A Generalized Focused Information Criterion for GMM Model and Moment Selection

Francis J. DiTraglia

University of Pennsylvania

March 15, 2013

Generalized Focused Information Criterion (GFIC)

Purpose

Simultaneous Model and Moment Selection for GMM Estimation

Main Idea

Choose model and moment conditions to yield minimum MSE estimator of user-specified target parameter even if mis-specified.

Related Work

- GMM Model and Moment Selection (Andrews & Lu, 2001)
- Focused Moment Selection Criterion (DiTraglia, 2013)
- Focused Information Criterion (Claeskens & Hjort, 2003)

Motivating Example: Dynamic Panel

(Large N, Small T)

Two Key Issues

1. Moment Selection
   - Which instruments to use?
   - Equivalently, what exogeneity assumption?
2. Model Selection
   - Which regressors to include?
   - How to specify the dynamics? (How many lags, etc.)

Motivating Example: Dynamic Panel

(Large N, Small T)

Bias–Variance Tradeoff

1. Moment Selection – Stronger Exogeneity Assumption
   - $\Rightarrow$ More and more relevant instruments hence lower variance
   - $\Rightarrow$ More ways for it to be violated, greater chance of endogenous instruments and associated bias
2. Model Selection – Include More Lags of y
   - $\Rightarrow$ "Safer" since there's a better chance we've captured the true dynamics, hence less chance of bias
   - $\Rightarrow$ Fewer time periods available for estimation so higher variance

Generalized Focused Information Criterion (GFIC)

Choose the Wrong Specification on Purpose

Choose a specification based on the Asymptotic MSE of associated estimator: tolerate small bias in exchange for reduction in variance.

Different Research Goal $\implies$ Different Criterion

Choose model and moment condition to yield minimum AMSE estimator of user-specified target parameter μ ("Focus")

Local Mis-specification

Local to zero asymptotics to yield a non-trivial bias-variance tradeoff in the limit.

Generalized Focused Information Criterion (GFIC)

- Underlying GMM Parameter Vector: $\beta' = (\gamma', \theta')$
  - Always Estimate "Protected Parameters" θ
  - Consider setting "Nuisance Parameters" γ equal to γ_0
- Two Blocks of Moment Conditions: g, h
  - Block g correctly specified
  - Block h potentially mis-specified $\Rightarrow$ consider excluding
- Scalar Target Parameter $\mu = \varphi(\theta, \gamma)$

Which, if any, of the parameters γ should we estimate and which of the moment conditions should we use to produce a minimum AMSE estimator of μ?

Dynamic Panel Example

True Data Generating Process

$$y_{it} = \gamma y_{it-1} + \theta x_{it} + \eta_i + v_{it}$$

1. True DGP has dynamics
2. Correlated individual effects $\eta_i \implies$ estimate in differences
3. Regressor x_{it} is predetermined but not strictly exogenous
4. Stationarity

Goal: Estimate θ with minimum MSE.

Dynamic Panel Example

Suppose out target parameter is θ

$$\mathbb{E}\left[\begin{pmatrix} y_{i,t-2} \\ x_{i,t-1} \\ x_{it} \end{pmatrix} (\Delta y_{it} - \gamma \Delta y_{i,t-1} - \theta \Delta x_{it})\right] = \begin{bmatrix} 0 \\ 0 \\ -\sigma_{xv} \end{bmatrix}$$

Model Selection

Should we set $\gamma = 0$ (exclude the lag) to gain an extra time period?

Moment Selection

Should we use x_{it} as an instrument for period t?

GFIC Asymptotics: Local Mis-specification

Let $\{Z_{ni}\}_{i=1}^{n}$ be a triangular array of random vectors such that

$$\mathbb{E}\begin{bmatrix} g(Z_{ni}, \gamma_n, \theta_0) \\ h(Z_{ni}, \gamma_n, \theta_0) \end{bmatrix} = \begin{bmatrix} 0 \\ \tau_n \end{bmatrix}$$

where

$$\gamma_n = \gamma_0 + \delta/\sqrt{n}$$
$$\tau_n = \tau/\sqrt{n}$$

and δ, τ are unknown constant vectors (possibly zero).

