A Generalized Focused Information Criterion for GMM Model and Moment Selection

Francis J. DiTraglia

University of Pennsylvania

September 20, 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.

Some Related Work

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

Key Features of GFIC

Select "Wrong" Specification on Purpose

- Minimize MSE rather than search for "true" specification.
- Tolerate some bias to reduce variance.

Focused Selection

- MSE of user-specified target parameter μ
- Different Research Goal $\Rightarrow$ Different Criterion

Local Mis-specification

- Asymptotic MSE to approximate finite sample MSE
- Local asymptotics $\Rightarrow$ bias-variance tradeoff in the limit

GFIC Model & Moment Selection Framework

Parameters

- Always estimate "protected" parameters θ
- Consider setting "nuisance" parameters γ equal to constant γ_0

Moment Conditions

- *g* correctly specified provided we estimate γ
- *h* possibly mis-specified even if we estimate γ

Scalar Target Parameter

- Want minimum MSE estimator of $\mu = \phi(\theta, \gamma)$

GFIC Asymptotics: Local Mis-specification

Triangular Array DGP (Only a Device!)

$$E\begin{bmatrix} g(Z_{ni}, \gamma_0 + \delta/\sqrt{n}, \theta_0) \\ h(Z_{ni}, \gamma_0 + \delta/\sqrt{n}, \theta_0) \end{bmatrix} = \begin{bmatrix} 0 \\ \tau/\sqrt{n} \end{bmatrix}$$

δ Controls Model Mis-specification

- Restriction $\gamma = \gamma_0$ *false* for finite n unless $\delta = 0$
- Model mis-specification disappears in the limit

τ Controls Moment Mis-specification

- MCs in h are invalid for finite n unless $\tau = 0$
- Moment mis-specification disappears in the limit

Notation for Model and Moment Selection

Model Selection – Which Elements of γ to estimate?

- Parameters $\beta = (\theta, \gamma)$
- Model Selection Vector b

Moment Selection – Which MCs to use in Estimation?

- Full set of moment conditions $f = (g, h)$
- Moment Selection Vector c

Putting Them Together

- Candidate Specification (b, c)
- Set of all candidates $\mathcal{BC}$

Overview of GFIC Derivation

Step 1 – Asymptotic Normality of GMM Estimator $\widehat{\beta}(b, c)$

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

Step 2 – Asymptotic Normality of Target Parameter $\widehat{\mu}(b, c)$

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

Step 3 – GFIC $= \widehat{\text{AMSE}}\left(\widehat{\mu}(b, c)\right)$

- Substitute asymptotically unbiased estimator $\widehat{B}$ of B, consistent estimators of everything else.

Estimating δ, τ – Overview

Why is this difficult?

- Local mis-specification $\Rightarrow$ no consistent estimators of δ, τ
- Can construct asymptotically unbiased estimators
- Actually need to estimate $B = \begin{bmatrix} \tau\tau' & \tau\delta' \\ \delta\tau' & \delta\delta' \end{bmatrix}$

How and when can we proceed?

- $\widehat{\beta}_v = (\widehat{\theta}_v, \widehat{\gamma}_v)$ estimates *all* parameters using g *only*
- Plug $\widehat{\beta}_v$ into sample analogue of h to estimate $\tau/\sqrt{n}$
- Use $(\widehat{\gamma}_v - \gamma_0)$ to estimate $\delta/\sqrt{n}$
- Bias correction to get asymptotically unbiased estimator of B

Estimating δ, τ – Details

Limit Distribution of Bias Parameter Estimators

$$\begin{bmatrix} \widehat{\delta} \\ \widehat{\tau} \end{bmatrix} = \sqrt{n} \begin{bmatrix} (\widehat{\gamma}_v - \gamma_0) \\ h_n(\widehat{\beta}_v) \end{bmatrix} \to_d \begin{bmatrix} \delta \\ \tau \end{bmatrix} + \Psi\ N(0, \Omega)$$

- Both Ψ and Ω can be estimated consistently!

Asymptotically Unbiased Estimator of B

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

Using the GFIC for Selection

- Calculate $\widehat{\text{AMSE}}(\widehat{\mu}(b,c))$ for each $(b,c) \in \mathcal{BC}$
- Choose the specification with the lowest AMSE estimate.
- Expression for $\widehat{\text{AMSE}}$ is complicated but easy to compute.

