Omega(t), \text{ pointwise in } t \in \mathbb{R}^k,
$$

regardless whether $H_0$ is true or not.

Assuming, similar to White (1981), that

$$
\text{Assumption 5.3. For } t \neq 0, \Omega(t) \text{ is nonsingular,}
$$

it follows now that under $H_0$,

$$
\hat{H}_n(t) = n \left( \hat{\theta}_n(t) - \hat{\theta}_n \right)' \hat{\Omega}_n(t)^{-1} \left( \hat{\theta}_n(t) - \hat{\theta}_n \right) \overset{d}{\rightarrow} \chi_m^2 \quad (5.4)
$$

pointwise in $t \neq 0$, whereas by Lemma 5.1 and under $H_1$,

$$
\hat{H}_n(t)/n \overset{p}{\rightarrow} \kappa(t) = (\theta_0(t) - \theta_0)' \Omega(t)^{-1} (\theta_0(t) - \theta_0) > 0 \text{ a.e.} \quad (5.5)
$$

These results carry over if we replace $t$ by a random drawing $\tilde{t}$ from an absolutely continuous distribution on $\mathbb{R}^k$:

$$
\hat{H}_n(\tilde{t}) = n \left( \hat{\theta}_n(\tilde{t}) - \hat{\theta}_n \right)' \hat{\Omega}_n(\tilde{t})^{-1} \left( \hat{\theta}_n(\tilde{t}) - \hat{\theta}_n \right) \overset{d}{\rightarrow} \chi_m^2 \text{ under } H_0,
$$

$$
\hat{H}_n(\tilde{t})/n \overset{p}{\rightarrow} (\theta_0(\tilde{t}) - \theta_0)' \Omega(\tilde{t})^{-1} (\theta_0(\tilde{t}) - \theta_0) > 0 \text{ a.s. under } H_1.
$$

This is the main result in my unpublished paper Bierens (1987).
Finally, let us confine \( t \) to a compact subset \( T \) of \( \mathbb{R}^k \backslash \{0\} \). Then it can be shown, similar to (1.3), that under \( H_0 \) the \( o_p(1) \) term in (5.3) is uniform in \( T \), and that similar to (1.4),

\[
C_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_j \left( A(t)^{-1} \nabla f(X_j, \theta_0) \exp(t'\Phi(X_j)) - A(0)^{-1} \nabla f(X_j, \theta_0) \right) \tag{5.6}
\]

\( \Rightarrow C(t) \)

on \( T \), where \( C(t) \) is an \( \mathbb{R}^m \)-valued zero mean Gaussian process on \( T \), with matrix-valued covariance function

\[
E[C(t_1)C(t_2)'] = A(t_1)^{-1} B(t_1 + t_2) A(t_2)^{-1} - A(t_1)^{-1} B(t_1) A(0)^{-1} - A(0)^{-1} B(t_2) A(t_2)^{-1} + A(0)^{-1} B(0) A(0)^{-1}.
\]

Moreover, it is not hard to verify that under Assumption A in B90,

\[
p \lim_{n \to \infty} \Omega_n(t) = \Omega(t) \text{ uniformly on } T.
\]

Consequently, under \( H_0 \),

\[
\tilde{H}_n(t) \Rightarrow H(t) = C(t)'\Omega(t)^{-1}C(t) \tag{5.7}
\]

on \( T \), and thus by the continuous mapping theorem,

\[
\sup_{t \in T} \tilde{H}_n(t) \xrightarrow{d} \sup_{t \in T} H(t),
\]

whereas under \( H_1 \),

\[
p \lim_{n \to \infty} \sup_{t \in T} \tilde{H}_n(t)/n = \sup_{t \in T} (\theta_0(t) - \theta_0)'\Omega(t)^{-1}(\theta_0(t) - \theta_0) > 0.
\]

It is now obvious that the procedure in Theorems 4 and 5 also applies to \( \tilde{H}_n(t) \), yielding a consistent Hausman test. In particular, similar to Theorem 4.1 the following results hold.

**Theorem 5.1.** Let Assumptions 5.1-5.3 and the conditions in Theorem 5 in B90 be satisfied, with \( T \) a compact subset of \( \mathbb{R}^k \backslash \{0\} \). Let \( \tilde{t}_i, i = 0, 1, \ldots, K_n \) be random
drawings from an absolutely continuous distribution with support $T$, and let $\hat{H}_n(t)$ be defined by (5.4). Let $\hat{t} = \arg \max_{t \in \{\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_{\kappa_n}\}} \hat{H}_n(t)$ and

$$\tilde{t} = \tilde{t}_0 \text{ if } \hat{H}_n(\tilde{t}) - \hat{H}_n(\tilde{t}_0) \leq \gamma n^\rho, \quad \tilde{t} = \hat{t} \text{ if } \hat{H}_n(\tilde{t}) - \hat{H}_n(\tilde{t}_0) > \gamma n^\rho.$$ 

Then under $H_0$, $\hat{H}_n(t) \xrightarrow{d} \chi^2_m$, whereas under $H_1$, $\lim_{n \to \infty} \hat{H}_n(\tilde{t})/n = \sup_{t \in T} \kappa(t) > 0$, where $\kappa(t)$ is defined by (5.5).