GFIC Asymptotics: No Mis-specification in the Limit

The limiting Law Z of the triangular array $\{Z_{ni}\}_{i=1}^{n}$ satisfies

$$\mathbb{E}\left[\begin{array}{c} g(Z,\gamma_0,\theta_0) \\ h(Z,\gamma_0,\theta_0) \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \end{array}\right]$$

In other words, all the moment conditions are valid in the limit and the parameter restriction $\gamma = \gamma_0$ holds.

Local Mis-specification for Dynamic Panel Example

$$\mathbb{E}\left[\begin{pmatrix} y_{i,t-2} \\ x_{i,t-1} \\ x_{it} \end{pmatrix}\left(\Delta y_{it} - \left(\delta/\sqrt{n}\right)\Delta y_{i,t-1} - \theta \Delta x_{it}\right)\right] = \begin{bmatrix} 0 \\ 0 \\ \tau/\sqrt{n} \end{bmatrix}$$

In the Limit

1. No dynamics
2. x_{it} is a valid instrument for period t

Notation

Sample Analogue of Moment Conditions

$$f_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} f(Z_{ni}, \gamma, \theta) = \left[\begin{array}{c} g_n(\beta) \\ h_n(\beta) \end{array}\right] = \left[\begin{array}{c} n^{-1}\sum_{i=1}^{n} g(Z_{ni}, \gamma, \theta) \\ n^{-1}\sum_{i=1}^{n} h(Z_{ni}, \gamma, \theta) \end{array}\right]$$

PSD Weighting Matrix

$$\widetilde{W} = \left[\begin{array}{cc} \widetilde{W}_{gg} & \widetilde{W}_{gh} \\ \widetilde{W}_{hg} & \widetilde{W}_{hh} \end{array}\right]$$

Estimators

Each model/moment selection pair $(b, c) \in \mathcal{BC}$ defines a GMM estimator

$$\widehat{\beta}(b, c) = \underset{\beta^{(b)} \in \mathbf{B}^{(b)}}{\arg\min} \left[\Xi_c f_n\left(\beta^{(b)}, \gamma_0^{(-b)}\right)\right]' \left[\Xi_c \widetilde{W} \Xi_c'\right] \left[\Xi_c f_n\left(\beta^{(b)}, \gamma_0^{(-b)}\right)\right]$$

Under local mis-specification, *each* of these yields a consistent estimator of θ. Estimators based on an incorrect specification, however, show a bias in their limiting distributions.

More Notation

$$F = \begin{bmatrix} \nabla_{\gamma'} g\left(Z, \gamma_0, \theta_0\right) & \nabla_{\theta'} g\left(Z, \gamma_0, \theta_0\right) \\ \nabla_{\gamma'} h\left(Z, \gamma_0, \theta_0\right) & \nabla_{\theta'} h\left(Z, \gamma_0, \theta_0\right) \end{bmatrix}$$

$$F = \begin{bmatrix} F_\gamma & F_\theta \end{bmatrix} = \begin{bmatrix} G_\gamma & G_\theta \\ H_\gamma & H_\theta \end{bmatrix} = \begin{bmatrix} G \\ H \end{bmatrix}$$

$$\Omega = Var \begin{bmatrix} g(Z, \gamma_0, \theta_0) \\ h(Z, \gamma_0, \theta_0) \end{bmatrix} = \begin{bmatrix} \Omega_{gg} & \Omega_{gh} \\ \Omega_{hg} & \Omega_{hh} \end{bmatrix}$$

N.B. These expressions involve the limiting random variable Z rather than Z_{ni} so expectations are taken with respect to a distribution for which all MCs have expectation zero at (γ_0, θ_0).

Theorem (Asymptotic Distribution)

$$\sqrt{n}\left(\widehat{\beta}(b,c) - \beta_0^{(b)}\right) \to_d -K(b,c)\Xi_c \left(\left[\begin{array}{c} \mathcal{N}_g \\ \mathcal{N}_h \end{array} \right] + \left[\begin{array}{c} 0 \\ \tau \end{array} \right] - F_\gamma \delta \right)$$

where $\beta_0^{(b)\prime} = \left(\theta_0, \gamma_0^{(b)}\right)$,

$$K(b,c) = \left[F(b,c)' W_c F(b,c)\right]^{-1} F(b,c)' W_c$$

and

$$\left[\begin{array}{c} \mathcal{N}_g \\ \mathcal{N}_h \end{array} \right] \sim N\left(\left[\begin{array}{c} 0 \\ 0 \end{array} \right], \left[\begin{array}{cc} \Omega_{gg} & \Omega_{gh} \\ \Omega_{hg} & \Omega_{hh} \end{array} \right] \right)$$

