Martingale
Research
Corporation

Identifying Probability Modeling Flaws using Generalized Information Matrix Tests

Richard M. Golden[1], Steven S. Henley[2,3], Halbert White, and T. Michael Kashner[3,4]

University of Texas at Dallas, Behavioral and Brain Sciences[1]
Martingale Research Corporation[2]
Loma Linda University, School of Medicine[3]
Loma Linda VA Medical Center, Center for Advanced Statistics in Education[4]

Presented at the 50th Annual Meeting of the Society for Mathematical Psychology, European Mathematical Psychology Group, 15th Annual Meeting of the International Conference on Cognitive Modelling

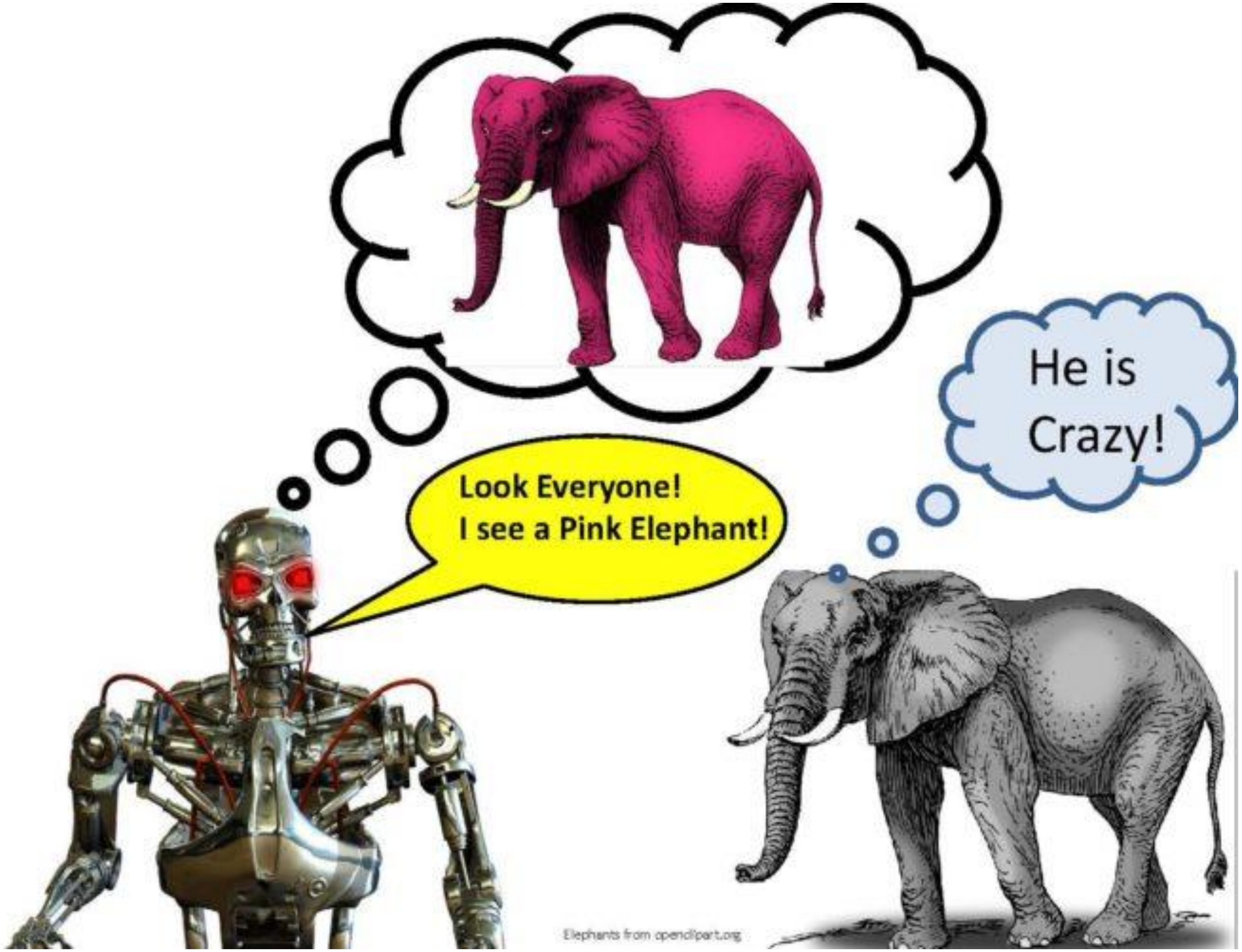
University of Warwick, UK (July 22-July 25, 2017)

This research was supported by grants from the National Institute of General Medical Sciences (NIGMS) (R43GM114899, PI: S.S. Henley; R43GM106465, PI: S.S. Henley), the National Institute of Mental Health (NIMH) (R43MH105073, PI: S.S. Henley), the National Cancer Institute (NCI) (R44CA139607, PI: S.S. Henley), and the National Institute on Alcohol Abuse and Alcoholism (NIAAA) (R43/R44AA013768, PI: S.S. Henley; R43/R44AA013351, PI: S.S. Henley) **under the Small Business Innovation Research (SBIR) Program.**

Recently published in “Econometrics, November 2016”

Alternative Talk Title:

How to Determine when your Model is Hallucinating!



Definitions

- **Correctly Specified Model:**
  - A model is a set of probability distributions.

Definitions

- **Correctly Specified Model:**
  – A model is a set of probability distributions.
  – A correctly specified model contains the data-generating distribution.

Definitions

- **Correctly Specified Model:**
  - A model is a set of probability distributions.
  - A correctly specified model contains the data-generating distribution.

- **Misspecified Model:**
  Model that is not correctly specified.

Definitions

- **Correctly Specified Model:**
  - A model is a set of probability distributions.
  - A correctly specified model contains the data-generating distribution.
- **Misspecified Model:**
  Model that is not correctly specified.
- **Model Fit** (not Goodness-of-Fit): Magnitude of residual error, prediction accuracy, etc.

