

A Unified Robust Inference Framework for Univariate and Multivariate Nonlinear Statistical Models

Sreenivasulu Arigela¹, M. Venkataramanaiah², B. Sarojamma³

¹Research Scholar, Department of Statistics, S.V University, Tirupati

²Professor (Rtd), Department of Statistics, S.V University, Tirupati.

³Professor and HoD, Department of Statistics, S.V University, Tirupati.

Corresponding author : saroja14397@gmail.com

Abstract

Nonlinear regression models underpin quantitative work across the biological, physical, pharmacological, and economic sciences, yet the inferential machinery most practitioners rely on—ordinary nonlinear least squares (NLS) together with the classical asymptotic covariance—rests on assumptions that empirical data routinely violate. When errors depart from normality, exhibit non-constant variance, are serially or cross-sectionally dependent, or are contaminated by a handful of outliers and high-leverage design points, NLS estimates lose efficiency, the associated standard errors become misleading, and Wald and likelihood-ratio decisions can be badly miscalibrated. This paper develops a Unified Robust Nonlinear Inference (URNI) framework that couples high-breakdown robust estimation with heteroscedasticity- and autocorrelation-consistent covariance estimation and resampling calibration inside a single, internally coherent procedure. Estimation uses an MM-type strategy that begins from a high-breakdown S-estimator and refines it with a redescending M-step, augmented by adaptive weights that jointly address residual outlyingness, design leverage, variance heterogeneity, and dependence. Inference is based on a robust sandwich covariance, with a heteroscedasticity- and autocorrelation-consistent variant, and is calibrated in finite samples by a wild bootstrap implemented through a fast robust reweighting device. The construction is extended to multivariate nonlinear responses with a general error covariance, yielding stacked estimating equations, a multivariate sandwich covariance, and joint cross-response tests. The theoretical development records assumptions and proof outlines for consistency, asymptotic normality, boundedness of the influence function, breakdown behaviour, and covariance and bootstrap validity, and separates results that are proved from those left as conjectures for later empirical study. No simulation or data analysis is reported; the contribution is methodological and theoretical.

Keywords: Nonlinear regression; robust estimation; MM-estimators; sandwich covariance; wild bootstrap; influence function; multivariate nonlinear models; heteroscedasticity

1. Introduction

Many scientific relationships are intrinsically nonlinear in their parameters. Exponential growth and decay laws describe radioactive processes and pharmacokinetic clearance; the logistic and Gompertz curves capture bounded growth of populations, tumours, and technology adoption; the Michaelis–Menten equation governs enzyme kinetics; and sigmoidal dose–response models are the workhorses of toxicology and pharmacology. In each case the mean response is a smooth but nonlinear function of a low-dimensional parameter vector whose components carry direct scientific meaning—a rate constant, an asymptote, a half-saturation concentration, an inflection point. Because these quantities are the object of inference, the reliability of the standard errors and tests attached to them matters as much as the point estimates themselves.

The dominant estimation principle for such models is nonlinear least squares, whose large-sample behaviour was placed on rigorous footing by Jennrich (1969), Malinvaud (1970), and Wu (1981), and later synthesised in the monographs of Gallant (1987), Bates and Watts (1988), and Seber and Wild (1989). Under independent, homoscedastic, and ideally Gaussian errors, NLS is consistent, asymptotically normal, and efficient, and the familiar covariance $\hat{\sigma}^2(F'F)^{-1}$ supports valid Wald and likelihood-ratio inference. The difficulty is that the conditions guaranteeing these properties are demanding, and the estimator inherits from ordinary least squares an acute fragility whose consequences are amplified by curvature of the response surface.

Several distinct problems arise in practice, often simultaneously. Non-normal, heavy-tailed error distributions inflate the variance of NLS and undermine the normal reference distributions used for testing. Heteroscedasticity, endemic in growth and dose–response data where variability scales with the mean, leaves the point estimator consistent but renders the classical covariance inconsistent, so that reported standard errors are systematically wrong (White, 1980). Serial or spatial dependence has a similar effect on the covariance and additionally distorts efficiency. A single gross outlier, or a small cluster of them, can dominate the sum of squared residuals and pull the fitted curve far from the bulk of the data; high-leverage design points—extreme predictor configurations that are common when concentrations or doses are spaced geometrically—exert disproportionate influence and can mask their own presence. Model misspecification, near-collinearity of the columns of the Jacobian in over-parameterised mechanistic models, and small sample sizes each further degrade the accuracy of first-order asymptotic approximations (White, 1981). Finally, when several responses are measured on the same units and are correlated, treating each response with a separate univariate fit discards information and precludes coherent joint inference.

