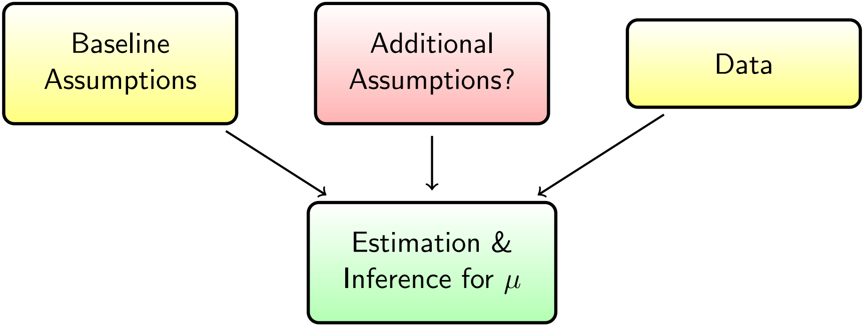
A Generalized Focused Information Criterion for GMM Model and Moment Selection

Francis J. DiTraglia

University of Pennsylvania

December 6, 2013

Generalized Focused Information Criterion (GFIC)



1. False Assumptions on Purpose: Min MSE
2. Focused Choice of Assumptions: User-specified μ
3. Asymptotics: Local Mis-specification
4. Averaging, Inference post-selection

Example – Dynamic Panel

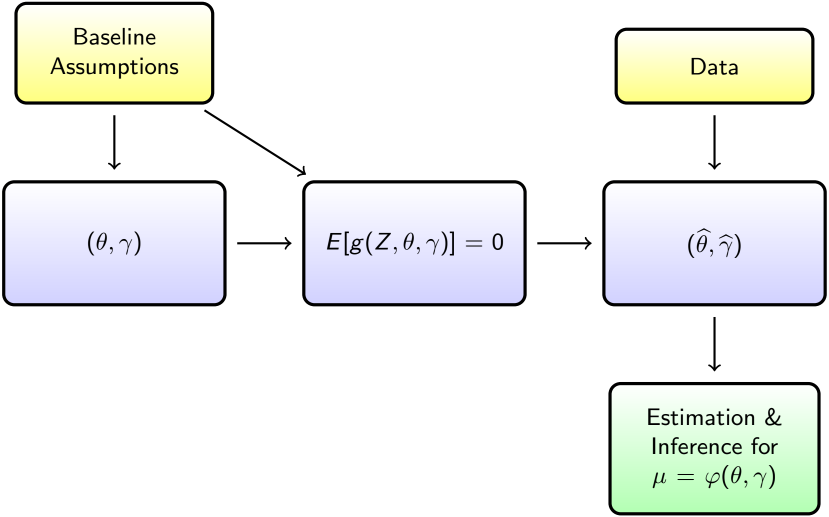
Data Generating Process

$$y_{it} = \alpha_1 y_{it-1} + \ldots + \alpha_p y_{it-p} + \beta x_{it} + \eta_i + v_{it}$$

Potential Target Parameters μ

- $\mu = \beta/(1 - \alpha_1 - \ldots - \alpha_p)$
- $\mu = \beta$
- $\mu = \alpha_1$

GMM Framework



Example – Dynamic Panel

$$\Delta v_{it} = \Delta y_{it} - (\alpha_1 \Delta y_{it-1} + \ldots + \alpha_p \Delta y_{it-p} + \beta \Delta x_{it})$$

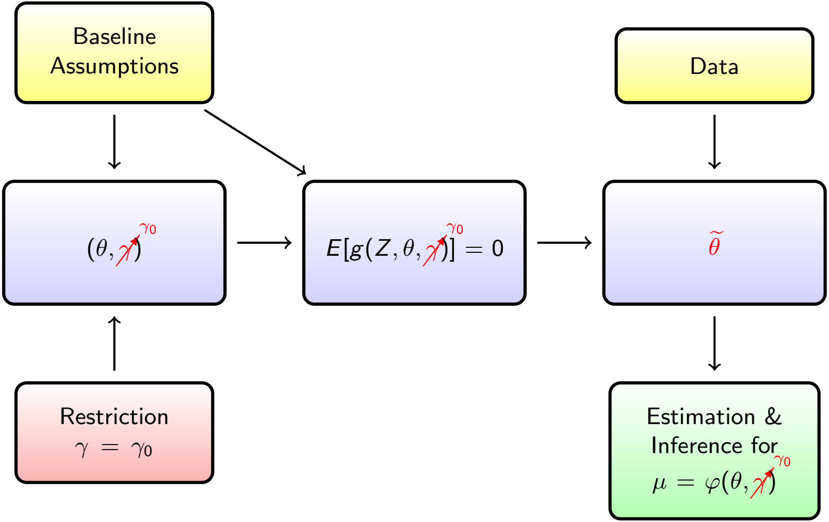
Baseline Assumptions

- $E[\Delta v_{it} y_{it-k}] = 0$ for $k \geq 2$
- $E[\Delta v_{it} x_{it-k}] = 0$ for $k \geq 1$