Model Fit is different from Model Specification

$$y = 2x_1 + 3x_2 + 1 + \varepsilon^3, \quad \varepsilon \sim N\left(0, \sigma_0^2\right)$$ **DGP**

Model Fit is different from Model Specification

$$y = 2x_1 + 3x_2 + 1 + \varepsilon^3, \quad \varepsilon \sim N\left(0, \sigma_0^2\right)$$ **DGP**

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_0 + \varepsilon, \quad \varepsilon \sim N\left(0, \sigma^2\right)$$ **Linear Model**

Model Fit is different from Model Specification

$$y = 2x_1 + 3x_2 + 1 + \varepsilon^3, \quad \varepsilon \sim N\left(0, \sigma_0^2\right)$$ **DGP**

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_0 + \varepsilon, \quad \varepsilon \sim N\left(0, \sigma^2\right)$$ **Linear Model**

- Model is misspecified in error term, yet
- Prediction/Residual Error depends on σ_0^2

Misspecification Detection is Important for Mathematical Psychology

- Goal is to model biological/behavioral systems that have testable assumptions…. **Need methods for testing!**

Misspecification Detection is Important for Mathematical Psychology

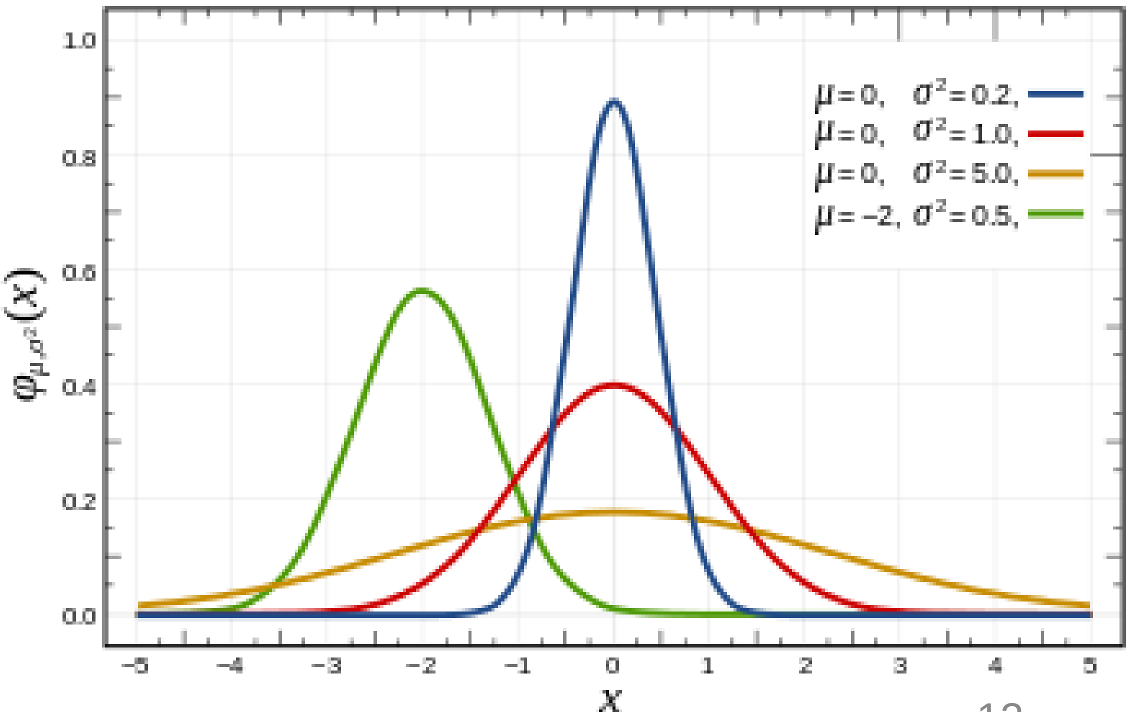
- Goal is to model biological/behavioral systems that have testable assumptions…. **Need methods for testing!**

- **Interpretable Parameters Require Correct Model Specification!**

Class of Probability Models

- Independent and Identically Distributed
- Smooth Probability Models
  (e.g., General Linear Models; Markov Fields; Hierarchical Linear Models; nonlinear regression)
- Local Identifiability (one or more strict local minimizers)
- Expectations Exist

$$\left\{ f\left(\mathbf{x};\boldsymbol{\theta}\right) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \right\}$$



MLEs (Maximum Likelihood Estimates)

Given probability model: $\left\{ f\left(\mathbf{x};\boldsymbol{\theta}\right) : \boldsymbol{\theta} \in \Theta \right\}$

- MLE of Parameters make observed data most likely
- MLE is random vector converging to local minimizer of expected value of negative normalized log-likelihood function
- MLE random variable has covariance matrix which can be estimated in 2 ways.

Two Different Methods for Estimating Covariance Matrix of MLE

Method 1 (using 2nd Derivatives):

$$\hat{\mathbf{A}}_n \equiv -\frac{1}{n}\sum_{i=1}^{n} \nabla^2 \log f\left(\mathbf{x}_i;\hat{\boldsymbol{\theta}}_n\right), \quad \hat{\mathbf{A}}_n \to \mathbf{A}^*,$$

$$\mathbf{C}^* = (1/n)\left(\mathbf{A}^*\right)^{-1}$$

Method 2 (Using 1st derivatives):

$$\hat{\mathbf{B}}_n \equiv \frac{1}{n}\sum_{i=1}^{n} \nabla \log f\left(\mathbf{x}_i;\hat{\boldsymbol{\theta}}_n\right)\left(\nabla \log f\left(\mathbf{x}_i;\hat{\boldsymbol{\theta}}_n\right)\right)^T, \quad \hat{\mathbf{B}}_n \to \mathbf{B}^*,$$

$$\mathbf{C}^* = (1/n)\left(\mathbf{B}^*\right)^{-1}$$

Information Matrix Equality

IM Equality:

Correctly Specified Model $\rightarrow$ $\mathbf{A}^* = \mathbf{B}^*$

Information Matrix Equality and **The Big Idea**

$\textbf{IM Equality}: \text{Correctly Specified Model} \rightarrow \mathbf{A}^* = \mathbf{B}^*$

$\textbf{Contrapositive}: \mathbf{A}^* \neq \mathbf{B}^* \rightarrow \text{Misspecified Model}$

Big Idea (White, 1982):