Robust statistics offers principled remedies for the contamination problems, beginning with the M-estimators of Huber (1964, 1973) and the influence-function and breakdown-point apparatus of Hampel (1974) and Hampel, Ronchetti, Rousseeuw, and Stahel (1986). High-breakdown regression estimators—least median and least trimmed squares (Rousseeuw, 1984), S-estimators (Rousseeuw and Yohai, 1984), and the MM-estimators of Yohai (1987) that combine a 50%

10.48047/jocaaa.2026.35.05.26

breakdown point with high Gaussian efficiency—are now standard for the linear model, and their extension to nonlinear regression has been studied by Stromberg and Ruppert (1992), Stromberg (1993), and Fasano, Maronna, Sued, and Yohai (2012). In parallel, the econometric literature produced covariance estimators that are consistent under heteroscedasticity (White, 1980; MacKinnon and White, 1985; Cribari-Neto, 2004) and under both heteroscedasticity and autocorrelation (Newey and West, 1987; Andrews, 1991), while resampling methods (Efron, 1979; Wu, 1986; Mammen, 1993) provide finite-sample calibration when asymptotic approximations are unreliable. These strands, however, have largely developed in isolation, and their combination in the nonlinear setting—where robustness, covariance estimation, and resampling all interact with response-surface curvature—has received comparatively little unified treatment.

The objective of this paper is to formulate and analyse a single framework that brings robust estimation, robust covariance estimation, and resampling calibration together for both univariate and multivariate nonlinear models, and to record its theoretical properties carefully, distinguishing what can be proved from what remains to be checked empirically. The specific contributions are as follows.

1. We define the Unified Robust Nonlinear Inference (URNI) framework, which integrates an MM-type high-breakdown estimator, adaptive weights that simultaneously downweight residual outliers and leverage points and accommodate variance heterogeneity and dependence, a robust sandwich covariance with a heteroscedasticity- and autocorrelation-consistent variant, and a wild-bootstrap calibration built on a fast robust reweighting device.
2. We extend the framework to multivariate nonlinear models with a general error covariance, developing a robust Mahalanobis-type objective, stacked estimating equations, a multivariate sandwich covariance, and joint tests across responses and across shared parameters.
3. We assemble the supporting large-sample theory—consistency, asymptotic normality, the influence function and its boundedness, breakdown behaviour, and consistency of the covariance and bootstrap procedures—as propositions with assumptions and proof outlines, citing the results on which each rests.
4. We give explicit algorithms with precise convergence criteria and an analysis of computational cost relative to NLS, clarifying the overhead introduced by robustness and resampling and how the fast robust bootstrap mitigates it.

We make no claim of priority for the individual ingredients, several of which are classical; the intended value is in their coherent assembly, the unified univariate–multivariate treatment, and the careful separation of established from conjectural properties.

2. Review of Literature

2.1 Nonlinear least squares

The asymptotic theory of NLS was established by Jennrich (1969), who proved strong consistency and asymptotic normality of the estimator under a uniform convergence condition on the empirical sum-of-squares surface, and by Malinvaud (1970), who gave complementary consistency arguments. Wu (1981) characterised the design conditions necessary and sufficient for the existence of consistent estimators in the nonlinear model, sharpening the earlier results. Practical and geometric perspectives were developed by Gallant (1975, 1987), by Ratkowsky (1983), whose treatment emphasised parameterisation and the behaviour of particular growth models, and in the influential texts of Bates and Watts (1988) and Seber and Wild (1989). The behaviour of NLS when the mean function is misspecified was analysed by White (1981), who showed that the estimator converges to a well-defined pseudo-true value and characterised the consequences for inference; this line connects nonlinear estimation to the quasi-maximum-likelihood viewpoint.

2.2 Numerical estimation

Because the NLS normal equations are nonlinear, estimation is iterative. The Gauss–Newton method linearises the mean function about the current iterate, and its damped and trust-region modifications improve global behaviour: Levenberg (1944) and Marquardt (1963) introduced the additive damping that bears their names, and the modern trust-region theory was developed by Moré (1978) and Moré and Sorensen (1983) and is presented comprehensively by Nocedal and Wright (2006). Quasi-Newton updates of BFGS type were proposed independently by Broyden (1970), Fletcher (1970), Goldfarb (1970), and Shanno (1970), and the numerical treatment of unconstrained optimisation and nonlinear systems is codified by Dennis and Schnabel (1983). The reliability of curvature-based confidence regions in nonlinear models was quantified by Bates and Watts (1980), whose relative curvature measures separate intrinsic from parameter-effects nonlinearity and thereby indicate when linear approximations to the inference are trustworthy.