Corollary

$\sqrt{n}\left(\widehat{\mu}(b,c)-\mu_n\right)$ *converges in distribution to*

$$-\nabla_\beta \varphi_0' \Xi_b' K(b,c) \Xi_c \left(\begin{bmatrix} \mathscr{N}_g \\ \mathscr{N}_h \end{bmatrix} + \begin{bmatrix} 0 \\ \tau \end{bmatrix} - F_\gamma \delta \right) - \nabla_\gamma \varphi_0' \delta$$

where $\varphi_0 = \varphi(\gamma_0, \theta_0)$*,* $\mu_n = \phi(\theta_0, \gamma_n)$*.*

- AMSE$\left(\widehat{\mu}\right)$ comes as immediate consequence of this result
- Usual estimators of K, etc. consistent under local mis-spec.
- The problem is τ, δ

How and When Can We Estimate τ and δ?

No consistent estimators exist under local mis-spec. but we can construct asymptotically unbiased estimators provided:

1. There are enough moment conditions in g to identify the full parameter vector $\rightarrow$ Valid Estimator $\widehat{\beta}_v = \left(\widehat{\gamma}_v, \widehat{\theta}_v\right)'$.
2. It is possible to evaluate h_n, sample analogue of "suspect" MCs, at $\widehat{\beta}_v$. (This is usually trivial.)

Estimating δ

Corollary (Asymptotic Distribution of Valid Estimator)

$$\sqrt{n}\left(\widehat{\beta}_v - \beta_0\right) = \sqrt{n}\left(\begin{array}{c} \widehat{\gamma}_v - \gamma_0 \\ \widehat{\theta}_v - \theta_0 \end{array}\right) \rightarrow_d \left[\begin{array}{c} \delta \\ 0 \end{array}\right] - K_v \mathscr{N}_g$$

where $K_v = \left[G' W_{gg} G\right]^{-1} G' W_{gg}$ *and* $W_{gg} = plim_{N\to\infty} \widetilde{W}_{gg}$.

This immediately provides asymptotically unbiased estimator of δ, namely $\widehat{\delta} = \sqrt{n}(\widehat{\gamma}_v - \gamma_0)$ since γ_0 is known and $\mathscr{N}_g$ is mean-zero.

Estimating τ

Lemma (Asymptotically Unbiased Estimator of τ)

$$\widehat{\tau} = \sqrt{n} h_n \left(\widehat{\beta}_v \right) \rightarrow_d \tau - H K_v \mathscr{N}_g + \mathscr{N}_h$$

where $K_v = \left[G' W_{gg} G \right]^{-1} G' W_{gg}$.

This results gives asymptotically unbiased estimator of τ since $\mathscr{N}_g$ and $\mathscr{N}_h$ mean zero.

But AMSE Requires *Squared Bias*

Rewriting the Expression from Above:

$$\text{BIAS}\left(\widehat{\mu}\left(b,c\right)\right)^2 = \nabla_\beta \varphi_0' M(b,c) \begin{bmatrix} \tau\tau' & \tau\delta' \\ \delta\tau' & \delta\delta' \end{bmatrix} M(b,c)' \nabla_\beta \varphi_0$$

where

$$M(b,c) = \Xi_b' K(b,c) \Xi_c \begin{bmatrix} -G_\gamma & 0 \\ -H_\gamma & I \end{bmatrix} + \begin{bmatrix} I_r & 0_{r\times q} \\ 0_{p\times r} & 0_{s\times q} \end{bmatrix}$$

Problem

Although $(\widehat{\delta}, \widehat{\tau})$ are asymptotically unbiased estimators of (δ, τ), $\widehat{\delta\delta'}$ is not an asymptotically unbiased estimator of $\delta\delta'$ and $(\widehat{\tau\tau'}, \widehat{\tau\delta'})$ are not asymptotically unbiased estimators of $(\tau\tau', \tau\delta')$.