### 5.4. Power under \( \sqrt{n} \)-local alternatives

Under the local alternative (4.1) the empirical process $C_n(t)$ in (5.6) converges weakly to $\gamma(t) + C(t)$, where $C(t)$ is the same as before, and $\gamma(t)$ it the mean function:

$$\gamma(t) = E \left[ g(X) \left( A(t)^{-1} \nabla f(X, \theta_0) \exp(t' \Phi(X)) - A(0)^{-1} \nabla f(X, \theta_0) \right) \right]$$

$$= A(t)^{-1} E \left[ \nabla f(X, \theta_0) \left( \nabla f(X, \theta_0) \right) \exp(t' \Phi(X)) \right] A(0)^{-1} E \left[ g(X) \nabla f(X, \theta_0) \right]$$

$$- A(0)^{-1} E \left[ \nabla f(X, \theta_0) \nabla f(X, \theta_0) \right] A(0)^{-1} E \left[ g(X) \nabla f(X, \theta_0) \right]$$

$$+ E \left[ U_g \left( A(t)^{-1} \nabla f(X, \theta_0) \exp(t' \Phi(X)) - A(0)^{-1} \nabla f(X, \theta_0) \right) \right]$$

$$= A(t)^{-1} E \left[ U_g \nabla f(X, \theta_0) \exp(t' \Phi(X)) \right]$$

where $U_g$ is defined by (4.3). Note that by (4.4), $\gamma(0) = 0$, and that similar to Lemma 5.1, the set

$$\{ t \in \mathbb{R}^k : \gamma(t) = 0 \}$$

has Lebesgue measure zero and is nowhere dense, provided that

**Assumption 5.5.** $\Pr[\nabla f(X, \theta_0) = 0 \mid E(U_g|X) \neq 0] < 1$.

Now the result (5.7) becomes

$$\hat{H}_n(t) \Rightarrow H_L(t) = (C(t) + \gamma(t))' \Omega(t)^{-1} (C(t) + \gamma(t))$$

on $T$, hence

$$\sup_{t \in T} \hat{H}_n(t) \xrightarrow{d} \sup_{t \in T} H_L(t)$$

Therefore, under the local alternative (4.1) the test statistic $\hat{H}_n(\tilde{t})$ is asymptotically equivalent to $\hat{H}_n(\tilde{t}_0)$.
Also in this case the power of the test is nontrivial. To see this, note that for \( t \in T \) such that \( \gamma(t) \neq 0 \),

\[
\begin{align*}
H_L(t) &= (V(t) + \delta(t))(V(t) + \delta(t)) = \sum_{i=1}^{m} (V_i(t) + \delta_i(t))^2, \text{ where} \\
V(t) &= (V_1(t), V_2(t), ..., V_m(t))' = \Omega(t)^{-1/2}C(t) \sim N_m(0, I_m), \\
\delta(t) &= (\delta_1(t), \delta_2(t), ..., \delta_m(t))' = \Omega(t)^{-1/2}\gamma(t) \neq 0
\end{align*}
\]

Then for \( c > 0 \),

\[
\Pr[H_L(t) > c] = \Pr \left[ \sum_{i=1}^{m} (V_i(t) + \delta_i(t))^2 > c \right]
\]

\[
= \Pr \left[ (V_1(t) + \delta_1(t))^2 > c - \sum_{i=2}^{m} (V_i(t) + \delta_i(t))^2 \right. \\
\left. \quad \text{and} \sum_{i=2}^{m} (V_i(t) + \delta_i(t))^2 \leq c \right]
\]

\[
+ \Pr \left[ (V_1(t) + \delta_1(t))^2 > c - \sum_{i=2}^{m} (V_i(t) + \delta_i(t))^2 \right. \\
\left. \quad \text{and} \sum_{i=2}^{m} (V_i(t) + \delta_i(t))^2 > c \right]
\]

\[
= \Pr \left[ (V_1(t) + \delta_1(t))^2 > c - \sum_{i=2}^{m} (V_i(t) + \delta_i(t))^2 \right. \\
\left. \quad \sum_{i=2}^{m} (V_i(t) + \delta_i(t))^2 \leq c \right] \\
\times \Pr \left[ \sum_{i=2}^{m} (V_i(t) + \delta_i(t))^2 \leq c \right. \\
\left. \quad + \Pr \left[ \sum_{i=2}^{m} (V_i(t) + \delta_i(t))^2 > c \right. \right]
\]

