AN ADAPTIVE, RATE-OPTIMAL TEST OF A PARAMETRIC MODEL AGAINST A NONPARAMETRIC ALTERNATIVE

by

Joel L. Horowitz
Department of Economics,
Northwestern University

INTRODUCTION

- Problem: Test parametric model of a conditional mean function against nonparametric alternative

- Model

$$Y_i = f(X_i) + \varepsilon_i; \quad i = 1, 2, ...$$

  - Y_i = Scalar random variable

  - $\{X_i\}$ = Sequence of d-dimensional, distinct, non-stochastic design points

  - $\{\varepsilon_i\}$ = Sequence of unobserved, independent random variables with means of zero.

- Null hypothesis H_0: $f(X_i) = F(X_i, \theta)$ for all i and some finite-dimensional θ

- Alternative hypothesis H_1: There is no θ such that $f(X_i) = F(X_i, \theta)$ for all i.

AIM

- Develop a test that:
  - Is consistent against alternative models whose distance from the parametric model converges to zero as rapidly as possible as sample size, n, increases
  - Has other good power properties
  - Does not require *a priori* knowledge of smoothness of alternative model

BACKGROUND

- Many tests are already available.
  - Some compare nonparametric estimator of $f(\cdot)$ with parametric estimator $F(\cdot,\theta_n)$ (e.g., Härdle & Mammen 1993 and many others)
  - Other tests do not require nonparametric estimation of f (e.g., Andrews 1997)
- Asymptotic power often investigated through sequence of local alternative models:

  $$f_n(x) = F(x,\theta_1) + \rho_n g(x)$$

  for some g, θ_1, and $\rho_n \to 0$ as $n \to \infty$.

  - Many tests that compare a nonparametric estimator of f with a parametric estimator have non-trivial power only against local alternatives for which $\rho_n \to 0$ more slowly than $n^{-1/2}$
  - Test of Andrews (1997) has non-trivial power against local alternatives for which $\rho_n \propto n^{-1/2}$
  - Latter tests seem to dominate former in terms of power, but this appearance is misleading.

WHY $\rho_n \propto n^{-1/2}$ IS MISLEADING

- Local alternatives $f_n(x) = F(x, \theta_1) + \rho_n g(x)$ are well-known but this class is too small.
  - Talk describes with good power against a class of alternatives that is not too small
- The problem: If $\rho_n \propto n^{-1/2}$, then no test can be consistent uniformly over reasonable classes of functions g (e.g., functions that are s times continuously differentiable)
  - Power of any test of H_0 against sequence of local alternatives $f_n(x) = F(x, \theta_1) + n^{-1/2} g_n(x)$ is probability that test rejects correct H_0 for some sequence of smooth functions $\{g_n\}$.
  - Practical consequence: Any test of H_0 for which $\rho_n \propto n^{-1/2}$ has low finite-sample power against certain classes of smooth alternatives.
  - This is sense in which set of alternatives $f_n(x) = F(x, \theta_1) + \rho_n g(x)$ is too small.
- Problem can be overcome through use of minimax approach to asymptotic local power

THE MINIMAX APPROACH

- Permits set of alternatives to consist of entire smoothness class (e.g., Hölder or Sobolev ball).
  - We use Hölder class
- Minimax approach assumes that f belongs to smoothness class B (e.g., Hölder class).
  - B separated from null-hypothesis $\{F(\cdot,\theta):\theta \in \Theta\}$ by distance r_n that converges to 0 as $n \to \infty$.
  - *Optimal rate of testing* is find fastest rate at which r_n can approach 0 while permitting consistent testing **uniformly** over B.
- Test is consistent uniformly over B if

$$\lim_{n\to\infty} \inf_{f\in B} \boldsymbol{P}(H_0 \text{ is rejected against } f) = 1.$$

  - Optimal rate of testing is fastest rate at which r_n can approach 0 while maintaining this relation.
- Optimal rate for smoothness class with s bounded derivatives is $n^{-2s/(4s+d)}$ if s is known *a priori* and $\left(n^{-1}\sqrt{\log\log n}\right)^{2s/(4s+d)}$ otherwise.