Joint Distribution of $(\widehat{\delta}, \widehat{\tau})$

$$\begin{bmatrix} \widehat{\delta} \\ \widehat{\tau} \end{bmatrix} = \sqrt{n} \begin{bmatrix} (\widehat{\gamma}_v - \gamma_0) \\ h_n(\widehat{\beta}_v) \end{bmatrix} \to_d \begin{bmatrix} \delta \\ \tau \end{bmatrix} + \Psi \begin{bmatrix} \mathcal{N}_g \\ \mathcal{N}_h \end{bmatrix}$$

where

$$\Psi = \begin{bmatrix} -K_v^\gamma & \mathbf{0} \\ -HK_v & I \end{bmatrix}$$

Each of the quantities in the matrix pre-multiplying $(\mathcal{N}_g', \mathcal{N}_h')'$ is consistently estimable under local mis-specification, as is the variance matrix of $(\mathcal{N}_g', \mathcal{N}_h')'$.

Bias Correction

Provided that $\widehat{\Psi}$ and $\widehat{\Omega}$ are consistent estimators of Ψ and Ω,

$$\widehat{B} = \begin{bmatrix} \widehat{\tau}\widehat{\tau}' & \widehat{\tau}\widehat{\delta}' \\ \widehat{\delta}\widehat{\tau}' & \widehat{\delta}\widehat{\delta}' \end{bmatrix} - \widehat{\Psi}\widehat{\Omega}\widehat{\Psi}'$$

is an asymptotically unbiased estimator of the squared bias matrix

$$\begin{bmatrix} \tau\tau' & \tau\delta' \\ \delta\tau' & \delta\delta' \end{bmatrix}.$$

GFIC: Asymptotically Unbiased Estimator of AMSE

$$\text{GFIC}(b,c) = \widehat{\text{AVAR}}(b,c) + \widehat{\text{ABIAS}}^2(b,c)$$

$$\begin{aligned}
\widehat{\text{AVAR}}(b,c) &= \nabla_\beta \widehat{\varphi}_0' \Xi_b' \widehat{K}(b,c) \widehat{\Omega}_c \widehat{K}(b,c)' \Xi_b \nabla_\beta \widehat{\varphi}_0' \\
\widehat{\text{ABIAS}}^2(b,c) &= \nabla_\beta \widehat{\varphi}_0' \widehat{M}(b,c)\ \widehat{B}\ \widehat{M}(b,c) \nabla_\beta \widehat{\varphi}_0
\end{aligned}$$

$$\widehat{B} = \begin{bmatrix} \widehat{\tau}\widehat{\tau}' & \widehat{\tau}\widehat{\delta}' \\ \widehat{\delta}\widehat{\tau}' & \widehat{\delta}\widehat{\delta}' \end{bmatrix} - \widehat{\Psi}\widehat{\Omega}\widehat{\Psi}'$$

We choose the specification (b^*, c^*) that minimizes the value of the GFIC over the candidate set $\mathcal{BC}$.

Simulation Study

Simple Example

True Data Generating Process

$$y_{it} = \gamma y_{it-1} + \theta x_{it} + \eta_i + v_{it}$$

1. True DGP has dynamics
2. Correlated individual effects η_i
3. Regressor x_{it} is predetermined but not strictly exogenous
4. Stationarity

Goal: Estimate θ with minimum MSE.

Consider Four Possible Specifications

1. LW – Assume x_{it} predetermined, include lagged y
2. LS – Assume x_{it} strictly exogenous, include lagged y
3. W – Assume x_{it} predetermined, exclude lagged y
4. S – Assume x_{it} strictly exogenous, exclude lagged y

W stands for "weak" – imposes weaker exogeneity assumption.