It follows similar to (4.6) that

\[
\Pr \left[ (V_1(t) + \delta_1(t))^2 > c - \sum_{i=2}^{m} (V_i(t) + \delta_i(t))^2 \right. \sum_{i=2}^{m} (V_i(t) + \delta_i(t))^2 \leq c
\]

\[
> \Pr \left[ V_1(t)^2 > c - \sum_{i=2}^{m} (V_i(t) + \delta_i(t))^2 \right. \sum_{i=2}^{m} (V_i(t) + \delta_i(t))^2 \leq c \quad \text{if } \delta_1(t) \neq 0
\]
hence
\[
\Pr \left[ H_L(t) > c \right] > \Pr \left[ V_1(t)^2 + \sum_{i=2}^{m} (V_i(t) + \delta_i(t))^2 > c \right] \text{ if } \delta_1(t) \neq 0
\]

Repeating this procedure for the latter probability, in the form
\[
\Pr \left[ (V_2(t) + \delta_2(t))^2 > c - V_1(t)^2 - \sum_{i=3}^{m} (V_i(t) + \delta_i(t))^2 \right],
\]
it follows that
\[
\Pr \left[ V_1(t)^2 + \sum_{i=2}^{m} (V_i(t) + \delta_i(t))^2 > c \right] > \Pr \left[ V_1(t)^2 + V_2(t)^2 + \sum_{i=3}^{m} (V_i(t) + \delta_i(t))^2 > c \right] \text{ if } \delta_2(t) \neq 0.
\]

Hence, by induction,
\[
\Pr \left[ H_L(t) > c \right] = \Pr \left[ (V(t) + \delta(t))' (V(t) + \delta(t)) > c \right]
\]
\[
> \Pr \left[ V(t)' V(t) > c \right] = \Pr \left[ H(t) > c \right] \text{ if } \gamma(t) \neq 0
\]

Summarizing, the following result has been shown.

**Theorem 5.2.** Under Assumption 5.5 and the conditions in Theorem 5.1 the Hausman-type test in the latter theorem has nontrivial asymptotic power against the local alternative (4.1).

### 6. Concluding Remarks

The crux in Theorems 4 and 5 in B90 and Theorem 4.1 in this addendum is the selection mechanism to decide whether the null hypothesis is true or not. In particular, in Theorem 4.1,

\[
I \left( \tilde{W}_n(t) - \tilde{W}_n(t_0) > \gamma n^\rho \right) \overset{p}{\to} \left\{ \begin{array}{ll}
0 & \text{under } H_0, \\
1 & \text{under } H_1.
\end{array} \right.
\]

with test statistic
\[
\hat{\gamma}_n = \left( 1 - I \left( \tilde{W}_n(t) - \tilde{W}_n(t_0) > \gamma n^\rho \right) \right) \tilde{W}_n(t_0) + \tilde{W}_n(t) I \left( \tilde{W}_n(t) - \tilde{W}_n(t_0) > \gamma n^\rho \right)
\]
This test is consistent because under $H_1$, $p \lim_{n \to \infty} \frac{\hat{W}_n(t)}{n} \in (0, \infty)$, whereas under $H_0$, $\hat{W}_n(t) = O_p(1)$.

Now let $Z$ be a random drawing from an arbitrary univariate continuous distribution, for example the standard exponential distribution, and consider the ”test statistic”

$$\bar{\Upsilon}_n = \left(1 - I \left( \hat{W}_n(t) - \hat{W}_n(t_0) > \gamma n^\rho \right) \right) Z + \hat{W}_n(t)I \left( \hat{W}_n(t) - \hat{W}_n(t_0) > \gamma n^\rho \right)$$

Then under $H_0$, $\bar{\Upsilon}_n \overset{d}{\to} Z$, whereas under $H_1$, $\bar{\Upsilon}_n/n \overset{p}{\to} p \lim_{n \to \infty} \frac{\hat{W}_n(t)}{n} = \sup_{t \in T} \eta(t) \in (0, \infty)$. Thus, the right-sided test $\bar{\Upsilon}_n$ is consistent against any global deviation from the null hypothesis and is pivotal. However, the asymptotic null distribution $Z$ in this case is arbitrary and is therefore irrelevant. What really matters is the result (6.1). Therefore, in essence the test in Theorem 4.1 is equivalent to the decision rule

Accept $H_0$ if $I \left( \hat{W}_n(t) - \hat{W}_n(t_0) > \gamma n^\rho \right) = 0$,

Accept $H_1$ if $I \left( \hat{W}_n(t) - \hat{W}_n(t_0) > \gamma n^\rho \right) = 1$,

and similarly for the tests in Theorems 4 and 5 in B90. Replacing $\hat{W}_n$ by $\hat{H}_n$ the same applies to the Hausman-type test in Theorem 5.1. This is the very reason why I have never followed-up on the approach in B90.

Finally, note that Hausman (1978) only proves that under $H_0$ the variance matrix $V_1 - V_0$ in (5.1) is positive semidefinite, and even if it is nonsingular, there is no guarantee that under $H_1$ its estimate converges in probability to a positive definite matrix. The same applies to White’s (1981) version of the Hausman test. Therefore, Assumption 5.3 may not hold.

References