OBJECTIVE OF TALK

- Develop test that has optimal rate of testing uniformly over Hölder classes and does not require *a priori* knowledge of s, order of differentiability of f.

  - Test called *adaptive* and *rate optimal* because it adapts to unknown s and achieves optimal rate of testing.

- Test that achieves optimal rate of testing has advantage of being sensitive to alternatives uniformly over a smoothness class whose distance from H_0 converges to 0 at fastest possible rate

  - These classes contain sequences of smooth alternatives against which existing tests are inconsistent.

  - In practice, this means that there are smooth alternatives against which these tests have much lower finite-sample power than does a test that achieves the optimal rate of testing.

AVOIDING A POTENTIAL DRAWBACK OF THE MINIMAX APPROACH

- A test that achieves optimal rate uniformly over B is necessarily oriented toward alternatives in B that are most extreme and hardest to detect.
  - These functions have narrow peaks or valleys whose widths decrease with increasing n (high-frequency alternatives)
  - Test that is oriented toward such alternatives may have low power against functions that are less extreme
- To guard against this problem, we develop test that is consistent against local alternatives $f_n(x) = F(x,\theta_1) + \rho_n g(x)$ whenever $\rho_n \geq Cn^{-1/2}\sqrt{\log\log n}$ for some finite $C > 0$.
  - Tests of Andrews (1997) and others are consistent against such alternatives whenever $\rho_n \to 0$ more slowly than $n^{-1/2}$.
  - In terms of consistency, is essentially no penalty paid for adaptiveness and rate optimality.

OUTLINE OF REMAINDER OF TALK

- Test statistic and Monte Carlo method for finding critical values
- Theorems giving properties of test under H_0 and various forms of H_1
- Monte Carlo experiments that
  - Illustrate numerical performance of test
  - Compare its finite-sample power with powers of some existing tests

IDEA OF TEST STATISTIC

- Similar to Härdle-Mammen (1993) statistic
  - Based on distance between kernel nonparametric estimator of f and kernel-smoothed parametric estimator
  - Compute distance with many different values of bandwidth of kernel smoother
  - Reject H_0 if distance obtained with any of the bandwidths is too large
  - Rate-optimal and adaptive properties of the test arise from use of many bandwidths
- Parametric model:
  - θ_0 = True value of θ if H_0 is true. Under H_0, $\boldsymbol{E}(Y_i) = F(X_i, \theta_0)$
  - θ_n = Estimator of θ that is $n^{1/2}$- consistent if H_0 is true
  - Under H_1, $n^{1/2}(\theta_n - \theta^*) = O_p(1)$ as $n \to \infty$ for some θ^*
- Assume that ε has finite, nonzero variance $\sigma^2(X_i)$

KERNEL SMOOTHER AND DISTANCE MEASURE

- For kernel K and bandwidth h, define $K_h(x) = K(x/h)$ and

- $w_h(X_i, X_j) = \dfrac{K_h(X_i - X_j)}{\sum_{k=1}^{n} K_h(X_i - X_k)}$ (weights)

- $a_{ij,h} = \sum_{k=1}^{n} w_h(X_k - X_i) w_h(X_k - X_j)$

- Nonparametric estimator:

$$f_h(X_i) = \sum_{j=1}^{n} w_h(X_i - X_j) Y_j$$

- Smoothed parametric estimator:

$$F_h(X_i, \theta_n) = \sum_{j=1}^{n} w_h(X_i - X_j) f(X_j, \theta_n)$$

- Distance between the two:

$$S_n(\theta_n) = \sum_{i=1}^{n} [f_h(X_i) - F_h(X_i, \theta_n)]^2$$

TEST STATISTIC

- Centered, Studentized form of $S_h(\theta_n)$:

$$T_h = \left(S_h(\theta_n) - \hat{N}_h\right) \Big/ \hat{V}_h$$

where

$$\hat{N}_h = \sum_{i=1}^{n} a_{ij,h} \sigma_n^2(X_i),$$

$$\hat{V}_h^2 = 2 \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij,h}^2 \sigma_n^2(X_i) \sigma_n^2(X_j),$$

and $\sigma_n^2(X_i)$ = estimator of $\sigma^2(X_i)$ that is consistent under both H_0 and H_1.