Anderson & Hsiao–esque 2SLS Estimators (1982)

Difference to Remove Correlated Individual Effects

LW Moment Conditions:

$$\mathbb{E}\left[\begin{pmatrix} y_{i,t-2} \\ x_{i,t-1} \end{pmatrix} (\Delta y_{it} - \gamma \Delta y_{i,t-1} - \theta \Delta x_{it})\right] = 0, \text{ for } t = 3, \ldots, T$$

LS Adds the Moment Conditions:

$$\mathbb{E}\left[x_{it} (\Delta y_{it} - \gamma \Delta y_{i,t-1} - \theta \Delta x_{it})\right] = 0, \text{ for } t = 3, \ldots, T$$

Only the LW conditions are correct

Anderson & Hsiao–esque 2SLS Estimators (1982)

Difference to Remove Correlated Individual Effects

W Moment Conditions:

$$\mathbb{E}\left[x_{i,t-1}\left(\Delta y_{it} - \theta \Delta x_{it}\right)\right] = 0, \text{ for } t = 2, \ldots, T$$

S Adds the Moment Conditions:

$$\mathbb{E}\left[x_{it}\left(\Delta y_{it} - \theta \Delta x_{it}\right)\right] = 0, \text{ for } t = 2, \ldots, T$$

None of these moment conditions are correct

Simulation Setup

c.f. Andrews & Lu (2001)

- $y_{i0} = 0$, mean of stationary distribution
- For $t = 1, \ldots, T$

$$y_{it} = \gamma y_{it-1} + \theta x_{it} + \eta_i + v_{it}$$

$$\begin{bmatrix} x_i \\ \eta_i \\ v_i \end{bmatrix} \sim \text{iid} \left(\begin{bmatrix} 0_T \\ 0 \\ 0_T \end{bmatrix}, \begin{bmatrix} I_T & \sigma_{x\eta}\iota_T & \sigma_{xv}\Gamma \\ \sigma_{x\eta}\iota'_T & 1 & 0'_T \\ \sigma_{xv}\Gamma' & 0_T & I_T \end{bmatrix} \right)$$

- Γ such that $\mathbb{E}[x_{it}v_{it-1}] = \sigma_{xv}$ but $\mathbb{E}[x_{it}v_{is}] = 0, s \neq t-1$
- $\theta = 0.5$, $\sigma_{x\eta} = 0.2$
- Vary γ and σ_{xv} over a grid

Intuition

If violation of assumptions is small enough, we can obtain a lower MSE by using an incorrect specification: additional time period/one less parameter to estimate/more instruments could lead to a decrease in variance that outweighs the increase in bias.

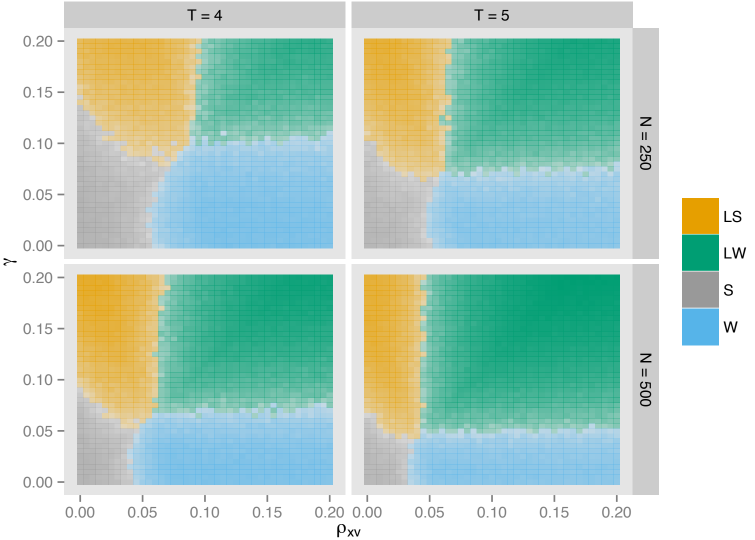

Figure : Minimum RMSE Specification at each combination of parameter values. Shading gives RMSE relative to second best specification.

