Nearly Optimal Tests when a Nuisance Parameter is Present Under the Null Hypothesis

Graham Elliott, Ulrich K. Müller and Mark Watson
UCSD and Princeton University

October 18, 2012

Motivation

- Recent interest in non-standard inference problems

  1. Weak instruments
  2. Inference involving local-to-unity regressors
  3. Moment inequalities
  4. Regressor selection problems

- How to construct tests with well-defined optimality property?

This paper

- Deals with generic non-standard testing problem
- Derives set of bounds on weighted average power of any valid test
- Suggests algorithm that numerically determines test with weighted average power close the bound
- Derives nearly optimal tests in six non-standard problems

Literature

- Power bound closely related to Minimax Theorem of classical decision theory

  ⇒ discussed and employed in weak instrument problem by Andrews, Moreira and Stock (2008)

- Numerical determination of optimal decision rules and tests

  ⇒ Kempthorne (1987), Sriananthakumar and King (2006), Chiburis (2009)

Example: Nuisance Parameter with Known Sign

- Bivariate normal regression model with non-negative coefficient on control variable z_i

$$y_i = x_i\beta + z_i\delta + \varepsilon_i, \quad \varepsilon_i \sim iid\mathcal{N}(0, \sigma^2), \ \sigma^2 \text{ known}$$

leads via sufficiency argument to testing problem

$$\begin{pmatrix} \hat{\beta} \\ \hat{\delta} \end{pmatrix} = Y = \begin{pmatrix} Y_\beta \\ Y_\delta \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \beta \\ \delta \end{pmatrix}, \Sigma \right)$$

$$H_0 \ : \ \beta = 0, \ \delta \geq 0 \quad \text{vs} \quad H_1 : \beta \neq 0, \ \delta \geq 0.$$

- Can normalize $V[Y_\beta] = V[Y_\delta] = 1$, so problem is effectively indexed by scalar $\rho = \text{Cov}(Y_\beta, Y_\delta)$.

Example: Nuisance Parameter with Known Sign

- Testing problem

$$Y = \begin{pmatrix} Y_\beta \\ Y_\delta \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \beta \\ \delta \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

$$H_0 \quad : \quad \beta = 0,\, \delta \geq 0 \quad \text{vs} \quad H_1 : \beta \neq 0,\, \delta \geq 0.$$

arises more generally (after suitably normalizations) as limiting problem in LeCam's Limits of Experiment theory in LAN model with partial knowledge of a nuisance parameter.

- Parameter of interest is β. Presence of nuisance parameter δ makes both null and alternative hypothesis composite. How to construct optimal test?

Outline

1. Introduction

2. Approximate Least Favorable Distributions: Theory

3. Approximate Least Favorable Distributions: Implementation

4. Applications:

   (a) Nuisance parameter with known sign

   (b) Break date

   (c) Set-identified parameter

   (d) Regressor selection

   (e) Mean of AR(1) with coefficient possibly close to one

Generic Problem

- We observe single observation $Y \in S$ with density $f_\theta(y)$ wrt ν, where $\theta \in \Theta \in \mathbb{R}^k$. Want to test

$$H_0 : \theta \in \Theta_0 \quad \text{against} \quad H_1 : \theta \in \Theta_1 \tag{1}$$

where $\Theta_0 \cap \Theta_1 = \varnothing$ and Θ_0 is not a singleton, so that the null hypothesis is composite.

- Tests are $S \mapsto [0, 1]$ functions, where $\varphi(y)$ indicates rejection probability conditional on $Y = y$.

  If $0 < \varphi(y) < 1$ for some y, then test is randomized.

  Test is of level α if $\sup_{\theta \in \Theta_0} E_\theta[\varphi(Y)] = \sup_{\theta \in \Theta_0} \int \varphi(y) f_\theta(y) d\nu \leq \alpha$.

Weighted Average Power

- Typical, no uniformly most powerful test

- Focus on weighted average power for given weight function F on Θ_1

$$\mathrm{WAP}(\varphi) = \int \left(\int \varphi f_\theta d\nu \right) dF(\theta)$$

- By Fubini's Theorem, WAP is equivalently $\mathrm{WAP}(\varphi) = \int \varphi \left(\int f_\theta dF(\theta) \right) d\nu$, so that testing problem effectively becomes

$$\begin{aligned} H_0 &: \text{ the density of } Y \text{ is } f_\theta,\ \theta \in \Theta_0 \\ H_{1,F} &: \text{ the density of } Y \text{ is } h = \int f_\theta dF(\theta) \end{aligned}$$

- Choice of F matters

Power Bounds

- Testing problem

$$H_0 \ : \ \text{the density of } Y \text{ is } f_\theta,\ \theta \in \Theta_0$$
$$H_{1,F} \ : \ \text{the density of } Y \text{ is } h = \int f_\theta dF(\theta).$$

- **Lemma:** Let φ be any level α test of H_0 against $H_{1,F}$. For any probability distribution Λ, let φ_Λ be the Neyman-Pearson level α test of

$$H_\Lambda : \text{the density of } Y \text{ is } \int f_\theta d\Lambda(\theta)$$

against $H_{1,F}$. Then φ_Λ is at least as powerful as φ.

- **Proof**: Since φ is of level α under H_0, it is also a valid level α test of H_Λ against $H_{1,F}$. But by assumption, φ_Λ is the best level α test in this problem, so its power is at least as high.

- Least favorable distribution Λ^{**}: $\varphi_{\Lambda^{**}}$ is of level α under H_0.

Two Uses for Upper Bounds on Power

1. Compare power bound to power of an *ad hoc* test that is known to control size under H_0. If the power of the *ad hoc* is close to the bound, then it is close to optimal (cf. Müller and Watson (2009)).

2. Use numerical methods to find powerful test. Power bound can tell us when to stop searching.

Approximately Least Favorable Distributions

- Neyman-Pearson tests of H_0 against $H_{1,F}$ are of the form (with continuously distributed LR statistic)

$$\varphi_\Lambda(y) = \begin{cases} 1 \text{ if } h(y) > \text{cv} \int f_\theta(y) d\Lambda(\theta) \\ 0 \text{ if } h(y) < \text{cv} \int f_\theta(y) d\Lambda(\theta) \end{cases}$$

- **Definition:** An ε-*ALFD* is a probability distribution Λ^* on Θ_0 satisfying

  (i) the Neyman-Pearson test with $\Lambda = \Lambda^*$ and $\text{cv} = \text{cv}^*$, φ_{Λ^*}, is of level α under H_{0,Λ^*}, and has power $\bar{\pi}$ against $H_{1,F}$;

  (ii) there exists $\text{cv}^{*\varepsilon} > \text{cv}^*$ such that the test with $\Lambda = \Lambda^*$ and $\text{cv} = \text{cv}^{*\varepsilon}$, $\varphi^\varepsilon_{\Lambda^*}$, is of level α under H_0, and has power of at least $\bar{\pi} - \varepsilon$ against $H_{1,F}$.

