HAR Inference:
Recommendations for Practice

Eben Lazarus, Harvard
Daniel Lewis, Harvard
Mark Watson, Princeton & NBER
James H. Stock, Harvard & NBER

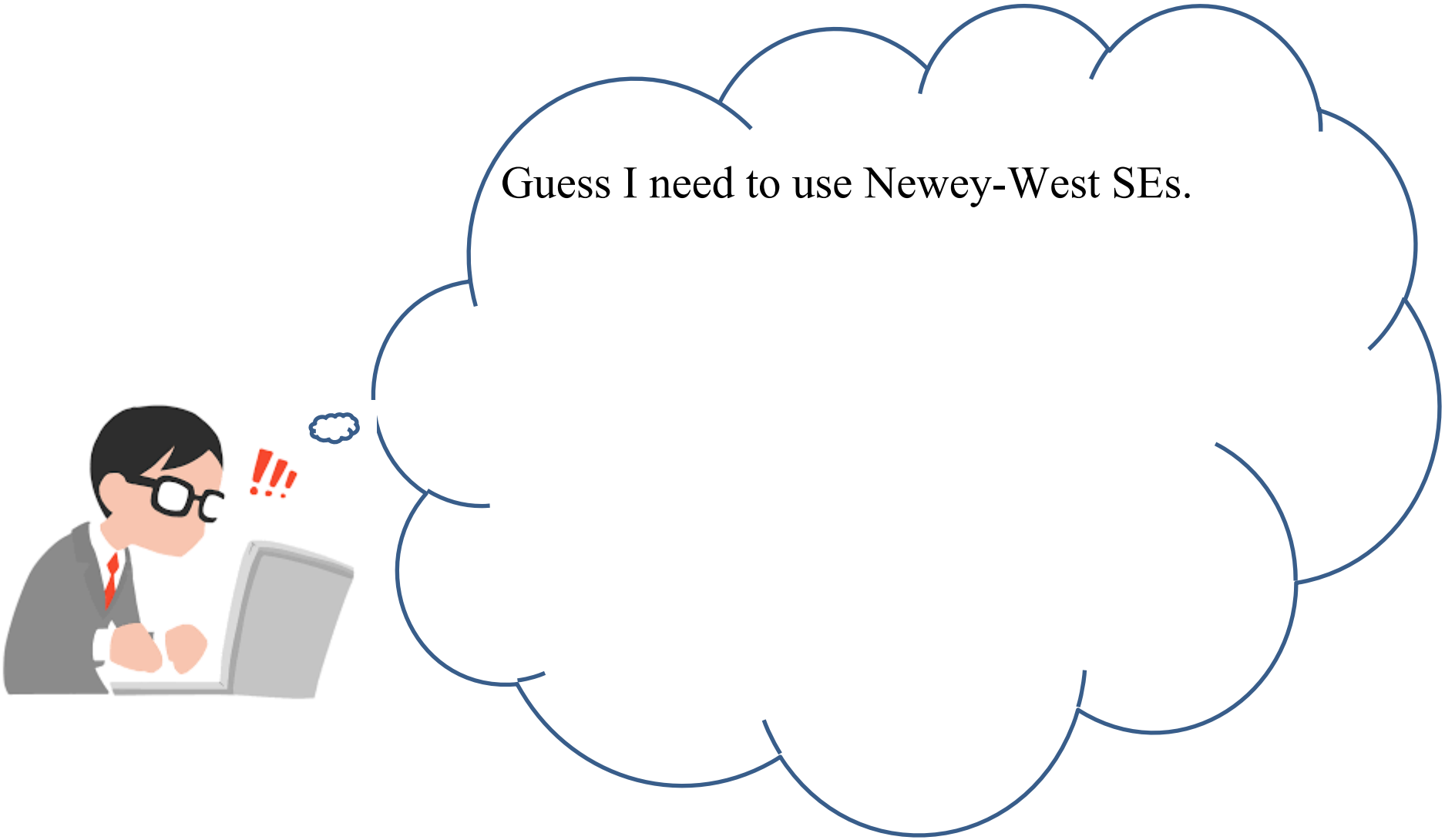
JBES Invited Session

Sunday, Jan. 7, 2018, 1-3 pm
Marriott Philadelphia Downtown, Independence Ballroom II

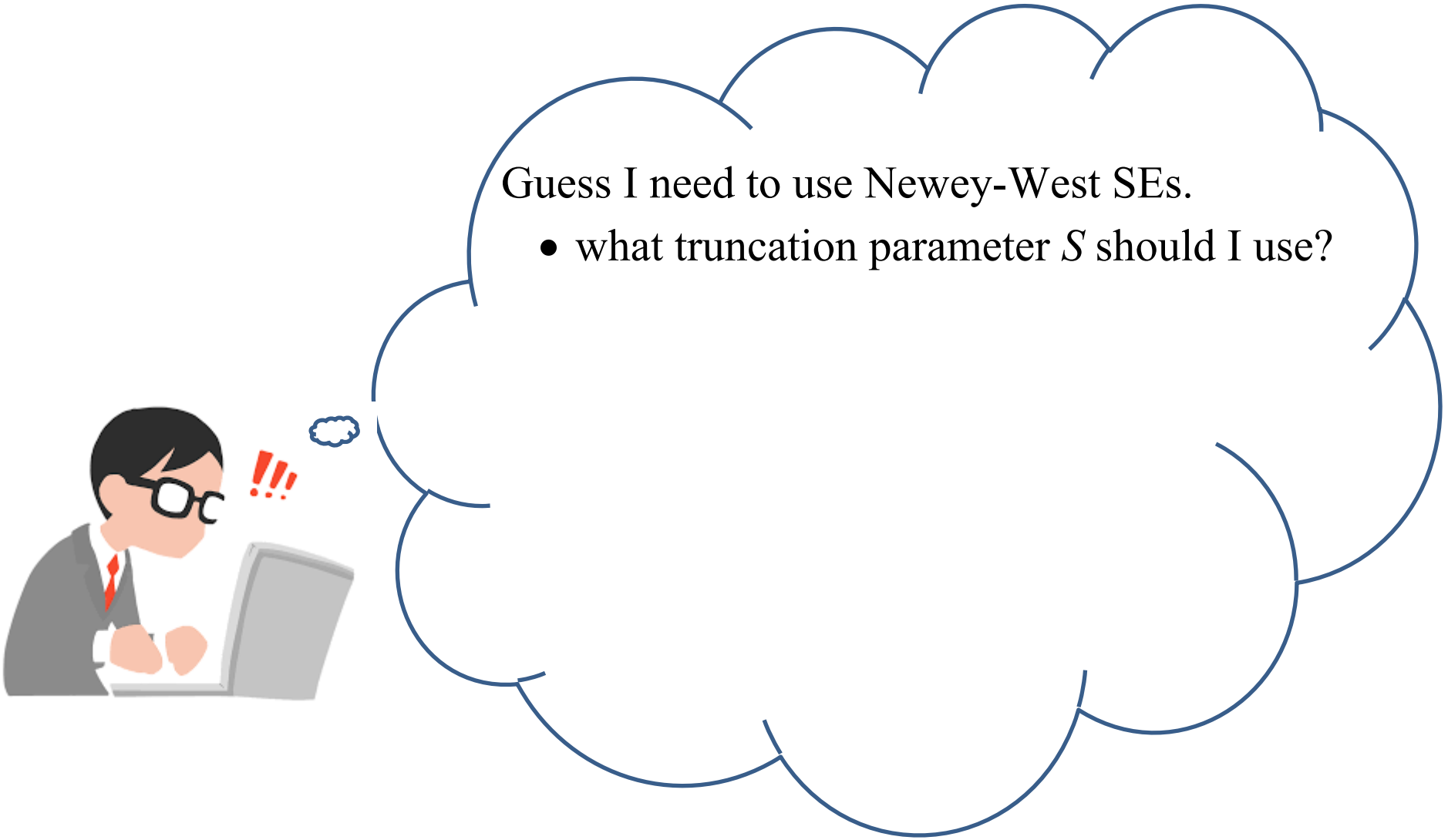
The Heteroskedasticity- and Autocorrelation Robust Inference (HAR) problem

Shoot, my error term is serially correlated!

The HAR problem



The HAR problem



The HAR problem

Guess I need to use Newey-West SEs.

- what truncation parameter *S* should I use?
- what critical value?

The HAR problem

Guess I need to use Newey-West SEs.

- what truncation parameter *S* should I use?
- what critical value?
- what about all those hard papers by Vogelsang, Müller, Sun, and others*?

* Kiefer, Vogelsang, Bunzel (2000), Velasco and Robinson (2001), Kiefer and Vogelsang (2002, 2005), Jansson (2004), Phillips (2005), Müller (2007, 2014), Sun, Phillips, & Jin (2008), Ibragimov and Müller (2010), Sun (2011, 2013, 2014a, 2014b), Gonçalves & Vogelsang (2011), Zhang and Shao (2013), Pötscher and Preinerstorfer (2016, 2017),…

The HAR problem

Guess I need to use Newey-West SEs.

- what truncation parameter *S* should I use?
- what critical value?
- what about all those hard papers by Vogelsang, Müller, Sun, and others*?
- nobody uses them anyway…

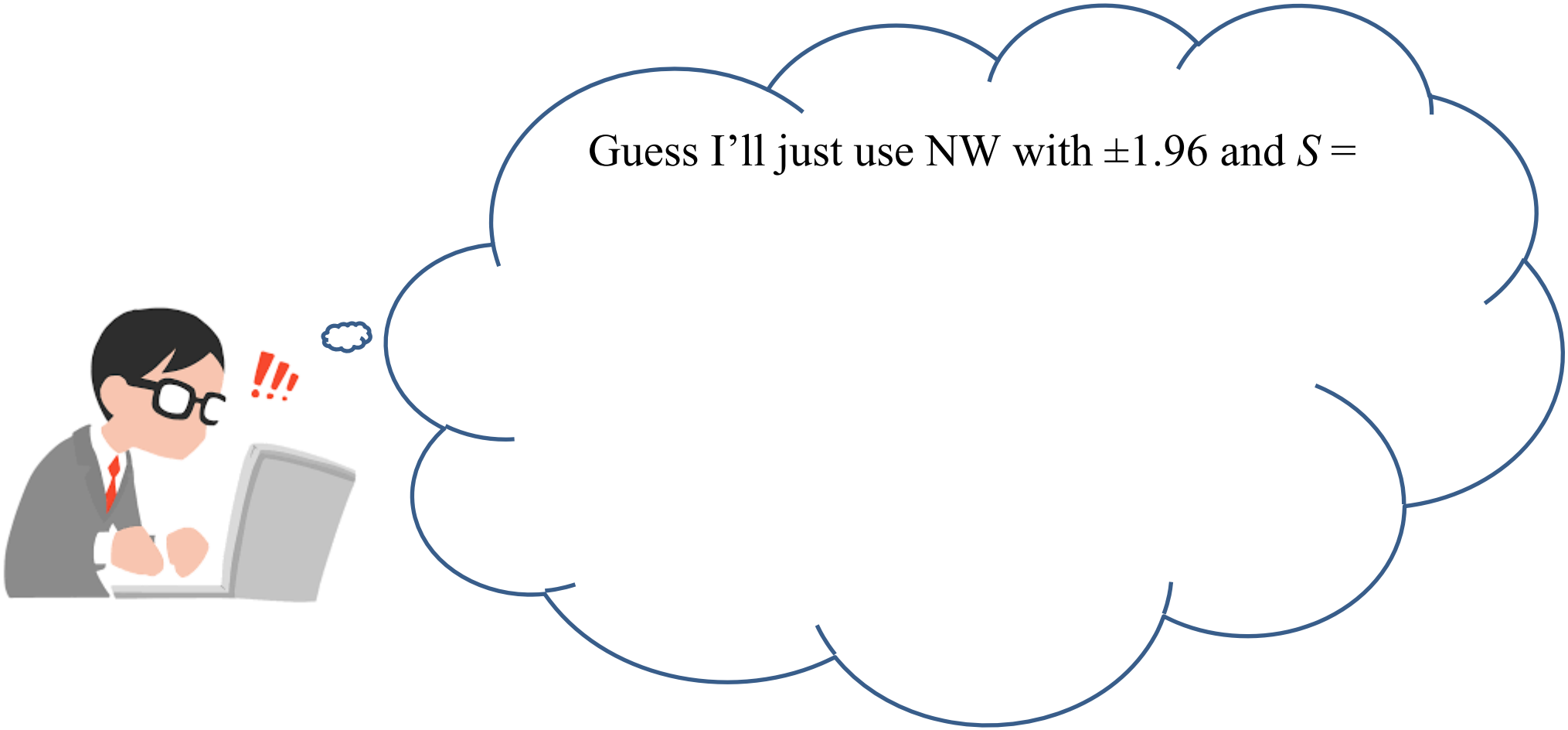
The HAR problem

Guess I need to use Newey-West SEs.

- what truncation parameter *S* should I use?
- what critical value?
- what about all those hard papers by Vogelsang, Müller, Sun, and others*?
- nobody uses them anyway…

  and the referees never complain…

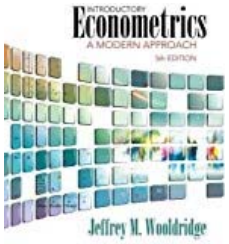
The HAR problem

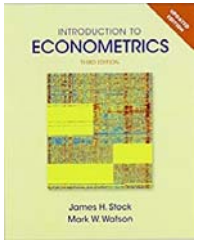


The HAR problem

Guess I'll just use NW with ±1.96 and S =

$4(T/100)^{2/9}$ or $0.75T^{1/3}$

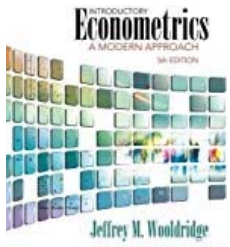


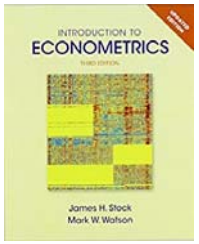


The HAR problem

Guess I'll just use NW with ±1.96 and S =

$4(T/100)^{2/9}$ or $0.75T^{1/3}$





hmm… they give the same answer…
must be OK…

20 years of research says: Bad idea.

Rejection rates of HAR tests with nominal level 5% ($b = S/T$)

$y_t = \beta_0 + \beta_1 x_t + u_t$, x_t & u_t Gaussian AR(1), $\rho_x = \rho_u = 0.7^{1/2}$, $T = 200$

Estimator	Truncation rule for b	Critical values	Null imposed?	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$
NW	$0.75T^{-2/3}$	N(0,1)	No	**0.079**	**0.105**	**0.164**
NW	$1.3T^{-1/2}$	fixed-b (nonstandard)	No	0.067	0.080	0.107
EWP	$1.95T^{-2/3}$	fixed-b (t_ν)	No	0.063	0.074	0.100
NW	$1.3T^{-1/2}$	fixed-b (nonstandard)	Yes	0.057	0.062	0.073
EWP	$1.95T^{-2/3}$	fixed-b (t_ν)	Yes	0.052	0.056	0.066
Theoretical bound based on Edgeworth expansions for the Gaussian location model						
NW	$1.3T^{-1/2}$	fixed-b (nonstandard)	No	0.054	0.058	0.067
EWP	$1.95T^{-2/3}$	fixed-b (t_ν)	No	0.052	0.056	0.073

HAR inference: Newey-West/Andrews approach

$$y_t = \beta' x_t + u_t,\ E(u_t|x_t) = 0,\ (x_t, y_t) \text{ stationarity, } T \text{ observations}$$

- HAR SEs are needed for testing when $z_t = x_t u_t$ is possibly serially correlated and heteroskedastic:

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Sigma_{XX}^{-1} \Omega \Sigma_{XX}^{-1}), \text{ where}$$

$$\Omega = \sum_{j=-\infty}^{\infty} \Gamma_j = 2\pi S_z(0),\ S_z = \text{spectral density of } z_t.$$

- HAR inference entails estimating Ω, the long-run variance (LRV) of z_t
- Kernel estimator:

$$\hat{\Omega}^{SC} = \sum_{j=-(T-1)}^{T-1} k\left(\frac{|j|}{S_T}\right)\hat{\Gamma}_j,\ \ \hat{\Gamma}_j = \frac{1}{T}\sum_{t=1}^{T} \hat{z}_t \hat{z}_{t-j}'$$

  - $\hat{z}_t = x_t \hat{u}_t$, ***k*(.)** = kernel; S = truncation parameter; ***b*** = S/T
  - Standard S choice motivated by minimizing MSE($\hat{\Omega}^{SC}$) (Andrews (1991), Newey-West (1994)) – trades off bias^2 and variance. For Bartlett kernel, $S_T \propto T^{1/3}$. QS is asymptotically optimal.

Literature post NW & Andrews: 4 lessons

1. NW/Andrews inference has large size distortions (Den Haan & Levin (1994, 1997))

2. The culprit is bias of $\hat{\Omega}$, and the solution is using larger S
   - Kiefer, Vogelsang, Bunzel (2000)
   - Velasco & Robinson (2001), Sun, Phillips, & Jin (2008) show this using small-b expansion of rejection rates: size depends on bias($\hat{\Omega}$) & var($\hat{\Omega}$)

3. But larger S reduces quality of normal approximation under null – so use Kiefer-Vogelsang "fixed-b" critical values. Fixed-b critical values provide a higher order refinement (Jansson (2004), Sun (2014)).

4. Even after using fixed-b critical values, there remains the question of which kernel to use, and how to choose the bandwidth constant at the new optimal rate
   - The literature doesn't resolve this size/power tradeoff – so there are no modern, widely accepted, concrete guidelines for practice.

This paper

Goal

To provide concrete recommendations for HAR inference that incorporate the past 20 years of research and improve upon standard practice

Ground rules

1. Time series regression: $y_t = \beta' x_t + u_t$, $E(u_t|x_t) = 0$, (x_t, y_t) stationarity
2. Focus on moderate dependence: $|\rho_z| \leq 0.7$, where ρ_z is AR(1) parm
   - Actually, this parameterizes our bound, which is on the normalized second derivative of the spectral density, $\omega^{(2)} = s_z''(0) / s_z(0)$
   - If x_t, u_t are independent AR(1), then $\rho_z = \rho_x \rho_u$
3. LRV estimator must be psd with known fixed-b asymptotic distⁿ ($b = S/T$)
   - No VARHAC
   - No bootstrap
   - No non-psd kernels with adjustments for psd cases
4. Tests should be asymptotically efficient (so, sequences of b)
5. Ideally, standard critical values (no nonstandard tables)

This paper: methods

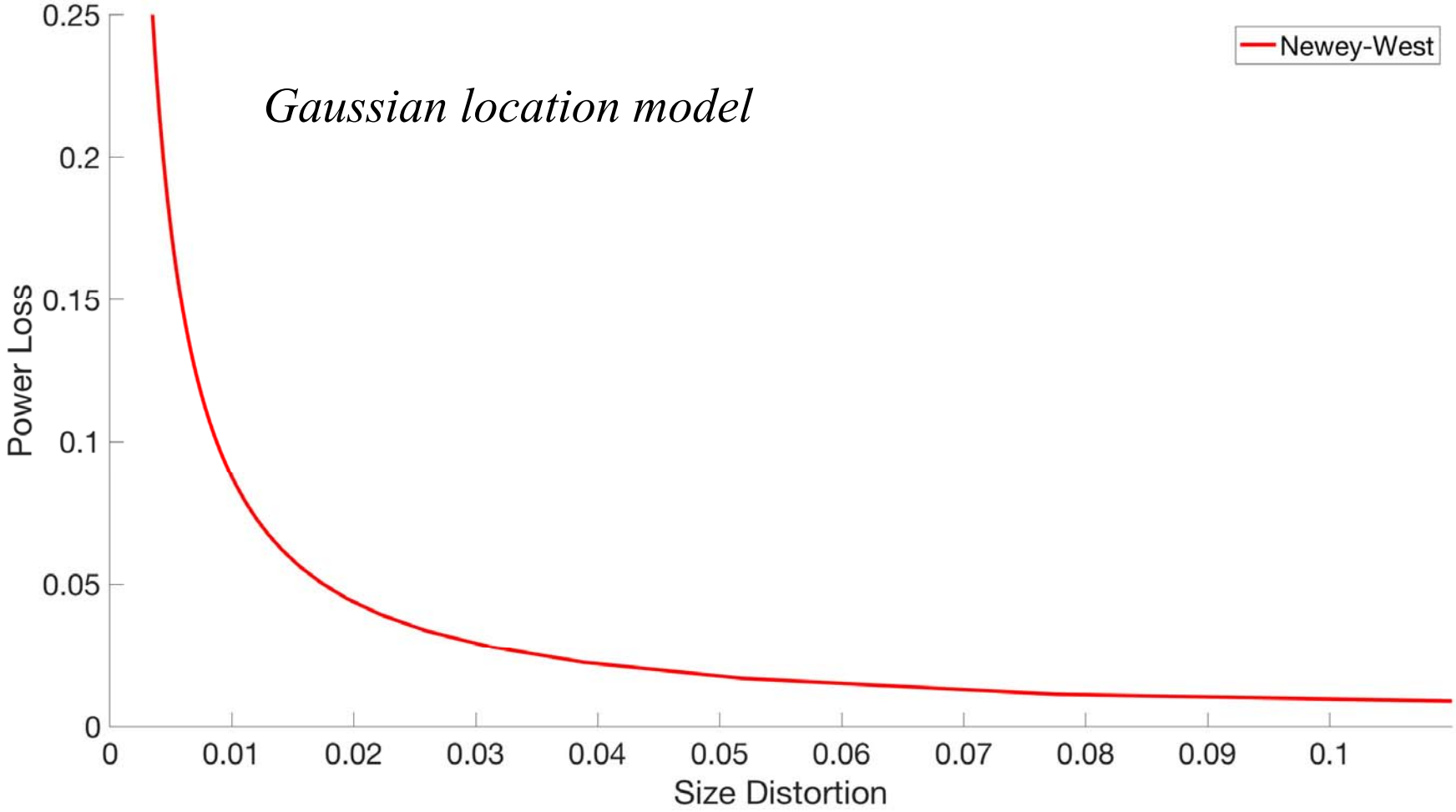
1. Use recent results in Lazarus, Lewis, and Stock (2017) on size-power tradeoff in **Gaussian location model** (no x's) to guide kernel choice.
   a. LLS results are based on Edgeworth expansion literature – Sun (multiple), Sun, Phillips, Jin (2008)
   b. All tests evaluated using fixed-b critical values
   c. LLS provide theoretical tradeoffs & frontier for psd kernels.