2.3 Robust estimation

Robust estimation begins with Huber (1964), who introduced M-estimators of location and the minimax variance criterion, and Huber (1973), who extended the asymptotics to regression; the results are consolidated in Huber (1981). Hampel (1974) formalised the influence function as a measure of local sensitivity, and the influence-function approach was developed at book length by Hampel et al. (1986). The redescending biweight loss of Beaton and Tukey (1974) provided a widely used bounded-influence objective. High-breakdown estimation was inaugurated by Rousseeuw (1984) with least median and least trimmed squares, followed by the S-estimators of Rousseeuw and Yohai (1984) and the MM-estimators of Yohai (1987), which attain both a 50% breakdown point and high efficiency at the Gaussian model; the field is surveyed by Rousseeuw

10.48047/jocaaa.2026.35.05.26

and Leroy (1987) and by Maronna, Martin, and Yohai (2006). For nonlinear regression specifically, Stromberg and Ruppert (1992) analysed breakdown, Stromberg (1993) addressed computation of high-breakdown nonlinear estimators, Sakata and White (2001) developed S-estimation for nonlinear models with dependent and heterogeneous observations, and Fasano et al. (2012) established continuity, consistency, and asymptotic normality of nonlinear S- and MM-functionals. Tabatabai and Argyros (1993) proposed robust estimation and testing procedures for general nonlinear models. The iteratively reweighted least-squares device that underlies practical M-estimation was studied by Holland and Welsch (1977), and the interplay of transformation and weighting for variance heterogeneity was treated by Carroll and Ruppert (1988).

2.4 Robust variance estimation

The sandwich form of the covariance for estimating equations originates with Huber (1967), who analysed maximum-likelihood behaviour under nonstandard conditions. White (1980) introduced the heteroscedasticity-consistent estimator now labelled HC0; MacKinnon and White (1985) proposed the finite-sample corrections HC1–HC3, introduced precisely because HC0, though asymptotically valid, tends to understate standard errors in small samples, with HC3 approximating the jackknife estimator through a $(1 - h_{ii})^2$ denominator; and Cribari-Neto (2004) proposed the leverage-sensitive HC4. Comparing these variants, Long and Ervin (2000) found HC0 unreliable in samples of a few hundred observations or fewer and recommended HC3 as the default in such cases. Under dependence, Newey and West (1987) gave a positive-semidefinite heteroscedasticity- and autocorrelation-consistent estimator using a Bartlett kernel, and Andrews (1991) developed the general kernel theory and data-driven bandwidth selection. In the clustered and longitudinal setting, Liang and Zeger (1986) introduced generalised estimating equations with a robust working-correlation covariance, and Kauermann and Carroll (2001) examined the efficiency cost of sandwich estimation and its finite-sample coverage.

2.5 Resampling-based inference

Efron (1979) introduced the bootstrap; the methodology is developed in Efron and Tibshirani (1993), Hall (1992), and Davison and Hinkley (1997). For regression, Freedman (1981) analysed the residual bootstrap, while heteroscedasticity-robust resampling was achieved through the wild bootstrap of Wu (1986), Liu (1988), and Mammen (1993), whose refinements Davidson and Flachaire (2008) studied in detail. Practical guidance on bootstrap hypothesis testing—in particular the importance of resampling a pivotal, studentised statistic—was given by Hall and Wilson (1991), and the theory of bootstrap confidence intervals was reviewed by DiCiccio and Efron (1996). For robust estimators the naive bootstrap is both unstable and expensive; Salibián-Barrera and Zamar (2002) introduced the fast and robust bootstrap, which resamples a linearised reweighted representation and so avoids recomputing the estimator on every replicate, and Salibián-Barrera, Van Aelst, and Willems (2008) extended and consolidated the approach.

2.6 Multivariate nonlinear models

Systems of related regressions were formalised by Zellner (1962), whose seemingly unrelated regression (SUR) estimator exploits cross-equation error correlation to gain efficiency; the nonlinear analogue is treated by Gallant (1987). Robust and high-breakdown methods for multivariate and system regressions were developed subsequently: Koenker and Portnoy (1990) studied M-estimation of multivariate regressions, Bilodeau and Duchesne (2000) introduced affine-equivariant S-estimators for the SUR model, and Rousseeuw, Van Aelst, Van Driessen, and Agulló (2004) developed robust multivariate regression with fast computation. The multivariate location–scatter problem that these methods build upon was addressed by the S-estimators of Lopuhaä (1989), the minimum covariance determinant estimator and its fast algorithm (Rousseeuw and Van Driessen, 1999), and the study of uniqueness of S- and M-functionals under non-elliptical distributions by Tatsuoka and Tyler (2000).

