

Large-Sample Hypothesis Testing in Estimating Equations and M-Estimation: Wald, Rao-Score, and Distance-Type Tests

A Synthesized Review of Theoretical Equivalence, Model Adequacy, Software Implementations, and Targeted Literature

Gemini Notebook

September 26, 2026

Abstract

In large-sample statistical theory based on estimating equations (e.g., Generalized Estimating Equations [GEE], quasi-likelihood, and general M -estimation), evaluating model adequacy or testing nested parameter restrictions $H_0 : \gamma = 0$ presents distinct challenges compared to classical maximum likelihood estimation. Because no true joint likelihood function generally exists, standard Likelihood Ratio Tests (LRT) and Analysis of Deviance are undefined or invalid unless information equality holds. This paper synthesizes the formulation, asymptotic behavior, and practical software implementations of the three main large-sample test families: the Wald-type test, the Generalized (Rao) Score-type test, and Distance-type / Quasi-Deviance tests. While the Wald and Generalized Score tests are asymptotically equivalent under model misspecification ($A \neq B$) via empirical sandwich variance estimation, the raw distance test is not. Sandwich-based eigenvalue adjustments (e.g., Rotnitzky–Jewell rescaling) are required to restore chi-square asymptotics and equivalence. Finally, we provide a targeted, annotated bibliography highlighting specific monograph chapters that treat estimating-equation variants of these tests for applied and theoretical statisticians.

1 Introduction and Theoretical Framework

Consider an independent sample of n observations $X_1, X_2, \dots, X_n$. In M -estimation and Generalized Estimating Equations [1], parameter estimation proceeds via an estimating function vector $U_n(\theta) = \frac{1}{n} \sum_{i=1}^n \psi(X_i; \theta) = 0$, where $\theta = (\beta^T, \gamma^T)^T \in \mathbb{R}^p$ represents the parameter vector. Here, $\beta \in \mathbb{R}^{p-q}$ denotes nuisance parameters under a restricted null model $H_0 : \gamma = 0$, while $\gamma \in \mathbb{R}^q$ represents additional parameters added to test model adequacy.

Under standard regularity conditions, the estimator $\hat{\theta}$ satisfies large-sample asymptotic normality:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V(\theta_0)), \quad V(\theta) = A(\theta)^{-1}B(\theta)A(\theta)^{-T} \quad (1)$$

where $A(\theta) = E \left[-\frac{\partial \psi(X; \theta)}{\partial \theta^T} \right]$ is the sensitivity matrix (bread) and $B(\theta) = E [\psi(X; \theta)\psi(X; \theta)^T]$ is the variability matrix (meat).

When information equality holds (i.e., $A(\theta) = B(\theta)$), such as in correctly specified likelihoods, the asymptotic variance simplifies to $A(\theta)^{-1}$. However, in general estimating equations where working correlation or variance structures may be misspecified, $A(\theta) \neq B(\theta)$, necessitating empirical sandwich estimators $\hat{V}_n = \hat{A}_n^{-1}\hat{B}_n\hat{A}_n^{-T}$ to maintain valid asymptotic inference.

2 Large-Sample Test Families for Model Adequacy

To test the adequacy of a restricted model $H_0 : \gamma = 0$ against an expanded model $H_1 : \gamma \neq 0$, three principal test construction methods exist.

2.1 The Wald-Type Test (W)

The Wald test requires fitting the full, unconstrained model H_1 to obtain the estimate $\hat{\theta} = (\hat{\beta}^T, \hat{\gamma}^T)^T$. Extracting the submatrix of the sandwich variance corresponding to γ , denoted $\hat{V}_{\gamma\gamma}$, the Wald statistic is:

$$W = \hat{\gamma}^T \left[\hat{V}_{\gamma\gamma} \right]^{-1} \hat{\gamma} \quad (2)$$

Under H_0 , $W \xrightarrow{d} \chi_q^2$ as $n \rightarrow \infty$, regardless of whether $A = B$.

