Nearly Minimal Weighted Risk Unbiased Estimation

Ulrich K. Müller and Yulong Wang
Princeton University

November 2015

Motivation

- Interest in non-standard problems
  - Weak instruments
  - Time series with local-to-unity autoregressive roots
  - Moment inequalities
  - Etc.
- Much recent work on in some sense efficient hypothesis tests/confidence intervals
- Good point estimators?

Properties of Good Estimators

- Low risk relative to some loss function

  ⇒ But uniformly minimal risk estimator essentially never exists

- Invariance

  ⇒ But not necessarily applicable, or only partially so

- Weighted risk minimizing estimators

  ⇒ General approach, leads to Bayes solution, but dependence on weight function/prior

- Unbiasedness

Unbiasedness as Reasonable Constraint?

- No decision theoretic justification

- Limits and disciplines dependence on weight function/prior

- Can be of direct interest

  $\Rightarrow$ Want to use estimator on many data sets and average

  $\Rightarrow$ Repeated applications, such as Value at Risk estimator

- Revealed preference

  $\Rightarrow$ Very large literature on bias and bias reducing techniques

AR(1) Bias Literature

- Analysis of bias: Hurwicz (1950), Marriott and Pope (1954), Kendall (1954), White (1961), Sawa (1978), Phillips (1977, 1978), Tanaka (1983), Shaman and Stine (1988), Le Breton and Pham (1989), Phillips and Yu (2009), Yu (2012)

- Alternative estimators: Quenouille (1956), Orcutt and Winokur (1969), Andrews (1993), Andrews and Chen (1994), Park and Fuller (1995), MacKinnon and Smith (1998), Cheang and Reinsel (2000), Roy and Fuller (2001), Hayakawa (2006), Crump (2008), Phillips and Han (2008), Han, Phillips and Sul (2011), Phillips (2012)

Other Applications

- MA(1)/TVP: Sargan and Bhargava (1983), Tanaka (1984), Shephard and Harvey (1990), Shephard (1993), Cordeiro and Klein (1994), Davis and Dunsmuir (1996), Stock and Watson (1998), Bao and Ullah (2007), Demos and Kyriakopoulou (2008)

- Predictive Regression: Mankiw and Shapiro (1986), Stambaugh (1986, 1999), Cavanagh, Elliott and Stock (1995), Kothari and Shanken (1997), Amihud and Hurvich (2004), Eliasz (2005), Jansson and Moreira (2006), Chen and Deo (2009)

- AR(1) Forecasting: Phillips (1979), Stock (1996), Kemp (1999), Gospodinov (2002), Turner (2004), Elliott (2006)

This Paper

- Seek weighted average risk minimizing estimators under unbiasedness constraint

- Leads to Lagrangian with unknown multipliers

- Use numerical techniques to approximate solution

  $\Rightarrow$ Lower bound on weighted risk of all nearly unbiased estimators

  $\Rightarrow$ Identify nearly weighted risk minimizing nearly unbiased estimator

  $\Rightarrow$ "Estimation counterpart" to Elliott, Müller and Watson (2015)

Two Noteworthy Special Cases

1. Quadratic loss and mean bias constraint

   $\Rightarrow$ Lagrange multipliers solve a positive definite quadratic programming problem

   $\Rightarrow$ Numerically very easy to implement

2. Median unbiased estimation without nuisance parameters

   $\Rightarrow$ Can invert median function to obtain exactly median unbiased estimator (Lehmann (1959))

   $\Rightarrow$ Contribution here is determination of right statistic to invert: obtain nearly weighted risk minimizing *exactly* median unbiased estimator

Plan of Presentation

1. Motivation and introduction

2. General approach

   - Lower bound on weighted risk of nearly unbiased estimators

   - Numerical approximation

   - Two special cases; AR(1) Application

3. Invariance

4. Further Applications
   Nearly noninvertible MA(1); Predictive regression; Forecast from AR(1) under quantile constraint; Minimum of Means

5. Conclusions

Set-up and Notation

- Observe $X \in \mathcal{X}$ with density f_θ relative to ν, $\theta \in \Theta$

- Estimand $\eta = h(\theta) \in H$, estimator $\delta : \mathcal{X} \mapsto H$

- Loss function $\ell : H \times \Theta \mapsto [0, \infty)$

- Risk $r(\delta, \theta) = E_\theta[\ell(\delta(X), \theta)] = \int \ell(\delta(x), \theta) f_\theta(x) d\nu(x)$

Bias and Weighted Risk

- For some $c : H \times \Theta \mapsto \mathbb{R}$, bias of δ is $b(\delta, \theta) = E_\theta[c(\delta(X), \theta)]$

  $\Rightarrow c(\eta, \theta) = \eta - h(\theta)$ yields mean bias $b(\delta, \theta) = E_\theta[\delta(X)] - h(\theta)$

  $\Rightarrow \mathbf{1}[\eta > h(\theta)] - \frac{1}{2}$ yields median bias $b(\delta, \theta) = P_\theta[\delta(X) > h(\theta)] - \frac{1}{2}$

- Weighted average risk $R(\delta, F) = \int r(\delta, \theta) dF(\theta)$ for given non-negative measure F

- Weighted risk minimal nearly unbiased estimator δ^* solves

$$\min_\delta R(\delta, F) \quad \text{s.t.}$$
$$|b(\delta, \theta)| \leq \varepsilon \text{ for all } \theta \in \Theta$$

Risk Unbiasedness

- Lehmann and Casella (1998): An estimator δ is *risk unbiased* if

$$E_{\theta_0}[\ell(\delta(X), \theta_0)] \leq E_{\theta_0}[\ell(\delta(X), \theta)]$$

for any $\theta_0, \theta \in \Theta$. Idea: estimate $\delta(x)$ is at least as close to the true value θ_0 on average as measured by ℓ as it is to any other value of θ.

- Easy to show:

  $\Rightarrow$ Under squared loss, risk unbiased estimators are mean unbiased

  $\Rightarrow$ Under absolute value loss, risk unbiased estimators are median unbiased

- We adopt these pairings in the following, but approach goes through for other choices

Weighted Risk Lower Bound

- Define weighted average bias $B(\delta, G) = \int b(\delta,\theta)dG(\theta)$ for some probability distribution G. Relax bias constraint $|b(\delta,\theta)| \le \varepsilon \; \forall \theta \in \Theta$ to

$$|B(\delta, G_i)| \le \varepsilon \text{ for all } i = 1, \ldots, m.$$

- Associated Lagrangian with $\lambda = ((\lambda_1^l, \lambda_1^u), \ldots, (\lambda_m^l, \lambda_m^u))$

$$L(\delta,\lambda) = R(\delta, F) + \sum_{i=1}^m \lambda_i^u (B(\delta, G_i) - \varepsilon) + \sum_{i=1}^m \lambda_i^l (-B(\delta, G_i) - \varepsilon)$$

$$= \int \Bigg(\int f_\theta(x) \ell(\delta(x), \theta) dF(\theta) + \sum_{i=1}^m \lambda_i^u \Big(\int f_\theta(x) c(\delta(x), \theta) dG_i(\theta) - \varepsilon \Big) + \sum_{i=1}^m \lambda_i^l \Big(-\int f_\theta(x) c(\delta(x),\theta) dG_i(\theta) - \varepsilon \Big) \Bigg) d\nu(x)$$

$\Rightarrow$ minimized by δ_λ that minimizes integrand for each x

Weighted Risk Lower Bound, ctd

$$L(\delta,\lambda) = R(\delta,F) + \sum_{i=1}^{m} \lambda_i^u (B(\delta,G_i) - \varepsilon) + \sum_{i=1}^{m} \lambda_i^l (-B(\delta,G_i) - \varepsilon)$$

Lemma: Suppose δ^* satisfies $|B(\delta^*,G_i)| \leq \varepsilon$ for $i = 1,\ldots,m$, and for arbitrary $\lambda \geq 0$ (that is, each element in λ is nonnegative), δ_λ minimizes $L(\delta,\lambda)$. Then $R(\delta^*,F) \geq L(\delta_\lambda,\lambda)$.