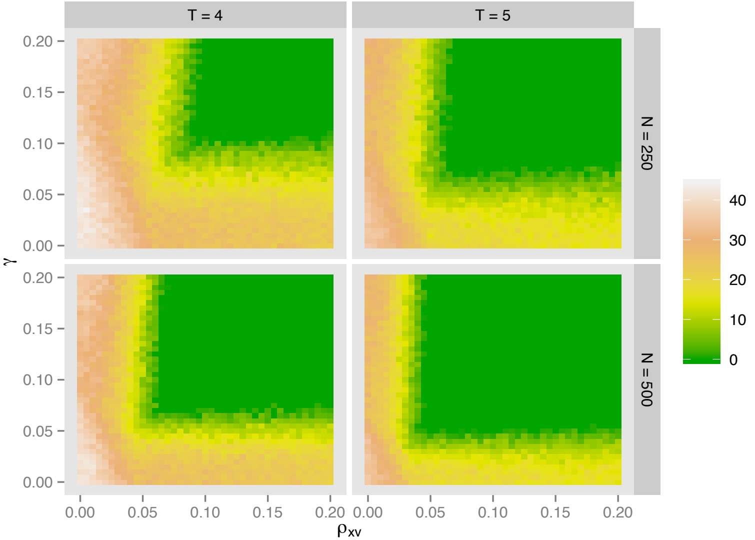

Figure : % RMSE Advantage of Best Specification (vs. LW)

Competing Procedure: Downward J-test

1. Use S unless J-test rejects.
2. If S rejected, use W unless J-test rejects.
3. If W rejected, use LS unless J-test rejects.
4. Only use LW if all others rejected.

Competing Procedure: Andrews & Lu (2001)

J-test Statistic Minus Penalty Term

BIC-Type	$J - (\lvert c\rvert - \lvert b\rvert)\log n$
AIC-Type	$J - 2(\lvert c\rvert - \lvert b\rvert)$
HQ-Type	$J - 2.01(\lvert c\rvert - \lvert b\rvert)\log\log n$

where $|b|$ is the number of parameters estimated, and $|c|$ the number of moment conditions used. We select the specification with the lowest value of the criterion.

	$N = 250$		$N = 500$	
	$T = 4$	$T = 5$	$T = 4$	$T = 5$
LW	19	10	13	7
LS	30	44	54	79
W	24	34	46	64
S	31	50	64	94
GFIC	17	13	15	10
J-test 10%	32	45	55	74
J-test 5%	31	47	57	79
GMM-BIC	32	48	62	87
GMM-HQ	32	46	57	77
GMM-AIC	31	39	47	57

Table : Average RMSE minus Pointwise Optimal (% points)

	$N = 250$		$N = 500$	
	$T = 4$	$T = 5$	$T = 4$	$T = 5$
LW	0	0	0	0
LS	42	81	94	154
W	49	88	105	158
S	48	92	107	171
GFIC	3	8	6	11
J-test 10%	43	78	91	140
J-test 5%	45	83	98	153
GMM-BIC	48	89	106	168
GMM-HQ	46	85	102	154
GMM-AIC	39	68	81	118

Table : Worst-case RMSE minus Minimax Optimal (% points)

Post Selection Inference / Model Averaging

Consider an estimator of the form

$$\widehat{\mu} = \sum_{(b,c)\in\mathcal{BC}} \widehat{\omega}(b,c)\widehat{\mu}(b,c)$$

where $\widehat{\omega}(b,c)$ is a set of data-dependent weights

Requirements for the Weights

Let $\widehat{\omega}(b, c)$ be a function of the data $Z_{n1}, \ldots, Z_{nn}$ and (b, c) satisfying

(a) $\sum_{(b,c)\in\mathcal{BC}} \widehat{\omega}(b, c) = 1$

(b) $\widehat{\omega}(b, c) \to_d \psi(\mathscr{N}, \delta, \tau | b, c)$ jointly for all $(b, c) \in \mathcal{BC}$ where ψ is a function of the normal random vector $\mathscr{N}$, the bias parameters δ and τ, and consistently estimable quantities only.

Covers GFIC, J-test, Andrews & Lu (2001), etc.

Limit Distribution of Averaging Estimator

Since the weights sum to one:

$$\sqrt{n}(\widehat{\mu} - \mu_n) = \sum_{(b,c)\in\mathcal{BC}} \widehat{\omega}(b,c)\sqrt{n}\left(\widehat{\mu}\left(b,c\right) - \mu_n\right)$$

and $\widehat{\omega}(b,c), \widehat{\mu}(b,c)$ converge *jointly* for all $(b,c) \in \mathcal{BC}$

$$\sqrt{n}\left(\widehat{\mu} - \mu_n\right) \to_d \Lambda(\tau,\delta)$$

where

$$\boxed{\Lambda(\tau,\delta) = -\nabla_\beta \varphi_0' \sum_{(b,c)\in\mathcal{BC}} \psi(\mathcal{N},\delta,\tau|b,c)\left\{\Xi_b' K(b,c)\Xi_c\mathcal{N} + M(b,c)\left[\begin{array}{c}\delta\\ \tau\end{array}\right]\right\}}$$