Detect Model Misspecification by testing

$$H_o : \mathbf{A}^* = \mathbf{B}^*$$

Full Information Matrix Test

(White, 1982, 1994)

$$\hat{\mathbf{A}}_n = \begin{bmatrix} 0.1 & 0.8 & 0.1 & 0.1 \\ 0.8 & 0.2 & 0.3 & \mathbf{0.4} \\ 0.1 & 0.3 & 0.3 & 0.5 \\ 0.1 & \mathbf{0.4} & 0.5 & 0.4 \end{bmatrix} \qquad \hat{\mathbf{B}}_n = \begin{bmatrix} 0.1 & 0.8 & 0.1 & 0.1 \\ 0.8 & 0.2 & 0.3 & \mathbf{0.2} \\ 0.1 & 0.3 & 0.3 & 0.5 \\ 0.1 & \mathbf{0.2} & 0.5 & 0.4 \end{bmatrix}$$

$$H_0 : \mathbf{A}^* = \mathbf{B}^*$$

- **Previous Logistic Regression Simulation Studies:**
  Poor Type 1 and Type 2 error rates
  (e.g., Aparicio & Villanua, 2001; Orme, 1990; Stomberg & White, 2000)

- **Possible Explanation:**
  - DF = K(K+1)/2 where K=number of parameters
  - Null hypothesis false if only one element is different

Generalized Information Matrix Test (GIMT) Hypothesis Function

Golden, Henley, White, and Kashner (2013, 2016)

$$H_o : \mathbf{s}\left(\mathbf{A}^*, \mathbf{B}^*\right) = \mathbf{0}_r$$

$$\text{if } \mathbf{A}^* = \mathbf{B}^*, \text{ then } H_o : \mathbf{s}\left(\mathbf{A}^*, \mathbf{B}^*\right) = \mathbf{0}_r$$

Generalized Information Matrix Test (GIMT) Hypothesis Function

Golden, Henley, White, and Kashner (2013, 2016)

$$H_o : \mathbf{s}\left(\mathbf{A}^*, \mathbf{B}^*\right) = \mathbf{0}_r$$

$$\text{if } \mathbf{A}^* = \mathbf{B}^*, \text{ then } H_o : \mathbf{s}\left(\mathbf{A}^*, \mathbf{B}^*\right) = \mathbf{0}_r$$

- Example 1: $H_o : \text{trace}\left(\mathbf{A}^*\right) = \text{trace}\left(\mathbf{B}^*\right)$

Generalized Information Matrix Test (GIMT) Hypothesis Function

Golden, Henley, White, and Kashner (2013, 2016)

$$H_o : \mathbf{s}\left(\mathbf{A}^*, \mathbf{B}^*\right) = \mathbf{0}_r$$

$$\text{if } \mathbf{A}^* = \mathbf{B}^*, \text{ then } H_o : \mathbf{s}\left(\mathbf{A}^*, \mathbf{B}^*\right) = \mathbf{0}_r$$

- Example 1: $H_o : \text{trace}\left(\mathbf{A}^*\right) = \text{trace}\left(\mathbf{B}^*\right)$
- Example 2: $H_o : \det\left(\mathbf{A}^*\right) = \det\left(\mathbf{B}^*\right)$

Generalized Information Matrix Test (GIMT) Hypothesis Function

Golden, Henley, White, and Kashner (2013, 2016)

$$H_o : \mathbf{s}\left(\mathbf{A}^*, \mathbf{B}^*\right) = \mathbf{0}_r$$

$$\text{if } \mathbf{A}^* = \mathbf{B}^*, \text{ then } H_o : \mathbf{s}\left(\mathbf{A}^*, \mathbf{B}^*\right) = \mathbf{0}_r$$

- Example 1: $H_o : \text{trace}\left(\mathbf{A}^*\right) = \text{trace}\left(\mathbf{B}^*\right)$
- Example 2: $H_o : \det\left(\mathbf{A}^*\right) = \det\left(\mathbf{B}^*\right)$
- Virtually an infinite number of "selection functions" can be defined corresponding to a virtually infinite number of GIMTs!

Directional GIMTs (Golden et al., 2013, 2016)

- Adjusted Classical GIMT

$$H_0 : \mathbf{s}\left(\mathbf{A}^*, \mathbf{B}^*\right) = \mathbf{T}\left(\mathbf{A}^* - \mathbf{B}^*\right) = 0$$

- Fisher Spectra GIMT

$$H_0 : \mathbf{s}\left(\mathbf{A}^*, \mathbf{B}^*\right) = diag\left(\left(\mathbf{A}^*\right)^{-1} \mathbf{B}^*\right) - \mathbf{1} = 0$$

- GAIC GIMT

$$H_0 : \mathbf{s}\left(\mathbf{A}^*, \mathbf{B}^*\right) = \log\left(\left(1/k\right) trace\left(\left(\mathbf{A}^*\right)^{-1} \mathbf{B}^*\right)\right) = 0$$

- GAIC Ratio GIMT

$$H_0 : \mathbf{s}\left(\mathbf{A}^*, \mathbf{B}^*\right) = \log\left(\frac{trace\left(\left(\mathbf{A}^*\right)^{-1} \mathbf{B}^*\right)}{trace\left(\left(\mathbf{B}^*\right)^{-1} \mathbf{A}^*\right)}\right) = 0$$

Nondirectional GIMTs

- Classical White (1982) Full Information Matrix Test

$$H_0 : \mathbf{s}\left(\mathbf{A}^*, \mathbf{B}^*\right) = \mathbf{A}^* - \mathbf{B}^* = 0$$

Nondirectional GIMTs

- Classical White (1982) Full Information Matrix Test

$$H_0 : \mathbf{s}\left(\mathbf{A}^*, \mathbf{B}^*\right) = \mathbf{A}^* - \mathbf{B}^* = 0$$

- Composite GAIC GIMT (Golden et al., 2016)

$$H_0 : \mathbf{s}\left(\mathbf{A}^*, \mathbf{B}^*\right) = \begin{cases} trace\left(\left(\mathbf{A}^*\right)^{-1} \mathbf{B}^*\right) = k \\ trace\left(\left(\mathbf{B}^*\right)^{-1} \mathbf{A}^*\right) = k \end{cases}$$

GIMT Selection Statistic Estimator

- The unobservable quantity $\mathbf{s}^* \equiv \mathbf{s}\left(\mathbf{A}^*, \mathbf{B}^*\right)$
- A consistent estimator of $\mathbf{s}^*$ is given by:

$$\hat{\mathbf{s}}_n \equiv \mathbf{s}\left(\hat{\mathbf{A}}_n, \hat{\mathbf{B}}_n\right)$$

GIMT Hypothesis Test Theorem

(Golden, Henley, White, Kashner, 2013, 2016)

- If $H_o : s\left(\mathbf{A}^*, \mathbf{B}^*\right) = 0$ is <u>true</u>, then $\hat{s}_n$ is asymptotically Gaussian with mean $s^* \equiv s\left(\mathbf{A}^*, \mathbf{B}^*\right)$ and variance $\sigma_s^* / \sqrt{n}$.