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:

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

LS Adds the Moment Conditions:

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

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

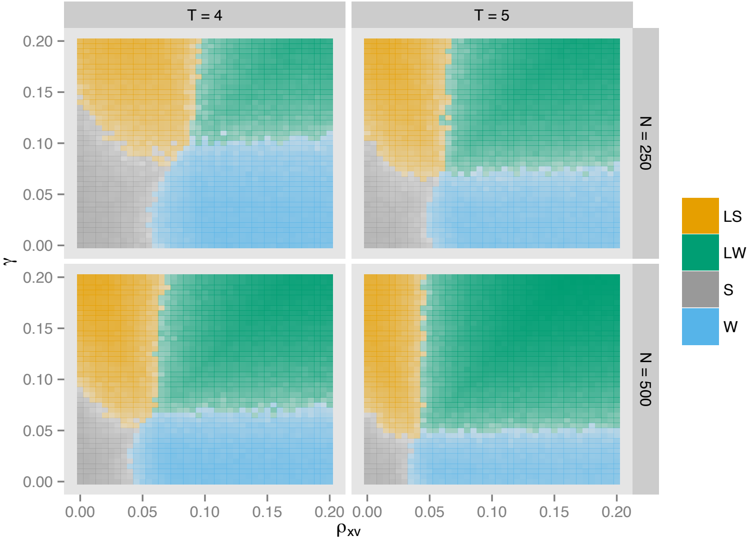

Figure: Minimum RMSE Specification at each combination of parameter values. Shading gives RMSE relative to second best specification.

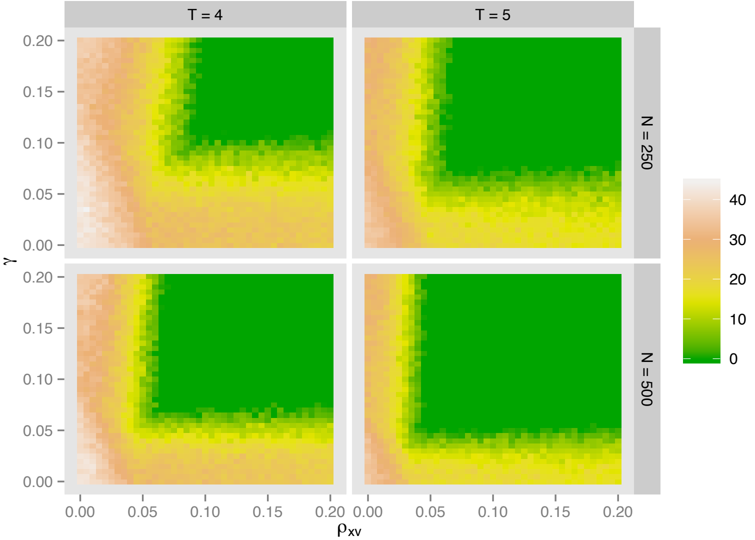

Figure: % RMSE Advantage of Best Specification (vs. LW)

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)

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.

Some Notation

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

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

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

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

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

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

Limit Distribution of Target Parameter Estimators

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

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

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

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

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

$$\begin{aligned} \Psi &= \begin{bmatrix} -K_v^{\gamma} & \mathbf{0} \\ -HK_v & I \end{bmatrix} \\ \mathscr{N} &\sim N(0, \Omega) \end{aligned}$$

Key Point: Joint Convergence

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

- 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...

Requirements for the Weights

Weights Sum to 1

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

Joint Convergence

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

Limit Function ψ

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

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

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)

"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

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

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

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

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 and Future Work

This Paper

Simulations for post-selection inference and averaging in progress.

Other Projects Underway

Risk-based model selection and averaging using local-asymptotics:

- Combining OLS and IV Estimators
- "Change in Exogeneity" (with Otilia Boldea)
- "Covariate Choice in Treatment Assignment Problems" (with Debopam Battacharya)

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
- Works well in simulations
- Provides framework for model and moment averaging
- Valid post-selection confidence intervals

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