Parameters

- Full Parameter Vector $(\alpha_1, \alpha_2, \ldots, \alpha_p, \beta)$
- Partition into (θ, γ) depends on μ

Additional Assumption – Parameter Restriction



Example – Dynamic Panel

$$\Delta v_{it} = \Delta y_{it} - (\alpha_1 \Delta y_{it-1} + \ldots + \alpha_p \Delta y_{it-p} + \beta \Delta x_{it})$$

Dynamics $\mu = \alpha_1$

$\gamma = \gamma_0 \iff \beta = \alpha_2 = \ldots = \alpha_p = 0$

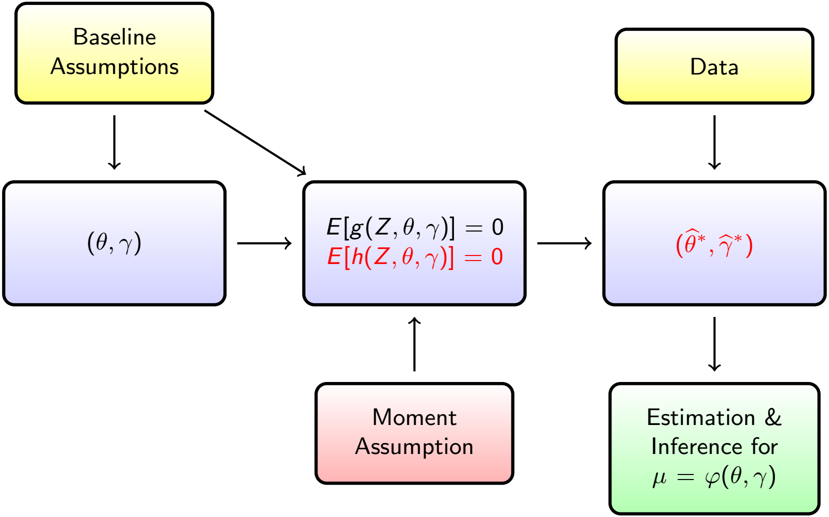
SR Effect $\mu = \beta$

$\gamma = \gamma_0 \iff \alpha_1 = \ldots = \alpha_p = 0$

LR Effect $\mu = \beta/(1 - \alpha_1 - \ldots - \alpha_p)$

$\gamma = \gamma_0 \iff \alpha_2 = \ldots = \alpha_p = 0$

Additional Assumption – Moment Conditions



Example – Dynamic Panel

$$\Delta v_{it} = \Delta y_{it} - (\alpha_1 \Delta y_{it-1} + \ldots + \alpha_p \Delta y_{it-p} + \beta \Delta x_{it})$$

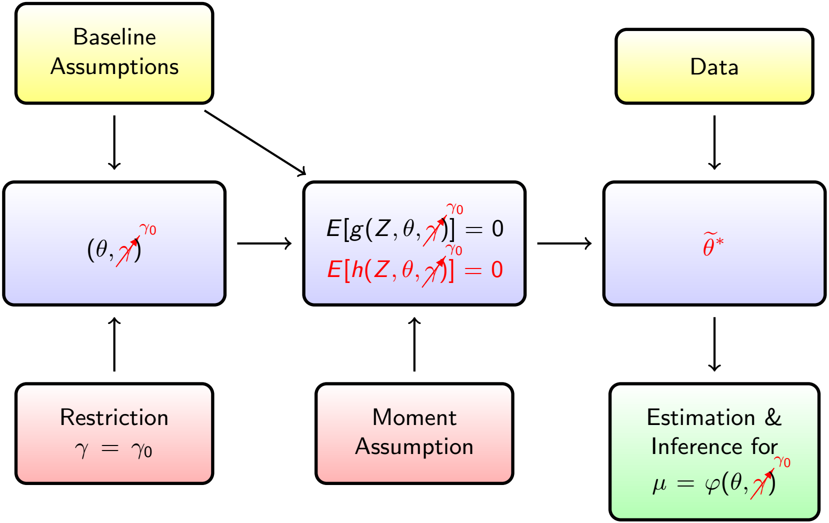
Baseline Assumptions

- $E[\Delta v_{it} y_{it-k}] = 0$ for $k \geq 2$
- $E[\Delta v_{it} x_{it-k}] = 0$ for $k \geq 1$