Proof: For any δ, $L(\delta_\lambda,\lambda) \leq L(\delta,\lambda)$ by definition of δ_λ, so in particular, $L(\delta_\lambda,\lambda) \leq L(\delta^*,\lambda)$. Furthermore, $L(\delta^*,\lambda) \leq R(\delta^*,F)$, since $\lambda \geq 0$ and δ^* satisfies $|B(\delta^*,G_i)| \leq \varepsilon$. Combining these inequalities yields the result.

$\Rightarrow$ Holds a fortiori for any estimator δ^* satisfying $|b(\delta^*,\theta)| \leq \varepsilon$ for all $\theta \in \mathbf{\Theta}$

$\Rightarrow$ Smallest lower bound obtained by $\lambda^* \geq 0$ such that $|B(\delta_{\lambda^*},G_i)| \leq \varepsilon$ and complementarity slackness conditions hold

Nearly Weighted Risk Minimal Unbiased Estimator

- Aim: For given $\varepsilon_B > 0$ and $\varepsilon_R > 0$, seek $\hat{\delta}$ so that
  - $|b(\hat{\delta}, \theta)| \leq \varepsilon_B$ for all $\theta \in \Theta_0$
  - $R(\hat{\delta}, F) \leq (1 + \varepsilon_R) R(\delta, F)$ for any δ satisfying $|b(\delta, \theta)| \leq \varepsilon_B$

Numerical Approach

1. Discretize Θ by point masses or other distributions G_i, $i = 1, \ldots, m$.

2. Obtain approximately optimal Lagrange multipliers $\hat{\lambda}$ for

$$\min_{\delta} R(\delta, F) \quad \text{s.t. } |B(\delta, \theta)| \leq \varepsilon_B \text{ for } i = 1, \ldots, m$$

   and associated risk lower bound $\underline{R} = L(\delta_{\hat{\lambda}}, \hat{\lambda})$.

3. Obtain an approximate solution $(\hat{\varepsilon}^*, \hat{\delta}^*)$ to the problem

$$\min_{\varepsilon \geq 0, \delta} \varepsilon \text{ s.t. } R(\delta, F) \leq (1 + \varepsilon_R)\underline{R} \text{ and } |B(\delta, G_i)| \leq \varepsilon \text{ for } i = 1, \ldots, m$$

   and check whether $\hat{\delta}^*$ satisfies $|b(\hat{\delta}, \theta)| \leq \varepsilon_B$. If it doesn't, go back to Step 1 and use a finer discretization. If it does, $\hat{\delta} = \hat{\delta}^*$ has the desired properties by Lemma 1.

Switching Estimators

- Problem: Numerical approach for unbounded parameter spaces?

- Most non-standard problems "become standard" as $\varsigma(\theta) \to \infty$ for appropriate choice of ς (cf. EMW), and good unbiased estimator δ_S available

  - In local-to-unity problem, OLS estimator normal as $\varsigma(\theta) = \theta \to \infty$ (Phillips 1987, Chan and Wei 1987)

  - In weak instruments, concentration parameter $\to \infty$ recovers good properties of MLE

- Proposed solution: consider estimators of the switching form

$$\hat{\delta}(x) = \chi(x)\delta_S(x) + (1 - \chi(x))\delta_N(x)$$

for $\chi(x) = \mathbf{1}[\hat{\varsigma}(x) > \kappa_\chi]$ and estimator $\hat{\varsigma}(x)$ of $\varsigma(\theta)$, and numerically determine δ_N.

Switching Estimators, ctd

- Determine δ_N to ensure overall unbiasedness of $\hat{\delta}$

  $\Rightarrow$ Component of bias $b(\hat{\delta}, \theta)$ now inherited from δ_S, unresponsive to Lagrange multipliers that determine δ_N

  $\Rightarrow$ feasibility in general depends on χ and δ_S

- Under switching, good properties of $\hat{\delta}$ for large $\varsigma(\theta)$ inherited from δ_S

  $\Rightarrow$ weighting function F can focus on truly problematic region

  $\Rightarrow$ avoids potential uniformity issues as $\varsigma(\theta) \to \infty$ for appropriately chosen δ_S (cf. Mikusheva (2008), Phillips (2014))

- Don't impose switching for calculation of risk bound

Special Case 1: Squared Loss and Mean Bias

- For given x, Lagrangian

$$\int f_\theta(x)\ell(\delta(x),\theta)dF(\theta)+\sum_{i=1}^{m}(\lambda_i^u-\lambda_i^l)\int f_\theta(x)c(\delta(x),\theta)dG_i(\theta)$$

is quadratic in $\delta(x)$, and optimal $\delta(x)$ is linear in $\tilde{\lambda}_i=\frac{1}{2}(\lambda_i^l-\lambda_i^u)$

- Risk becomes quadratic form in $\tilde{\lambda}$, and (weighted average) bias is linear function of $\tilde{\lambda}$

  $\Rightarrow$ Determination of $\tilde{\lambda}$ is linearly constrained quadratic programming with positive definite objective

  $\Rightarrow$ Easy and fast numerical solution even for very large m

- Same structure also under switching

Example: AR(1)

- Observe

$$y_t = \rho y_{t-1} + \varepsilon_t, t = 1, \ldots, T$$

where $\varepsilon_t \sim iid\mathcal{N}(0, 1)$ and $y_0 = 0$. Under $\rho = \rho_T = 1 - \theta/T$, limit experiment is observation of O-U process

$$dX(s) = -\theta X(s)ds + dW(s)$$

with $X(0) = 0$ and W a standard Wiener process

- Density is $f_\theta(x) = \exp[-\frac{1}{2}\theta(x(1)^2 - 1) - \frac{1}{2}\theta^2 \int_0^1 x(s)^2 ds]$

- OLS/MLE estimator $\delta_{OLS}(x) = -\frac{1}{2}(x(1)^2 - 1)/\int_0^1 x(s)^2 ds$

AR(1) Bias and Risk Normalization

- Phillips (1987), Chan and Wei (1987): As $\theta \to \infty$, $\delta_{OLS}(X) \overset{a}{\sim} \mathcal{N}(\theta, 2\theta)$

- Let $n(\theta) = \sqrt{2\theta + 10}$, and define
  - Normalized mean bias: $b_n(\delta, \theta) = (E_\theta[\delta(X)] - \theta)/n(\theta)$

    $\Rightarrow$ corresponds to $c(\eta, \theta) = (\eta - \theta)/n(\theta)$
  - Normalized risk: $r_n(\delta, \theta) = E_\theta[(\delta(X) - \theta)^2]/n(\theta)^2$

    $\Rightarrow \int r_n(\delta, \theta) dF_n(\theta) = R(\delta, F)$ with $dF(\theta) = dF_n(\theta)/n(\theta)^2$