- Λ^* not necessarily a good approximation to least favorable distribution Λ^{**}, but by Lemma, $\varphi^\varepsilon_{\Lambda^*}$ has power within ε of the bound.

Numerical Determination of the ALFD

- Discretize the problem by specifying distributions Ψ_i on Θ_0, $i = 1, \cdots, M$

- Let J_N be a subset of N of the M baseline indices, $J_N \subset \{1, 2, \cdots, M\}$, and consider first the simpler problem where it is known that Y is drawn from $f_i = \int f_\theta d\Psi_i(\theta)$, $i \in J_N$ under the null

  – NP test φ_N^* is described by cv* and $p_i^* \geq 0$ with $\sum_{i \in J_N} p_i^* = 1$.

  – $\int \varphi_N^* f_i d\nu \leq \alpha$ for $i \in J_N$ and $\int \varphi_N^* f_i d\nu < \alpha$ only if $p_i^* = 0$.

  $\Rightarrow$ Translate these conditions into a numerical nonlinear optimization problem

- Algorithm seeks J_N so that the corresponding test $\varphi_N^{\varepsilon *}$ with slightly larger critical value cv$^{*\varepsilon}$ has null rejection probability below α under H_0

  $\Rightarrow$ feasibility and magnitude of N depend on problem and Ψ_i

Switching to Standard Tests

- In appropriate parameterization, nonstandard problem typically approaches a standard problem as nuisance parameter δ becomes large, $||\delta|| \to \infty$.
  - In weak instrument problem, large concentration parameter implies that instruments are "almost" strong
  - Large local-to-unity parameter implies that standard stationary theory "almost" applies
  - etc.

Switching to Standard Tests ctd

- Focus on tests of the form

$$\varphi_{D,S,\chi}(y) = (1 - \chi(y))\varphi_D(y) + \chi(y)\varphi_S(y)$$

  with

  - $\chi \mapsto \{0, 1\}$ is a "switching rule" (such as $\chi(y) = \mathbf{1}[||\hat{\delta}|| > K]$)
  - φ_S is a "Standard" test
  - φ_D is the test for the "Difficult" part of the parameter space

- Positive nuisance parameter example: $\varphi_S(y) = \mathbf{1}[|y_\beta| > 1.96]$, $\chi(y) = \mathbf{1}[y_\delta > 6]$

- Optimality now conditional on "switching rule" as described by χ and φ_S, that is find WAP test maximizing over φ_D.

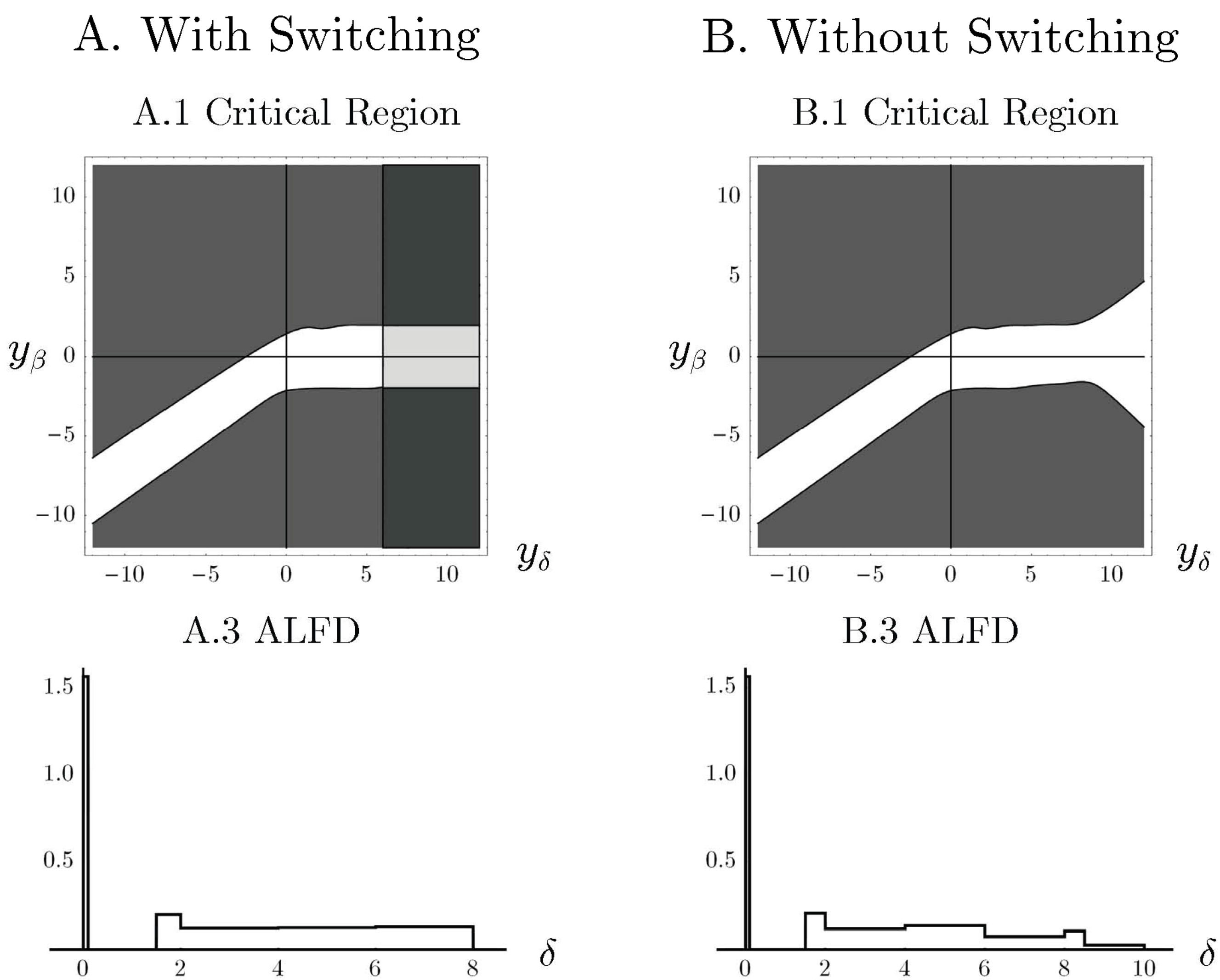
Choice of Weighting Function F

- With switching, F only needs to measure performance in genuinely non-standard part of problem