Additional Assumptions

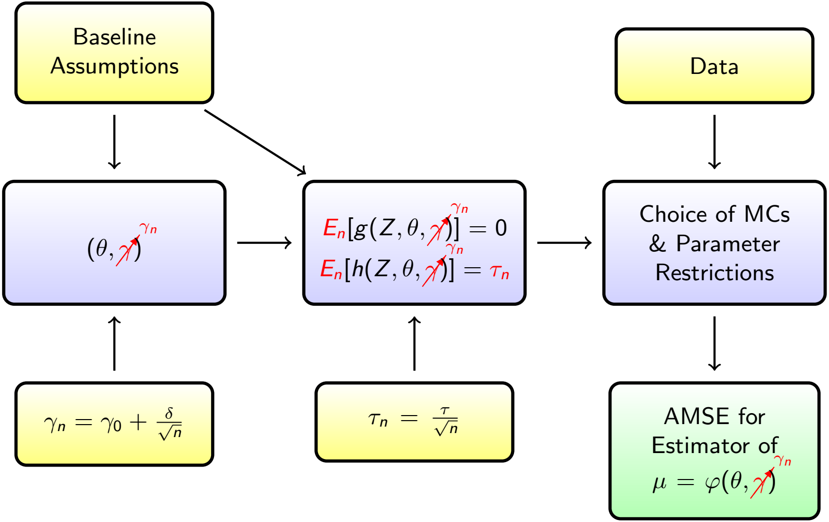
- Strict Exogeneity
- Initial Conditions
- Homoskedasticity

Both at Once



Baseline Assumptions
Data
$(\theta, \cancel{\gamma}^{\gamma_0})$
$E[g(Z, \theta, \cancel{\gamma}^{\gamma_0})] = 0$
$E[h(Z, \theta, \cancel{\gamma}^{\gamma_0})] = 0$
$\widetilde{\theta}^*$
Restriction
$\gamma = \gamma_0$
Moment Assumption
Estimation & Inference for
$\mu = \varphi(\theta, \cancel{\gamma}^{\gamma_0})$

GFIC Asymptotics – Local Mis-Specification



Overview of GFIC Derivation

Asymptotic Normality of GMM Estimator

- Biased unless γ estimated, no MCs from h used
- Smaller variance if γ set to γ_0, MCs from h used

Asymptotic Normality of Target Parameter $\widehat{\mu}(b, c)$

- Inherits bias-variance tradeoff from $\widehat{\beta}(b, c)$
- AMSE $(\widehat{\mu}(b, c))$ depends on $B = \left[\begin{array}{cc} \tau\tau' & \tau\delta' \\ \delta\tau' & \delta\delta' \end{array}\right]$

GFIC = Asymptotically Unbiased Estimator of $\widehat{\text{AMSE}}\,(\widehat{\mu}(b, c))$

- Asymptotically unbiased estimator $\widehat{B}$ of B
- Select (b, c) to minimize GFIC

Some Notation

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

$$F = \left[\begin{array}{cc} F_\gamma & F_\theta \end{array} \right] = \left[\begin{array}{cc} G_\gamma & G_\theta \\ H_\gamma & H_\theta \end{array} \right] = \left[\begin{array}{c} G \\ H \end{array} \right]$$

$$\Omega = Var \left[\begin{array}{c} g(Z, \gamma_0, \theta_0) \\ h(Z, \gamma_0, \theta_0) \end{array} \right] = \left[\begin{array}{cc} \Omega_{gg} & \Omega_{gh} \\ \Omega_{hg} & \Omega_{hh} \end{array} \right]$$

These expressions are evaluated *in the limit* where all MCs have expectation zero at (γ_0, θ_0).