Non-normal limit distribution that depends on (δ,τ)

Suppose (δ, τ) Known

(i) For each $j = 1, 2, \ldots, J$, generate $\mathcal{N}_j \sim N(0, \widehat{\Omega})$

(ii) For each for $j = 1, 2, \ldots, J$ set

$$\Lambda_j(\tau, \delta) = -\nabla_\beta \widehat{\varphi}_0' \sum_{(b,c) \in \mathcal{BC}} \widehat{\psi}(\mathcal{N}_j, \delta, \tau | b, c) \left\{ \Xi_b' \widehat{K}(b, c) \Xi_c \mathcal{N}_j + \widehat{M}(b, c) \left[\begin{array}{c} \delta \\ \tau \end{array} \right] \right\}$$

(iii) Using $\{\Lambda_j(\delta, \tau)\}_{j=1}^J$, calculate $\widehat{a}(\delta, \tau)$, $\widehat{b}(\delta, \tau)$ such that

$$\mathbb{P}\left\{ \widehat{a}(\delta, \tau) \leq \Lambda(\delta, \tau) \leq \widehat{b}(\delta, \tau) \right\} = 1 - \alpha$$

Accounting for Estimated (δ, τ)

Let $R(\alpha_1)$ be a $(1-\alpha_1) \times 100\%$ confidence region for (δ, τ).

1. For each $(\delta, \tau) \in R(\alpha_1)$ construct a confidence interval

$$\mathbb{P}\left\{\widehat{a}(\delta,\tau) \leq \Lambda(\delta,\tau) \leq \widehat{b}(\delta,\tau)\right\} = 1-\alpha_2$$

   using the simulation procedure from the previous slide.

2. Define

$$\begin{aligned} \widehat{a}_{min}(\widehat{\delta},\widehat{\tau}) &= \min_{(\delta,\tau)\in R(\alpha_1)} \widehat{a}(\delta,\tau) \\ \widehat{b}_{max}(\widehat{\delta},\widehat{\tau}) &= \max_{(\delta,\tau)\in R(\alpha_1)} \widehat{b}(\delta,\tau) \end{aligned}$$

3. The following CI has asymptotic coverage of *at least* $1-(\alpha_1+\alpha_2)$

$$\text{CI}_{sim} = \left[\widehat{\mu} - \frac{\widehat{b}_{max}(\widehat{\delta},\widehat{\tau})}{\sqrt{n}}, \quad \widehat{\mu} - \frac{\widehat{a}_{min}(\widehat{\delta},\widehat{\tau})}{\sqrt{n}}\right]$$

Extensions/Future Work

- More on inference/averaging
- Risk functions besides MSE
- Covariate Choice in Treatment Assignment Problems (with Debopam Battacharya)

Supplementary Material

Average RMSE	$N = 250$		$N = 500$	
	$T = 4$	$T = 5$	$T = 4$	$T = 5$
LW	0.073	0.057	0.051	0.040
LS	0.079	0.074	0.070	0.066
W	0.075	0.069	0.066	0.061
S	0.080	0.077	0.074	0.072
GFIC	0.071	0.058	0.052	0.041
Downward J-test (10%)	0.080	0.074	0.070	0.065
Downward J-test (5%)	0.080	0.075	0.071	0.067
GMM-BIC	0.080	0.076	0.073	0.069
GMM-HQ	0.080	0.075	0.071	0.066
GMM-AIC	0.080	0.071	0.066	0.058

Worst-Case RMSE	$N = 250$		$N = 500$	
	$T = 4$	$T = 5$	$T = 4$	$T = 5$
LW	0.084	0.064	0.059	0.045
LS	0.120	0.116	0.115	0.113
W	0.125	0.120	0.122	0.115
S	0.125	0.123	0.122	0.121
GFIC	0.087	0.069	0.063	0.049
Downward J-test (10%)	0.120	0.114	0.113	0.107
Downward J-test (5%)	0.122	0.117	0.117	0.113
GMM-BIC	0.125	0.121	0.122	0.119
GMM-HQ	0.123	0.118	0.120	0.113
GMM-AIC	0.117	0.107	0.107	0.097