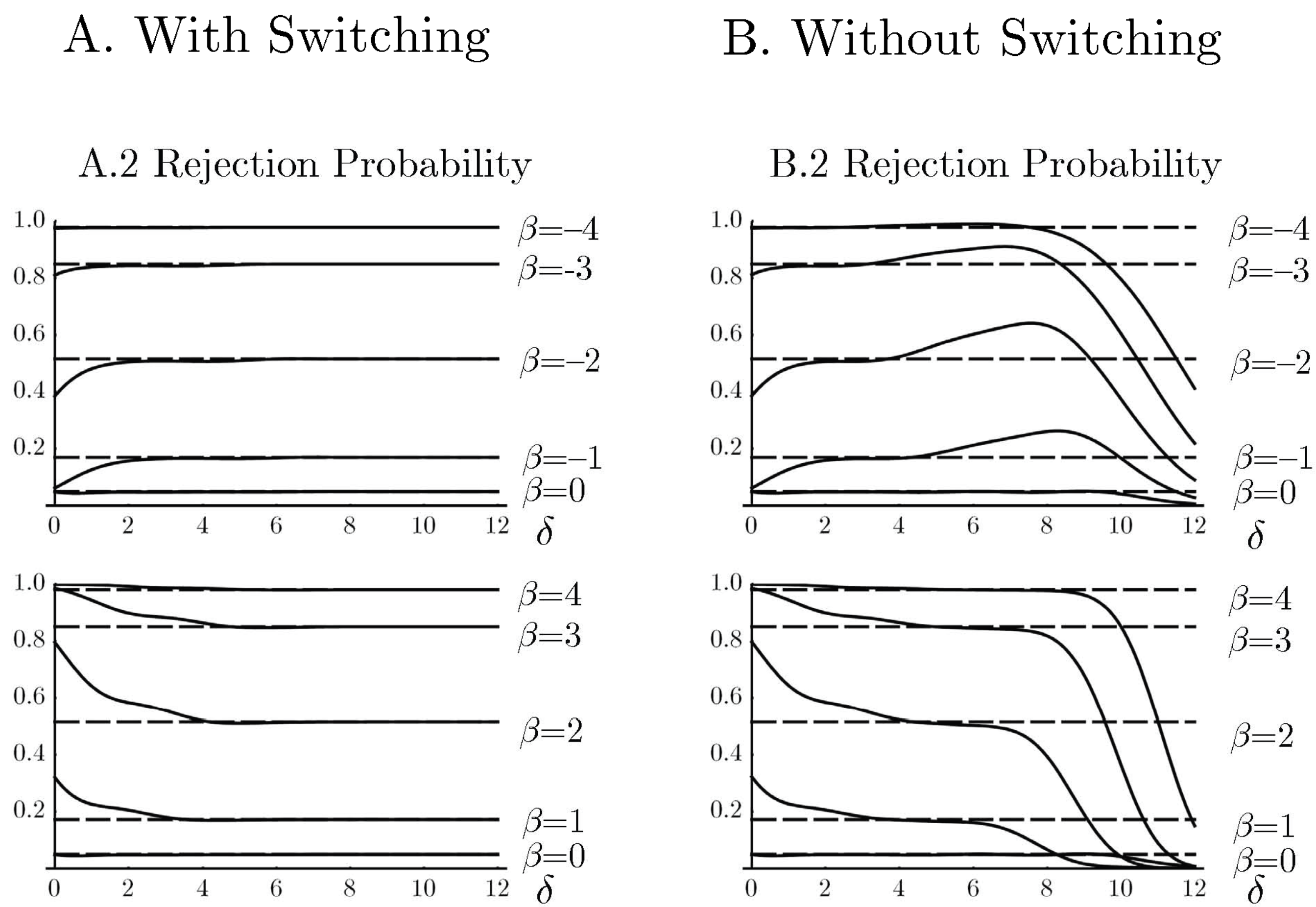
- Our choice of F is guided by
  - ensure smooth transition of critical region across switching boundary
  - in two-sided problems that are symmetric in standard portion, but equal weight on both sides also in non-standard portion
  - focus on alternatives where good 5% level tests achieve power of approximately 50% (cf. King (1988))

$\Rightarrow$ in positive nuisance parameter problem, F is uniform on $\delta \in [0, 8]$, with equal mass on the two points $\beta \in \{-2, 2\}$.

Positive Nuisance Parameter Problem, $\rho = 0.7$



Positive Nuisance Parameter Problem, $\rho = 0.7$



Dashed lines: Power of standard test $\varphi_S(y) = 1[|y_\beta| > 1.96]$

Inference about the Break Date

- Simplest model has

$$y_t = \mu + \mathbf{1}[t \geq \tau]d + \varepsilon_t, \quad \varepsilon_t \sim iid\mathcal{N}(0,1)$$

  and moderate (=contiguous) break magnitude arises as $T^{1/2}d \to \delta \in \mathbb{R}$.

- Limiting problem (after partial summing and invariance to translations) involves single Gaussian process observation G, where