Limit Distribution of GMM Estimators

$\sqrt{n}\left(\widehat{\beta}(b,c) - \beta_0^{(b)}\right)$ converges in distribution to

$$\boxed{-K(b,c)\Xi_c\left(\mathscr{N} + \begin{bmatrix} 0 \\ \tau \end{bmatrix} - F_\gamma \delta\right)}$$

$$\begin{aligned} K(b,c) &= \left[F(b,c)'W_cF(b,c)\right]^{-1} F(b,c)'W_c \\ \Xi_c &= \text{Moment Selection Matrix} \\ \mathscr{N} &\sim \mathrm{N}(0,\Omega) \end{aligned}$$

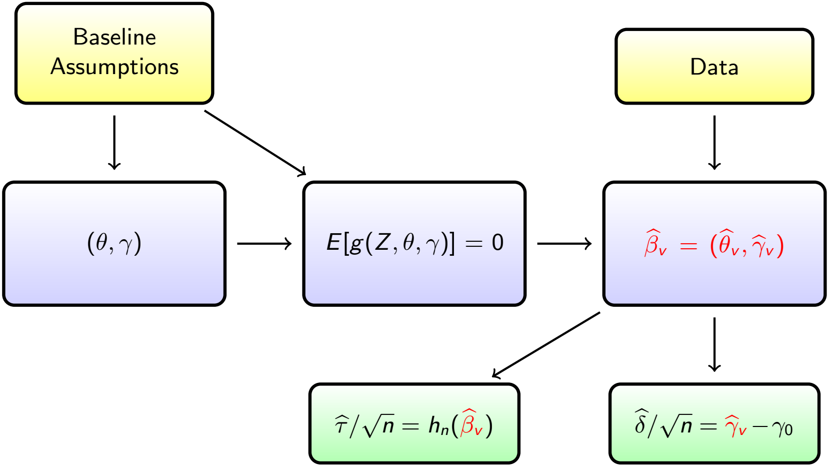
Limit Distribution of Target Parameter Estimators

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

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

$$\begin{array}{rcl} \mu & = & \varphi(\theta, \gamma) \\ \varphi_0 & = & \varphi(\gamma_0, \theta_0) \\ \mu_n & = & \varphi(\theta_0, \gamma_0 + \delta/\sqrt{n}) \\ \Xi_b & = & \text{Model Selection Matrix} \\ \Xi_c & = & \text{Moment Selection Matrix} \\ \mathscr{N} & \sim & N(0, \Omega) \end{array}$$

Estimating δ, τ



Estimating δ, τ – Details

Limit Distribution of Bias Parameter Estimators

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

- Both Ψ and $\Omega = Var(\mathscr{N})$ can be estimated consistently!

Asymptotically Unbiased Estimator of B

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

Valid Post-Selection Inference

Post Selection Estimator

Randomly Weighted Average of candidate estimators (0-1 weights).

Standard CIs are Invalid

Nonstandard limit distribution since weights are *data dependent*

What about consistent selection?

No *pointwise* effect on the limiting distribution, but the same is *not* true uniformly (Pötscher, 1991).

Post-Selection Inference via Model Average Estimators

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.

Key Point: Joint Convergence

$$\boxed{\sqrt{n}\left(\widehat{\mu}(b,c) - \mu_n\right) \text{ converge } \textit{jointly } \forall (b,c) \in \mathcal{BC} \textit{ along with } (\widehat{\delta}, \widehat{\tau})}$$

- Only source of randomness in the limit is $\mathscr{N}$
- Everything except δ and τ is consistently estimable.
- Just need to impose some conditions on the weights...

Limit Distribution of Averaging Estimator

Weights Sum to 1

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

Joint Convergence in Distribution

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

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

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

Generalized Focused Information Criterion

Purpose

Simultaneous Model and Moment Selection for GMM Estimation

Key Features

- Local mis-specification framework
- Estimator of AMSE of user-specified target parameter
- Focused Selection
- Select "wrong" specification on purpose
- Valid Post-Selection Inference

Didn't Discuss Today

- Model Averaging
- Simulation-based Procedure for Confidence Intervals
- Simulation Results

Supplementary Material

Simple Dynamic Panel Example – Large N, Small T

True Data Generating Process

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

- Dynamics unless $\gamma = 0$
- Correlated effects $\eta_i \Rightarrow$ first differences
- x_{it} predetermined but *not* strictly exogenous

Goal – Estimate θ with minimum MSE

- Model Selection Decision: set $\gamma = 0$?
- Moment Selection Decision: treat x_{it} as strictly exogenous?

Anderson & Hsiao–esque 2SLS Estimators (1982)

LW Moment Conditions:

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

LS Adds the Moment Conditions:

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

Only the LW conditions are correct

Anderson & Hsiao–esque 2SLS Estimators (1982)

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

Why Use an Incorrect Specification?