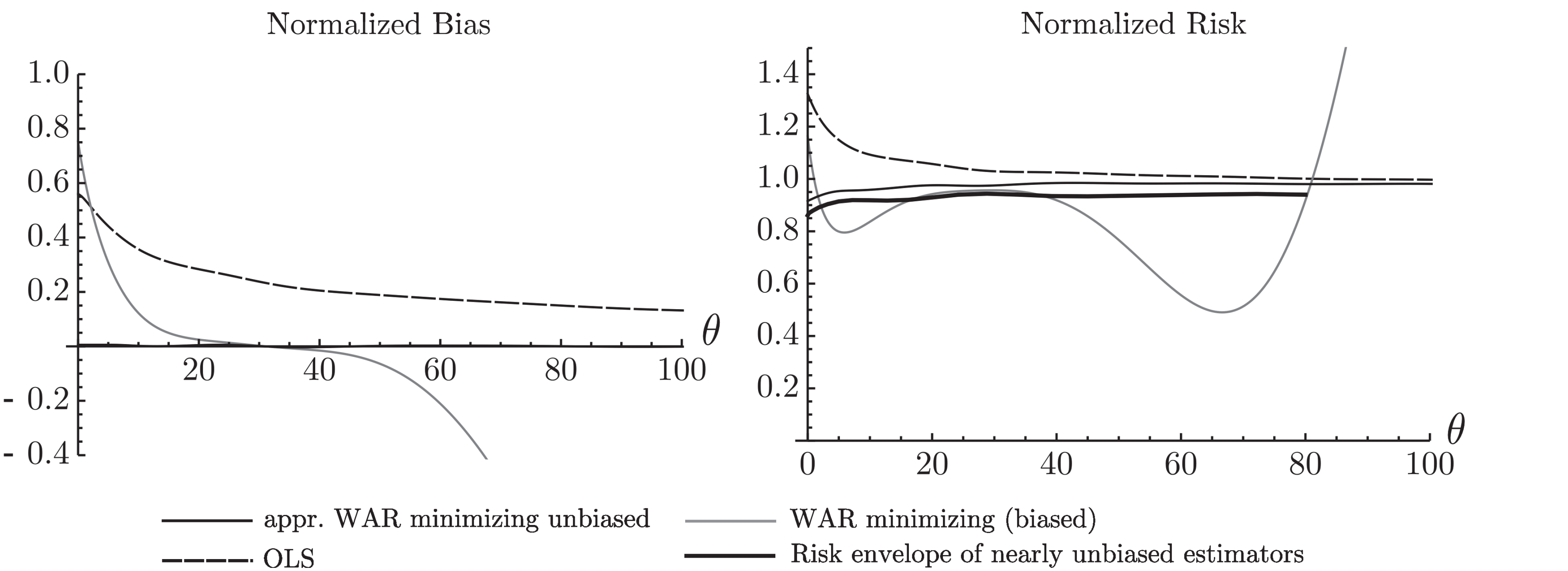
AR(1) Implementation

- F_n uniform on $\theta \in [0, 80]$
- G_i point masses on $\{0.0^2, 0.2^2, \ldots, 10^2\}$
- Switch to $\delta_S(x) = \delta_{OLS}(x) - 2$ if $\delta_{OLS}(x) > 40$

  $\Rightarrow \delta_S$ is local-to-unity limit of bias corrected $\hat{\rho}$ estimator
- Seek estimator with absolute normalized mean bias of less than $\varepsilon_B = 0.01$ and $\varepsilon_R = 0.01$

  $\Rightarrow$ need 10,000 MC draws before bias $\approx$ MC stderr

AR(1) Results



Normalized Bias
1.0
0.8
0.6
0.4
0.2
- 0.2
- 0.4
20
40
60
80
100
θ
Normalized Risk
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
20
40
60
80
100
θ
appr. WAR minimizing unbiased
OLS
WAR minimizing (biased)
Risk envelope of nearly unbiased estimators

Special Case 2: Median Unbiasedness Without Nuisance Parameters

- Estimand is $h(\theta) = \theta$, and $\Theta \subset \mathbb{R}$. Let $m_{\delta_B}(\theta)$ be the median function of δ_B, $P_\theta(\delta_B(X) \leq m_{\delta_B}(\theta)) = 1/2$ for all $\theta \in \Theta$.

- If $m_{\delta_B}(\theta)$ is one-to-one on Θ with inverse $m_{\delta_B}^{-1} : \mathbb{R} \mapsto \Theta$, then $\delta_U(x) = m_{\delta_B}^{-1}(\delta_B(x))$ is exactly median unbiased.

- But how to choose δ_B so that $m_{\delta_B}^{-1}(\delta_B(x))$ has low risk?

- Plausible candidate: nearly minimal weighted risk nearly median unbiased estimator

  $\Rightarrow$ Check by computing risk bound with $\varepsilon_B = 0$

AR(1) Implementation

- Absolute deviation loss normalized by $n(\theta) = \sqrt{2\theta + 10}$, $F_n \sim U[0, 80]$ as above

- G_i are quadratic b-splines, so that $\sum_{i=1}^{m} \lambda_i G_i(\theta)$ has differentiable density

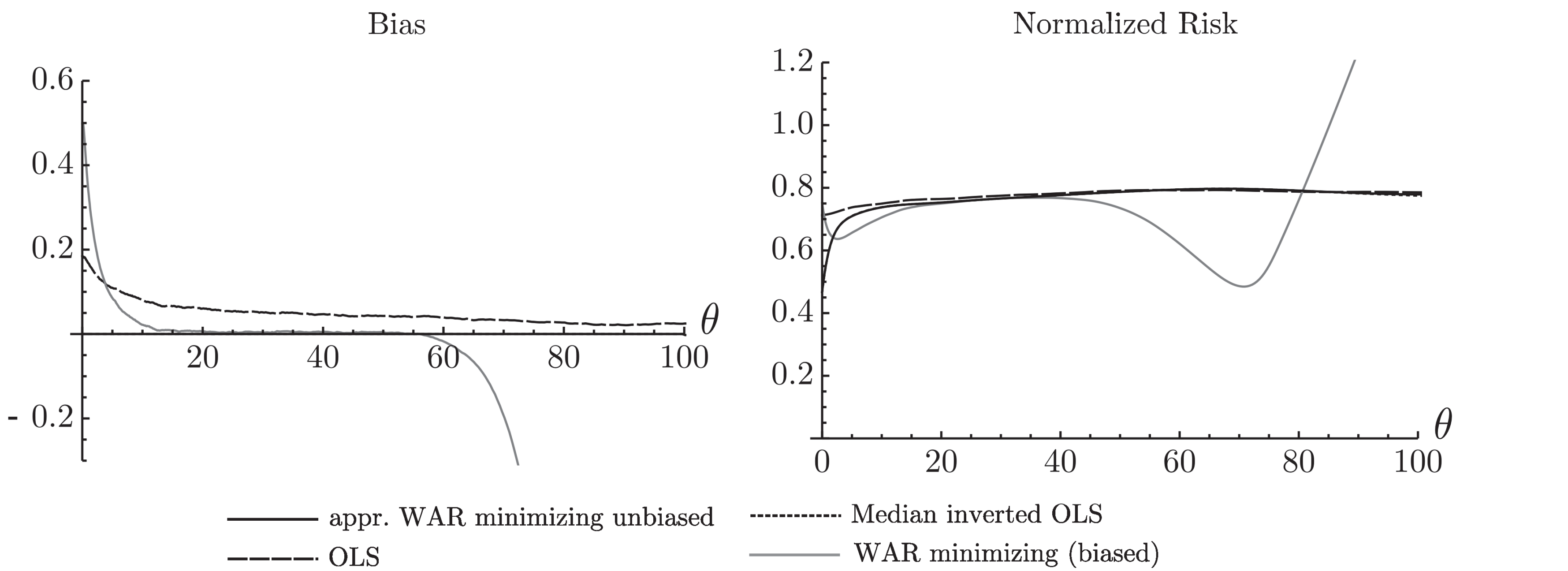
  $\Rightarrow$ ensures that $\delta_\lambda(x)$ is continuous function of x

- Switch to $\delta_S(x) = \delta_{OLS}(x) - 1$ (limit of bias corrected $\hat{\rho}$ using Phillips' (1977, 1978) Edgeworth expansions) when $\delta_{OLS}(x) > 40$

- Exactly median unbiased for $\theta \in [0, 80]$

- $\varepsilon_R = 0.01$

AR(1) Results



Bias
0.6
0.4
0.2
- 0.2
20
40
60
80
100
θ
Normalized Risk
1.2
1.0
0.8
0.6
0.4
0.2
0
20
40
60
80
100
θ
appr. WAR minimizing unbiased
OLS
Median inverted OLS
WAR minimizing (biased)

Invariance

- Three groups with action $a \in A$

$$g(X, a) \in \mathcal{X},\ \bar{g}(\theta, a) \in \Theta,\ \hat{g}(\eta, a) \in H.$$

- If all relevant aspects are invariant, i.e.

  density of $g(X, a)$ is $f_{\bar{g}(\theta,a)}$
  $h(\bar{g}(\theta, a)) = \hat{g}(h(\theta), a)\ \forall\ \theta \in \Theta,\ a \in A$
  $\ell(\eta, \theta) = \ell(\hat{g}(\eta, a), \bar{g}(\theta, a))\ \forall\ \eta \in H,\ \theta \in \Theta$ and $a \in A$
  $c(\eta, \theta) = c(\hat{g}(\eta, a), \bar{g}(\theta, a))\ \forall\ \eta \in H,\ \theta \in \Theta$ and $a \in A$

  then natural to consider estimators satisfying

$$\delta(g(x, a)) = \hat{g}(\delta(x), a) \text{ for all } x \in \mathcal{X} \text{ and } a \in A.$$