2. Posit a user loss function trading off size and power, to make a specific choice

3. Simulation to examine performance with:
   a. non-Gaussian errors
   b. Regression, with and without null imposed ("LM v. Wald")

Our argument, in pictures, for the Gaussian location model…

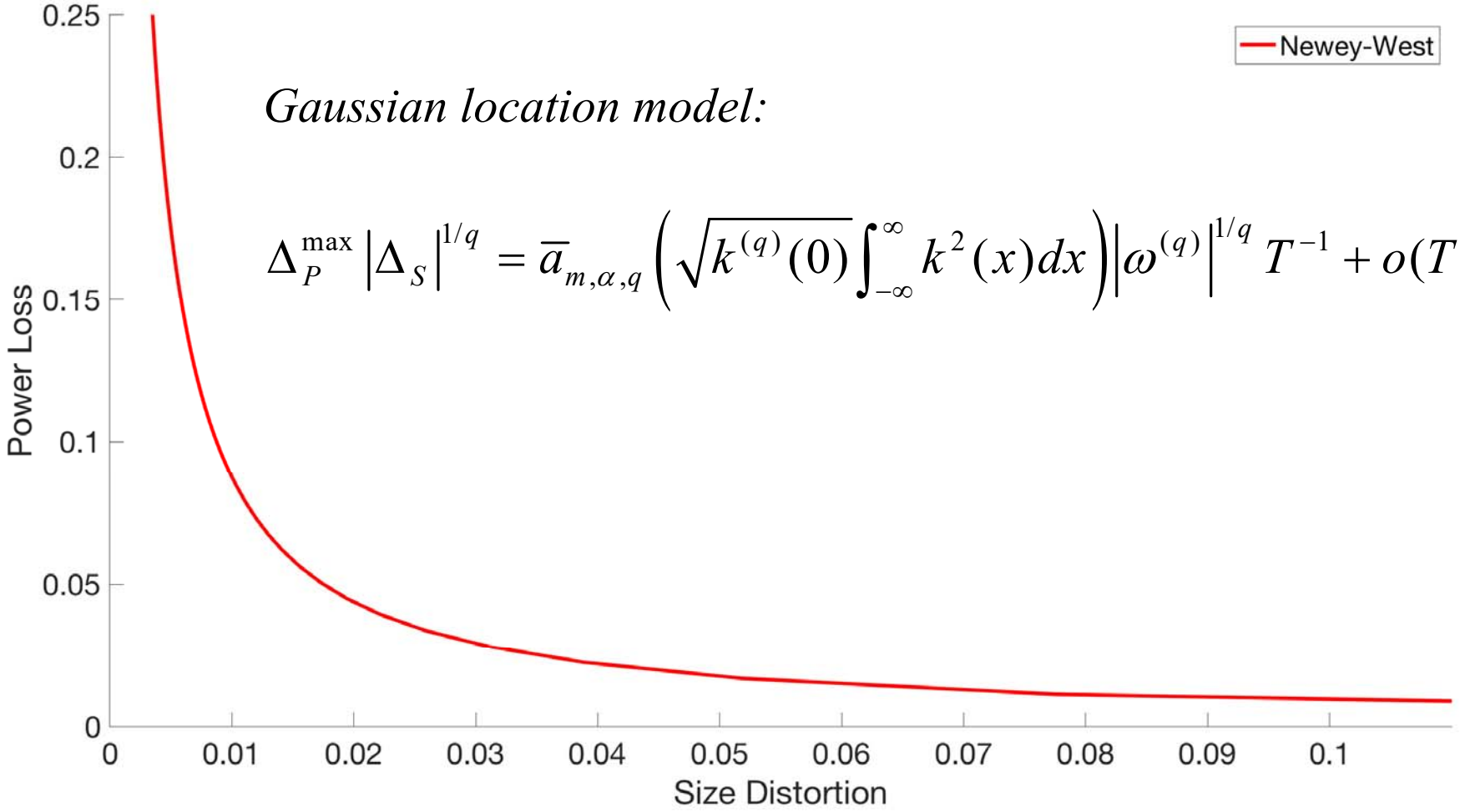
Theoretical size-power tradeoff, AR(1), $\rho_z = 0.7$, $T = 200$



x axis: Δ_S = |rejection rate| – 0.05 = size distortion

y axis: $\Delta_P^{\max}$ = maximum power loss, compared to oracle (Ω known) test with same second-order size

Theoretical size-power tradeoff, AR(1), $\rho_z = 0.7$, $T = 200$



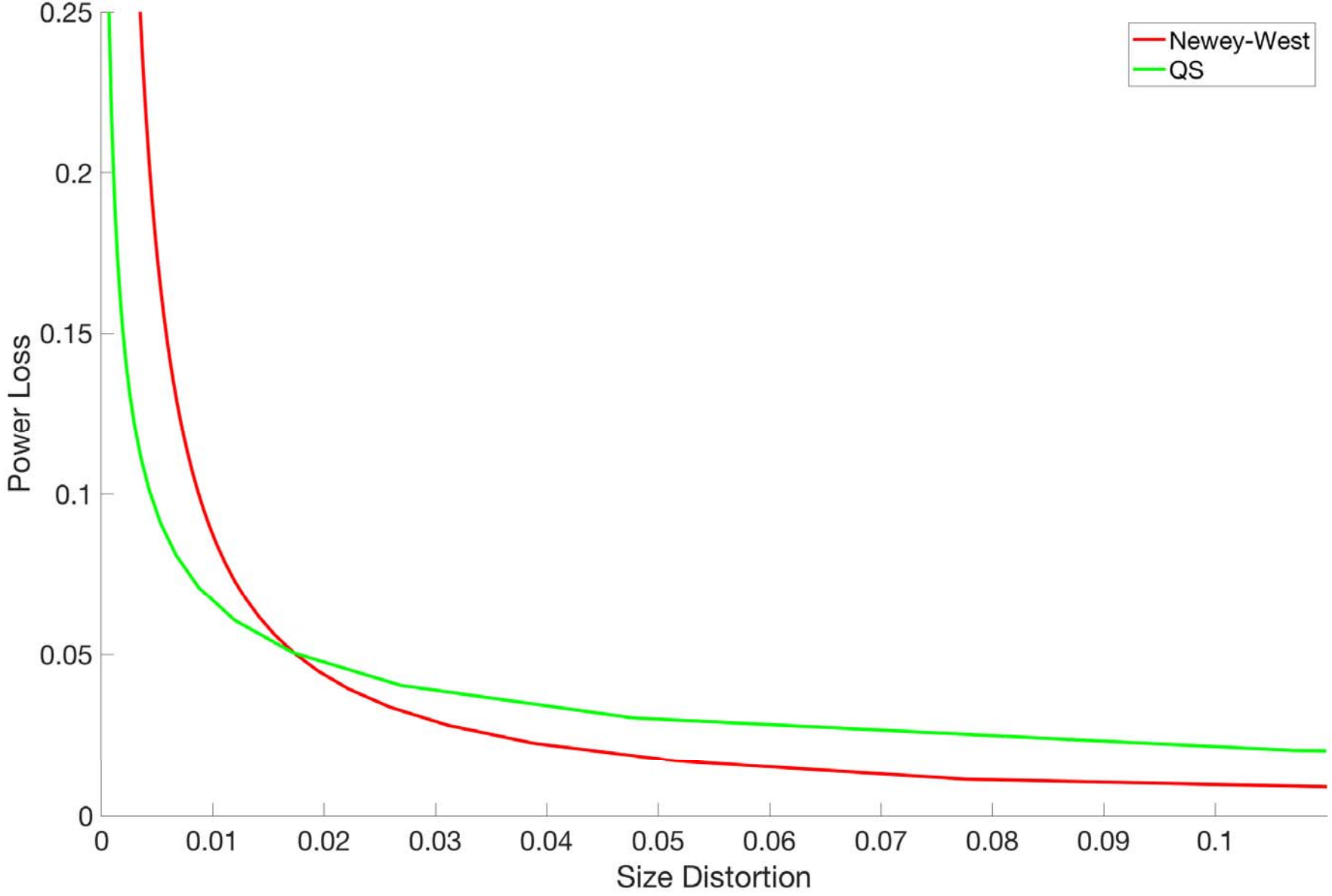
Gaussian location model:

$$\Delta_P^{\max} \left|\Delta_S\right|^{1/q} = \overline{a}_{m,\alpha,q} \left(\sqrt{k^{(q)}(0)} \int_{-\infty}^{\infty} k^2(x)dx \right) \left|\omega^{(q)}\right|^{1/q} T^{-1} + o(T^{-1}),$$

x axis: Δ_S = |rejection rate| – 0.05 = size distortion

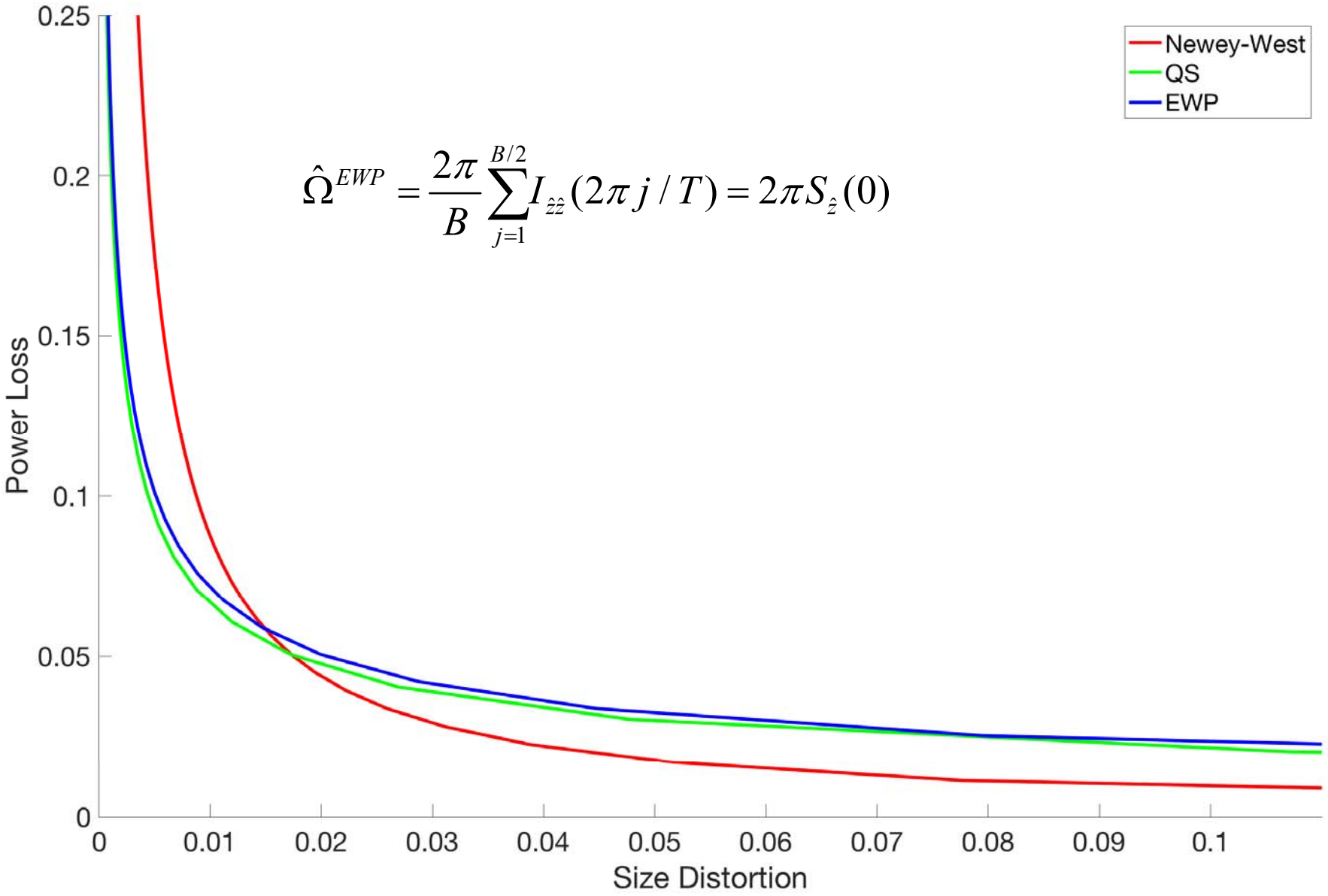
y axis: $\Delta_P^{\max}$ = maximum power loss, compared to oracle (Ω known) test with same second-order size

…and for QS (Epanechnikov) kernel

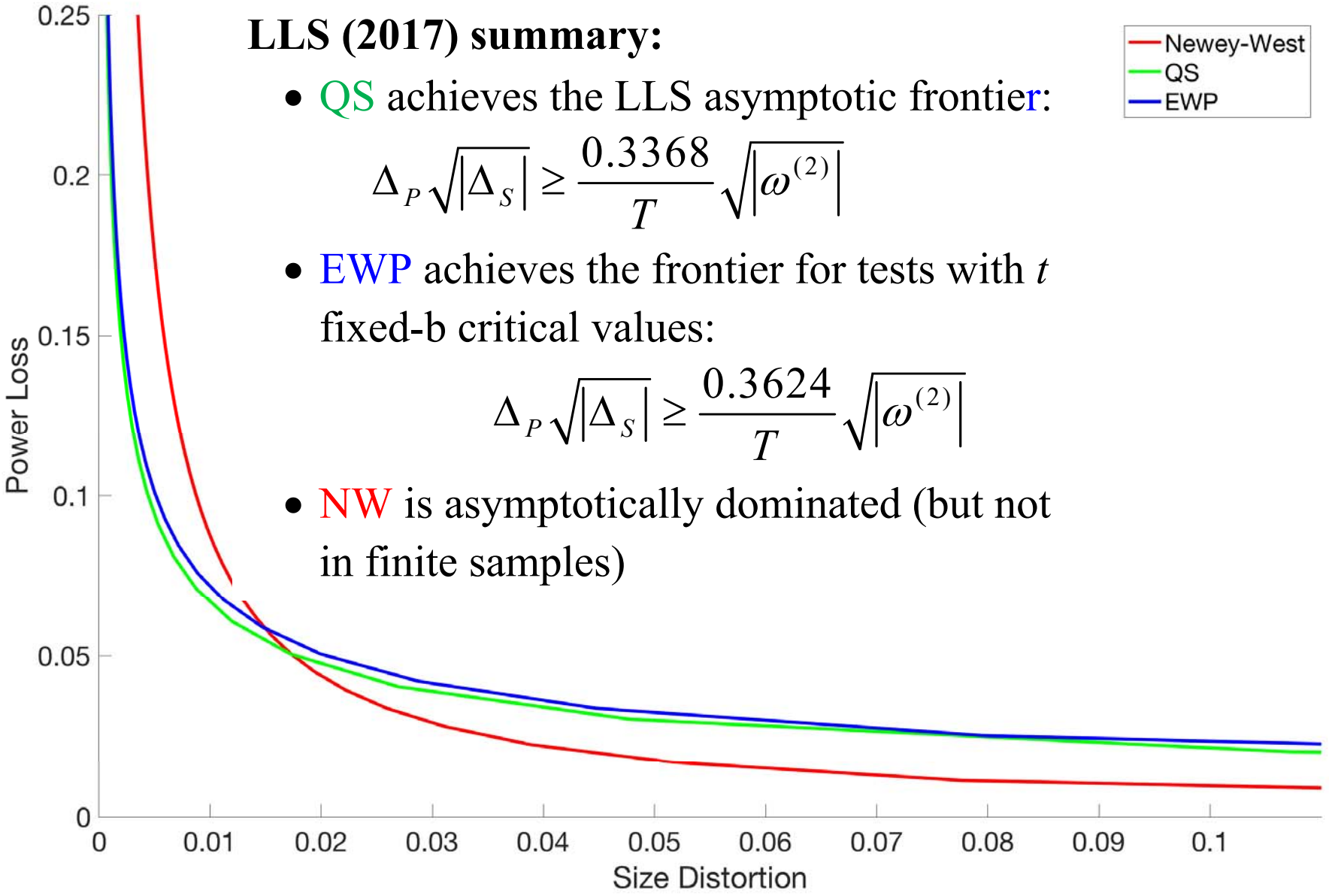


…and for Equal-weighted periodogram (EWP) kernel

$$\hat{\Omega}^{EWP} = \frac{2\pi}{B}\sum_{j=1}^{B/2} I_{\hat{z}\hat{z}}(2\pi j/T) = 2\pi S_{\hat{z}}(0)$$



…and for Equal-weighted periodogram (EWP) kernel



LLS (2017) summary:

- QS achieves the LLS asymptotic frontier:

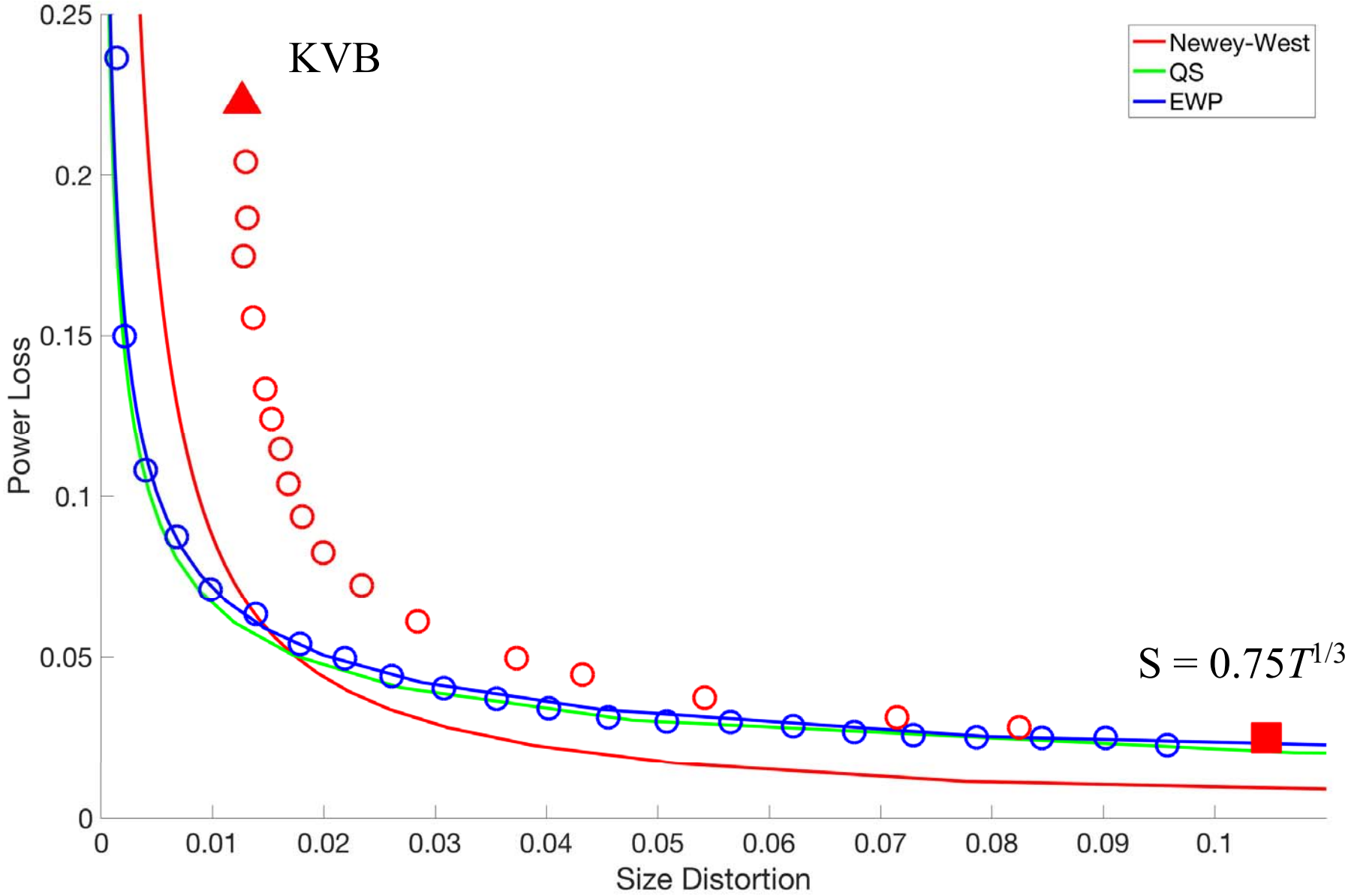
$$\Delta_P \sqrt{|\Delta_S|} \geq \frac{0.3368}{T} \sqrt{|\omega^{(2)}|}$$