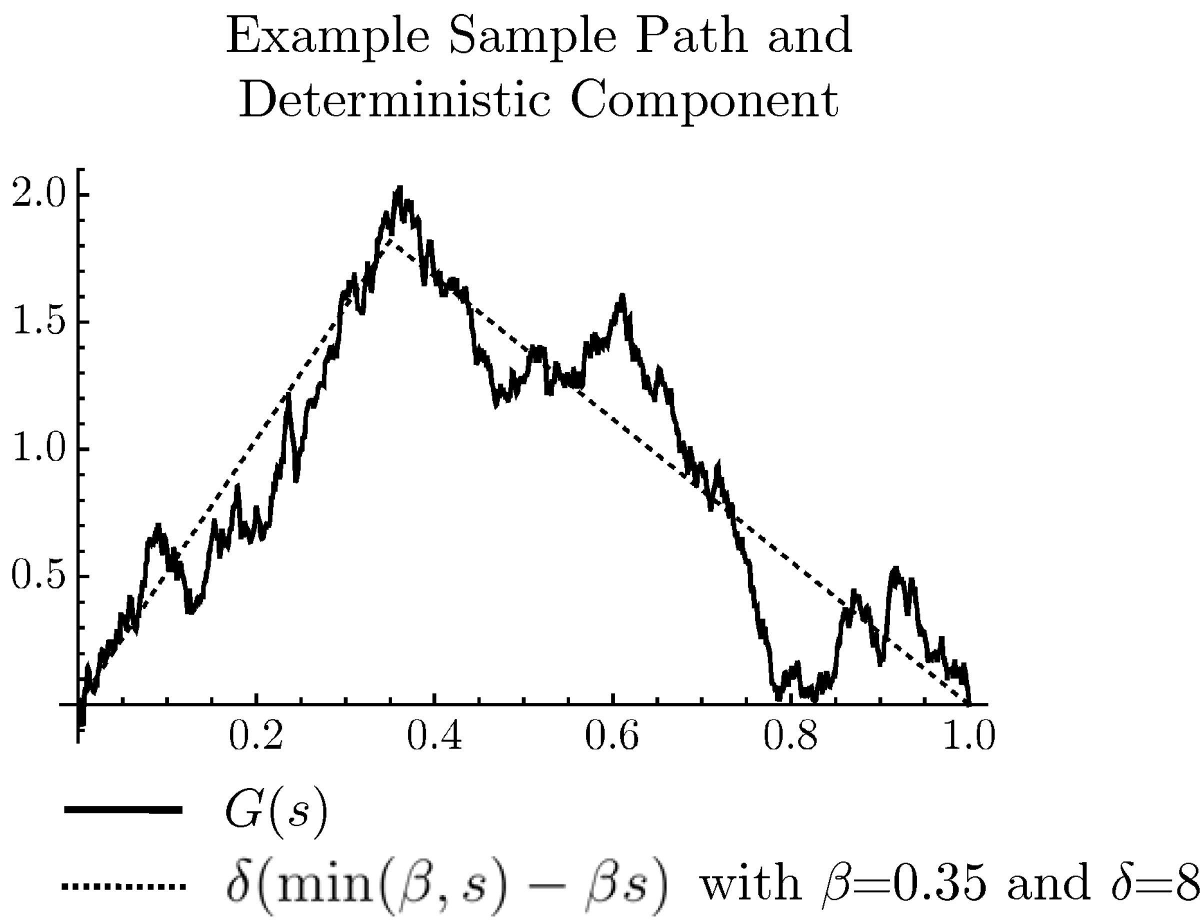
$$G(s) = W(s) - sW(1) - \delta(\min(\beta, s) - \beta s)$$

  W is a standard Wiener process and $\beta = \tau/T$.

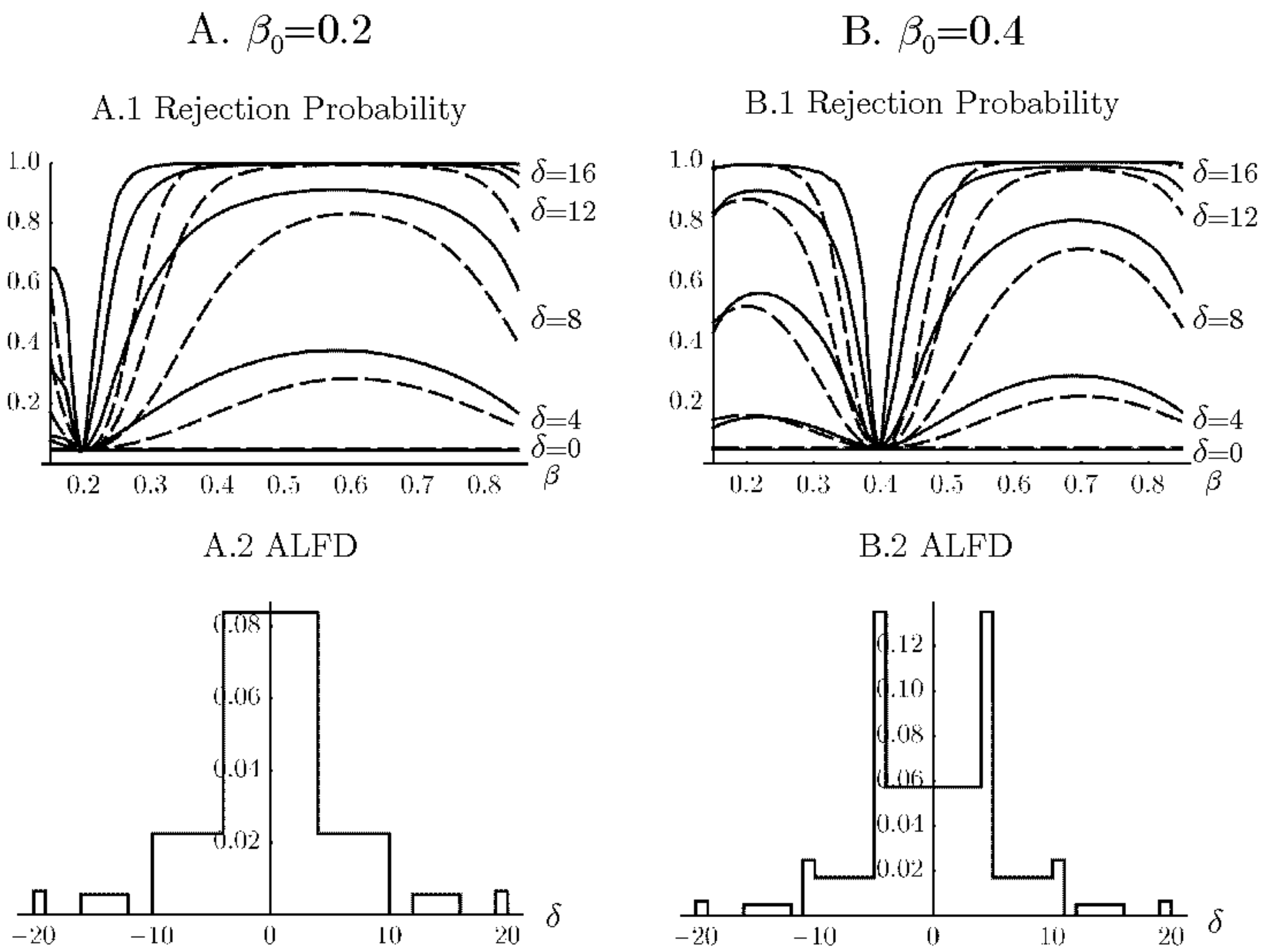
- Testing problem is $H_0 : \beta = \beta_0$, $\delta \in \mathbb{R}$ against $H_1 : \beta \neq \beta_0$, $\delta \in \mathbb{R}$.