Invariance, ctd

- Under invariance, can show

$$\begin{aligned} r(\delta,\theta) &= E_\theta[\ell(\delta(X),\theta)] \\ &= E_{\bar{M}(\theta)}[E_{\bar{M}(\theta)}[\ell\left(\delta(M(X)),\bar{g}(\bar{M}(\theta),O(X)^-)\right)|M(X)]] \end{aligned}$$

where $\bar{M}(\theta)$ and $M(x)$ are maximal invariants, $O(x)$ is orbit so that $x = g(M(x), O(x))$, and a^- is inverse of a. Similarly

$$\begin{aligned} b(\delta,\theta) &= E_\theta[c(\delta(X),\theta)] \\ &= E_{\bar{M}(\theta)}[E_{\bar{M}(\theta)}[c\left(\delta(M(X)),\bar{g}(\bar{M}(\theta),O(X)^-)\right)|M(X)]] \end{aligned}$$

- Same structure as problem above, with observation $X^* = M(X)$, parameter $\theta^* = M(\theta)$ and

$$\begin{aligned} \ell^*(\delta(x^*),\theta^*) &= E_{\theta^*}[\ell\left(\delta(X^*),\bar{g}(\theta^*,O(X)^-)\right)|X^* = x^*] \\ c^*(\delta(x^*),\theta^*) &= E_{\theta^*}[c\left(\delta(X^*),\bar{g}(\theta^*,O(X)^-)\right)|X^* = x^*]. \end{aligned}$$

AR(1) Example

- Suppose

$$\begin{aligned} y_t &= \mu + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t,\ u_0 = O_p(1) \end{aligned}$$

- With $\mu = mT^{1/2}$, limiting problem under $\rho = 1 - \eta/T$ is $X = m + J$, J is OU process with $J(0) = 0$

- $M(X) = X - X(0)$ is maximal invariant with same distribution as above

  $\Rightarrow$ previous results apply

Stationary AR(1) Example

- Suppose instead

$$\begin{aligned} y_t &= \mu + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t \\ u_0 &= \begin{cases} \frac{\varepsilon_0}{1-\rho^2} \text{ for } \rho < 1 \\ O_p(1) \text{ for } \rho \geq 1 \end{cases} \end{aligned}$$

- Under $\mu = mT^{1/2}$, limiting problem now is

$$X = m + J$$

where J is OU process with $J(0) \sim \mathcal{N}(0, (2\eta)^{-1})$ for $\eta > 0$ (and $J(0) = 0$ for $\eta = 0$)

- $X^* = M(X) = X - X(0)$ is maximal invariant with density (cf. Elliott (1999)) that only depends on η

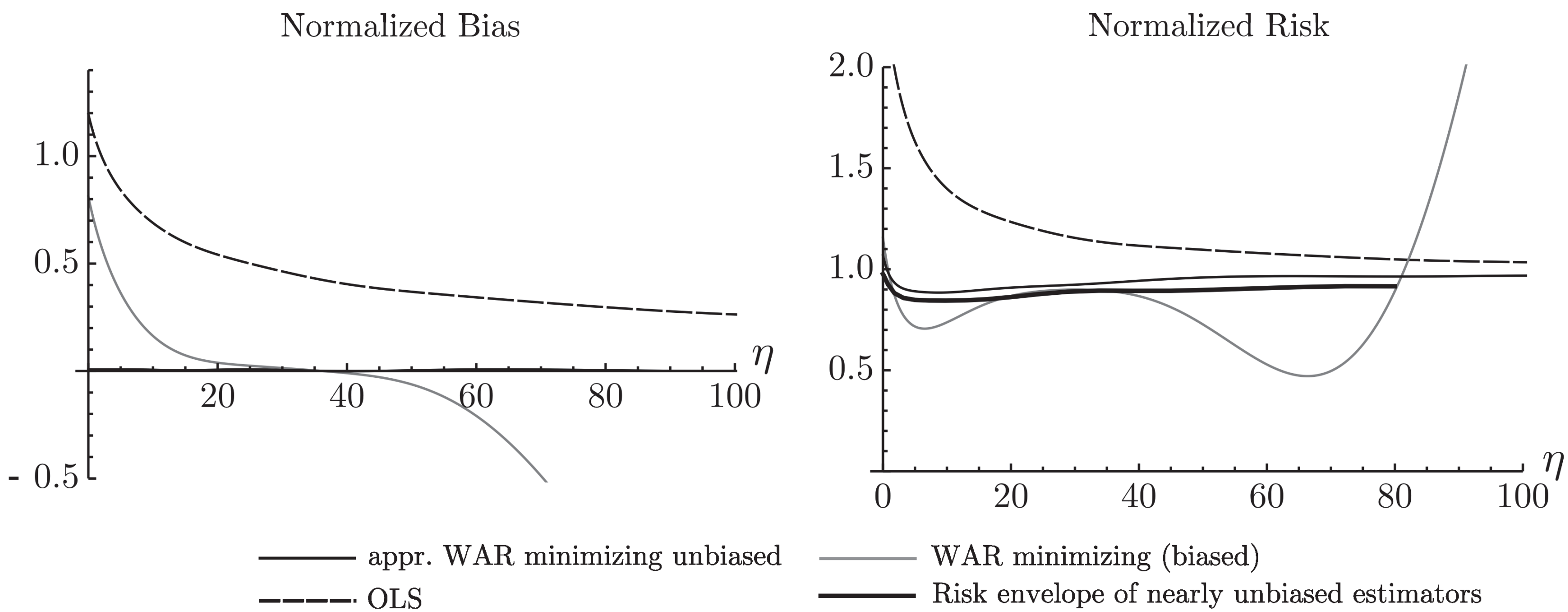
Stationary AR(1) Implementation

- Same normalization and weighting functions as in AR(1) example above

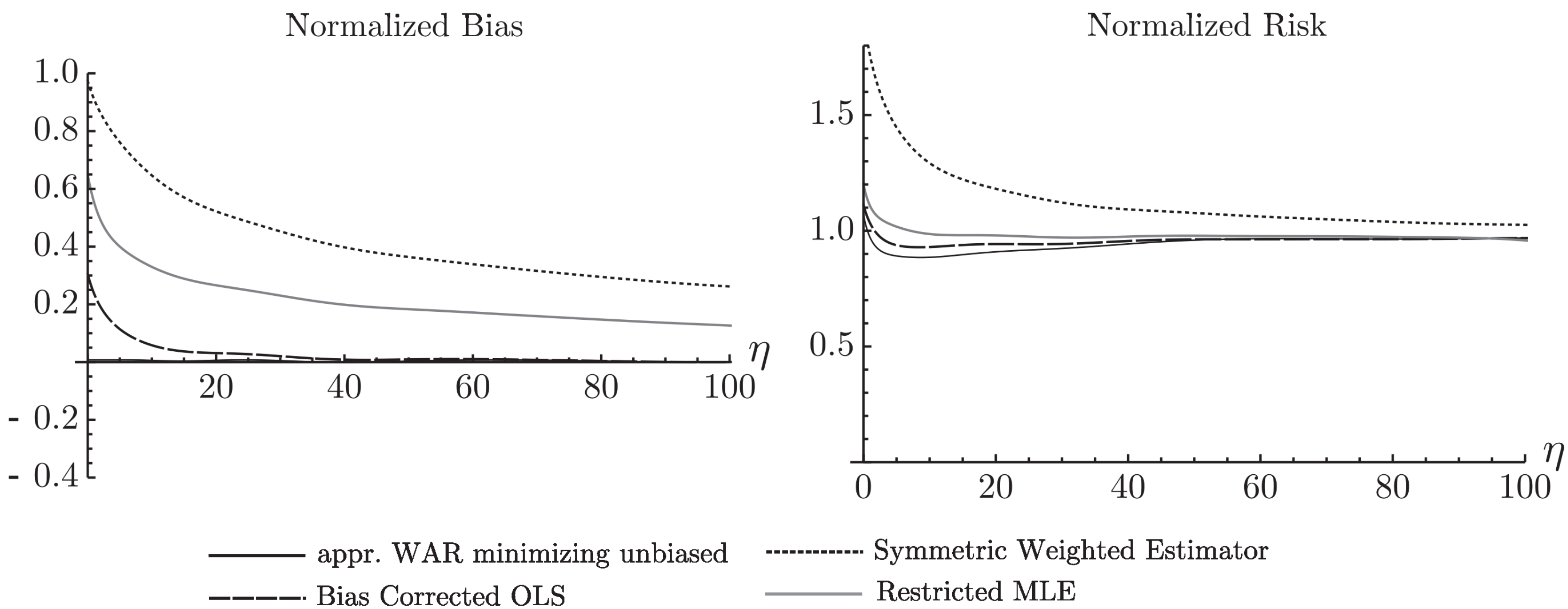
- If $\hat{\delta}_{OLS}(x) > 60$, switch to
  - $\hat{\delta}_{OLS}(x) - 4$ under mean bias constraints
  - $\hat{\delta}_{OLS}(x) - 3$ under median bias constraints