- EWP achieves the frontier for tests with *t* fixed-b critical values:

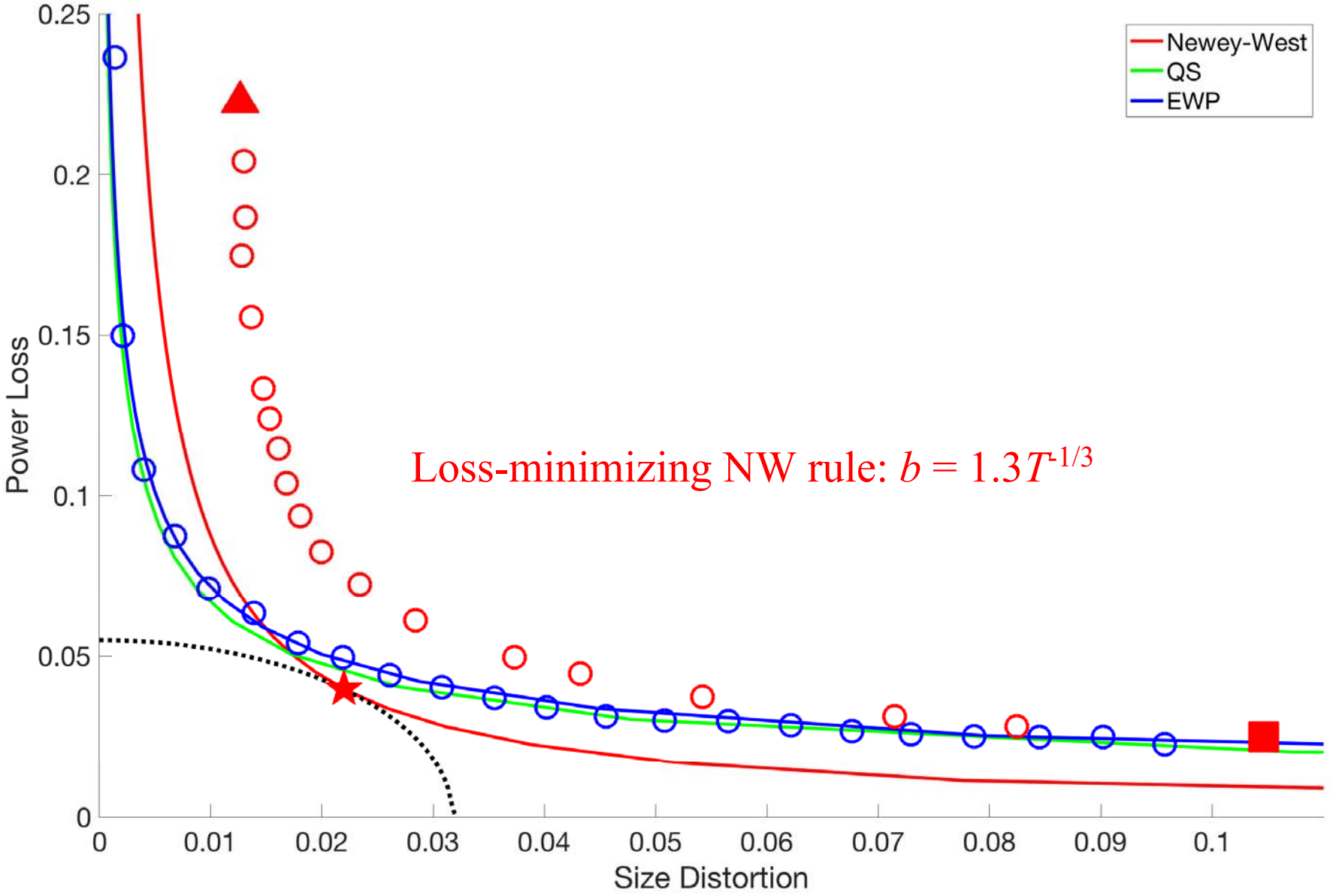
$$\Delta_P \sqrt{|\Delta_S|} \geq \frac{0.3624}{T} \sqrt{|\omega^{(2)}|}$$

- NW is asymptotically dominated (but not in finite samples)

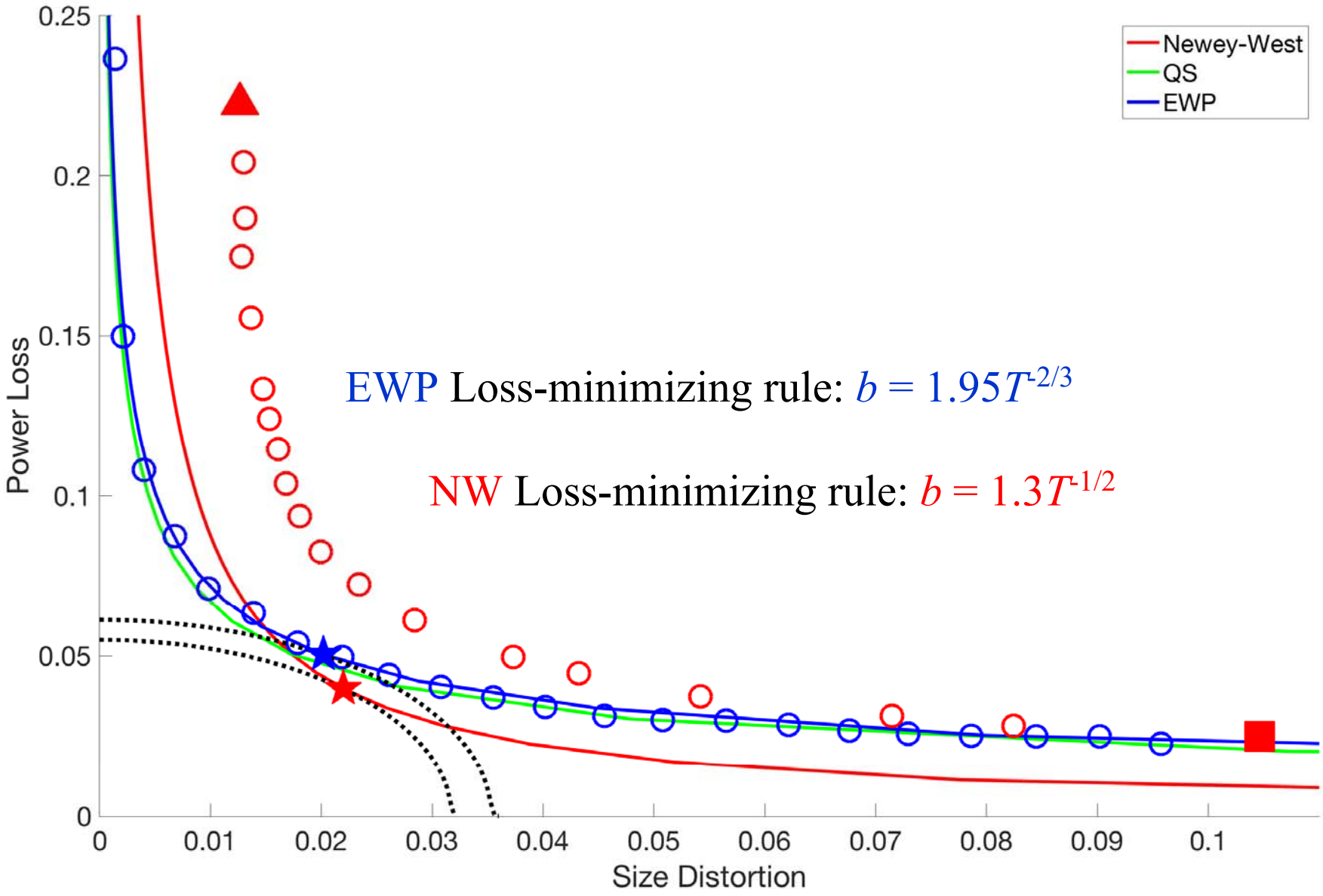
How about finite sample performance? $T = 200$, AR(1), 0.7



We propose using quadratic size/power loss to choose a point on the curve



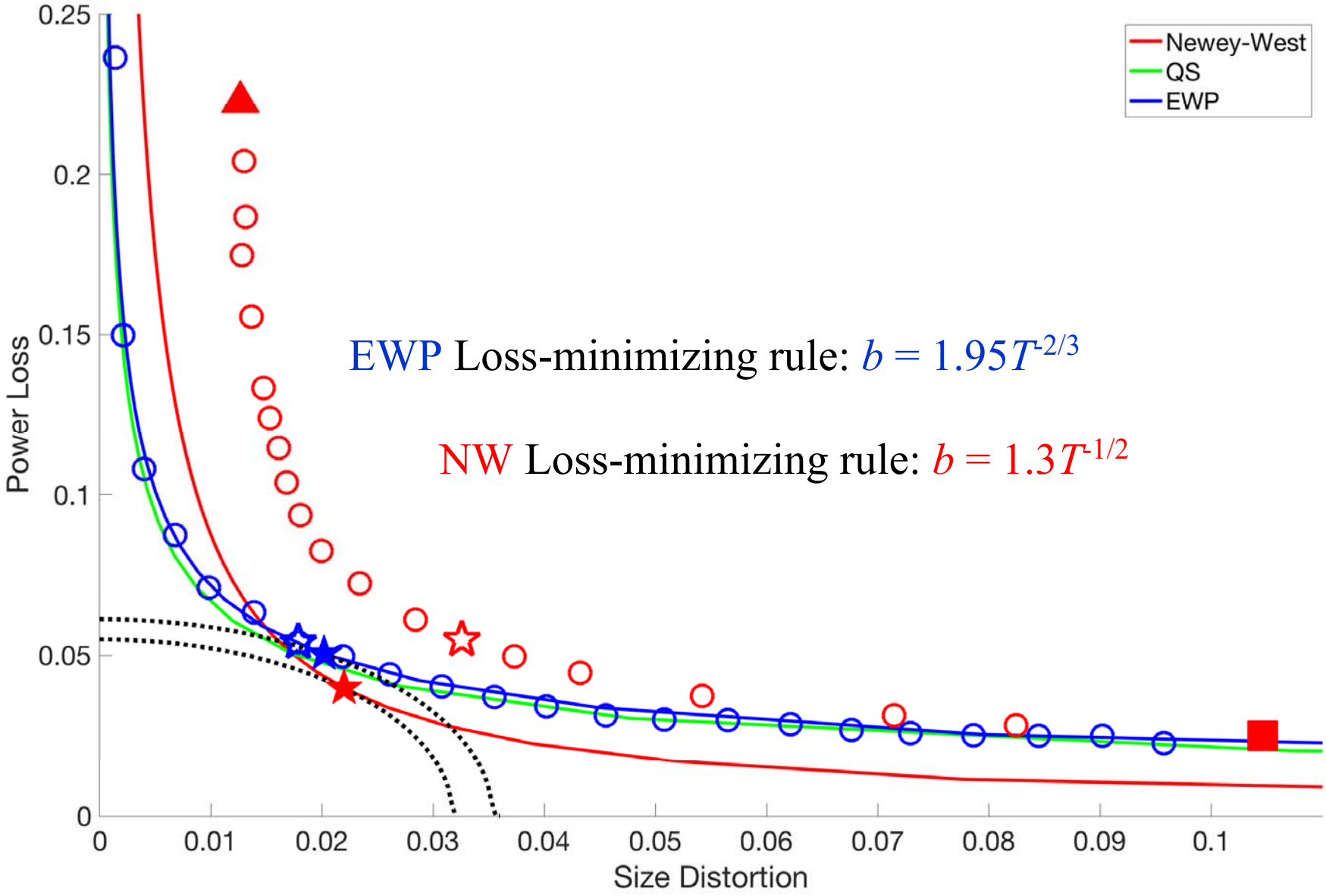
.. for both NW and EWP



EWP Loss-minimizing rule: $b = 1.95T^{-2/3}$

NW Loss-minimizing rule: $b = 1.3T^{-1/2}$

How do they perform in finite samples? (open stars = MC)



Odds & ends

- **Loss function**: $Loss = \kappa\left(\Delta_S\right)^2 + (1-\kappa)\left(\Delta_P^{\max}\right)^2.$

  We use $\kappa = 0.9$ and $\rho = 0.7$ for the rules.

Odds & ends

- **Loss function**: $Loss = \kappa\left(\Delta_S\right)^2 + (1-\kappa)\left(\Delta_P^{\max}\right)^2.$

  We use $\boldsymbol{\kappa = 0.9}$ and $\boldsymbol{\rho = 0.7}$ for the rules.

- **Spectral curvature**:

$$\omega^{(q)} = \sum_{j=-\infty}^{\infty} |\, j \,|^q \, \Gamma_j \Omega^{-1} \text{ (scalar case)}$$

$$\omega^{(2)} = -S_z''(0) / S_z(0)$$

AR(1) case: $\omega^{(1)} = 2\rho/(1-\rho^2)$ $\omega^{(2)} = 2\rho/(1-\rho)^2.$

Odds & ends

- **Loss function**: $Loss = \kappa\left(\Delta_S\right)^2 + (1-\kappa)\left(\Delta_P^{\max}\right)^2.$

  We use $\boldsymbol{\kappa = 0.9}$ and $\boldsymbol{\rho = 0.7}$ for the rules.

- **Spectral curvature**:

$$\omega^{(q)} = \sum_{j=-\infty}^{\infty} |\, j \,|^q \, \Gamma_j \Omega^{-1} \text{ (scalar case)}$$

$$\omega^{(2)} = -S_z''(0) / S_z(0)$$

AR(1) case: $\omega^{(1)} = 2\rho/(1-\rho^2)$ $\omega^{(2)} = 2\rho/(1-\rho)^2.$

- **Parzen characteristic exponent (q)** and generalized derivative:

$$k^{(q)}(0) = \lim_{x\to 0} \frac{1-k(x)}{|\, x \,|^q}, \text{ where } k = \text{kernel}$$

q = Parzen characteristic exponent = max q: $k^{(q)}(0) < \infty$

Odds & ends, ctd.

- We use **size-adjusted power** to allow apples-to-apples comparison
  - Adjust the critical values, so the test has same size under null
  - Standard for second order comparisons of tests (e.g. Rothenberg (1984)
  - How would you compare two tests with i.i.d. N(0,1) data: using median with ±1.645 or using mean with ±1.96?

Odds & ends, ctd.

- We use **size-adjusted power** to allow apples-to-apples comparison
  - Adjust the critical values, so the test has same size under null
  - Standard for second order comparisons of tests (e.g. Rothenberg (1984)
  - How would you compare two tests with i.i.d. N(0,1) data: using median with ±1.645 or using mean with ±1.96?

- Brillinger (1975) is first we know of to obtain fixed-*b* *t* distribution for EWP:

5.13.25 Under the conditions of Theorem 5.4.3 show that

$$\frac{c_X^{(T)} - c_X}{\sqrt{2\pi f_{XX^{(T)}}(0)/T}}$$

tends in distribution to a Student's t with $2m$ degrees of freedom. (This result may be used to set approximate confidence limits for c_X.)

Odds & ends, ctd.

- We use **size-adjusted power** to allow apples-to-apples comparison
  - Adjust the critical values, so the test has same size under null
  - Standard for second order comparisons of tests (e.g. Rothenberg (1984)
  - How would you compare two tests with i.i.d. N(0,1) data: using median with ±1.645 or using mean with ±1.96?

- Brillinger (1975) is first we know of to obtain fixed-*b* *t* distribution for EWP:

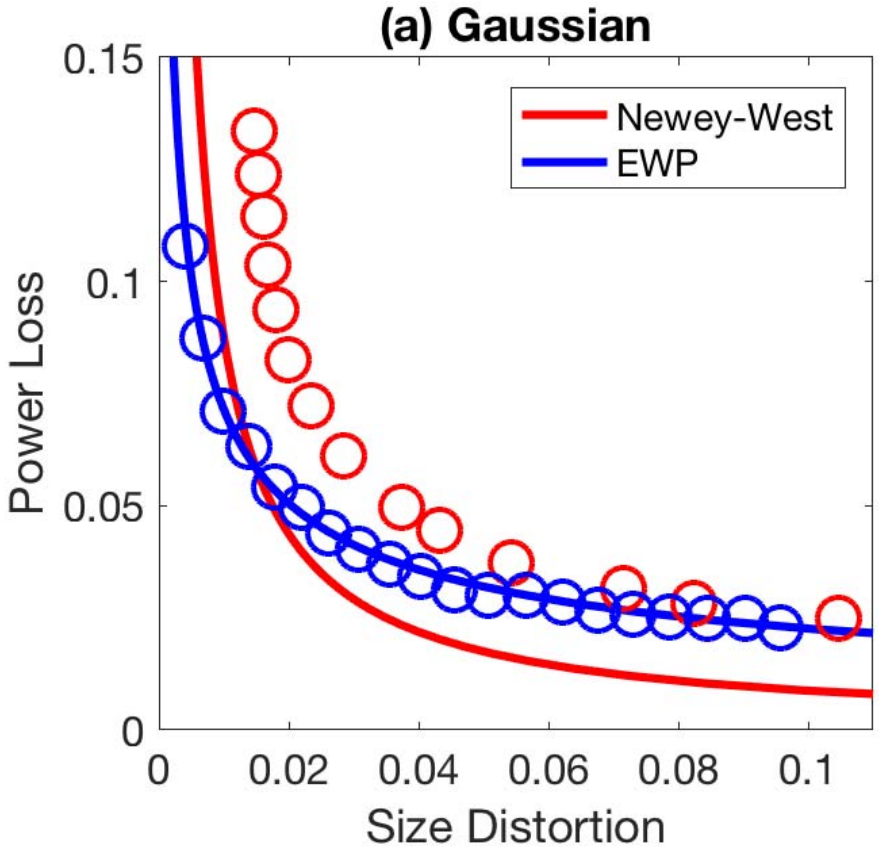
5.13.25 Under the conditions of Theorem 5.4.3 show that

$$\frac{c_X^{(T)} - c_X}{\sqrt{2\pi f_{XX^{(T)}}(0)/T}}$$

tends in distribution to a Student's t with $2m$ degrees of freedom. (This result may be used to set approximate confidence limits for c_X.)

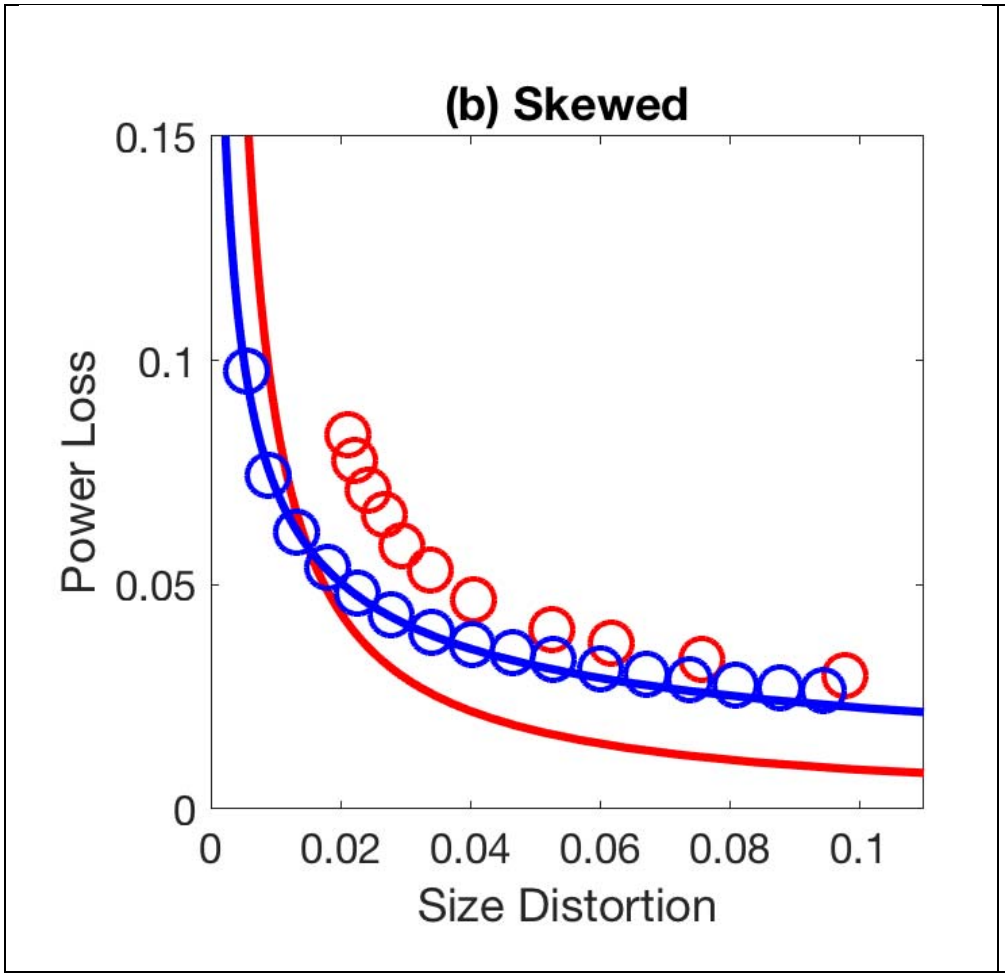
- The Edgeworth expressions are derived for the Gaussian location model – but they are a guide (we hope) to non-Gaussian location and regression.