- Weighting function F is uniform on $\beta \in [0.15, 0.85]$ and $\delta \sim \mathcal{N}(0, 100)$.

Sample Realization of $G(\cdot)$



Results for Inference about the Break Date



Dashed lines: Power of Elliott and Müller (2007) test

Set Identified Parameter

- Similar to Imbens and Manski (2004), Stoye (2009) and Hahn and Ridder (2011), we observe

$$Y = \begin{pmatrix} Y_l \\ Y_u \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_l \\ \mu_u \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

where $\mu_l \leq \mu_u$, and $\rho \in (-1, 1)$ is known.

- We want to test $H_0 : \mu = 0$, where

$$\mu_l \leq \mu \leq \mu_u$$

so that $[\mu_l, \mu_u]$ is identified set.

Set Identified Parameter ctd

- Reparametrize (μ_l, μ_u) in terms of $(\beta, \delta, \tau) \in \mathbb{R}^3$ as follows:
  - $\delta = \mu_u - \mu_l$ is length of identified set $[\mu_l, \mu_u]$,
  - β is distance of identified set $[\mu_l, \mu_u]$ from 0
  - $\tau = -\mu_l$.

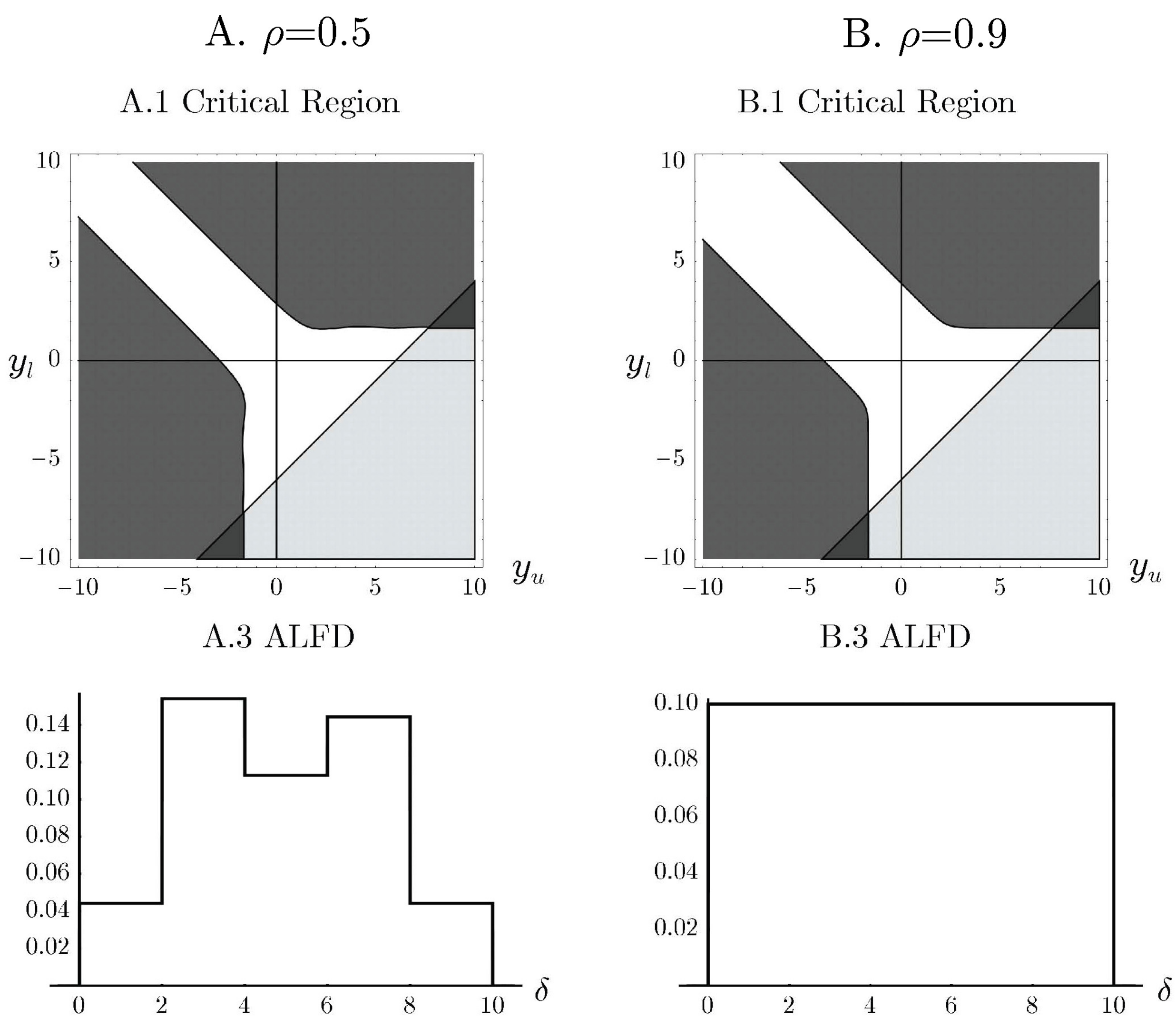
  $\Rightarrow$ Hypothesis testing problem becomes

$$H_0 : \beta = 0,\ \delta \geq 0,\ \tau \in [0, \delta] \quad \text{against} \quad H_1 : \beta > 0,\ \delta \geq 0.$$