  (local-to-unity limits of bias corrected estimators)

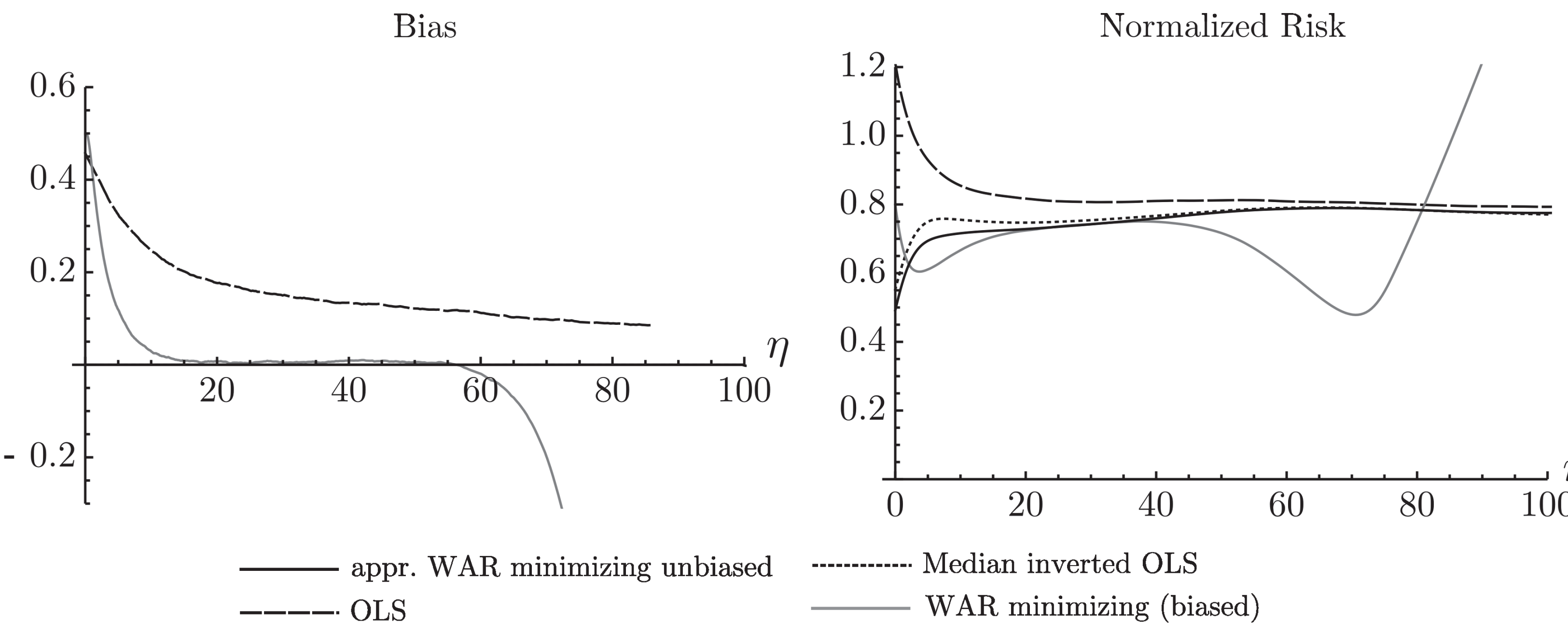
Stationary AR(1) Mean Unbiased



Stationary AR(1) Mean Unbiased



Stationary AR(1) Median Unbiased



MA(1) / Time Varying Parameters

- Canonical local-level model

$$y_t = \mu + \phi \sum_{s=1}^{t} \varepsilon_s + u_t, \;\; (\varepsilon_t, u_t) \sim iid(0, I_2)$$

  Interest in degree of time variation ϕ

- Under translation invariance, maximal invariant is mean-zero MA(1)

$$\Delta y_t = \phi \varepsilon_t + \Delta u_t$$

  and $\phi = 0$ corresponds to non-invertible MA(1)

- Well-known "pile-up" problem for MLE for small ϕ: $P(\hat{\phi} = 0 | \phi = 0) \approx 0.66$ (Sargan and Bhargava (1983))

MA(1) / Time Varying Parameters

- Under $\phi = \theta/T$ asymptotics $y_t = \mu + \frac{\theta}{T}\sum_{s=1}^{t}\varepsilon_s + u_t$ and

$$T^{-1/2}\sum_{t=1}^{\lfloor rT\rfloor}(y_t - \bar{y}) \Rightarrow X(r)$$
$$= \theta\left(\int_0^r W_\varepsilon(s)ds - r\int_0^1 W_\varepsilon(s)ds\right) + W_u(r) - rW_u(1)$$

- Density f_θ of X derived in Elliott and Müller (2005)

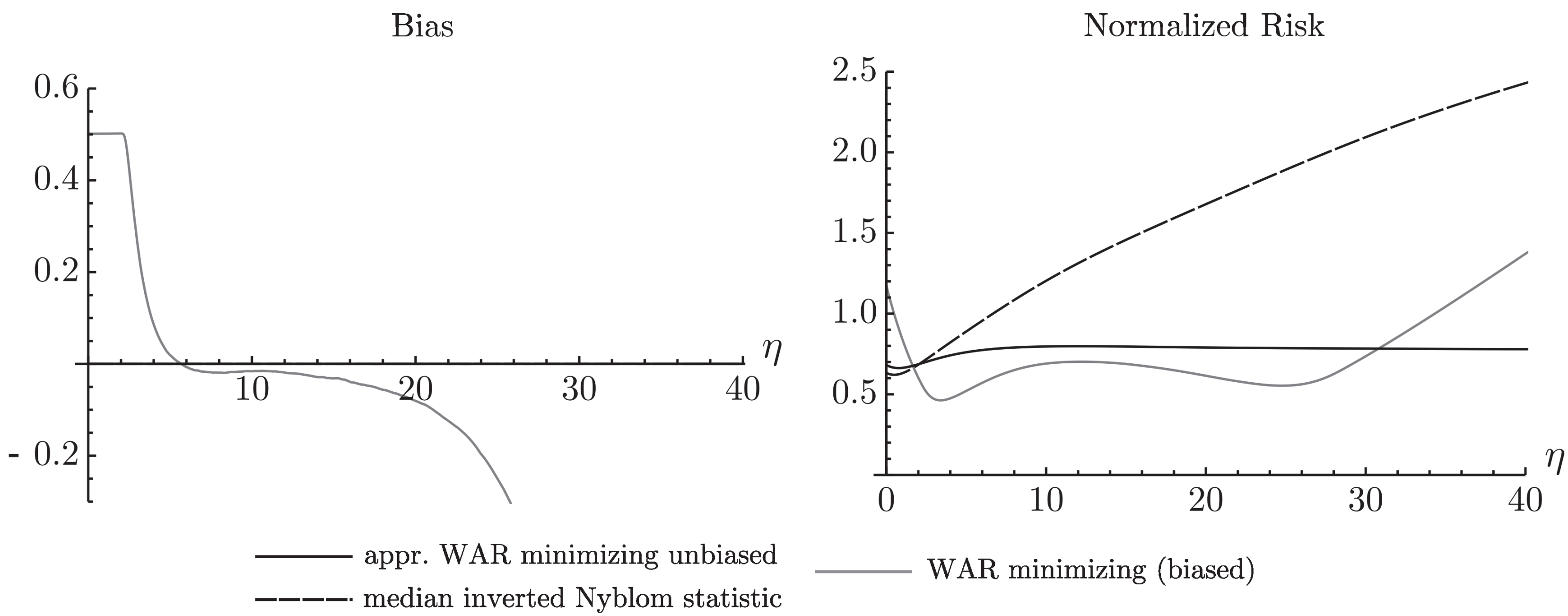
- Stock and Watson (1998) median-invert Nyblom (1989) statistic $\int_0^1 X(r)^2 dr$