GIMT Hypothesis Test Theorem

(Golden, Henley, White, Kashner, 2013, 2016)

- If $H_o : s\left(\mathbf{A}^*, \mathbf{B}^*\right) = 0$ is <u>true</u>, then $\hat{s}_n$ is asymptotically Gaussian with mean $s^* \equiv s\left(\mathbf{A}^*, \mathbf{B}^*\right)$ and variance $\sigma_s^* / \sqrt{n}$.

- If $H_o : s\left(\mathbf{A}^*, \mathbf{B}^*\right) = 0$ is <u>false</u>, then $\hat{s}_n \to \infty$ w.p.1.

GIMT Hypothesis Test Theorem

(Golden, Henley, White, Kashner, 2013, 2016)

- If $H_o : s\left(\mathbf{A}^*, \mathbf{B}^*\right) = 0$ is <u>true</u>, then $\hat{s}_n$ is asymptotically Gaussian with mean $s^* \equiv s\left(\mathbf{A}^*, \mathbf{B}^*\right)$ and variance $\sigma_s^* / \sqrt{n}$.

- If $H_o : s\left(\mathbf{A}^*, \mathbf{B}^*\right) = 0$ is <u>false</u>, then $\hat{s}_n \rightarrow \infty$ w.p.1.

GIMT for testing $H_o : s\left(\mathbf{A}^*, \mathbf{B}^*\right) = 0$ at 0.05 significance level

- Step 1. Compute $p_{obs} = 1 - \int_{-1.96}^{+1.96} \left(\left(\sigma_s \sqrt{2\pi} \right)^{-1} \exp\left[-s^2 / 2\sigma_s^2 \right] \right) ds$
- Step 2. If $p_{obs} < 0.05$, reject H_o; Else do not reject H_o.

Estimating the Variance of the GIMT Selection Statistic $\hat{s}_n$

- If $H_o : s\left(\mathbf{A}^*, \mathbf{B}^*\right) = 0$ is <u>true</u>, then $\hat{s}_n$ is asymptotically Gaussian with mean $s^* \equiv s\left(\mathbf{A}^*, \mathbf{B}^*\right)$ and variance $\sigma_s^* / \sqrt{n}$.

- Method 1 (Golden et al., 2013, 2016):

> The **variance** for either a scalar-valued or vector-valued selection statistic can be computed using an analytic formula which uses the **first**, **second**, and **third derivatives** of the log-likelihood function.

Estimating the Variance of the GIMT Selection Statistic $\hat{s}_n$

- If $H_o : s\left(\mathbf{A}^*, \mathbf{B}^*\right) = 0$ is <u>true</u>, then $\hat{s}_n$ is asymptotically Gaussian with mean $s^* \equiv s\left(\mathbf{A}^*, \mathbf{B}^*\right)$ and variance $\sigma_s^* / \sqrt{n}$

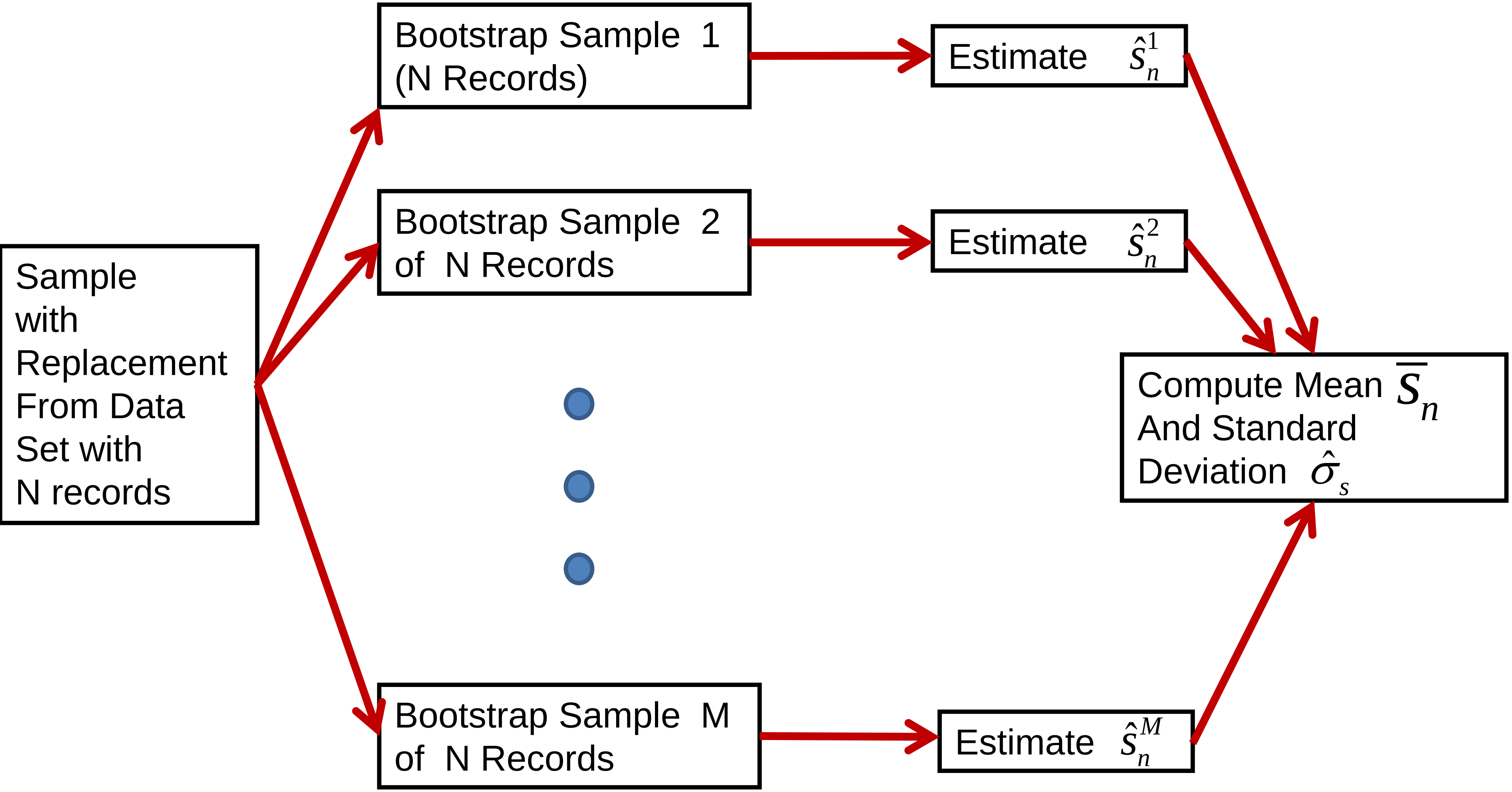
- Method 1 (Golden et al, 2013, 2016):