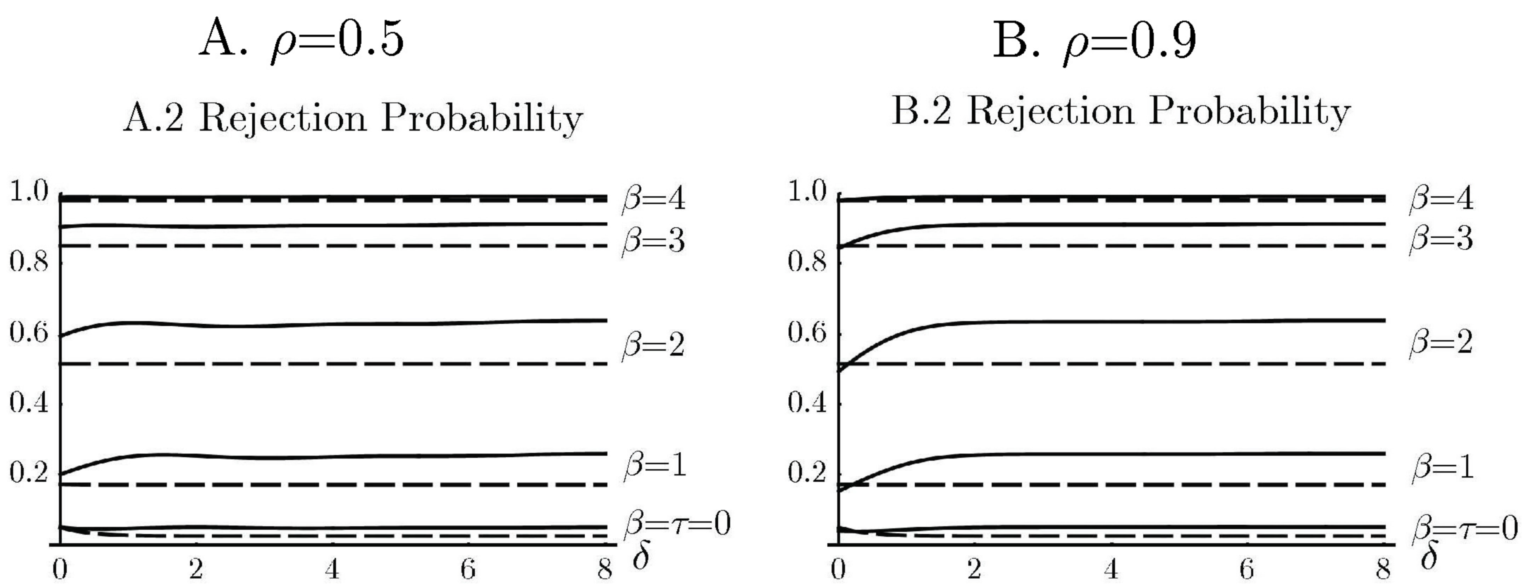
- Switch to $\varphi_S(y) = \mathbf{1}[y_l > 1.645 \text{ or } y_u < -1.645]$ according to $\chi(y) = \mathbf{1}[\hat{\delta} > 6]$, where $\hat{\delta} = Y_u - Y_l$.

- F is chosen to be uniform on $\delta \in [0, 8]$, with equal mass on the two points $\beta \in \{-2, 2\}$.

Results for Set Identified Parameter



Results for Set Identified Parameter



Dashed lines: Stoye's (2009) test $\varphi_{\mathsf{ST}}(y) = \mathbf{1}[y_l > 1.96 \text{ or } y_u < -1.96]$

Regressor Selection Problem

- Bivariate normal regression model, necessity of control variable z_i in doubt

$$y_i = x_i\beta + z_i\delta + \varepsilon_i, \quad \varepsilon_i \sim iid\mathcal{N}(0, \sigma^2), \ \sigma^2 \text{ known}$$

leads via sufficiency and suitably normalization to testing problem

$$\begin{pmatrix} \hat{\beta} \\ \hat{\delta} \end{pmatrix} = Y = \begin{pmatrix} Y_\beta \\ Y_\delta \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \beta \\ \delta \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

$$H_0 \ : \ \beta = 0, \delta \in \mathbb{R} \quad \text{vs} \quad H_1 : \beta \neq 0, \delta = 0$$

(weighting function F puts all mass at $\delta = 0$).

- Coefficient of "short" regression of y_i on x_i corresponds to $Y_\beta - \rho Y_\delta$.

- Known uniformity issues with data driven model selection (Leeb and Pötscher (2005), etc.)

One-sided Problem

- In one-sided problem

$$Y = \begin{pmatrix} Y_\beta \\ Y_\delta \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \beta \\ \delta \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

$$H_0 : \beta = 0,\ \delta \in \mathbb{R} \quad \text{vs} \quad H_1 : \beta = \beta_1 > 0,\ \delta = 0$$

  exact least favorably distribution Λ^{**} has point mass at $(\beta, \delta) = (0, -\rho\beta_1)$, leading to the test $\mathbf{1}[Y_\beta > \text{cv}]$.

- Analytical result: uniformly most powerful one-sided test rejects for large values of Y_β, that is uniformly best inference under size constraint corresponds to simply running the long regression.

Two-sided Problem

- Corresponding two-sided analytical result for

$$Y = \begin{pmatrix} Y_\beta \\ Y_\delta \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \beta \\ \delta \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