MA(1) / TVP Implementation

- Median unbiased estimation under absolute deviation loss
- $\varepsilon_B = 0.01$, $\varepsilon_R = 0.02$, $n(\theta) = \sqrt{\theta + 6}$, $F_n \sim U[0, 30]$
- Switch to $\delta_S(x) = \delta_{MLE}(x) + 1$ if $\delta_{MLE}(x) > 10$

  $\Rightarrow \delta_S$ is limit of bias corrected MLE (cf. Tanaka (1984))

MA(1) / TVP Results



Predictive Regression

- Model for observable $X = (\{y_t\}_{t=1}^T, \{x_t\}_{t=0}^T)$ under $\theta = (\mu_y, \mu_x, \beta, \rho)$

$$
\begin{aligned}
y_t &= \mu_y + \beta x_{t-1} + r\varepsilon_{xt} + \sqrt{1-r^2}\varepsilon_{yt}, \quad (\varepsilon_{xt}, \varepsilon_{yt})' \sim \mathcal{N}(0, I_2) \\
x_t &= \mu_x + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_{xt}, \quad u_0 \sim \mathcal{N}(0, (1-\rho^2)^{-1})
\end{aligned}
$$

- Invariance groups with action $(a_x, a_y, a_b) \in \mathbb{R}^3$

$$
\begin{aligned}
g(X, a) &= (\{y_t + a_y + a_b x_{t-1}\}, \{x_t + a_x\}) \\
\bar{g}(\theta, a) &= \theta + (a_y, a_x, a_b, 0) \\
\hat{g}(\beta, a) &= \beta + a_b
\end{aligned}
$$

$\Rightarrow$ Invariant δ are of the form

$$\delta(x) = \hat{\beta} + \delta(x^*)$$

where $x^* = (\{\hat{y}_t\}, \{\hat{x}_t\})$, $\hat{x}_t = x_t - \bar{x}$, $\hat{y}_t = y_t - \bar{y} - \hat{\beta} x_{t-1}$ and $\hat{\beta} = \frac{\sum \hat{x}_{t-1} y_t}{\sum \hat{x}_{t-1}^2}$

Predictive Regression

- From invariance formula, under squared loss and mean bias

$$\begin{aligned} \ell^*(\delta(x^*),\theta^*) &= E_{\theta^*}[(\delta^*(X^*)+\hat\beta)^2|X^*=x^*] \\ c^*(\delta(x^*),\theta^*) &= E_{\theta^*}[\delta^*(X^*)+\hat\beta|X^*=x^*] \end{aligned}$$

and calculation shows

$$\hat\beta|X^* \sim \mathcal{N}(r(\hat\rho_{OLS}-\rho),(1-r^2)/\sum_{t=1}^{T}\hat x_{t-1}^2)$$

$\Rightarrow$ Very similar to AR(1) estimation problem (cf. Stambaugh (1986))

$\Rightarrow$ Mean unbiased estimates of ρ lead to mean unbiased $\hat\beta$

$\Rightarrow$ Less obvious: Median unbiased estimator of β?

Predictive Regression

- Under local-to-unity $\rho = 1 - \gamma/T$, X and X^* are continuous time processes, and

$$\hat{\beta}|X^* \sim \mathcal{N}(r(\gamma - \hat{\gamma}_{OLS}), (1 - r^2)/\int \hat{J}_x(s)^2 ds)$$

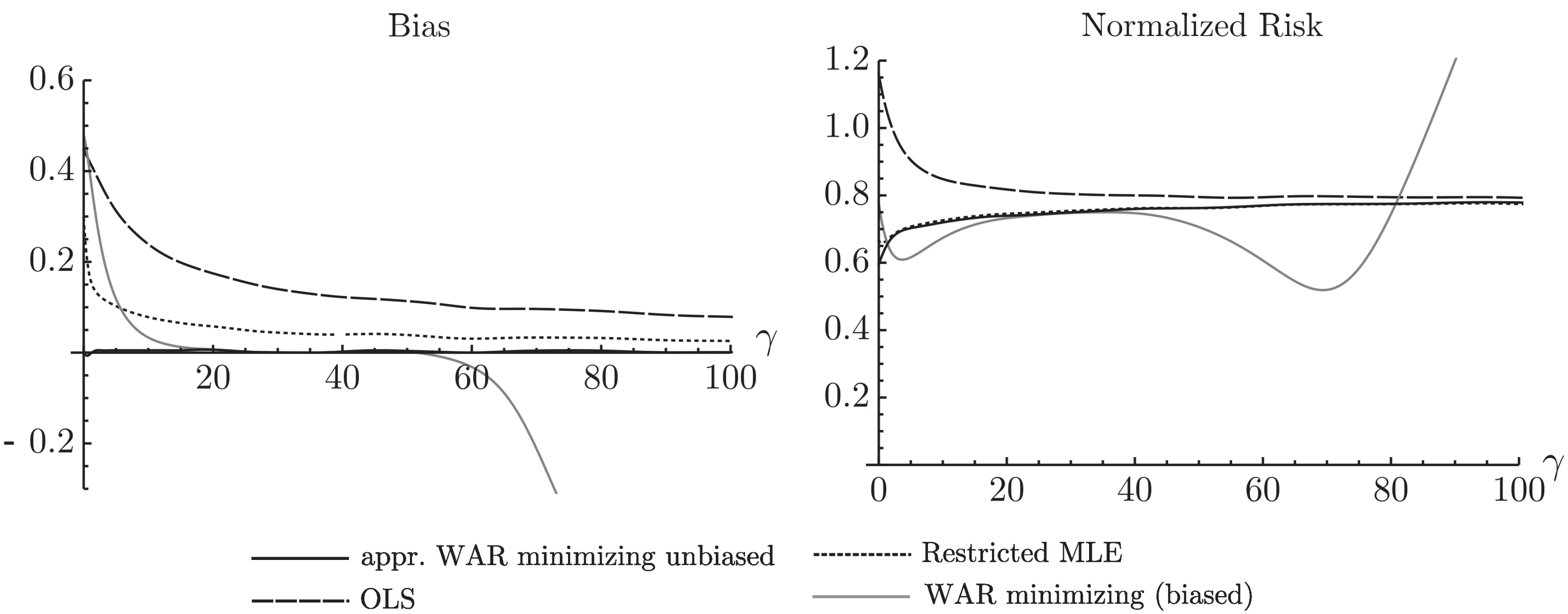
where $\hat{J}_x$ is demeaned OU process and $\hat{\gamma}_{OLS} = -\frac{1}{2}(\hat{J}_x(1)^2 - 1)/\int \hat{J}_x(s)^2 ds$

- Focus on absolute loss and median unbiasedness, $\gamma \in [0, \infty)$

- $F_n \sim U[0, 80]$ on γ, switch to $\delta_S(x^*) = 3r$ for $\hat{\gamma}_{OLS} > 60$

- $\varepsilon_B = 0.01$, $\varepsilon_R = 0.02$

Predictive Regression Results, $r = -0.95$



Quantile AR(1) Forecast

- Seek forecast δ from local-to-unity AR(1) $\{y_t\}_{t=1}^T = X$ and horizon $\lfloor hT \rfloor$ such that

$$P_\theta(y_{T+\lfloor hT \rfloor} > \delta(\{y_t\}_{t=1}^T)) = q \quad \text{for all } \theta$$