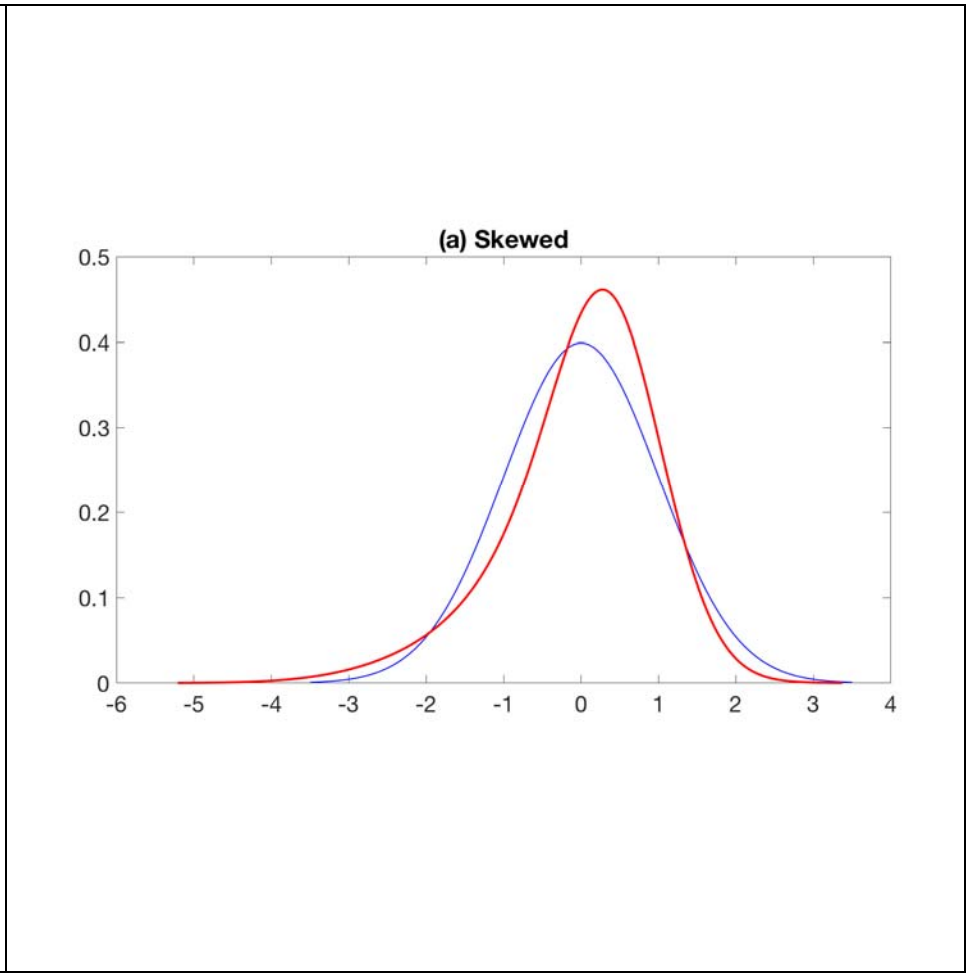
Size-power tradeoffs, location model, EWP and NW, various error distn's



AR(1), $\rho_z = 0.7$

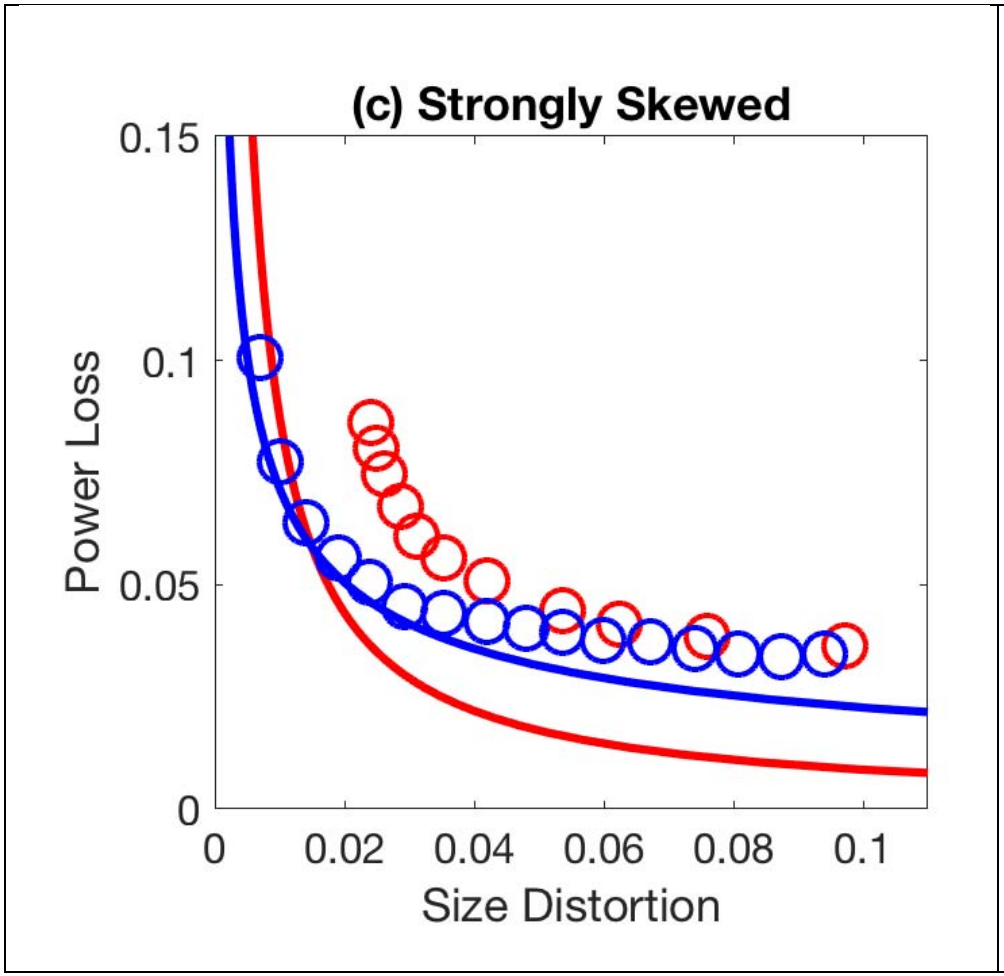
Size-power tradeoffs, location model, EWP and NW, various error distn's

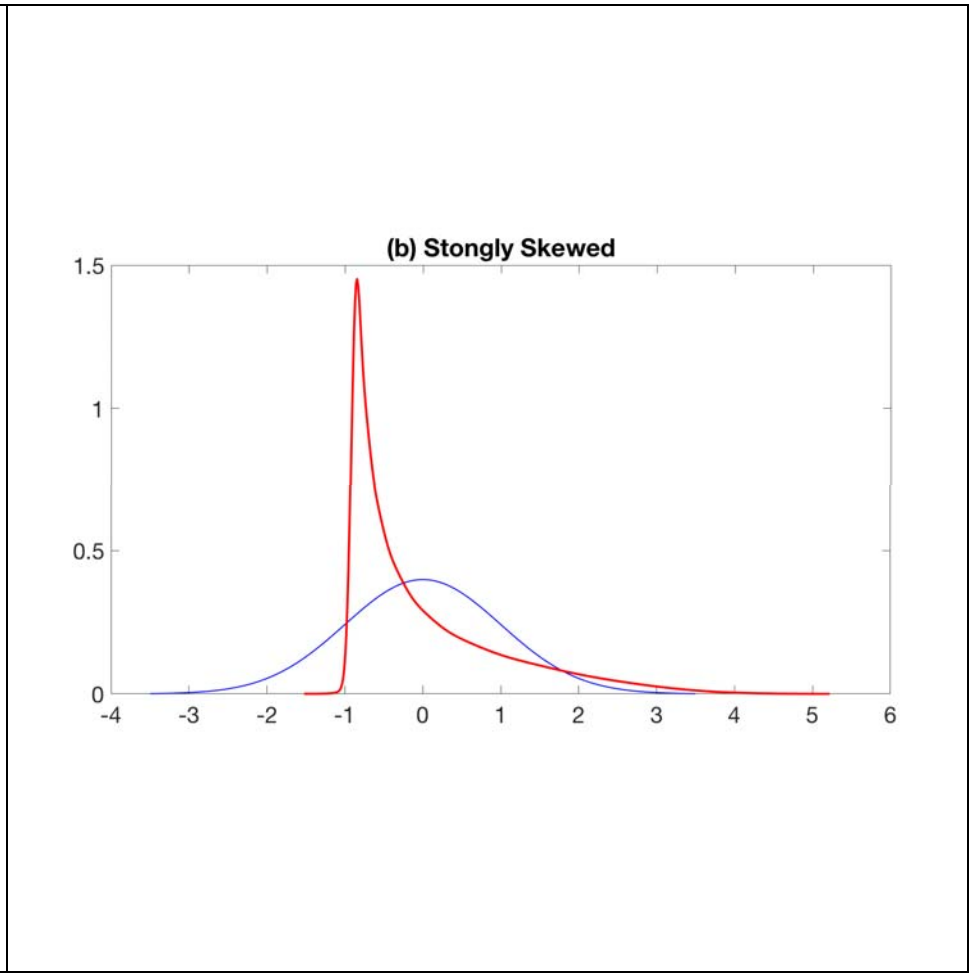




AR(1), $\rho_z = 0.7$

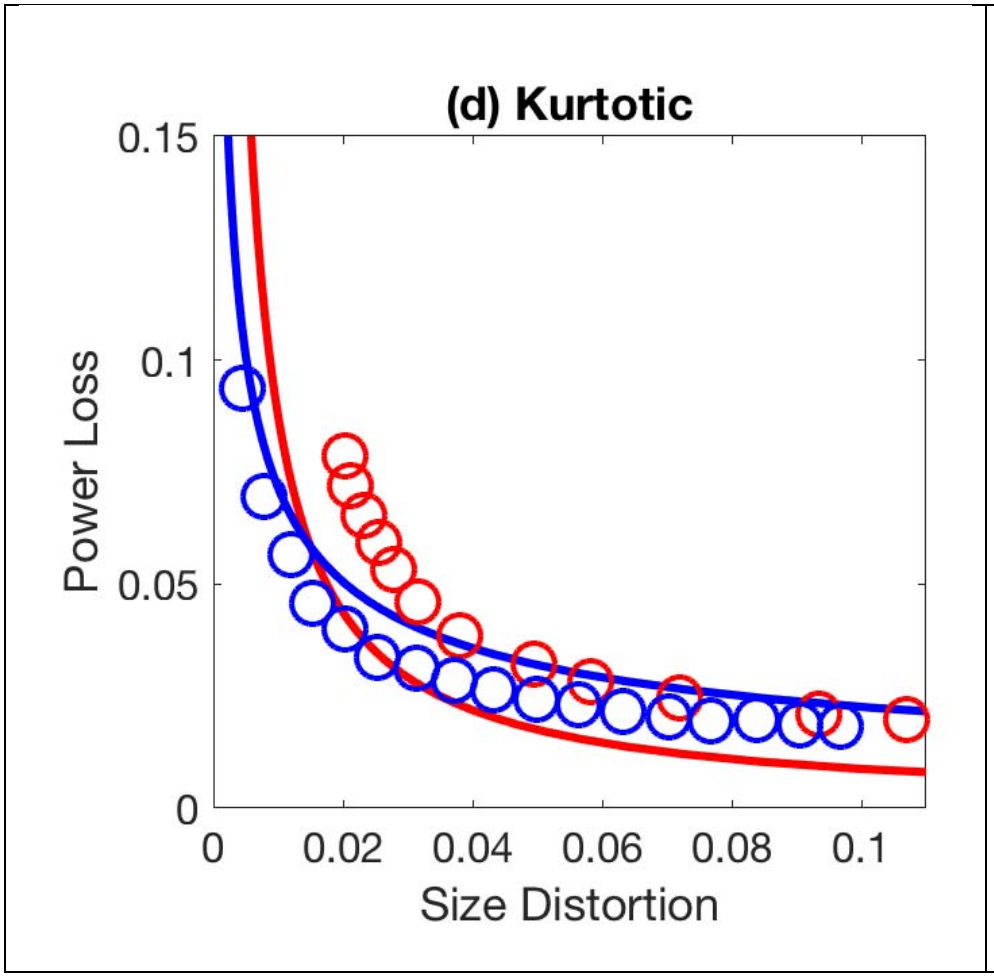
Size-power tradeoffs, location model, EWP and NW, various error distn's

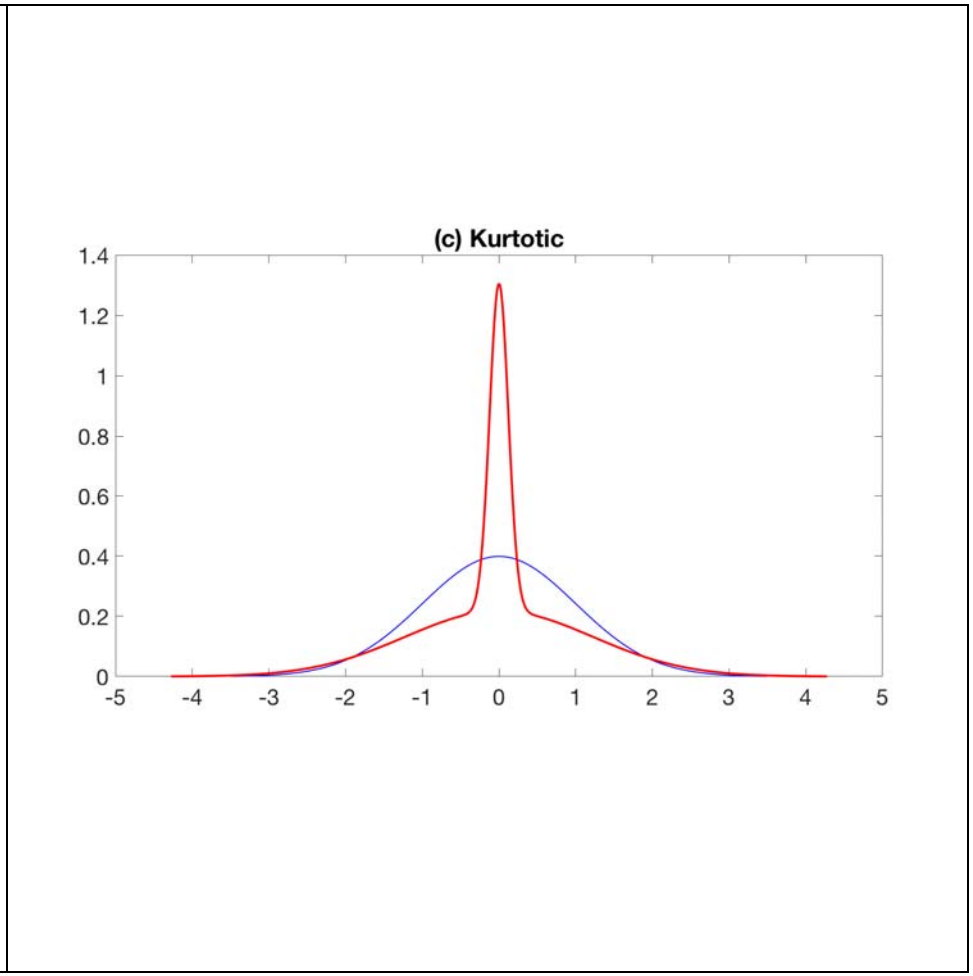




AR(1), $\rho_z = 0.7$

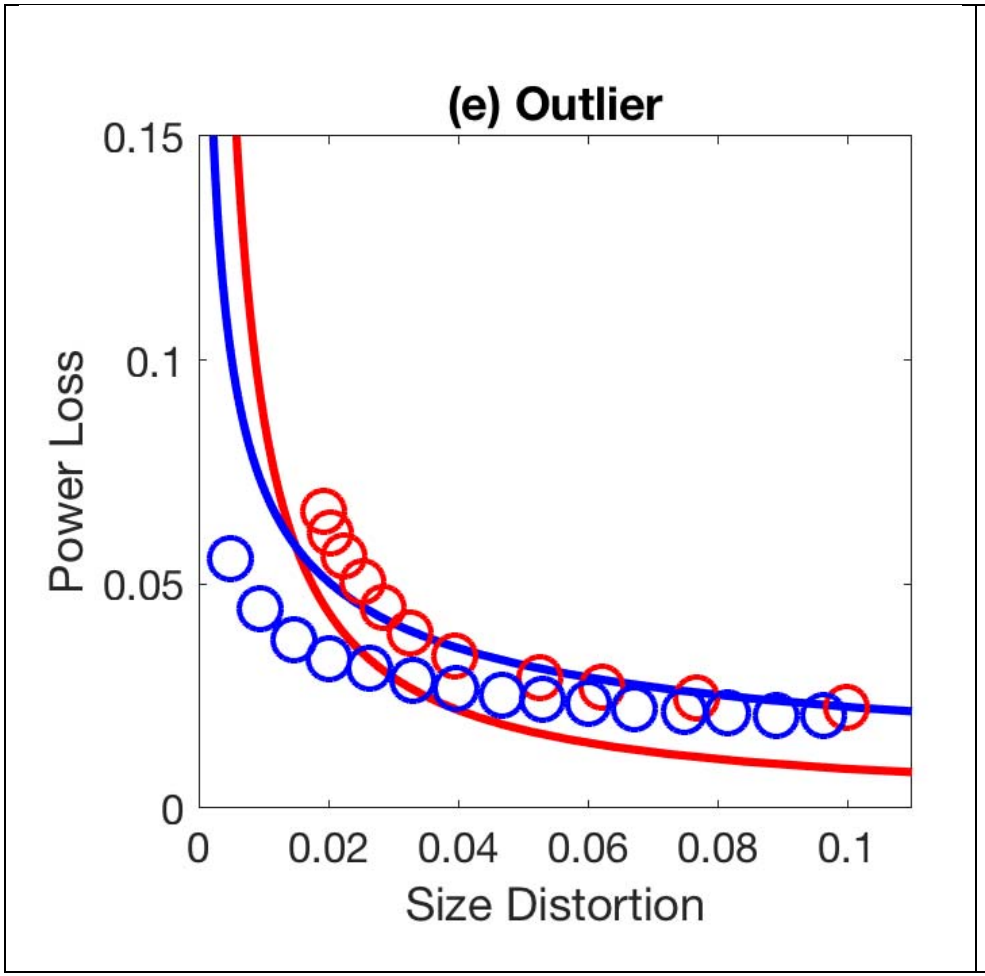
Size-power tradeoffs, location model, EWP and NW, various error distn's