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

Wrong Model

- γ small $\implies$ ignore dynamics
- Adds small bias
- Much lower variance: extra time period, fewer parameters

Invalid MCs

- $E[x_{it} v_{it-1}]$ small $\implies$ add x_{it} as instrument for period t
- Adds small bias
- Much lower variance: x_{it} is a strong instrument for Δx_{it}

Simulation Setup

Similar to Andrews & Lu (2001)

- $y_{i0} = 0$
- $y_{it} = \gamma y_{it-1} + 0.5 x_{it} + \eta_i + v_{it} \qquad (t = 1, \ldots, T)$

$$\begin{bmatrix} x_i \\ \eta_i \\ v_i \end{bmatrix} \sim \text{iid } N \left(\begin{bmatrix} 0_T \\ 0 \\ 0_T \end{bmatrix}, \begin{bmatrix} I_T & 0.2\iota_T & \sigma_{xv}\Gamma \\ 0.2\iota_T' & 1 & 0_T' \\ \sigma_{xv}\Gamma' & 0_T & I_T \end{bmatrix} \right)$$

- $E[x_{it} v_{it-1}] = \sigma_{xv}$ but $E[x_{it} v_{is}] = 0, s \neq t-1$

Vary γ and σ_{xv} over a grid

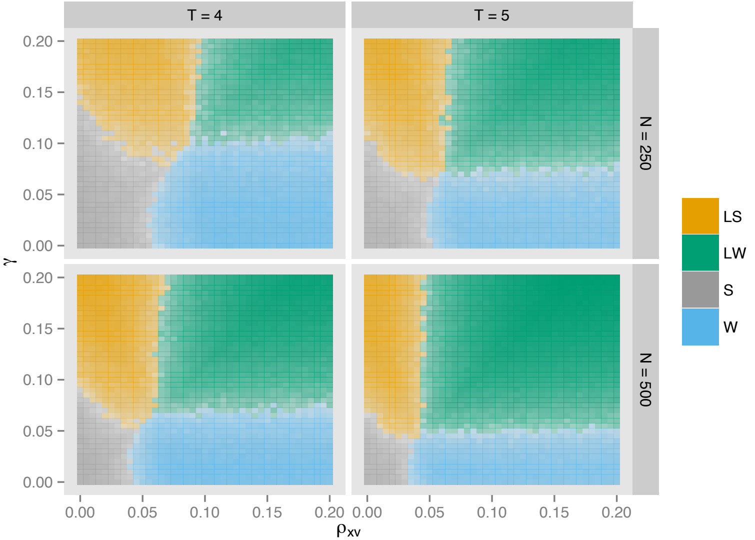


Figure: Minimum RMSE Specification at each combination of parameter values. Shading gives RMSE relative to second best specification.

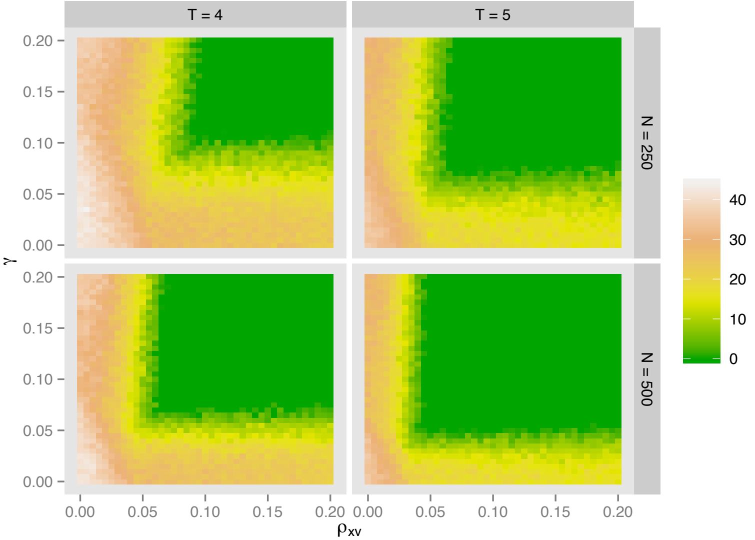


Figure: % RMSE Advantage of Best Specification (vs. LW)

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 Procedures: 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	$J - 2.01(\lvert c\rvert - \lvert b\rvert)\log\log n$

- $|b| = \#$ (parameters estimated)
- $|c| = \#$(MCs used)
- Select specification with *lowest* value of 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)

Requirements for the Weights

Weights Sum to 1

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

Joint Convergence

$\widehat{\omega}(b,c) \to_d \psi(\mathscr{N}, \delta, \tau | b, c)$ jointly for all $(b,c) \in \mathcal{BC}$

Limit Function ψ

Depends *only* on $\mathscr{N}, \delta, \tau$, and consistently estimable quantities.

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

“Bootstrapping the Limit Experiment”

Suppose δ and τ were 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

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

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

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

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* $1 - (\alpha_1 + \alpha_2)$

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