The variance for either a scalar-valued or vector-valued selection statistic can be computed using an analytic formula which uses the first, second, and third derivatives of the log-likelihood function.

- Method 2:

The variance for either a scalar-valued or vector-valued selection statistic can be computed using a Nonparametric (resampling) Bootstrap procedure.

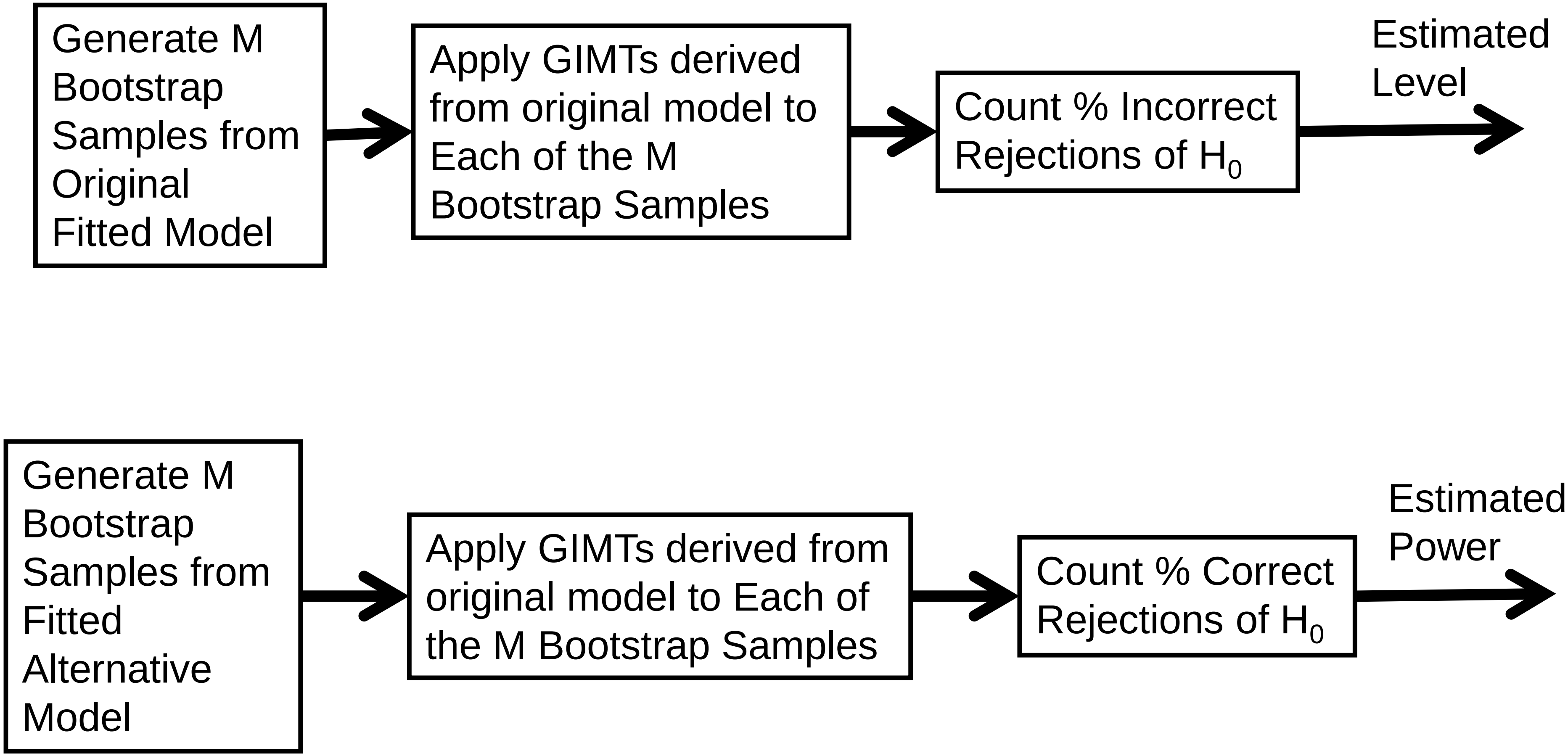
Nonparametric Bootstrap Resampling Method for Estimating Selection Statistic $\hat{S}_n$ Variance $\sigma_s^* / \sqrt{n}$



Parametric Bootstrap Simulation Studies

- Objective:
  Evaluate Quality of Derived Statistical Tests
  by Generating Data from Known
  Data Generating Process

Parametric Bootstrap Simulation Studies



GIMT Simulation Setup

- Data generated by randomly sampling x_1 on interval [-1,+1]

$$\log\left(\frac{p(y=1)}{p(y=0)}\right) = -2 + 4x_1 + 1.7x_1^{2} + 1.2x_1^{3}$$

- Correctly Specified Logistic Regression Model:

$$\log\left(\frac{p(y=1)}{p(y=0)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_1^{2} + \beta_3 x_1^{3}$$

- Misspecified Logistic Regression Model:

$$\log\left(\frac{p(y=1)}{p(y=0)}\right) = \beta_0 + \beta_1 x_1^3 + \beta_2 \sqrt{|x_1|} + \beta_3 x_2$$