- Bandwidths: $H_n = \{h = h_{\max} a^k : h \geq h_{\min},$ $k = 0, 1, 2, \ldots\}$, where $0 < h_{\min} < h_{\max}$, and $0 < a < 1$.

- Reject H_0 if T_h is sufficiently large for any $h \in H_n$.

- Thus, test statistic is

$$T^* = \max_{h \in H_n} T_h$$

HOW TO OBTAIN THE CRITICAL VALUE

- Exact α-level critical value. t_α^* solves $\boldsymbol{P}(T^* > t_\alpha^*) = \alpha$

- T^* is not asymptotically normal or asymptotically pivotal, so critical value cannot be obtained from standard tables or tabulated

- Can be shown, however, that t_α^* is determined by the variances $\sigma^2(X_i)$

  - Value of θ_0 and other features of distributions of the ε_i's do not affect critical value

- An asymptotic critical value, t_α, can be obtained as 1 - α quantile of distribution of T^* induced by model

$$Y_i^* = F(X_i, \theta_n) + \varepsilon_i^*,$$

where $\varepsilon_i^* \sim N[0, \sigma_n^2(X_i)]$.

  - t_α can be computed by Monte Carlo simulation

ESTIMATING $\sigma^2(X_i)$

- Need estimator that is consistent even if H_0 is false
- Case of homoskedastic ε's and one-dimensional X
  - Let $X_{(1)} < X_{(2)} < ... < X_{(n)}$ be ordered sequence of design points
  - Let $Y_{(1)}$, $Y_{(2)}$, ... be similarly ordered values of the Y_i's
  - Then estimate $\sigma^2 = \sigma^2(X_i)$ by

$$\sigma_n^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_{(i+1)} - Y_{(i)})^2$$

- Generalization to multi-dimensional, hetero-skedastic cases possible

REGULARITY CONDITIONS

- Parametric model satisfies standard smoothness conditions
- θ_n is $n^{1/2}$-consistent for θ_0 if H_0 is true, and $n^{1/2}(\theta_n - \theta)$ is bounded in probability if H_0 is false.
- Design points X_i are non-stochastic and are scaled so that $\|X_i\| \le 1$. For for each $h \in H_n$ and some finite C_1, $C_2 > 0$, $C_1 nh^{\boldsymbol{d}} \le M_h(X_i) \le C_2 nh^{\boldsymbol{d}}$, where $M_h(X_i)$ = no. of X_j's such that $\|X_j - X_i\| \le h$
- K is bounded, non-negative, and supported on $[-1,1]^d$. (K is not a higher-order kernel.)
- The ε_i's are independent. $\boldsymbol{E}(\varepsilon_i) = 0$, and satisfy certain conditions on moments through order $4+\delta$ for some $\delta > 0$.
- Bandwidths: $h_{\min} \ge n^{-\gamma}$ for some γ such that $0 < \gamma$ 1/2, and $h_{\max} = C_H(\log \log n)^{-1}$

PROPERTIES OF TEST

Behavior of Test when H_0 Is True

- Theorem 1 (Asymptotic validity of estimated critical value): If H_0 is true, then

$$\lim_{n \to \infty} \boldsymbol{P}(T^* > t_\alpha) = \alpha .$$

Consistency against a Fixed Alternative

- Measure distance between parametric family $\Im$ and f by

$$\rho(f, \Im) = \left\{ \inf_{\theta \in \Theta} n^{-1} \sum_{i=1}^{n} [f(X_i) - F(X_i, \theta]^2 \right\}^{1/2} .$$

- Theorem 2: If there is an n_0 such that $\rho(f, \Im) > c_\rho$ for all $n > n_0$ and some $c_\rho > 0$, then

$$\lim_{n \to \infty} \boldsymbol{P}(T^* > t_\alpha) = 1.$$

PROPERTIES OF TEST (cont.)