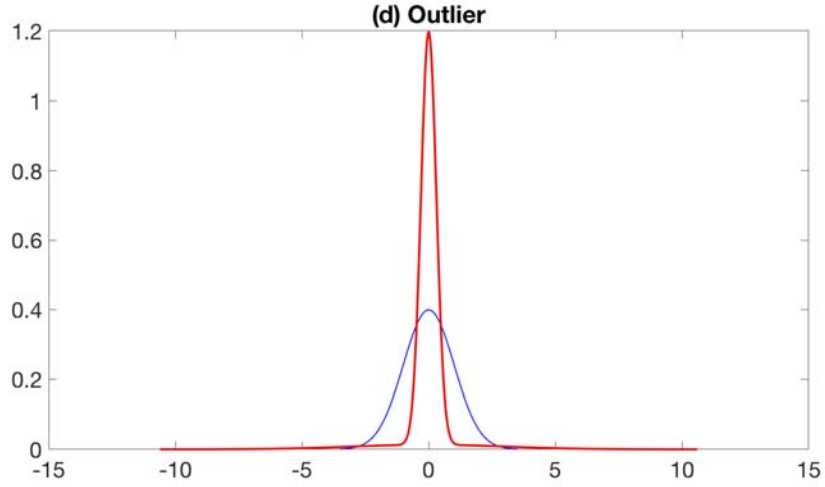




AR(1), $\rho_z = 0.7$

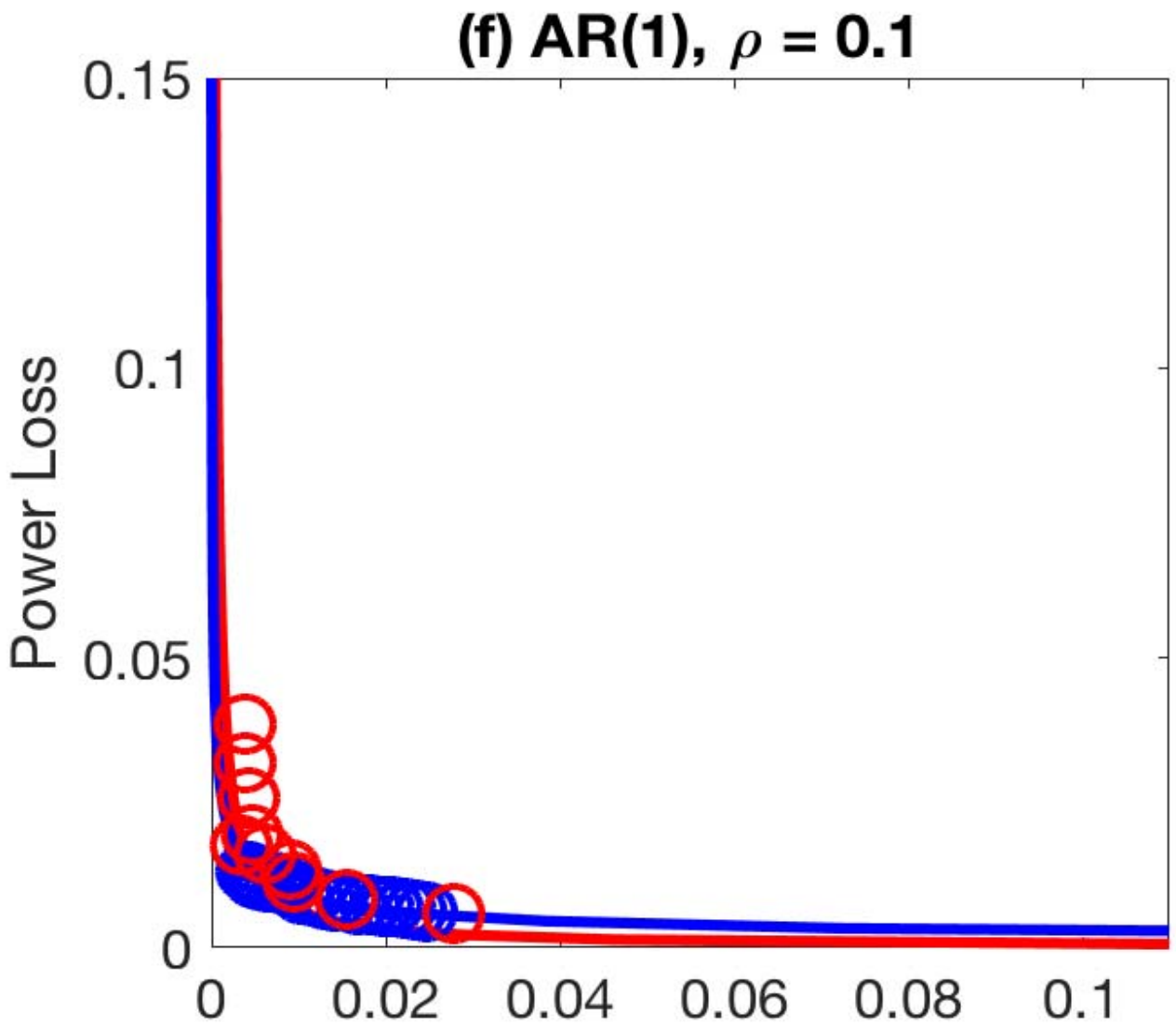
Size-power tradeoffs, location model, EWP and NW, various error distn's



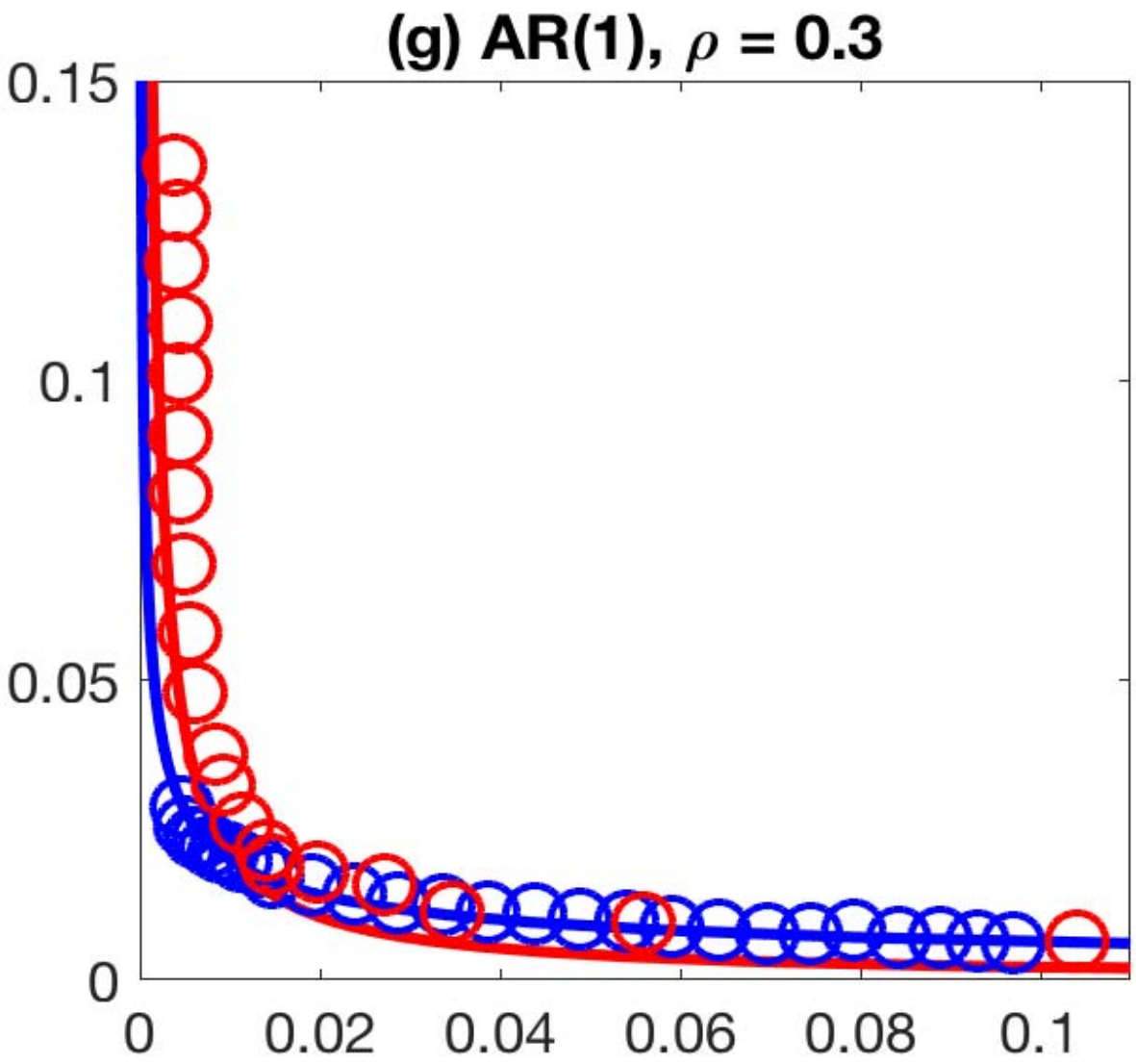


AR(1), $\rho_z = 0.7$

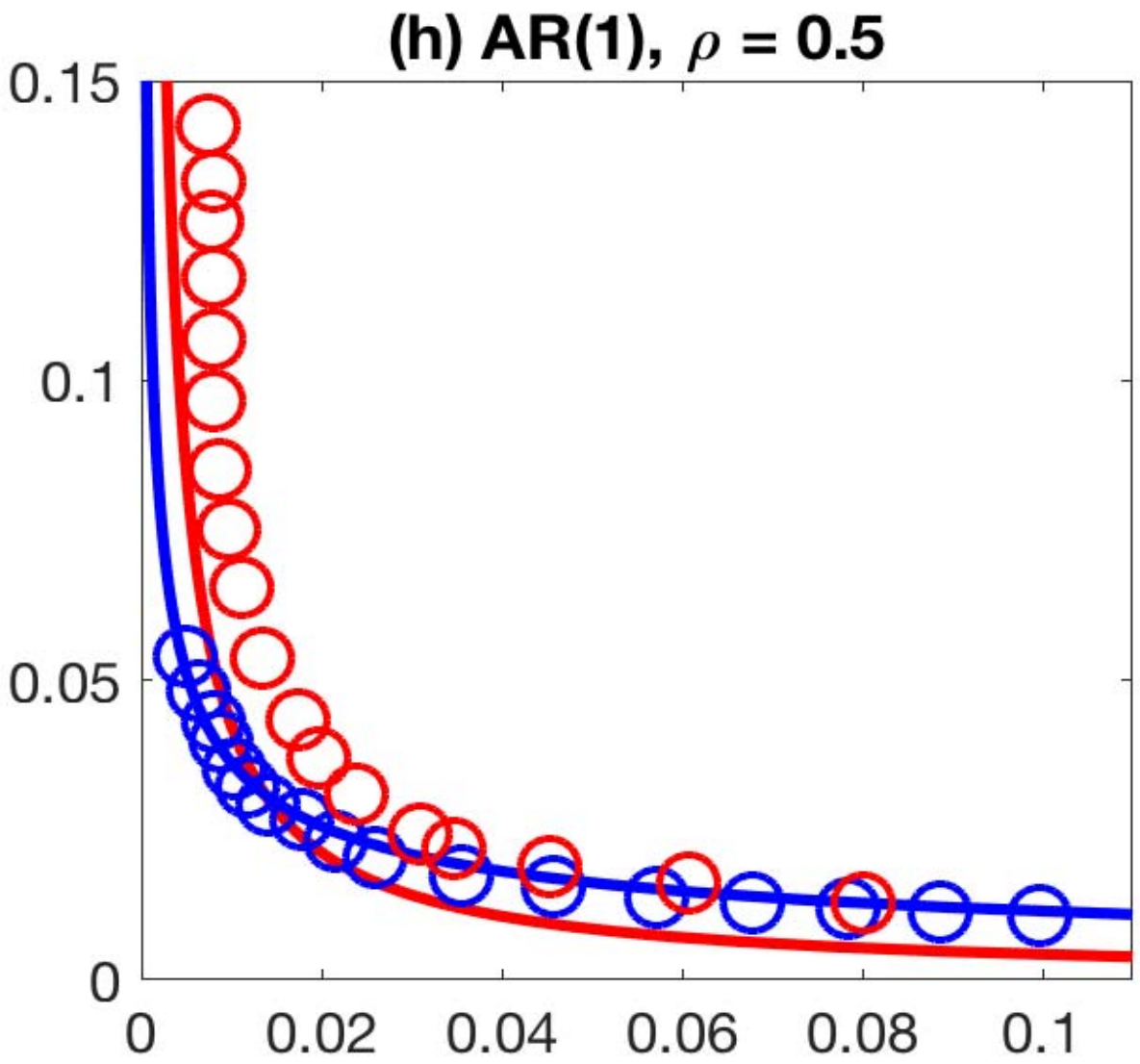
EWP and NW, Gaussian AR(1), various ρ, location model



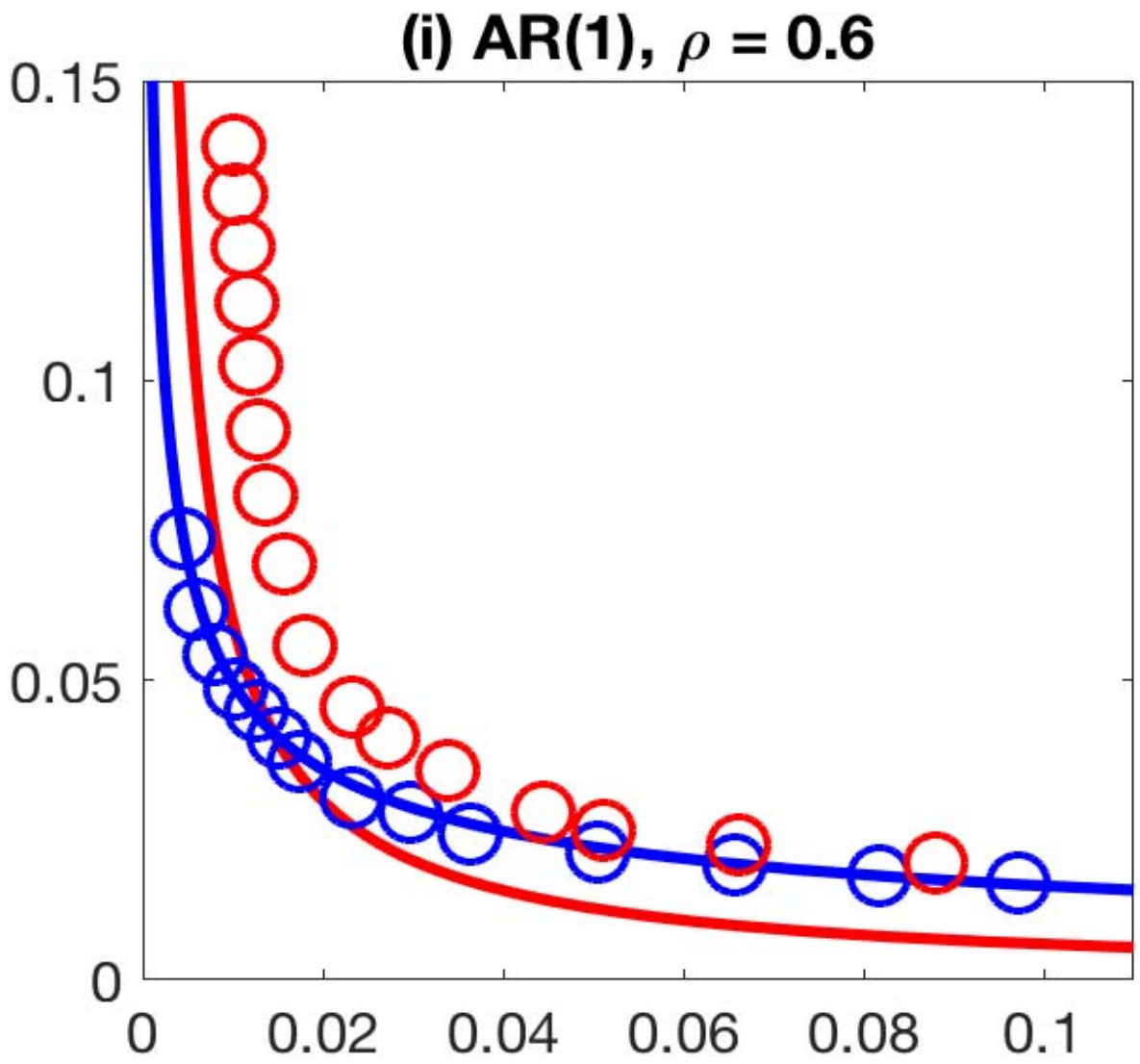
EWP and NW, Gaussian AR(1), various ρ, location model



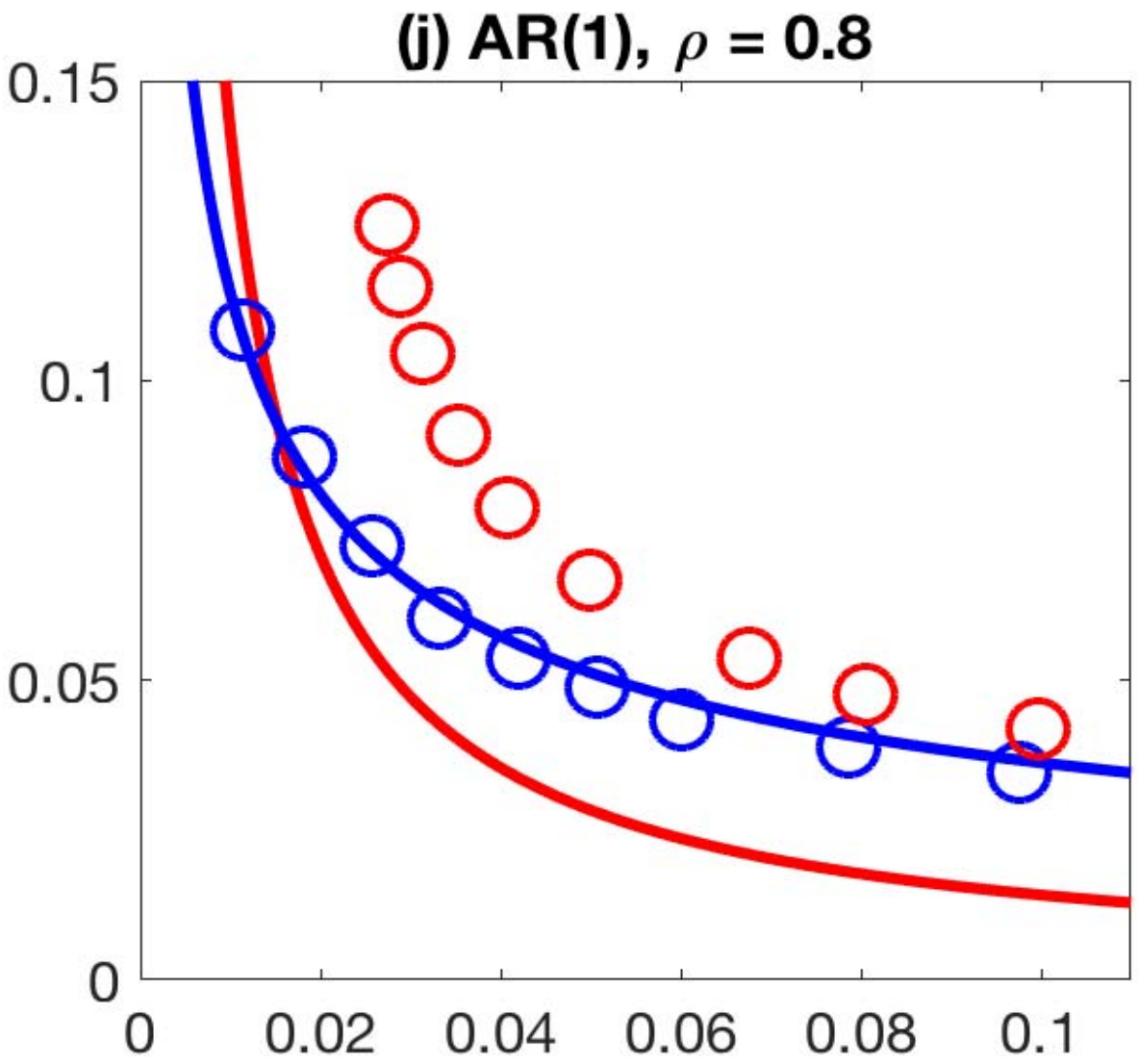
EWP and NW, Gaussian AR(1), various ρ, location model



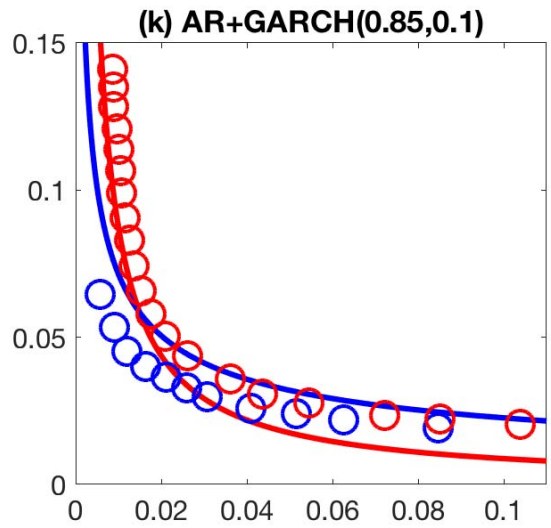
EWP and NW, Gaussian AR(1), various ρ, location model

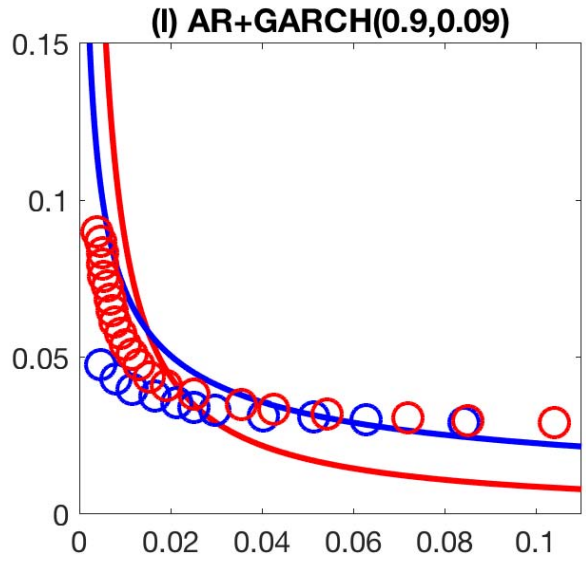


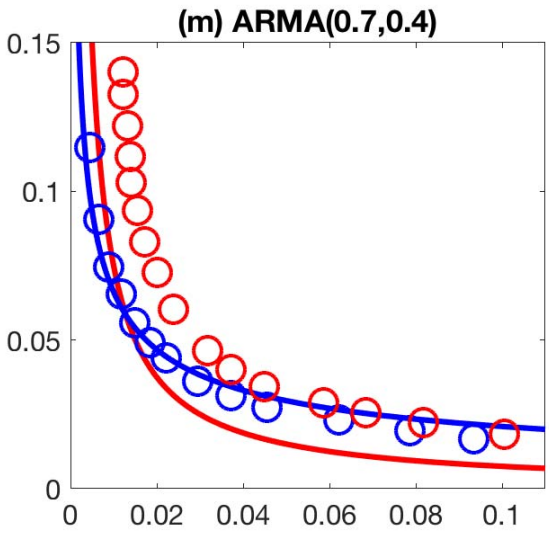
EWP and NW, Gaussian AR(1), various ρ, location model