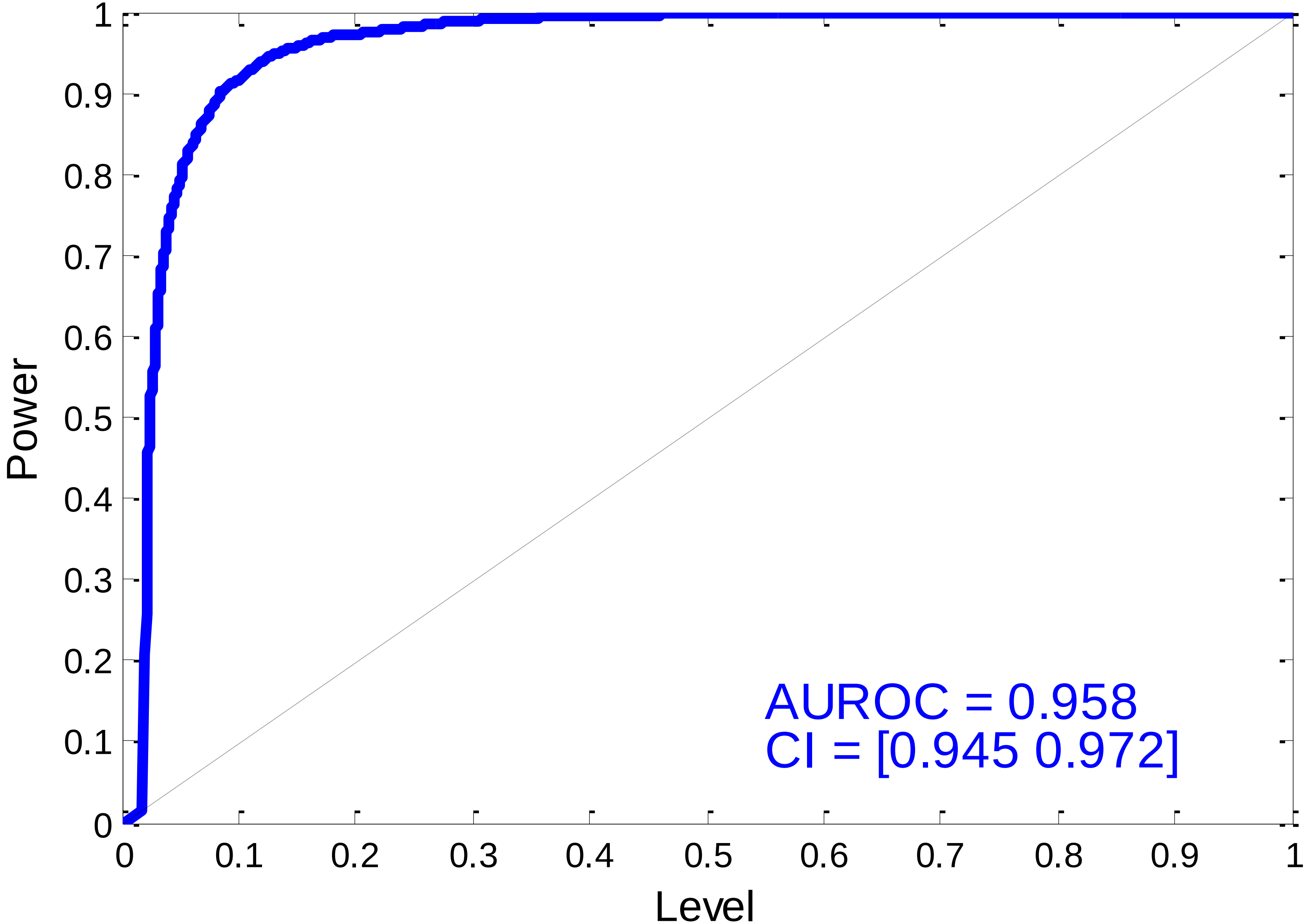
GIMT Level Performance

(Golden, Henley, White, and Kashner, 2016)

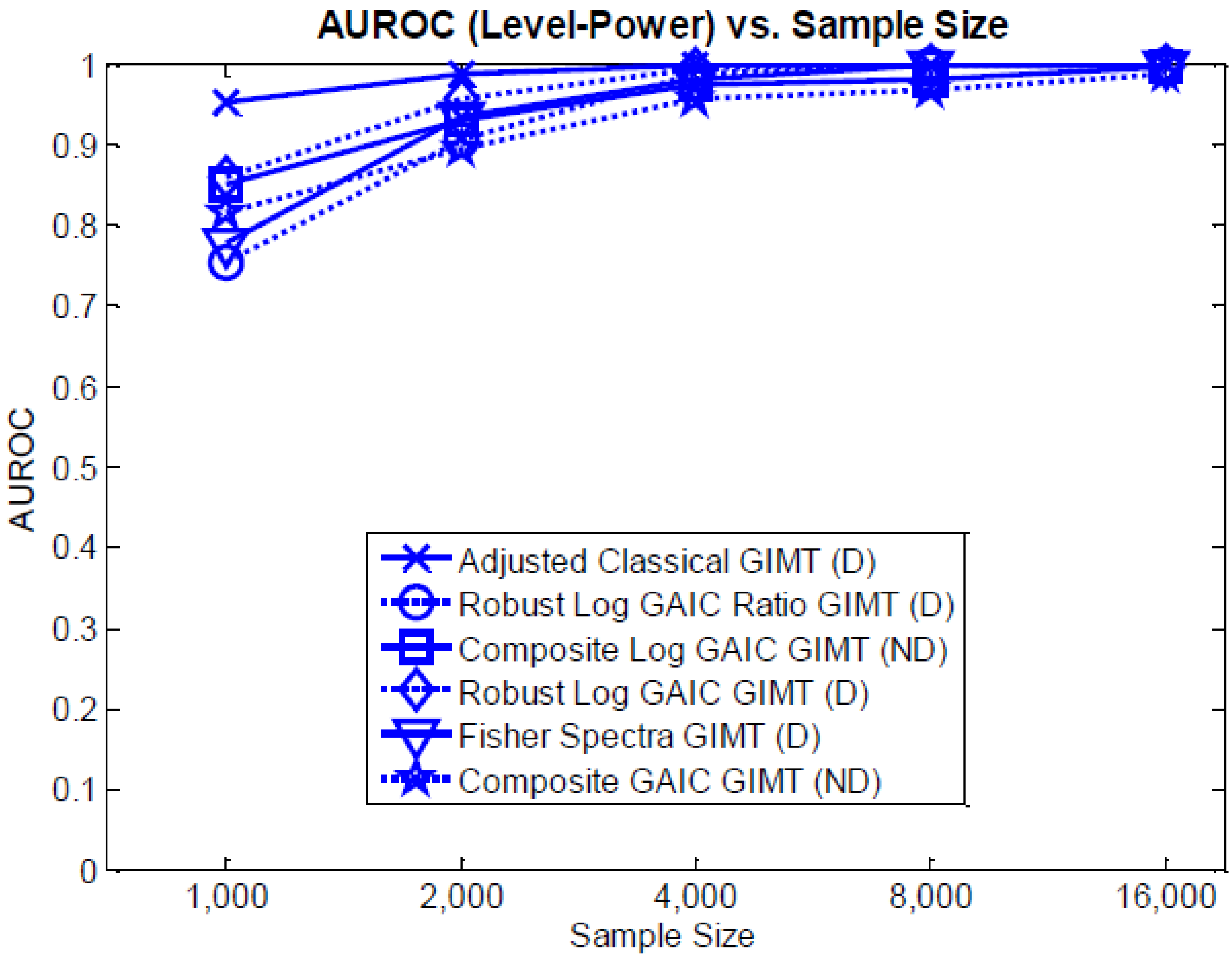
Table 1. Type 1 error performance of GIMTs using the analytic third derivative formula for pre-specified (nominal) significance levels: 0.01, 0.025, 0.05, and 0.10. Level performance for the directional GIMTs was better than level performance for the non-directional GIMTs. Bootstrap simulation standard errors are shown in parentheses. Computed values are for 10,000 simulated data samples for sample size $n = 16{,}000$. df = degrees of freedom.