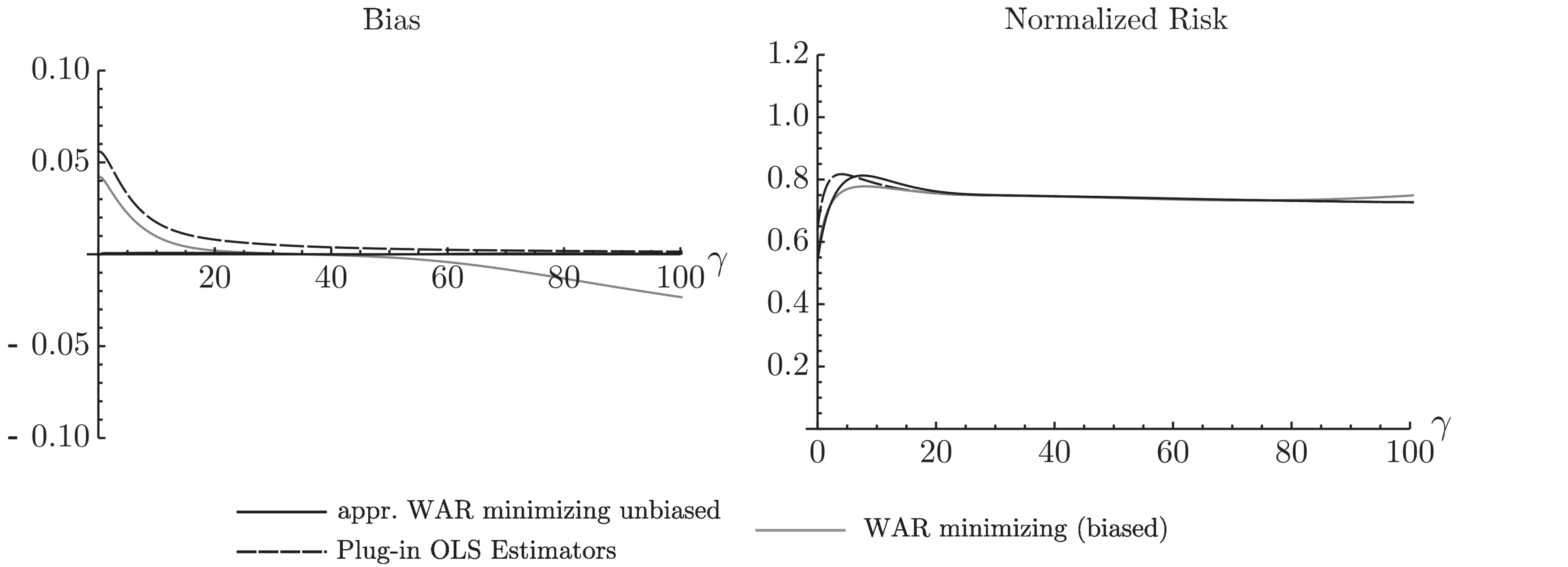
$\Rightarrow$ Can be rewritten as particular $b(\delta, \theta) = E_\theta[c(\delta(X), \theta)] = 0$

- Impose translation invariance

- Use quantile loss function

- $h = 0.2$, $q = 0.05$, $\varepsilon_B = 0.002$, $\varepsilon_R = 0.01$

5% AR(1) Forecast Results



Minimum of Means

- Stylized moment inequality problem

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, I_2\right)$$

  and interest in $h(\theta) = \min(\theta_1, \theta_2)$

- No mean unbiased estimator: Blumenthal and Cohen (1968), Hirano and Porter (2012)

- No median unbiased estimator: Fraser (1952)

- Nearly unbiased estimators?

Minimum of Means

- Impose translation equivariance

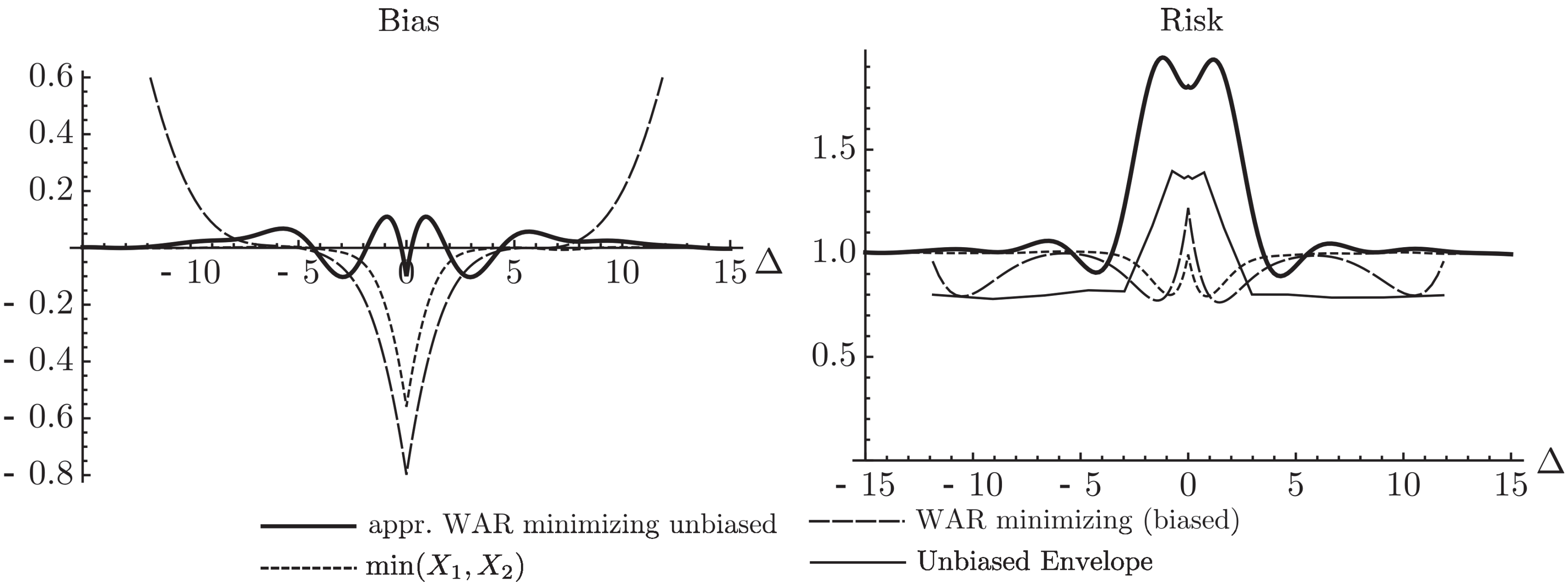
  $\Rightarrow$ Effective parameter becomes $\Delta = \theta_1 - \theta_2$

- Set F uniform on $\Delta \in [-12, 12]$

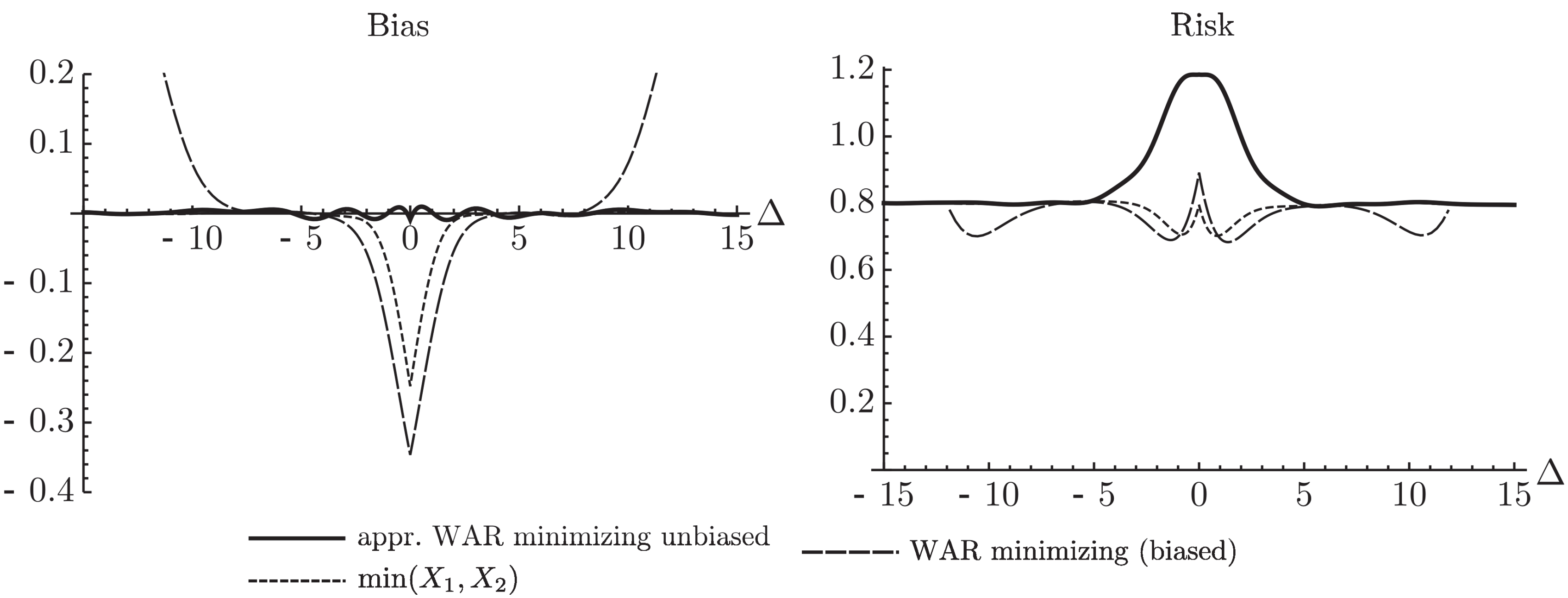
- Switch to $\delta_S = \min(X_1, X_2)$ if $|X_1 - X_2| > 10$