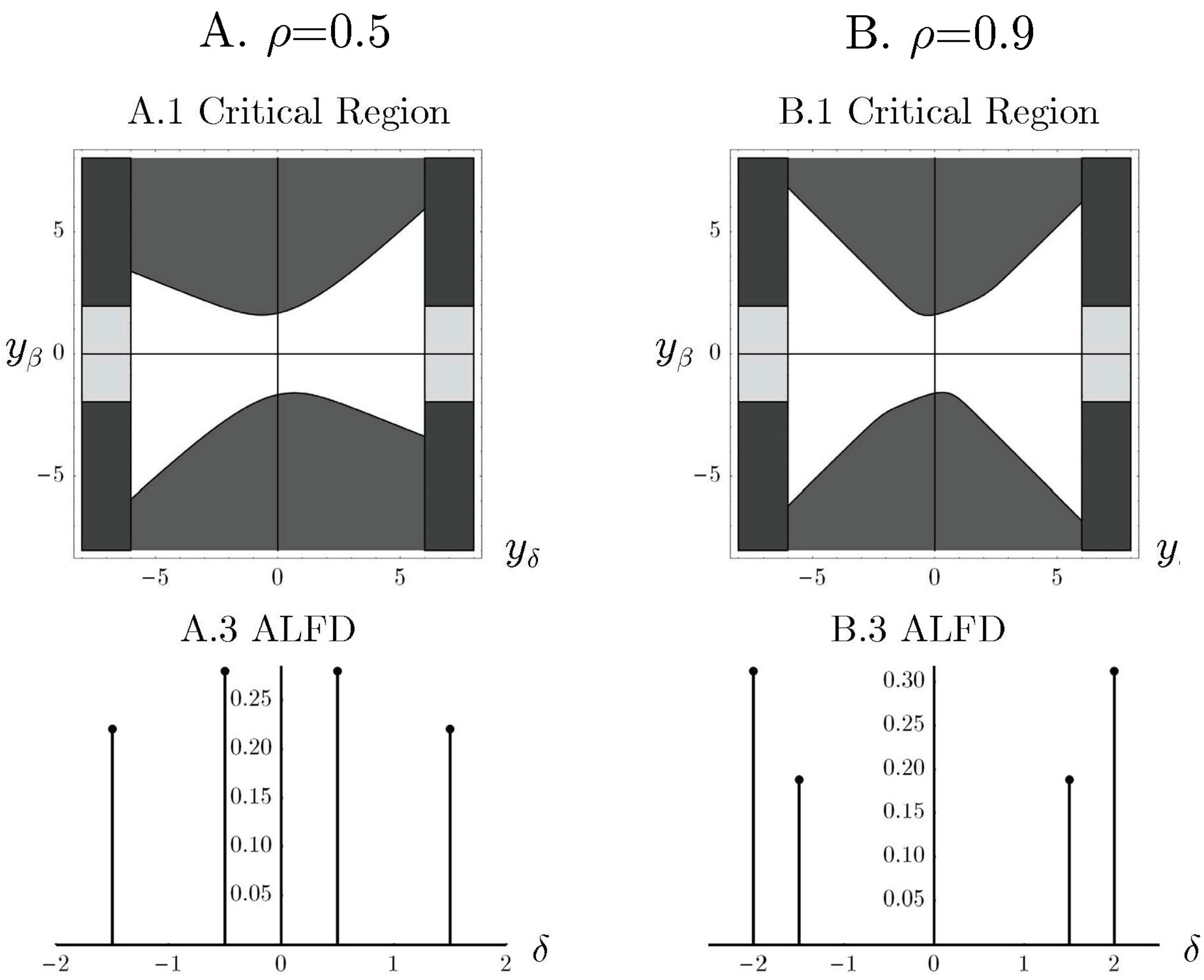
$$H_0 \ : \ \beta = 0,\ \delta \in \mathbb{R} \quad \text{vs} \quad H_1 : \beta \neq 0,\ \delta = 0$$

only holds under unbiasedness constraint. $\Rightarrow$ WAP maximizing (biased) test?

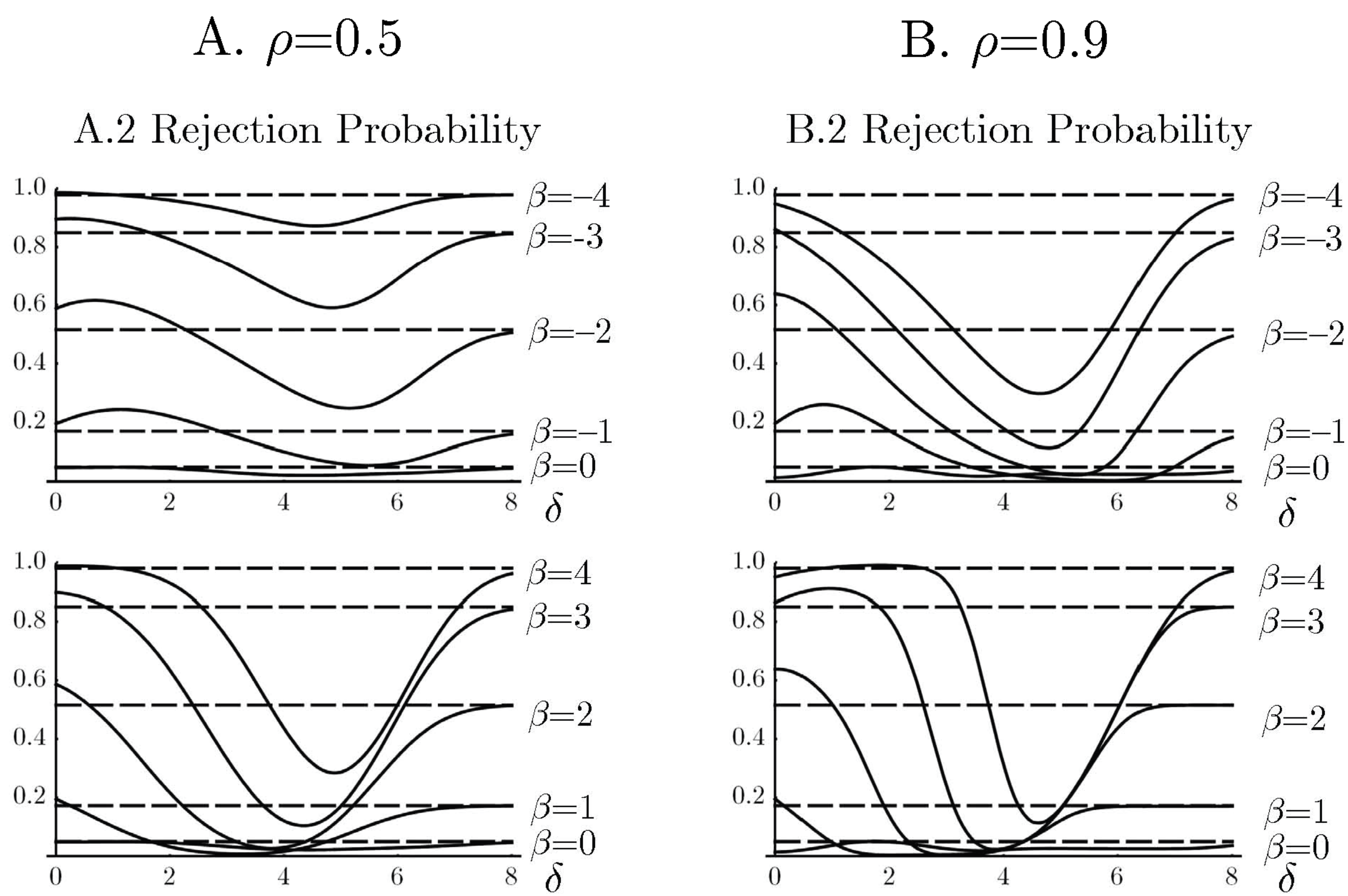
- Switch to $\varphi_S(y) = \mathbf{1}[|y_\beta| > 1.96]$ according to $\chi(y) = \mathbf{1}[|y_\delta| > 6]$.

- F puts equal mass at the two points $\beta \in \{-2, 2\}$.