2.2 The Generalized (Rao) Score-Type Test (S)

The Generalized Score test [2] requires fitting *only* the constrained null model H_0 , yielding $\tilde{\theta} = (\tilde{\beta}^T, 0^T)^T$. Evaluating the score component for γ at the constrained estimate, $U_{\gamma,n}(\tilde{\theta}) = \frac{1}{n} \sum_{i=1}^n \psi_{\gamma}(X_i; \tilde{\theta})$, the score statistic is:

$$S = n U_{\gamma,n}(\tilde{\theta})^T \left[\hat{V}_{\text{score}} \right]^{-1} U_{\gamma,n}(\tilde{\theta}) \quad (3)$$

where $\hat{V}_{\text{score}}$ accounts for the estimation of nuisance parameters β using sandwich components evaluated at $\tilde{\theta}$. Under H_0 , $S \xrightarrow{d} \chi_q^2$.

2.3 The Distance-Type / Pseudo-LRT Test (D)

When estimating equations stem from minimizing a contrast objective function $M_n(\theta) = \frac{1}{n} \sum_{i=1}^n \rho(X_i; \theta)$, the analogue to the Likelihood Ratio Test is the objective difference:

$$D = 2n \left[M_n(\tilde{\theta}) - M_n(\hat{\theta}) \right] \quad (4)$$

3 Asymptotic Equivalence and Sandwich Adjustments

3.1 Equivalence of Wald and Generalized Score Tests

Via a Taylor expansion of the estimating equations around θ_0 , it is straightforward to establish that under H_0 :

$$W - S = o_p(1) \quad (5)$$

Because both W and S incorporate empirical sandwich covariance matrices, both test statistics converge to the central χ_q^2 distribution and are **asymptotically equivalent** even when $A \neq B$.

3.2 Non-Equivalence of the Raw Distance Test when $A \neq B$

When model or correlation misspecification causes $A \neq B$, the **raw distance test** D is **not** asymptotically equivalent to W or S . Instead of a χ_q^2 limit, D converges in distribution to a weighted sum of independent chi-square variables:

$$D \xrightarrow{d} \sum_{j=1}^q \lambda_j Z_j^2, \quad Z_j \stackrel{iid}{\sim} N(0, 1) \quad (6)$$

where $\lambda_1, \dots, \lambda_q$ are eigenvalues of the matrix $A_{\gamma\gamma}^{-1} B_{\gamma\gamma}$.

3.3 The Adjusted Distance Test (D^*)

To restore asymptotic equivalence, the distance statistic must be rescaled using sandwich components [3]. Using a Satterthwaite or Rotnitzky–Jewell correction factor $\bar{\lambda} = \frac{1}{q} \sum_{j=1}^q \lambda_j$, the adjusted statistic:

$$D^* = \frac{D}{\bar{\lambda}} \xrightarrow{d} \chi_q^2 \quad \text{under } H_0 \quad (7)$$

With this adjustment, D^* becomes **asymptotically equivalent** to both W and S .

Table 1: Comparison of Large-Sample Test Families in Estimating Equations ($A \neq B$)

Test Family	Point of Evaluation	Null Limit ($A \neq B$)	Equivalence to Wald?
Wald Test (W)	Full Model ($\hat{\theta}$)	χ_q^2	Reference
Rao-Score Test (S)	Restricted Null ($\tilde{\theta}$)	χ_q^2	Asymptotically Equivalent
Raw Distance (D)	Both ($\tilde{\theta}$ and $\hat{\theta}$)	$\sum_{j=1}^q \lambda_j Z_j^2$	Not Equivalent
Adjusted Distance (D^*)	Both + Sandwich Matrix	χ_q^2	Asymptotically Equivalent

4 Software Implementations

In practice, applied practitioners cannot use traditional analysis of deviance when using GEE or general estimating equations because no joint log-likelihood exists. Standard implementations across major environments include:

- **R Environment:** The `glmttoolbox` package [4] provides comprehensive capabilities for comparing nested GEE models using robust Wald tests and generalized score tests. The `geepack` package [5] provides ‘`geeglm`’ object fitting with sandwich-based Wald inference.
- **SAS System:** PROC GENMOD with the REPEATED statement provides robust empirical sandwich variance estimates and supports Wald-type hypothesis tests (‘CONTRAST’ / ‘TYPE3’) as well as score statistics.
- **Stata:** The `xtgee` command fits GEE models and offers robust variance computation (`vce(robust)`), with linear hypothesis testing via post-estimation `test` syntax.