Consistency against "Conventional" Local Alternative

- Local alternative: $f_n(x) = F(x,\theta_1) + \rho_n g(x)$

- Assume that g satisfies conditions excluding

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} [f_n(X_i) - F(X_i, \theta_{n,0})]^2 \right\}^{1/2} = o(\rho_n)$$

for some sequence $\{\theta_{n,0}\} \in \Theta$.

  - Rate of convergence of f_n to parametric model is the same as rate of convergence of ρ_n to zero.

- Theorem 3: Let $\{f_n\}$ be sequence of local alternatives with $\rho_n \geq Cn^{-1/2}\sqrt{\log\log n}$ for some $C > 0$. Then

$$\lim_{n \to \infty} \boldsymbol{P}(T^* > t_\alpha) = 1.$$

UNIFORM CONSISTENCY AGAINST SMOOTH ALTERNATIVES

- Hölder class of models:
  - Let $j = (j_1, ..., j_d)$ be multi-index.
  - Define $|j| = \sum_{k=1}^{d} j_k$, $D^j f(x) = \dfrac{\partial^{|j|} f(x)}{\partial x_1^{j_1} \ldots \partial x_d^{j_d}}$.
  - Norm: $\|f\|_{H,s} = \sup_{x \in [-1,1]^d} \sum_{|j| \le s} |D^j f(x)|$
  - Smoothness class: $S(H,s) \equiv \{f : \|f\|_{H,s} \le C_F\}$ for some unknown $s \ge 2$ and $C_F < \infty$
- Define for some $s \ge 2$ and $C_a < \infty$

$$B_{H,n} = \left\{ f \in S(H,s) : \rho(f, \Im) \ge C_a \left(n^{-1} \sqrt{\log \log n}\right)^{2\boldsymbol{s}/(4\boldsymbol{s}+\boldsymbol{d})} \right\}$$

  - This is Hölder class whose distance from H_0 exceeds $C_a \left(n^{-1} \sqrt{\log \log n}\right)^{2\boldsymbol{s}/(4\boldsymbol{s}+\boldsymbol{d})}$

UNIFORM CONSISTENCY (cont.)

- Theorem 4: For all sufficiently large $C_a < \infty$,

$$\lim_{n \to \infty} \sup_{f \in B_{H,n}} \boldsymbol{P}(T^* > t_\alpha) = 1$$

AN EXAMPLE

- Gives parametric model and sequence of alternatives against which adaptive, rate-optimal test is consistent but other existing tests are not.

- Parametric model: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, where X_i's are symmetrical about 0 and $\varepsilon_i \sim N(0,\sigma^2)$ for all i

- Sequence of alternatives:

$$Y_i = X_i + \tau_n^4 \phi(X_i / \tau_n) + \varepsilon_i,$$

where $\varepsilon_i \sim N(0,\sigma^2)$, ϕ is standard normal density, and $\tau_n \propto \left(n^{-1}\sqrt{\log\log n}\right)^{1/9}$.

  - Sequence is contained in $B_{H,n}$ with $s = 2$

  - Distance between f_n and parametric model is

$$\rho(f_n, \Im) \propto \left(n^{-1}\sqrt{\log\log n}\right)^{4/9}$$

- Adaptive, rate-optimal test is consistent against this sequence.

MONTE CARLO EXPERIMENTS

- Null-hypothesis model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i; \quad i = 1, 2, ..., 250$$

where each X_i is sampled from $N(0,25)$ truncated at 5th and 95th percentiles.