Numerical Results

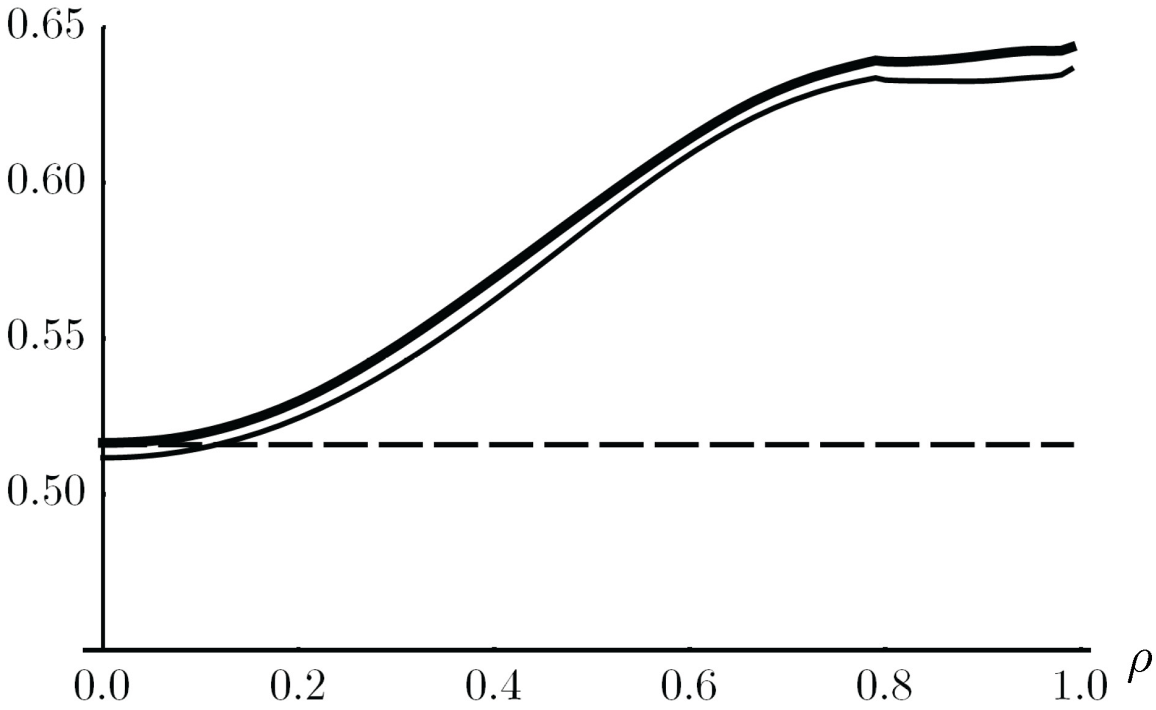


Numerical Results



Dashed line: Power of standard test $\varphi_S(y) = \mathbf{1}[|y_\beta| > 1.96]$

WAP as Function of ρ



Thick through line: upper bound

Thin through line: nearly optimal test

Dashed line: standard test $\varphi_S(y) = \mathbf{1}[|y_\beta| > 1.96]$

Inference about Mean of AR(1) Process

- We observe $\{y_t\}_{t=1}^T$, which is a stationary Gaussian AR(1) with mean μ and unknown coefficient $\rho \in [0,1)$. Optimal inference about μ?

- Under local-to-unity asymptotics $\rho = \rho_T = 1 - \delta/T$, asymptotically identical to nonstandard problem of inference about mean of stationary Ornstein-Uhlenbeck process

- Special case of optimal "HAC" test for specific assumption about auto-correlation structure (which is such that no consistent HAC estimator exists).

- Weighting function F is uniform on $\delta \in (0,80)$, and $\sqrt{T}\mu \sim \mathcal{N}(0, 9/(1-\rho)^2)$. Switch to standard HAC test if $\hat{\delta} > 50$.

Small Sample Results

	Mean of AR(1)				Regression			
	A91	AM92	KVB	φ_{Λ^*}	A91	AM92	KVB	φ_{Λ^*}
ρ	size							
0.00	0.05	0.05	0.05	0.05	0.06	0.06	0.05	0.06
0.70	0.10	0.07	0.06	0.07	0.10	0.08	0.07	0.08
0.90	0.17	0.11	0.09	0.08	0.18	0.13	0.11	0.08
0.95	0.26	0.15	0.13	0.08	0.26	0.16	0.15	0.06
0.98	0.44	0.30	0.23	0.06	0.37	0.18	0.22	0.05
	size adjusted power							
0.00	0.50	0.50	0.37	0.50	0.50	0.50	0.36	0.50
0.70	0.74	0.75	0.57	0.76	0.95	0.94	0.81	0.94
0.90	0.96	0.96	0.87	0.88	1.00	0.99	0.98	0.95
0.95	1.00	0.99	0.96	0.42	1.00	0.99	1.00	0.67
0.98	1.00	1.00	1.00	0.75	1.00	0.99	1.00	0.44

$T = 200$. 'Regression' has single AR(1) regressor and independent AR(1) disturbance, and includes a constant.

Decision Theoretic and Bayesian Interpretation

- Suppose a false rejection of H_0 induces loss 1, a false acceptance of H_F induces loss $L_F > 0$, and a correct decision has loss 0. Then Risk is

$$R(\theta, \varphi) = \mathbf{1}[\theta \in \Theta_0] \int \varphi f_\theta d\nu + L_F \mathbf{1}[\theta \in \Theta_1](1 - \int \varphi h d\nu)$$

  and the test $\varphi_{\Lambda^{**}}$ relative to the (unknown) least favorable distribution Λ^{**} minimizes $\sup_{\theta \in \Theta} R(\theta, \varphi)$ among all tests φ for the specific choice $L_F = \alpha/(1 - \pi^{**})$ with $\pi^{**} = \int \varphi_{\Lambda^{**}} h d\nu$.

- Approximately optimal test φ_{Λ^*} is correspondingly approximately minimax.

- Test $\varphi^{\varepsilon}_{\Lambda^*}$ corresponds to rejecting for large values of the Bayes factor with priors Λ^* on Θ_0 and F on Θ_1.

  $\Rightarrow$ Endogenous determination of Λ^* yields Bayes rule with attractive frequentist properties.

Conclusions

- General constructive method to obtain nearly optimal tests in the weighted average sense for nonstandard problems

- Numerical difficulties of checking size control if nuisance parameter dimension is larger than 2