Generalized Information Matrix Test (GIMT)	Test Type	$p = 0.01$	$p = 0.025$	$p = 0.05$	$p = 0.10$
Adjusted Classical (≤10 df)	Directional	0.0136 (0.0012)	0.0308 (0.0017)	0.0550 (0.0023)	0.1059 (0.0031)
Composite GAIC (2 df)	Non-Directional	0.0830 (0.0027)	0.1014 (0.0030)	0.1225 (0.0032)	0.1546 (0.0036)
Composite Log GAIC (2 df)	Non-Directional	0.0564 (0.0023)	0.0742 (0.0026)	0.0930 (0.0029)	0.1219 (0.0032)
Fisher Spectra (4 df)	Directional	0.0205 (0.0014)	0.0337 (0.0018)	0.0584 (0.0023)	0.1035 (0.0030)
Robust Log GAIC (1 df)	Directional	0.0185 (0.0013)	0.0360 (0.0018)	0.0618 (0.0024)	0.1144 (0.0031)
Robust Log GAIC Ratio (1 df)	Directional	0.0158 (0.0012)	0.0335 (0.0018)	0.0590 (0.0023)	0.1135 (0.0031)

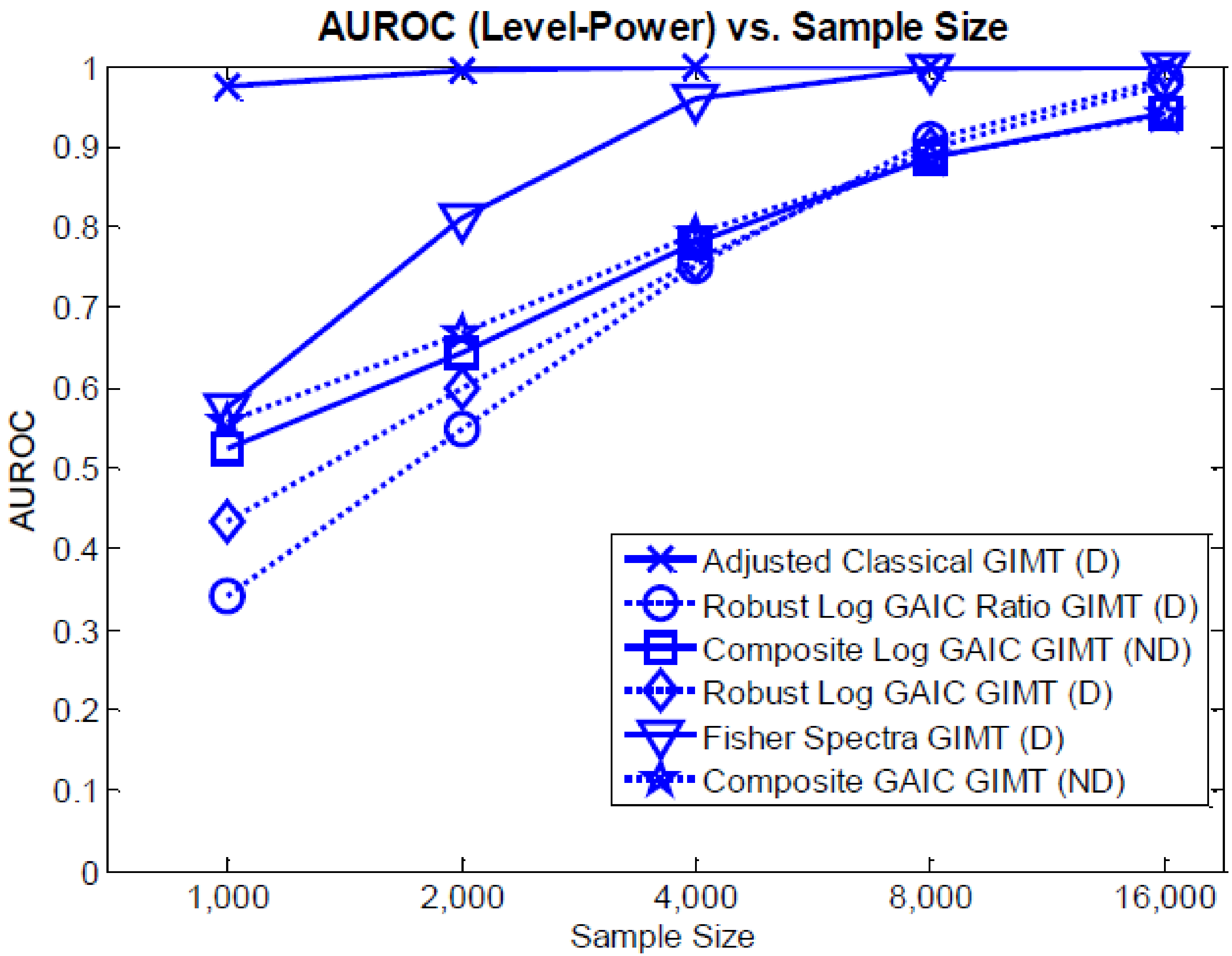
LEVEL-POWER CURVE
"Eigenspectrum" IMT Family
Log Eigenspectrum IMT, n = 1619