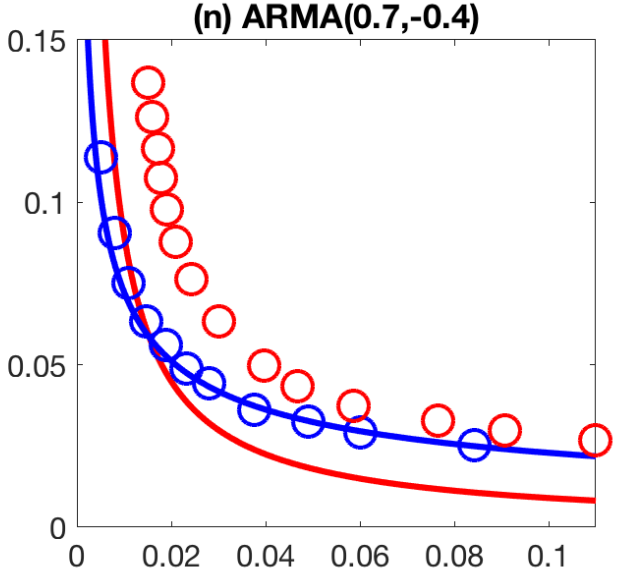


EWP and NW, GARCH and ARMA, location model









Digression: uniform size control for $|\omega| \leq \omega(0.7)$

A more standard approach is to control size uniformly (to 2nd order):

$$\sup_{\omega^{(q)} \leq \bar{\omega}^{(q)}} \Pr_0\left[F_T^* > \bar{c}_{m,T}^{\alpha}(b)\right] \leq \alpha + o(b) + o\left((bT)^{-q}\right).$$

- In the Gaussian location model, we can implement this because of formula for second-order adjustment to critical value (size-adjusted critical value):

- Doing so delivers a family of tests indexed by $\omega^{(q)}$. We select a test from this family by maximizing weighted average power:

$$\min_b \int_{\delta} \int_{|\rho| \leq \bar{\rho}} \Delta_P\left(\omega^{(q)}(\rho), \delta\right) d\Pi_\rho(\rho) d\Pi_\delta(\delta)$$

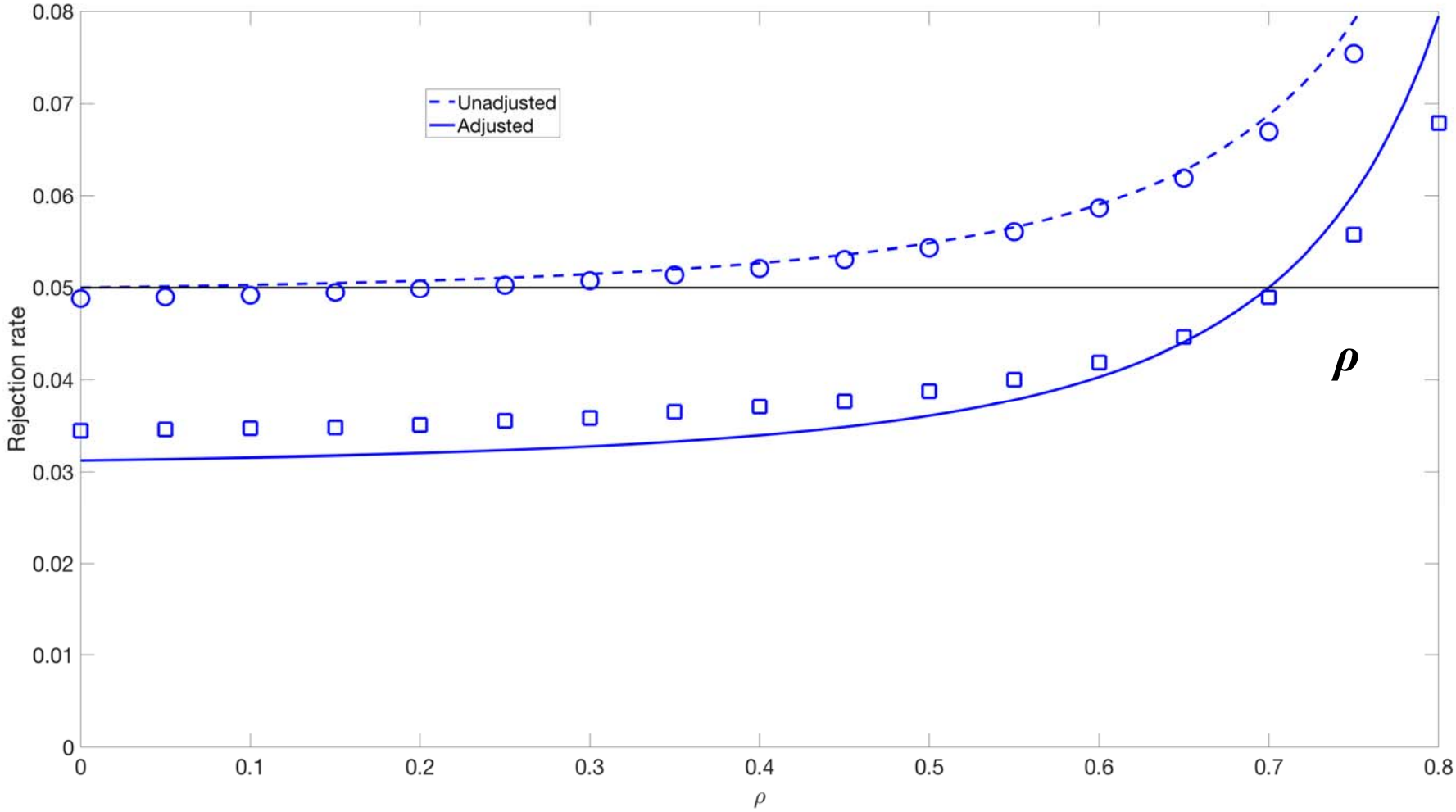
- For local alternatives and uniform weight over ρ: $|\rho| \leq 0.7$, the WAP tests are:

EWP: $b^{WAP,EWP} = 2.06T^{-2/3}$ and $\bar{c}_{m,T}^{\alpha}(b^{WAP,EWP}) = \left[1 + 6.04T^{-2/3}\right] F_{m,B^{WAP}-m+1}^{\alpha}$

NW: $b^{WAP,NW} = 2.19T^{-1/2}$ and $\bar{c}_{m,T}^{\alpha}(b^{WAP,NW}) = \left[1 + 1.26T^{-1/2}\right] c_m^{\alpha,NW}(b^{WAP,NW})$

Uniform approach: null rejection rates, EWP implementation, $T = 200$

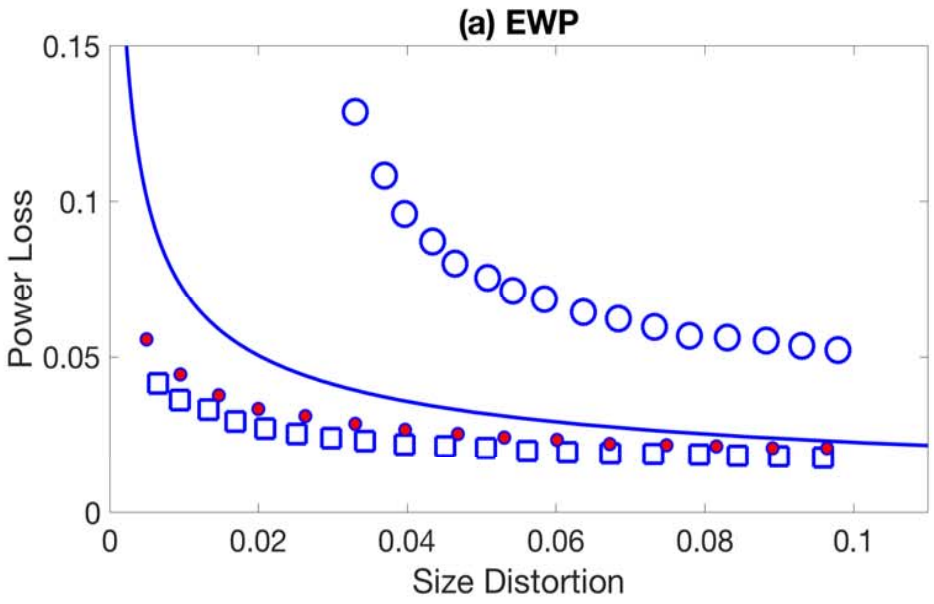
Uniform Local WAP test, adjusted and unadjusted, theory and MC: EWP

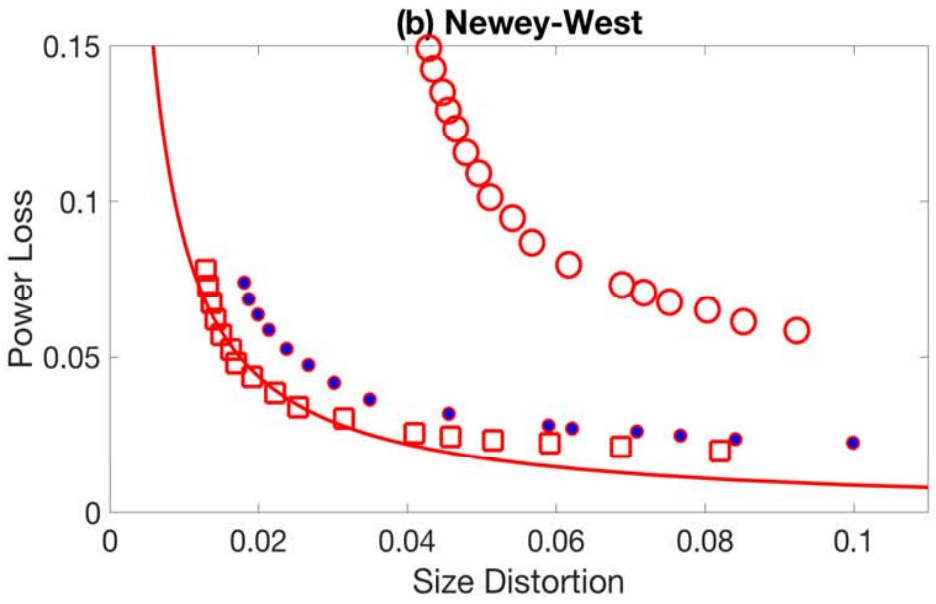


Warning! This approach relies on formulas for the second-order adjustment for the critical value – which are only available in the Gaussian location model

Regression case – single x

x_t, u_t are independent Gaussian AR(1) (so z_t non-Gaussian, heavy-tailed)





Squares: null imposed: $z_t = \tilde{x}_t(\tilde{y}_t - \beta_0 \tilde{x}_t)$, use $z_t - \overline{z}$

Circles: unrestricted: $\hat{z}_t = \tilde{x}_t(\tilde{y}_t - \hat{\beta}\tilde{x}_t)$, use $\hat{z}_t$

Solid: Gaussian location model, theory
Dots: Location model MC, Marron-Wand 5 (leptokurtic)

MC results, regression: different rules, restricted/unrestricted

Rejection rates of HAR tests with nominal level 5% ($b = S/T$)

$y_t = \beta_0 + \beta_1 x_t + u_t$, x_t & u_t Gaussian AR(1), $\rho_x = \rho_u = 0.7^{1/2}$, $T = 200$

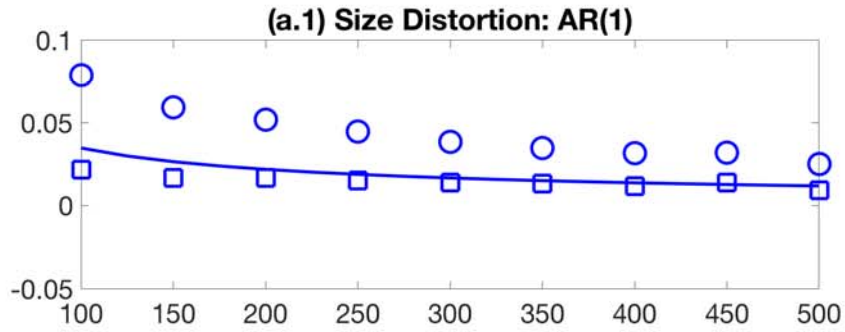
Estimator	Truncation rule for b	Critical values	Null imposed?	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$
NW	$0.75T^{-2/3}$	N(0,1)	No	**0.079**	**0.105**	**0.164**
NW	$1.3T^{-1/2}$	fixed-b (nonstandard)	No	0.067	0.080	0.107
EWP	$1.95T^{-2/3}$	fixed-b (t_v)	No	0.063	0.074	0.100
NW	$1.3T^{-1/2}$	fixed-b (nonstandard)	Yes	0.057	0.062	0.073
EWP	$1.95T^{-2/3}$	fixed-b (t_v)	Yes	0.052	0.056	0.066
Theoretical bound based on Edgeworth expansions for the Gaussian location model						
NW	$1.3T^{-1/2}$	fixed-b (nonstandard)	No	0.054	0.058	0.067
EWP	$1.95T^{-2/3}$	fixed-b (t_v)	No	0.052	0.056	0.073

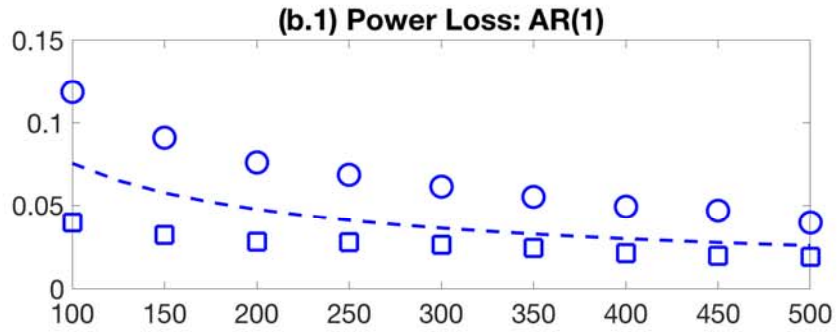
Note: x_t and u_t are independent Gaussian AR(1)'s, single regressor.

Proposed rules: size distortion and power loss curves as function of T

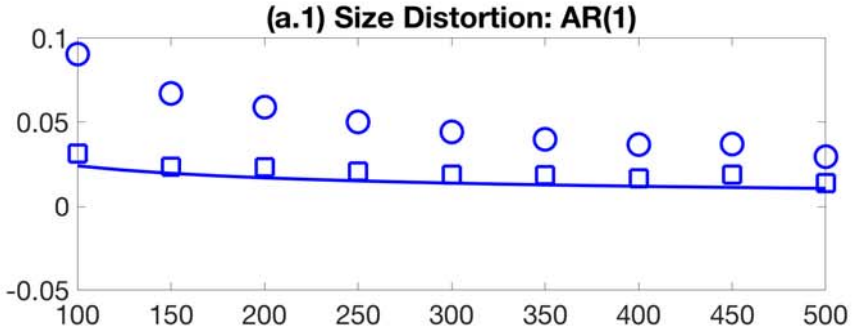
$y_t = \beta_0 + \beta_1 x_t + u_t$, various DGPs for x, u. x_t & u_t AR(1), $\rho = 0.7^{1/2}$

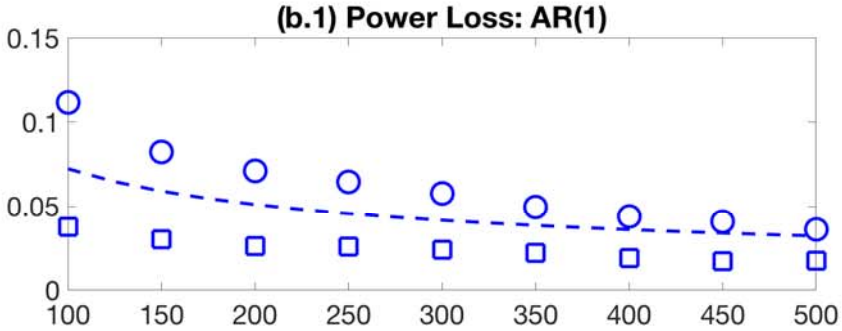
EWP





NW



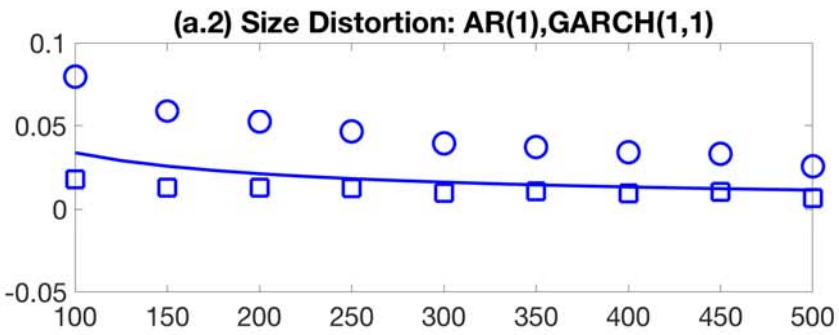


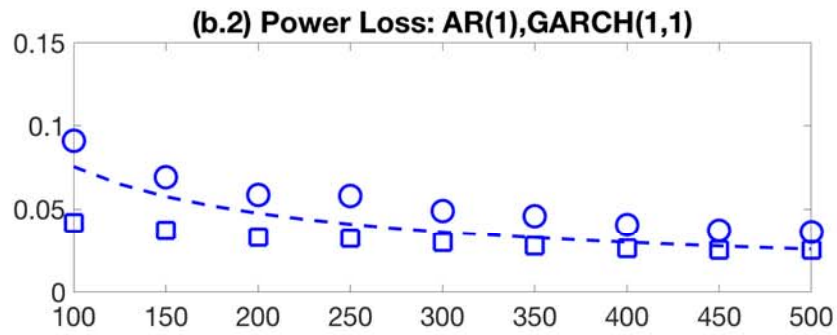
Circle = unrestricted Square = null imposed solid = Gaussian location theory

x_t & u_t AR(1), $\rho = 0.7^{1/2}$ + u_t is GARCH(0.85, 0.10)

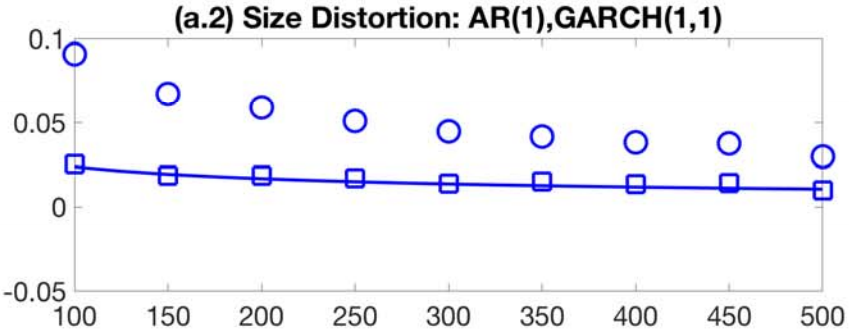
$y_t = \beta_0 + \beta_1 x_t + u_t$, various DGPs for x, u