5 Recommended Literature & Annotated Bibliography

While many advanced asymptotic texts detail general M -estimation asymptotics, applied practitioners and biostatisticians often struggle to find texts that specifically treat estimating-equation variants of Wald, score, and quasi-deviance tests. Below is a curated, annotated guide to targeted monograph chapters that address these exact methods.

5.1 Theoretical & Methodological Monographs

D. D. Boos & L. A. Stefanski (2013) — *Essential Statistical Inference: Theory and Methods*, Chapter 7 (M -Estimation and Estimating Equations)

This chapter provides the definitive theoretical exposition of estimating equations and generalized score tests. Building on Boos (1992) [2], it explicitly derives the generalized score test statistic S under nuisance parameter estimation, demonstrating why it maintains valid χ_q^2

asymptotics under sandwich covariance estimation without requiring full likelihood specifications.

P. McCullagh & J. A. Nelder (1989) — *Generalized Linear Models* (2nd ed.), Chapter 9 (Quasi-Likelihood Functions)

The foundational source for pseudo-LRT and analysis of quasi-deviance. McCullagh and Nelder define quasi-deviance $D(y; \hat{\mu}) = 2 \int_{\hat{\mu}}^y \frac{y-u}{V(u)} du$ as the direct estimating-equation analogue to likelihood deviance, providing the mathematical underpinning for quasi-likelihood ratio testing.

P. X.-K. Song (2007) — *Correlated Data Analysis: Modeling, Analytics, and Applications*, Chapters 3 & 4

Provides rigorous derivations of vector quasi-likelihood estimating equations, sandwich variance estimators, generalized score tests, and quadratic distance statistics specifically under correlated data settings where information equality fails ($A \neq B$).

5.2 Applied & Software-Oriented Textbooks

J. W. Hardin & J. M. Hilbe (2012) — *Generalized Estimating Equations* (2nd ed.), Chapters 4 & 5

The premier applied textbook dedicated to GEE. Chapters 4 and 5 explicitly detail why traditional likelihood ratio tests fail under GEE, detail the mechanics of robust Wald and generalized score tests, and explain Pan's Quasi-Likelihood Information Criterion (QIC) as the primary practical surrogate for deviance in GEE model selection.

G. M. Fitzmaurice, N. M. Laird, & J. H. Ware (2011) — *Applied Longitudinal Analysis* (2nd ed.), Chapters 12 & 13

An exceptionally clear guide for applied biostatisticians. It explicitly warns against using standard analysis of deviance in marginal models, provides intuitive derivations of robust Wald and score tests for nested GEE models, and illustrates their implementation in standard software.

References

- [1] Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13–22.
- [2] Boos, D. D. (1992). Generalized score tests for M -estimators and estimating equations. *The American Statistician*, 46(4), 327–333.
- [3] Rotnitzky, A., & Jewell, N. P. (1990). Hypothesis testing of regression parameters in semi-parametric generalized linear models for cluster correlated data. *Biometrika*, 77(3), 485–497.
- [4] Vanegas, L. H., Rondón, L. M., & Paula, G. A. (2023). Generalized Estimating Equations using the new R package glmtoolbox. *The R Journal*, 15(2), 105–133.
- [5] Højsgaard, S., Halekoh, U., & Yan, J. (2006). The R package geepack for generalized estimating equations. *Journal of Statistical Software*, 15(2), 1–11.
- [6] Boos, D. D., & Stefanski, L. A. (2013). *Essential Statistical Inference: Theory and Methods* (Chapter 7: M -Estimation and Estimating Equations). Springer Science & Business Media.
- [7] McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models* (2nd ed., Chapter 9: Quasi-Likelihood Functions). Chapman & Hall/CRC.

- [8] Hardin, J. W., & Hilbe, J. M. (2012). *Generalized Estimating Equations* (2nd ed., Chapters 4 & 5). Chapman & Hall/CRC.
- [9] Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied Longitudinal Analysis* (2nd ed., Chapters 12 & 13). John Wiley & Sons.
- [10] Song, P. X.-K. (2007). *Correlated Data Analysis: Modeling, Analytics, and Applications* (Chapters 3 & 4). Springer.
- [11] Zhang, H., Chu, Y., Liu, Y., & Roychoudhury, S. (2025). A robust score test in g-computation for covariate adjustment in randomized clinical trials leveraging different variance estimators via influence functions. *arXiv preprint arXiv:2503.13066*.