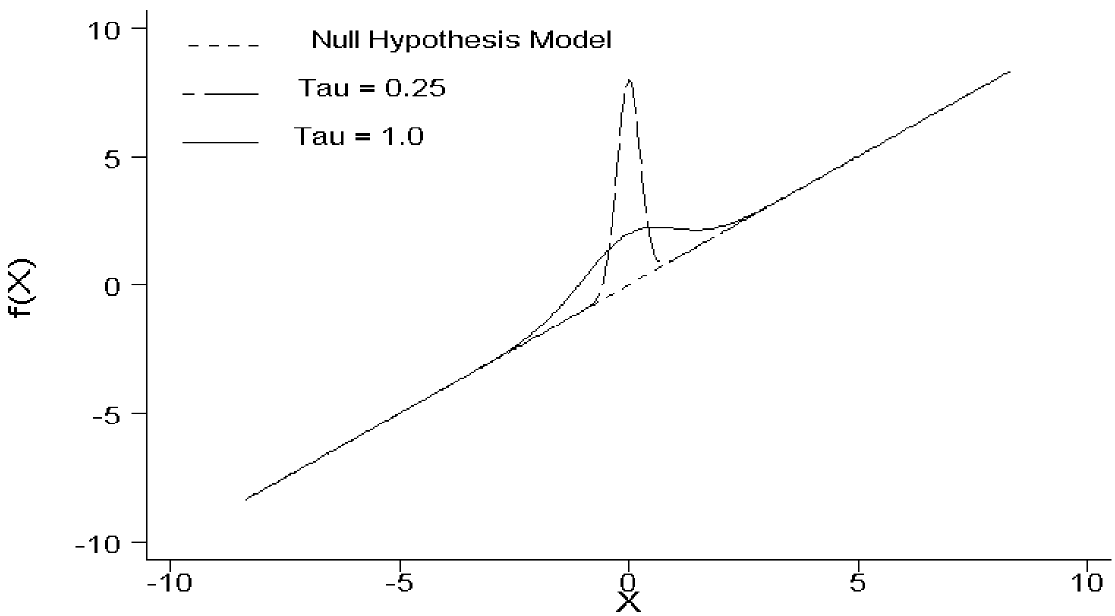
  - If H_0 is true, then $\beta_0 = \beta_1 = 1$
  - ε_i's sampled from three distributions: $N(0,4)$, 90-10 mixture of $N(0,1.56)$ and $N(0,25)$, and Type I extreme value distribution scaled to have variance of 4.

- Alternative models:

$$Y_i = 1 + X_i + (5/\tau)\phi(X_i/\tau) + \varepsilon_i$$

where ε_i's are sampled from one of foregoing distributions and $\tau = 1$ or $\tau = 0.25$.

- Compare power of new test with powers of tests of Härdle and Mammen (1993) and Andrews (1997).
- Expect power of new test higher than powers of others for $\tau = 0.25$ and similar to powers of others for $\tau = 1$.



Null and Alternative Models

RESULTS OF 1000 REPLICATIONS

- When H_0 is true, all tests have empirical rejection probabilities that are close to the nominal rejection probability of 0.05.
- Power of adaptive, rate-optimal test is much higher than powers of other tests when H_0 is false and $\tau = 0.25$.
- Power of adaptive, rate-optimal test is similar to that of Härdle-Mammen test but higher than that of Andrews' test when H_0 is false and $\tau = 1$.
- Results consistent with theory

CONCLUSIONS

- Test of parametric model of a conditional mean function against a nonparametric alternative
  - Adapts to the unknown smoothness of the alternative model
  - Is uniformly consistent against alternative models whose distance from the parametric model converges to 0 at the fastest possible rate
- Test is consistent (not uniformly) against "conventional" local alternatives whose distance from null hypothesis decreases at rate that is only slightly slower than $n^{-1/2}$.
  - This provides protection against situations in which power of new test is much lower than that of existing tests.