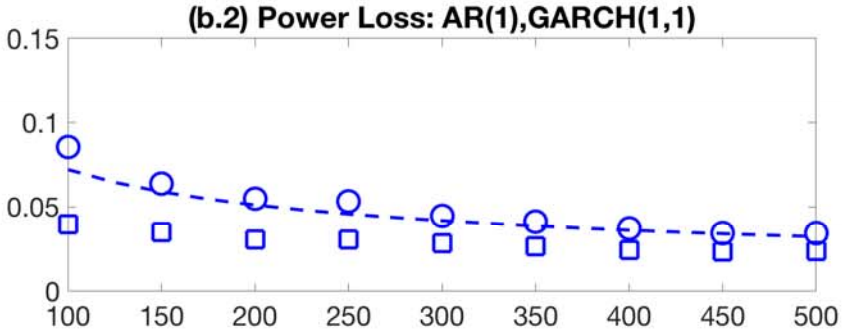
EWP





NW



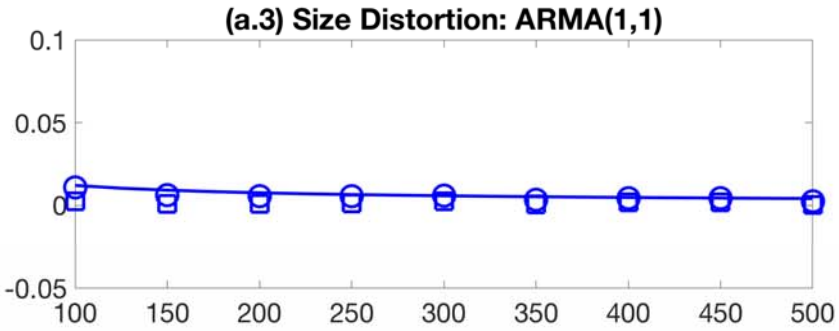


Circle = unrestricted Square = null imposed solid = Gaussian location theory

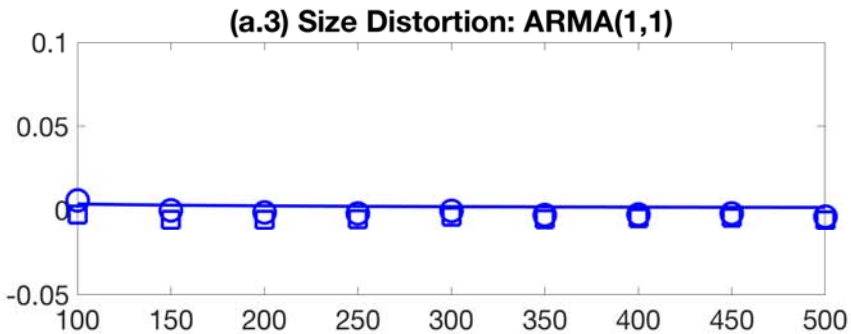
x_t & u_t ARMA(1,1): (1-0.8L)x_t = (1-0.7L)ε_t

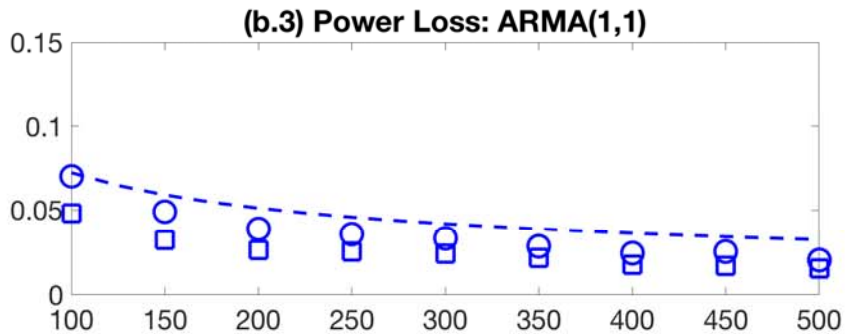
$y_t = \beta_0 + \beta_1 x_t + u_t$, various DGPs for x, u

EWP



NW



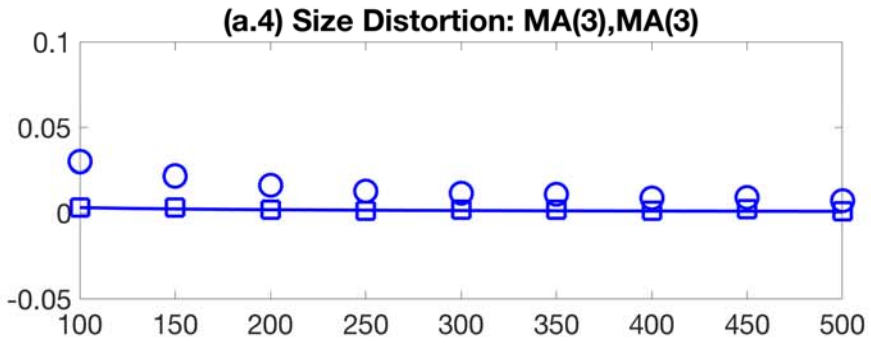


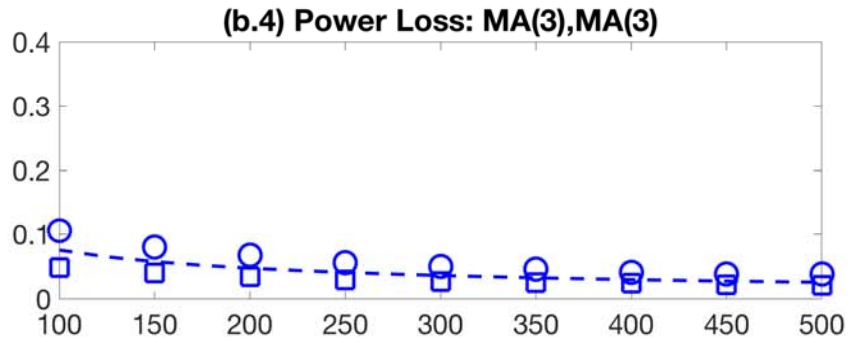
Circle = unrestricted Square = null imposed solid = Gaussian location theory

x_t & u_t independent MA(3): $x_t = (1 + L + L^2 + L^3)\varepsilon_t$

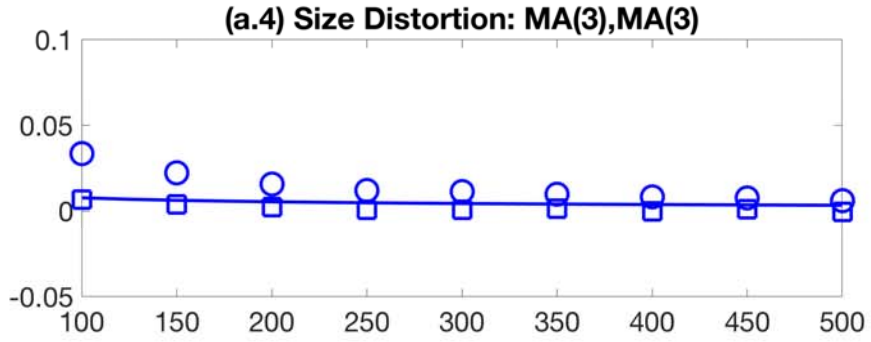
$y_t = \beta_0 + \beta_1 x_t + u_t$, various DGPs for x, u

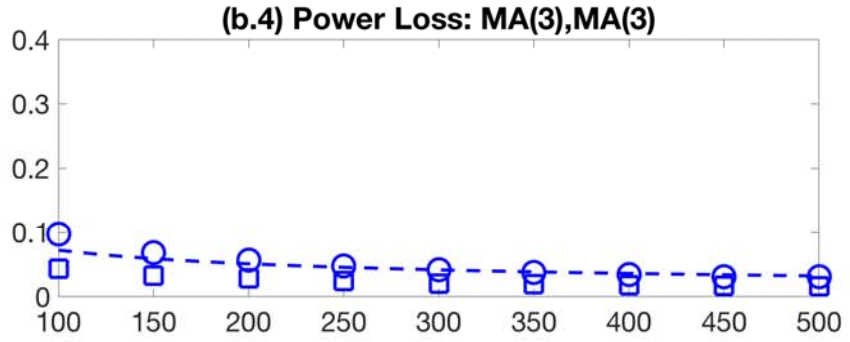
EWP





NW



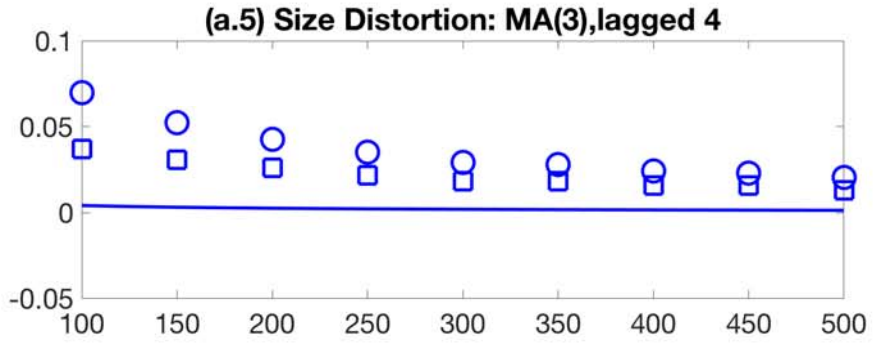


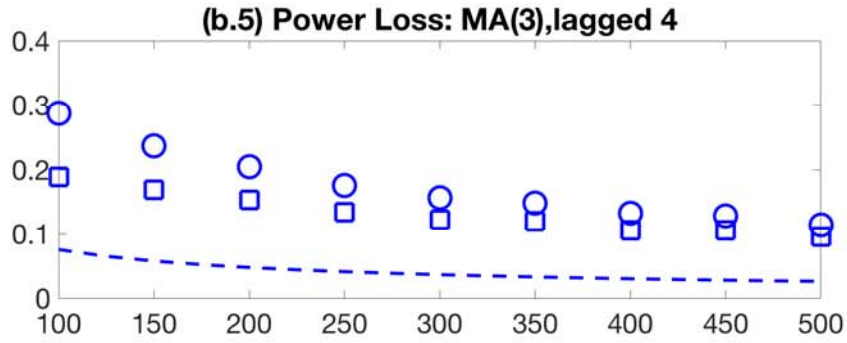
Circle = unrestricted Square = null imposed solid = Gaussian location theory

4-step ahead direct forecast: $u_t = (1 + L + L^2 + L^3)\varepsilon_t$, $x_t = u_{t-4}$

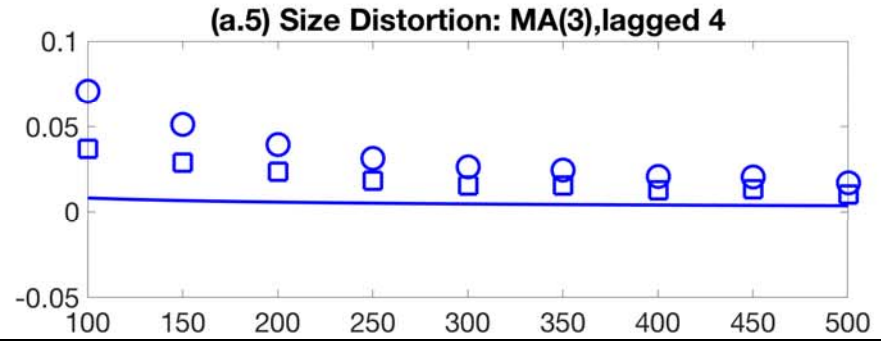
$y_t = \beta_0 + \beta_1 x_t + u_t$, various DGPs for x, u

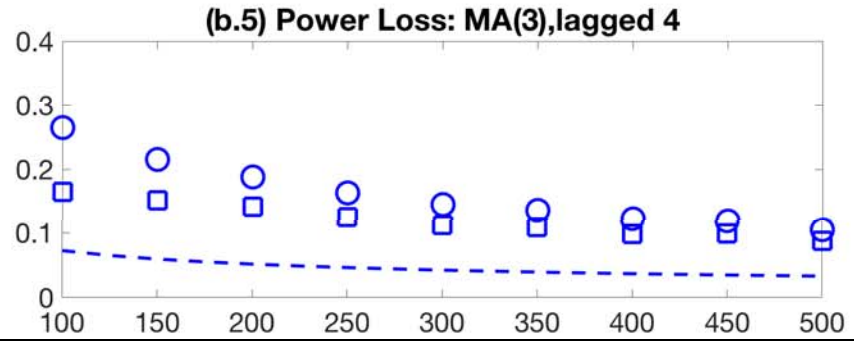
EWP





NW



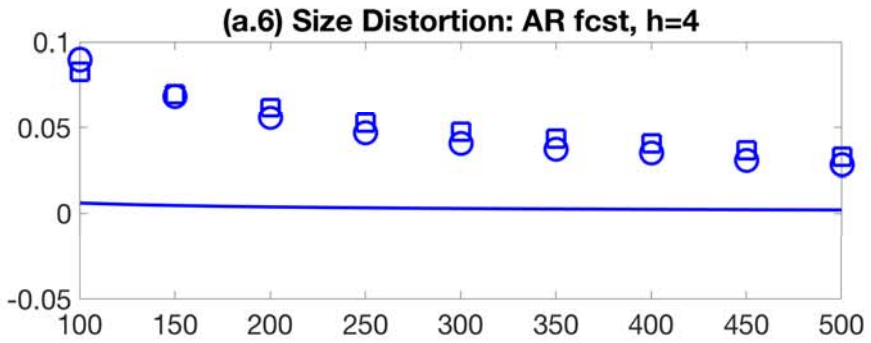


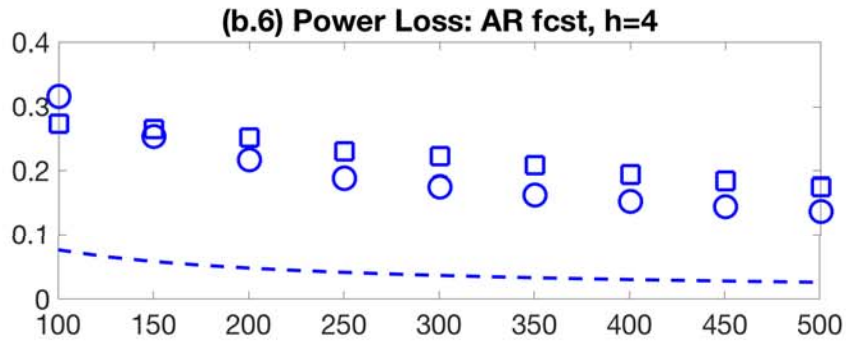
Circle = unrestricted Square = null imposed solid = Gaussian location theory

4-step ahead direct forecast: $(1-0.8L)u_t = \varepsilon_t$, $x_t = y_{t-4}$, $u_t = y_t - 0.8^4 x_t$

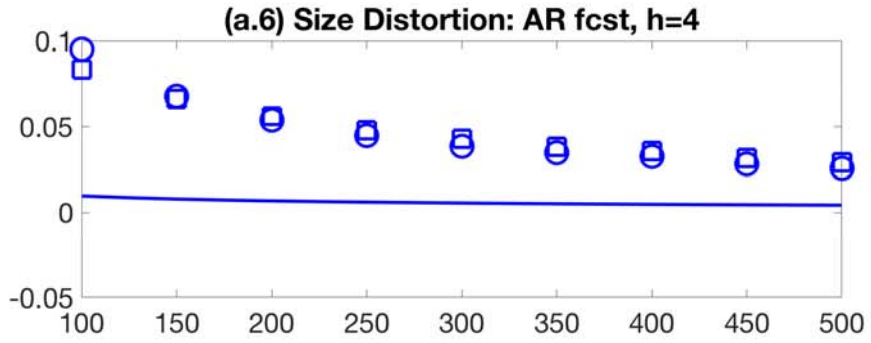
$y_t = \beta_0 + \beta_1 x_t + u_t$, various DGPs for x, u

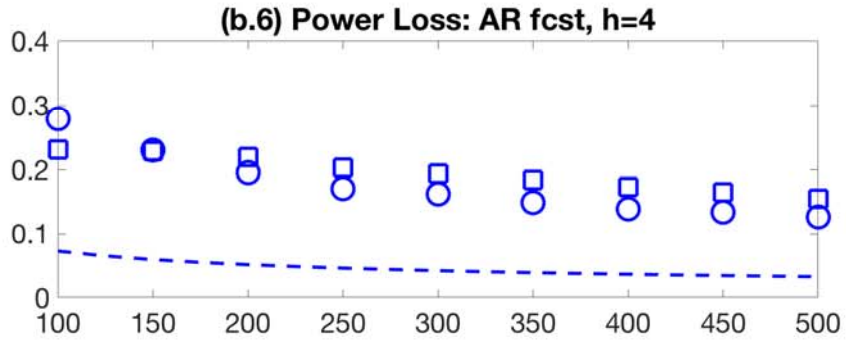
EWP





NW

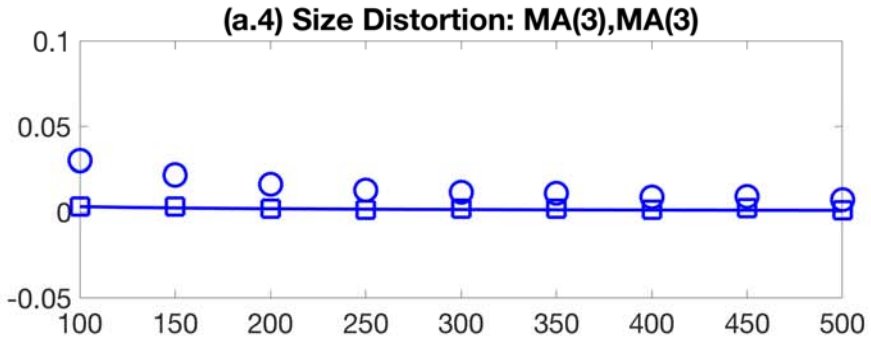


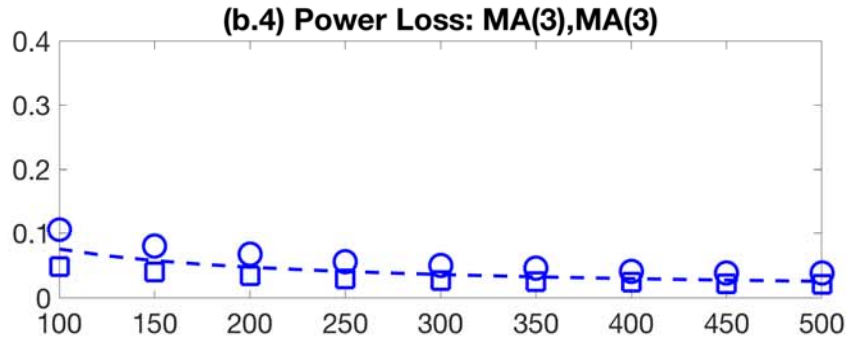


Circle = unrestricted Square = null imposed solid = Gaussian location theory

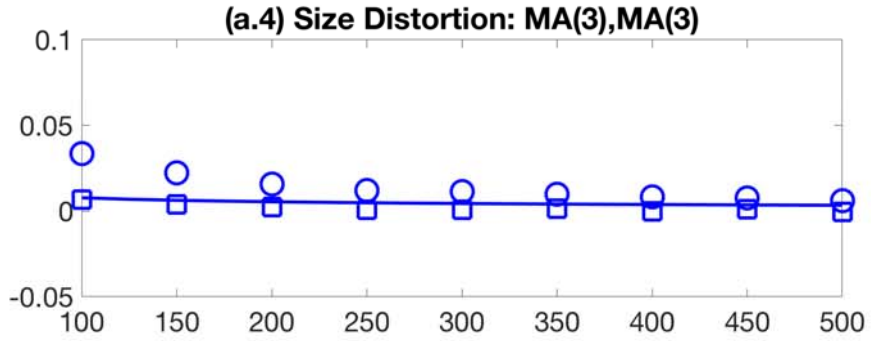
$y_t = \beta_0 + \beta_1 x_t + u_t$, various DGPs for x, u

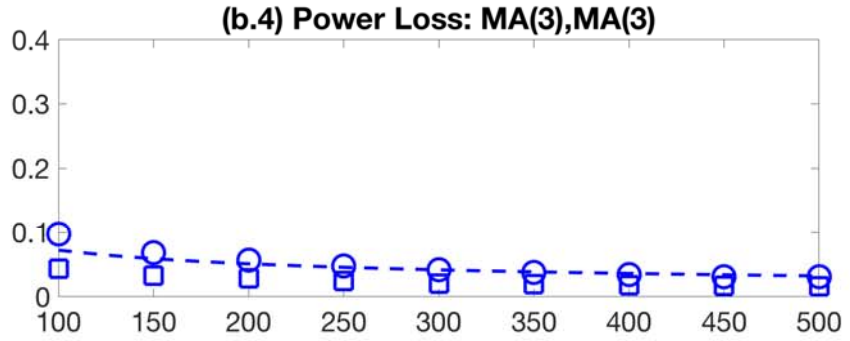
EWP





NW



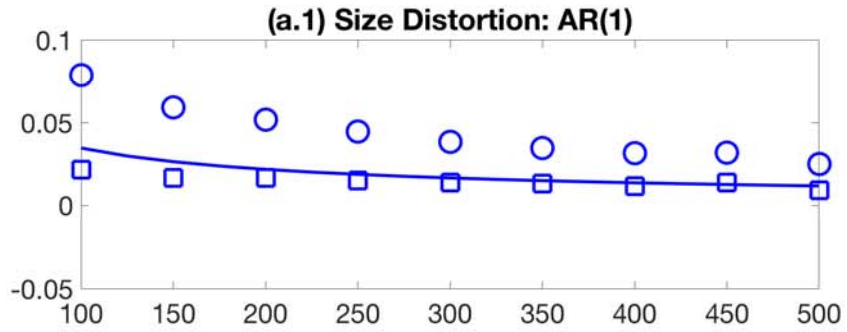


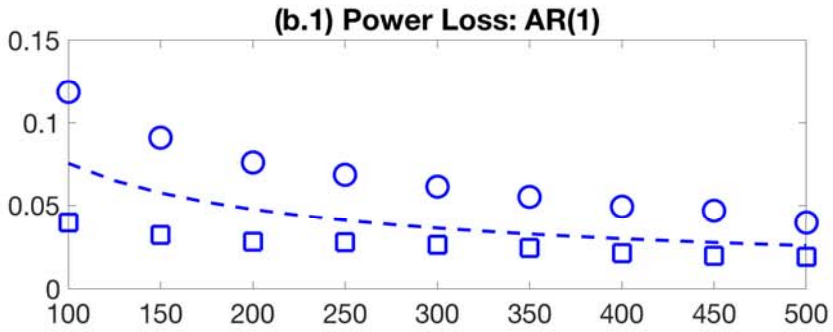
Circle = unrestricted Square = null imposed solid = Gaussian location theory