Analytic 3rd Derivative Formula Size-Power Results
(Golden, Henley, White, Kashner, 2016)



Lancaster-Chesher Formula Size-Power Results
(Golden, Henley, White, and Kashner, 2016)



Conclusions

- **Introduced a unified theory for specification testing applicable to most smooth parametric probability models**

Conclusions

- **Introduced a unified theory for specification testing applicable to most smooth parametric probability models**
- **GIMTs developed within this theory show good level and power performance**

Conclusions

- **Introduced a unified theory for specification testing applicable to most smooth parametric probability models**
- **GIMTs developed within this theory show good level and power performance**
- **Many types of model misspecification are possible --- Desirable to have a large variety of tests for assessing and identifying problems**

Publications

econometrics MDPI

Article

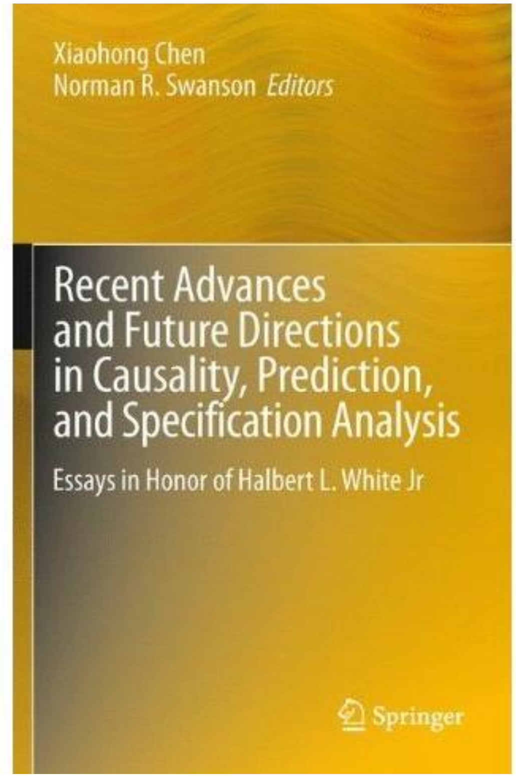
Generalized Information Matrix Tests for Detecting Model Misspecification

Richard M. Golden [1,*], Steven S. Henley [2,3,6], Halbert White [4,†] and T. Michael Kashner [3,5,6,7]

1 School of Behavioral and Brain Sciences, GR4.1, 800 W. Campbell Rd., University of Texas at Dallas, Richardson, TX 75080, USA
2 Martingale Research Corporation, 101 E. Park Blvd., Suite 600, Plano, TX 75074, USA; stevenh@martingale-research.com
3 Department of Medicine, Loma Linda University School of Medicine, Loma Linda, CA 92357, USA
4 Department of Economics, University of California San Diego, La Jolla, CA 92093, USA
5 Office of Academic Affiliations (10A2D), Department of Veterans Affairs, 810 Vermont Ave. NW (10A2D), Washington, DC 20420, USA; michael.kashner@va.gov
6 Center for Advanced Statistics in Education, VA Loma Linda Healthcare System, Loma Linda, CA 92357, USA
7 Department of Psychiatry, University of Texas Southwestern Medical Center at Dallas, Dallas, TX 75390, USA
* Correspondence: golden@utdallas.edu; Tel.: +1-972-883-2423
† Halbert White sadly passed away before this article was published.

Academic Editors: Kerry Patterson and Marc S. Paolella
Received: 29 December 2015; Accepted: 26 October 2016; Published: 15 November 2016

Download Open Access Econometrics Article from:
http://www.mdpi.com/2225-1146/4/4/46



Book Chapter:
New Directions in Information Matrix Testing: Eigenspectrum Tests (2013)
Richard M. Golden, Steven S. Henley, Halbert White, T. Michael Kashner

Or alternatively go to the blog:
www.learningmachines101.com
And visit link at end of Episode LM101-058 where *you can find a copy of this presentation*

References

- Golden RM, Henley SS, White Jr. H, Kashner TM. (2013). New Directions in Information Matrix Testing: Eigenspectrum Tests. in Causality, Prediction, and Specification Analysis: Recent Advances and Future Directions (*Festschrift Hal White Conference*), Norman Rasmus Swanson, editor, New York: Springer.
- Trivedi, MH, Greer TL, Church TS, Carmody TJ, Grannemann BD, Galper DI, Dunn AL, Earnest CP, Sunderajan P, Henley SS, Blair SN. (2011). Exercise as an augmentation treatment for nonremitted major depressive disorder. Journal of Clinical Psychiatry, 72(5), 677-684.
- Brakenridge SC, Phelan HA, Henley SS, Golden RM, Kashner TM, Eastman AE, Sperry JL, Harbrecht X, Moore EE, Cuscieri JM, Ronald V, Minei JP. Early blood product and crystalloid resuscitation: Risk association with multiple organ dysfunction after severe blunt traumatic injury. Journal of Trauma, Injury, Infection, and Critical Care, in press.

References (cont'd)

- Kashner TM, Henley SS, Golden RM, Byrne JM, Keitz SA, Cannon GW, Chang BK, Holland GJ, Aron DC, Muchmore EA, Wicker A, White Jr., HL. (2010). Studying the effects of ACGME duty hours limits on resident satisfaction: Results from VA Learners' Perceptions Survey. Academic Medicine, 85(7), 1130-1139.
- Kashner TM, Henley SS, Golden RM, Rush AJ, Jarrett RB. (2007). Assessing the preventive effects of cognitive therapy following relief of depression: A Methodological Innovation. Journal of Affective Disorders, 104, 251-261.
- White H. Maximum Likelihood Estimation of Misspecified Models. Econometrica 1982;50:1-25.
- White H. Estimation, inference, and specification analysis. Cambridge ; New York: Cambridge University Press; 1994.