- For mean bias, set $\varepsilon_B = 0.1$, and for median bias, $\varepsilon_B = 0.01$

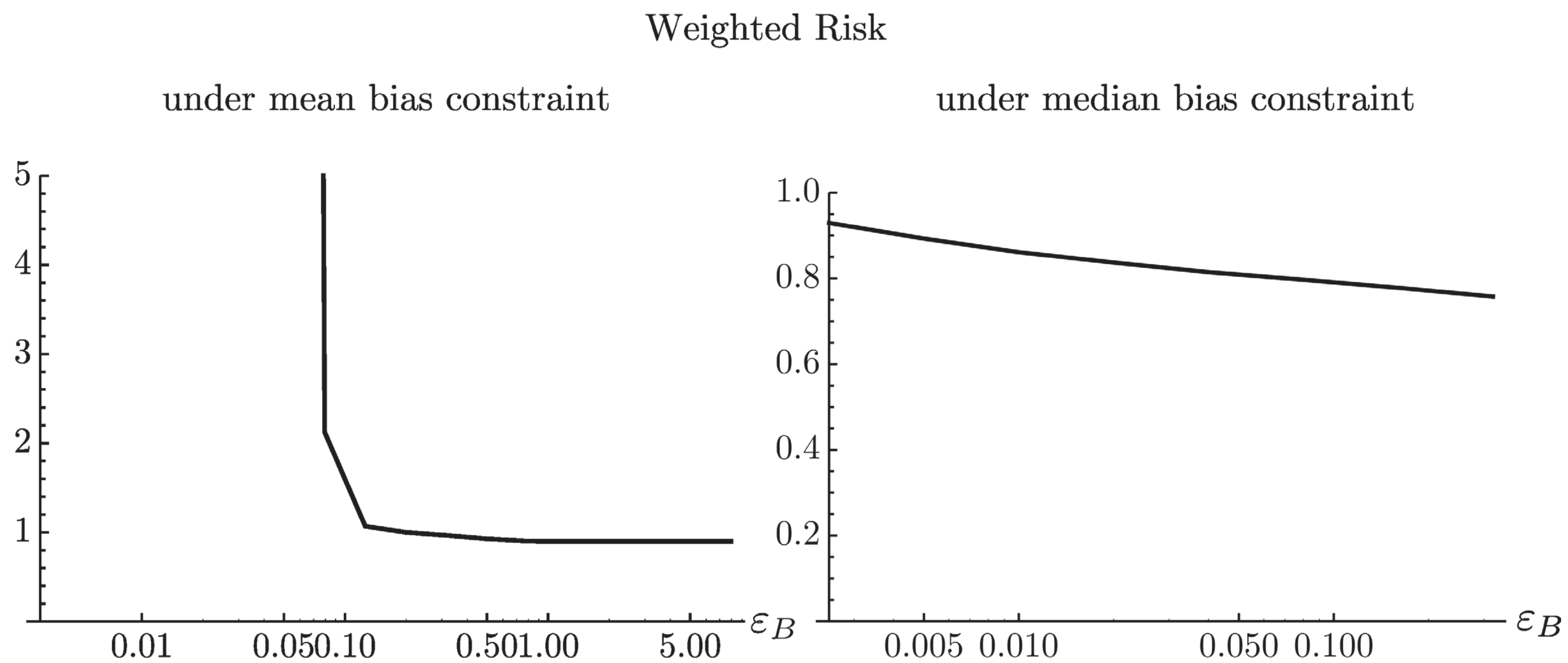
Minimum of Means: Mean Bias Results



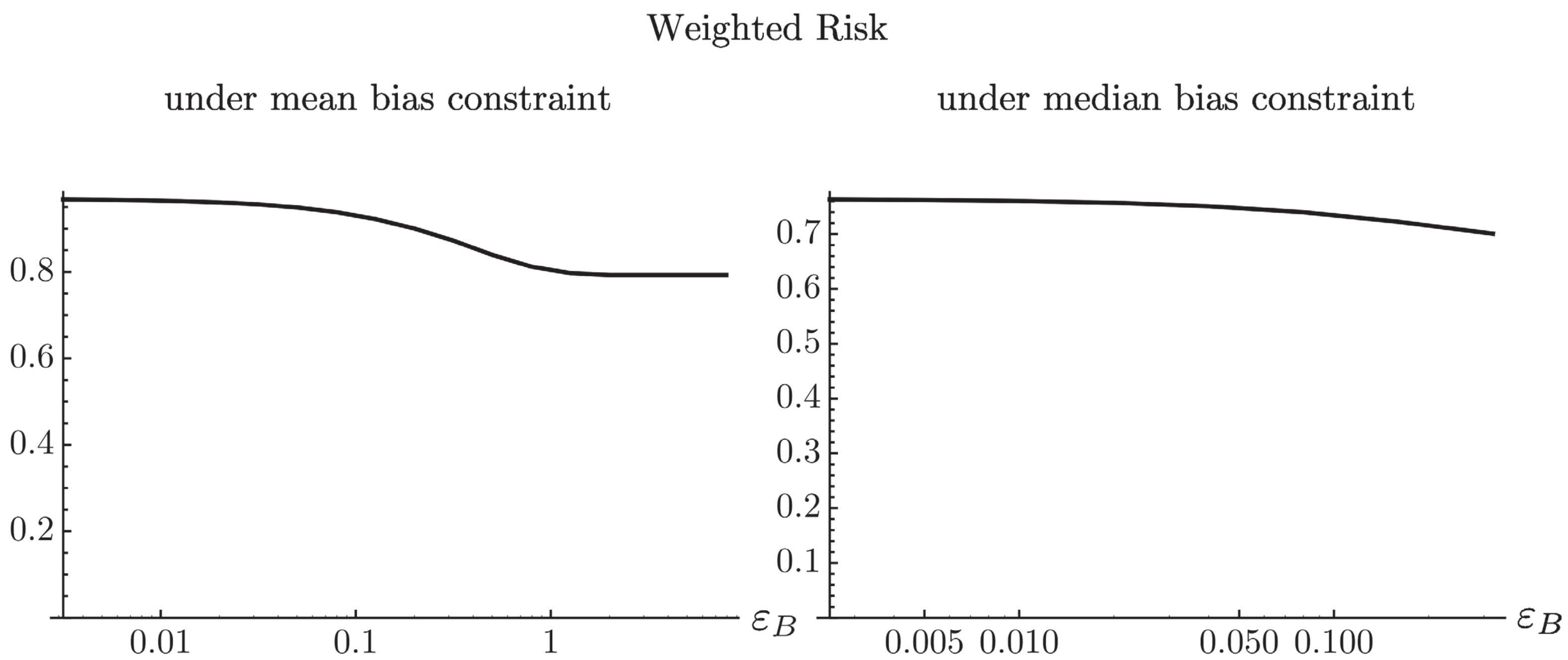
Minimum of Means: Median Bias Results



Minimum of Means



Comparison to Predictive Regression



Conclusions

- Systematic approach to determination of low risk (nearly) unbiased estimators in non-standard problems

- Computations only need to be performed once

- Entirely straightforward under quadratic loss and mean bias constraint

- Additional applications

  ⇒ Median unbiased estimation under weak instruments

  ⇒ Extreme Value at Risk forecast

  ⇒ Etc.?