Proposed rule: size distortion and power loss curves as function of T

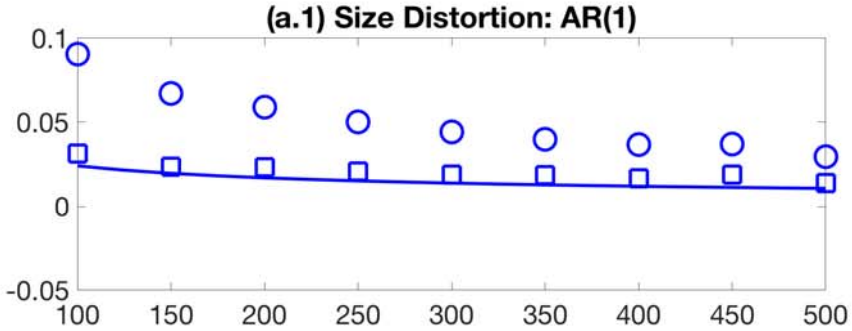
$y_t = \beta_0 + \beta_1 x_t + u_t$, various DGPs for x, u

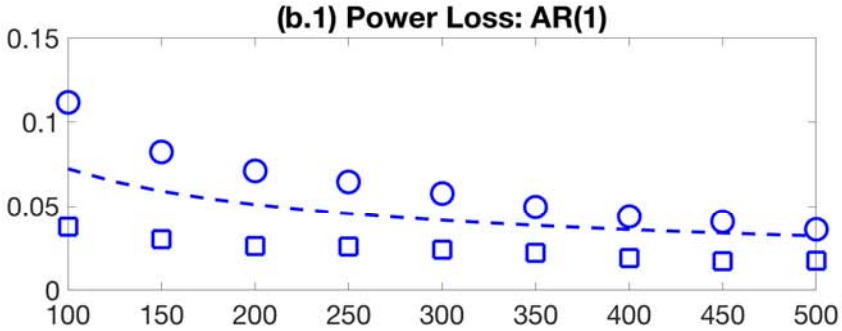
EWP





NW





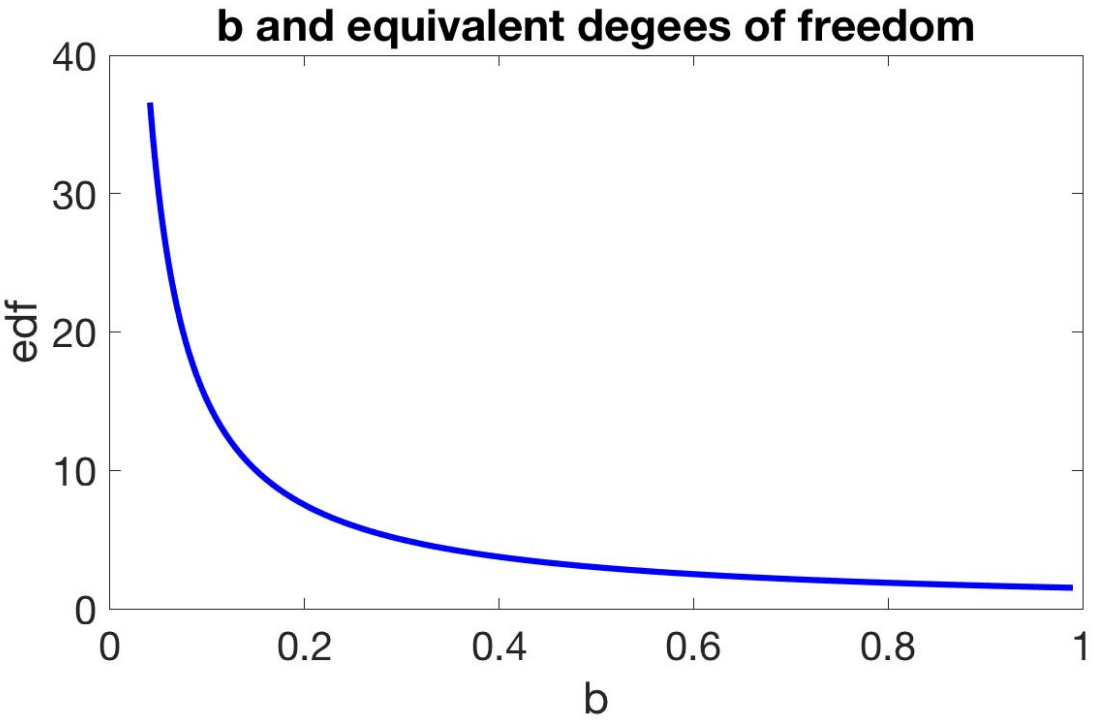
Circle = unrestricted Square = null imposed solid = Gaussian location theory

What about using t_v approximation to NW fixed-b critical values:

Fixed-b distributions are nonstandard but approximately t_v, where

$$v = \left(b \int_{-\infty}^{\infty} k^2(x) dx \right)^{-1}$$

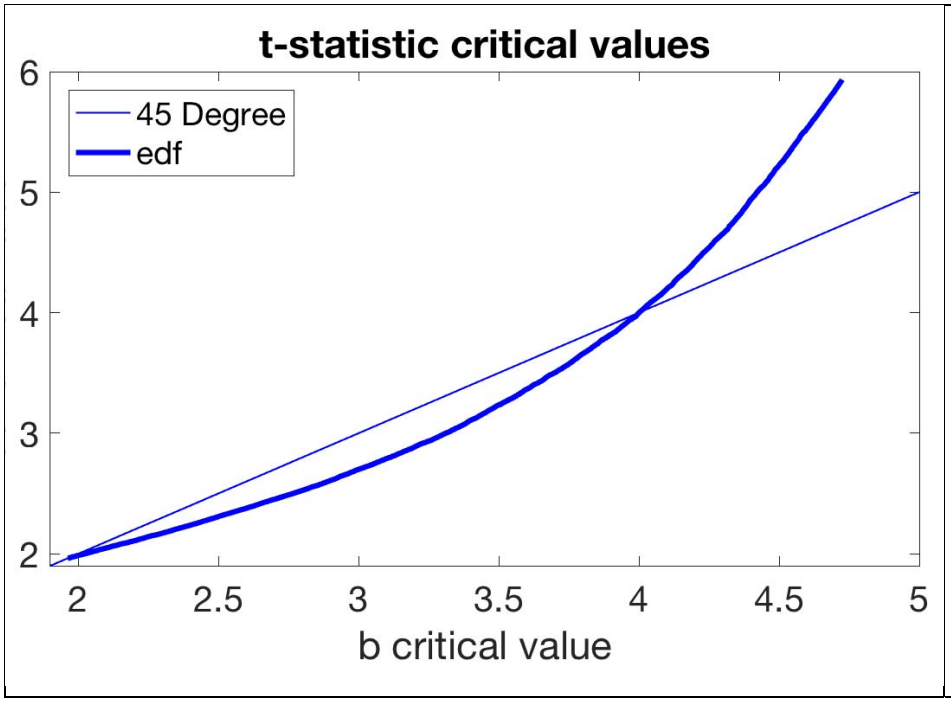
= "Tukey (1949) equivalent degrees of freedom"

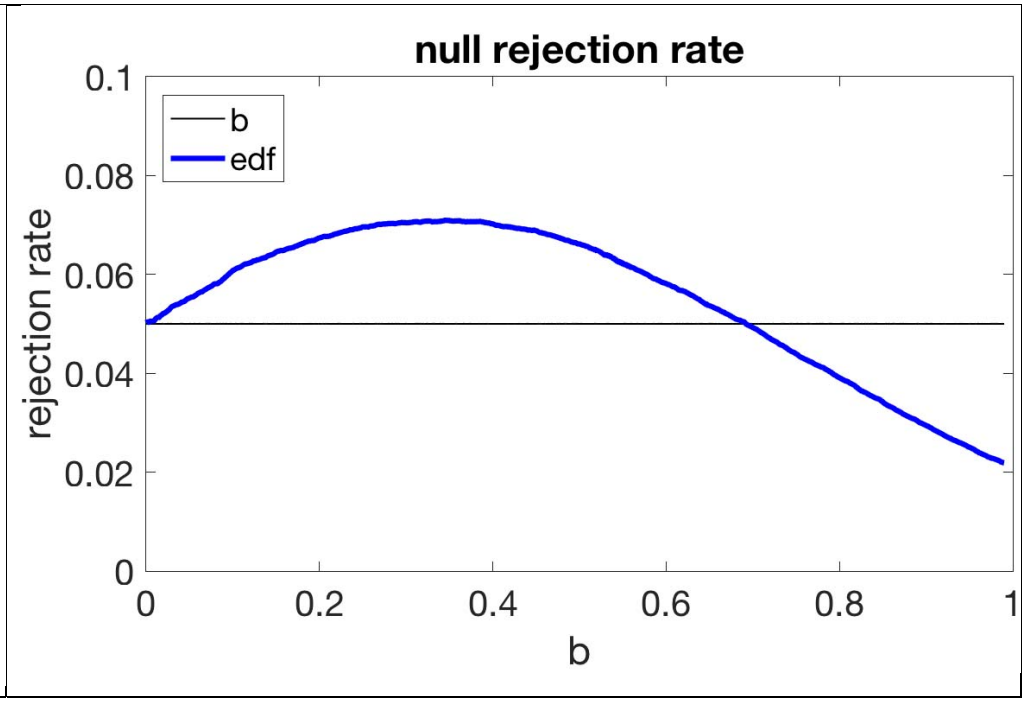


Sun (2014) extends the approximation to higher dimensions

What about using t_ν approximation to NW fixed-b critical values, ctd?

Unfortunately, at least for NW/Bartlett, the t approximation isn't great, at least in the region of interest ($\nu \approx$ 10-30, 5% critical value)





Summary

1. In Gaussian location model, EWP and NW tradeoffs cross in theory in typical sample sizes, but not in the MCs – NW does worse.

2. Using a size/power quadratic loss function with $\kappa = 0.9$ yields concrete rules:

   NW: $b = 1.3T^{-1/2}$ + nonstandard critical values

   EWP: $b = 1.95T^{-2/3}$ + t_v critical values, $v = b^{-1}$

3. In the location model, these rules work well for $|\rho_z| \leq 0.7$

4. In the regression model:
   a. The rules work well and as predicted theoretically for the restricted case

      Interesting side note: this is how weak-IV/AR would be calculated…

   b. Things deteriorate when β is estimated (the unrestricted case).

5. We also consider uniform size control by analytic size adjustment in the Gaussian location model, but have no basis for extending that to regression (not shown)

To do…

1. Stress test the rule – more simulations, especially empirically calibrated.

2. What is causing the distortion when β is estimated?

3. Multiple regression, tests of multiple restrictions & subset tests

4. Can the uniform approach (not shown) be extended to regression?

Additional Monte Carlo Results

Table 1a. HAR test rejection rates: tests of mean, AR(1), normal

(ρ, ϑ)	(0, 0)	(0.5, 0)	(0.7, 0)	(0.9, 0)	(0.95, 0)
NW-MSE(.5)	0.058	0.104	0.168	0.376	0.539
NW-CPE(.5)	0.070	0.088	0.118	0.230	0.374
NW-8	0.112	0.122	0.138	0.184	0.273
QS-MSE(.5)	0.065	0.078	0.115	0.267	0.433
QS-CPE(.5)	0.068	0.077	0.111	0.247	0.407
QS-plugin	0.062	0.082	0.133	0.307	0.532
QS-prewhiten	0.068	0.067	0.073	0.099	0.141
QS-8	0.109	0.113	0.123	0.165	0.258
KVB	--	--	--	--	--
fourier-8	0.085	0.086	0.101	0.145	0.250
fourier-16	0.066	0.072	0.094	0.203	0.356
cos-8	0.083	0.092	0.104	0.168	0.298
cos-16	0.067	0.073	0.098	0.227	0.400
Leg-8	0.084	0.096	0.114	0.168	0.269
Leg-16	0.068	0.085	0.111	0.219	0.364
IM-basis-8	0.084	0.094	0.112	0.173	0.277
IM-basis-16	0.068	0.083	0.115	0.236	0.391
IM-8	0.083	0.095	0.111	0.170	0.277
IM-16	0.066	0.082	0.111	0.226	0.374

Table 1b. HAR test rejection rates: tests of mean, AR(1), fixed-b

(ρ, ϑ)	(0, 0)	(0.5, 0)	(0.7, 0)	(0.9, 0)	(0.95, 0)
NW-MSE(.5)	0.050	0.092	0.155	0.360	0.528
NW-CPE(.5)	0.049	0.063	0.091	0.195	0.340
NW-8	0.048	0.058	0.074	0.126	0.225
QS-MSE(.5)	0.051	0.061	0.094	0.238	0.403
QS-CPE(.5)	0.051	0.059	0.086	0.217	0.373
QS-plugin	0.052	0.068	0.072	0.282	0.515
QS-prewhiten	0.051	0.051	0.057	0.078	0.120
QS-8	0.051	0.052	0.061	0.089	0.168
KVB	0.047	0.055	0.063	0.090	0.132
fourier-8	0.048	0.051	0.061	0.094	0.184
fourier-16	0.050	0.054	0.071	0.170	0.321
cos-8	0.049	0.052	0.061	0.111	0.229
cos-16	0.051	0.054	0.077	0.194	0.366
Leg-8	0.049	0.059	0.072	0.112	0.203
Leg-16	0.051	0.062	0.088	0.185	0.331
IM-basis-8	0.049	0.055	0.067	0.114	0.213
IM-basis-16	0.050	0.063	0.088	0.206	0.357
IM-8	0.049	0.056	0.068	0.112	0.210
IM-16	0.046	0.062	0.087	0.192	0.338

Table 2a. HAR tests of coefficient on a single stochastic regressor, VAR(1), fixed-b. NW MSE(.5), QS MSE-.5 use N(0,1) critical values

(ρ, ϑ)	(0, 0)	(0.5, 0)	(0.7, 0)	(0.9, 0)	(0.95, 0)
NW-MSE(.5)	0.062	0.089	0.127	0.279	0.408
NW-CPE(.5)	0.075	0.095	0.118	0.208	0.300
NW-8	0.051	0.065	0.081	0.145	0.208
QS-MSE(.5)	0.070	0.085	0.107	0.207	0.315
QS-CPE(.5)	0.054	0.068	0.087	0.173	0.267
QS-plugin	0.055	0.074	0.085	0.260	0.463
QS-prewhiten	0.055	0.065	0.076	0.114	0.142
QS-8	0.052	0.063	0.074	0.121	0.170
KVB	0.049	0.061	0.074	0.121	0.166
fourier-8	0.052	0.061	0.070	0.121	0.170
fourier-16	0.054	0.066	0.080	0.145	0.228
cos-8	0.048	0.062	0.075	0.125	0.175
cos-16	0.054	0.066	0.082	0.152	0.237
Leg-8	0.053	0.068	0.084	0.141	0.197
Leg-16	0.057	0.070	0.093	0.171	0.250
IM-basis-8	0.052	0.062	0.078	0.134	0.195
IM-basis-16	0.054	0.070	0.091	0.178	0.271

Table 6b, design (b). HAR test size: Tests of coefficient on x_t in cumulative h-step ahead regression, fixed-b critical values: VAR(1), $h = 8$.

(α, ϑ)	(0, 0)	(0.5, 0)	(0.7, 0)	(0.9, 0)	(0.95, 0)
NW-MSE(.5)	0.048	0.090	0.121	0.191	0.226
NW-CPE(.5)	0.050	0.074	0.087	0.134	0.159
NW-8	0.048	0.065	0.082	0.101	0.128
QS-MSE(.5)	0.048	0.070	0.081	0.130	0.153
QS-CPE(.5)	0.049	0.067	0.078	0.123	0.147
QS-plugin	0.048	0.084	0.157	0.346	0.461
QS-prewhiten	0.042	0.052	0.054	0.067	0.070
QS-8	0.050	0.063	0.073	0.104	0.124
KVB	0.048	0.067	0.073	0.102	0.125
fourier-8	0.051	0.064	0.072	0.105	0.125
fourier-16	0.054	0.069	0.080	0.119	0.141
cos-8	0.054	0.068	0.077	0.112	0.131
cos-16	0.055	0.071	0.082	0.123	0.146
Leg-8	0.069	0.086	0.097	0.136	0.157
Leg-16	0.071	0.092	0.106	0.150	0.179
IM-basis-8	0.049	0.067	0.076	0.111	0.136
IM-basis-16	0.052	0.073	0.088	0.132	0.157
IM-8	0.042	0.048	0.055	0.059	0.052
IM-16	0.046	0.052	0.056	0.057	0.052

Table 7, Predictive Regression design, $h = 4$.

(α, ρ)	(0.7, 0)	(0.7, 0.5)	(0.7, 0.7)	(0.7, 0.9)	(0.7, 0.95)
NW-MSE(.5)	0.094	0.102	0.099	0.113	0.115
NW-CPE(.5)	0.079	0.084	0.087	0.107	0.107
NW-8	0.070	0.076	0.080	0.096	0.098
QS-MSE(.5)	0.072	0.079	0.081	0.100	0.100
QS-CPE(.5)	0.073	0.078	0.081	0.099	0.101
QS-plugin	0.072	0.076	0.080	0.097	0.094
QS-prewhiten	0.061	0.064	0.068	0.085	0.088
QS-8	0.068	0.077	0.079	0.095	0.100
KVB	0.067	0.073	0.080	0.093	0.094
fourier-8	0.066	0.075	0.078	0.094	0.096
fourier-16	0.072	0.078	0.082	0.101	0.101
cos-8	0.070	0.079	0.082	0.096	0.100
cos-16	0.075	0.080	0.084	0.103	0.104
Leg-8	0.087	0.096	0.102	0.118	0.117
Leg-16	0.093	0.100	0.103	0.123	0.122
IM-basis-8	0.069	0.078	0.079	0.098	0.097
IM-basis-16	0.077	0.081	0.088	0.106	0.108
IM-8	0.049	0.199	0.358	0.588	0.645
IM-16	0.055	0.444	0.743	0.936	0.960

Table 11. HAR test size, design (f): Tests of coefficient on x_t in cumulative h-step ahead regression, with randomly drawn "flipped" macro time series: HP-detrended data, fixed-b critical values

Transformation	**Δ**	**Δ**	**HP**	**HP**
h	**4**	**8**	**4**	**8**
NW-MSE(.5)	0.121	0.144	0.272	0.312
NW-CPE(.5)	0.102	0.118	0.207	0.236
NW-8	0.076	0.083	0.126	0.117
QS-MSE(.5)	0.105	0.125	0.204	0.233
QS-CPE(.5)	0.105	0.123	0.195	0.223
QS-plugin	0.094	0.137	0.196	0.322
QS-prewhiten	0.091	0.091	0.128	0.106
QS-8	0.091	0.107	0.152	0.160
KVB	0.073	0.088	0.116	0.119
fourier-8	0.088	0.115	0.166	0.190
fourier-16	0.111	0.141	0.212	0.247
cos-8	0.104	0.135	0.190	0.222
cos-16	0.115	0.145	0.222	0.258
Leg-8	0.150	0.188	0.204	0.221
Leg-16	0.150	0.196	0.267	0.300
IM-basis-8	0.094	0.119	0.166	0.191
IM-basis-16	0.111	0.131	0.215	0.253
IM-8	0.033	0.032	0.035	0.033
IM-16	0.035	0.036	0.036	0.034

Split Sample (batch means) Estimator

Subsample tests have uncontrolled size with moderate persistence.

Monte Carlo: h-period returns predictive regression, ρ = correlation between (x_t, $returns_t$) innovations, $\alpha = x_t$ AR(1) coefficient.

Null rejection rates: $h = 4$, $T = 200$
All tests evaluated using fixed-b critical values

(α, ρ)	(0.7, 0)	(0.7, 0.5)	(0.7, 0.7)	(0.7, 0.9)	(0.7, 0.95)
NW-8	0.070	0.076	0.080	0.096	0.098
QS-8	0.068	0.077	0.079	0.095	0.100
fourier-8	0.066	0.075	0.078	0.094	0.096
IM-basis-8	0.069	0.078	0.079	0.098	0.097
IM-basis-16	0.077	0.081	0.088	0.106	0.108
IM-8	0.049	0.199	0.358	0.588	0.645
IM-16	0.055	0.444	0.743	0.936	0.960

Theory: subsampling effectively turns moderate persistence into high persistence, so the "predictive regression" problem arises in subsamples even if it isn't present in the full sample. LLSW (2017b).