2.7 Research gap

Three observations emerge from this body of work. First, robust point estimation and robust covariance estimation have advanced largely on separate tracks: high-breakdown nonlinear estimators are typically paired with asymptotic-variance formulas that assume symmetric, homoscedastic errors, while heteroscedasticity- and autocorrelation-consistent covariances are almost always applied to non-robust estimators. Their combination in nonlinear models, where the Jacobian is itself estimated and curvature affects both, is not standard. Second, univariate and multivariate nonlinear inference are usually treated apart, so that correlated multiple responses are handled by separate fits that forgo efficiency and cannot support joint tests. Third, the acute sensitivity of NLS to outliers and leverage in nonlinear settings, the demonstrated unreliability of classical standard errors under heteroscedasticity and dependence, and the well-documented inadequacy of first-order asymptotics in small samples together argue for resampling calibration—yet resampling is rarely integrated with robust nonlinear estimation because of its computational burden. The framework developed below is designed to close these gaps within a single construction.

3. Methodology

3.1 Models and notation

For the univariate model we observe pairs (y_i, x_i) , $i = 1, \dots, n$, and posit

$$y_i = f(x_i, \theta) + \varepsilon_i, \quad \theta \in \Theta \subseteq \mathbb{R}^p, \quad x_i \in \mathbb{R}^k, \quad (1)$$

where f is known and twice continuously differentiable in θ , and the ε_i are errors with zero conditional median. Write $r_i(\theta) = y_i - f(x_i, \theta)$ for the residual and

10.48047/jocaaa.2026.35.05.26

$$F_i(\theta) = \frac{\partial f(x_i, \theta)}{\partial \theta} \in \mathbb{R}^p, \quad F(\theta) = [F_1(\theta), \dots, F_n(\theta)]' \in \mathbb{R}^{n \times p}, \quad (2)$$

for the gradient and the $n \times p$ Jacobian.

For the multivariate model each unit yields an m -vector response,

$$\mathbf{Y}_i = \mathbf{f}(x_i, \theta) + \boldsymbol{\varepsilon}_i, \quad \mathbf{Y}_i \in \mathbb{R}^m, \quad \boldsymbol{\varepsilon}_i \sim (\mathbf{0}, \Sigma), \quad (3)$$

where $\mathbf{f} = (f_1, \dots, f_m)'$ collects the component mean functions, $\theta \in \mathbb{R}^p$ stacks the parameters of all m equations, and Σ is the $m \times m$ error covariance capturing dependence among responses on the same unit. Let θ_0 (respectively θ_0) denote the true parameter value.

3.2 Classical NLS and its covariance

Ordinary NLS minimises $S(\theta) = \sum_{i=1}^n r_i(\theta)^2$. The classical treatment (Jennrich, 1969; Wu, 1981) assumes: (A1) θ compact and θ_0 identified, so that the limiting objective is uniquely minimised at θ_0 ; (A2) $f(x, \cdot)$ continuous on θ and twice continuously differentiable in a neighbourhood of θ_0 ; (A3) errors independent and identically distributed with mean zero and variance σ^2 ; (A4) the limit $Q = \lim_{n \rightarrow \infty} n^{-1} F(\theta_0)' F(\theta_0)$ exists and is positive definite. The first-order condition is the system of p normal equations

$$F(\hat{\theta})' r(\hat{\theta}) = \sum_{i=1}^n F_i(\hat{\theta}) r_i(\hat{\theta}) = \mathbf{0}. \quad (4)$$

The Gauss–Newton iteration solves the linearised problem, giving the update

$$\theta^{(t+1)} = \theta^{(t)} + (F^{(t)'} F^{(t)})^{-1} F^{(t)'} r^{(t)}, \quad (5)$$

with $F^{(t)} = F(\theta^{(t)})$ and $r^{(t)} = r(\theta^{(t)})$. Under (A1)–(A4),

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(\mathbf{0}, \sigma^2 Q^{-1}), \quad (6)$$

estimated in finite samples by $\widehat{\text{Var}}(\hat{\theta}) = \hat{\sigma}^2 (F(\hat{\theta})' F(\hat{\theta}))^{-1}$ with $\hat{\sigma}^2 = (n - p)^{-1} \sum_{i=1}^n r_i(\hat{\theta})^2$. The covariance in (6) is valid only under (A3); it is inconsistent under heteroscedasticity or

10.48047/jocaaa.2026.35.05.26

dependence and is highly non-robust because both $\hat{\sigma}^2$ and the fitted Jacobian $F(\hat{\theta})$ are distorted by outliers and leverage.

3.3 The URNI framework

URNI replaces (4) and (6) with five interacting components.

Component 1: Robust MM-type objective. Given a robust residual scale $\hat{\sigma}$, define the M-estimator as a minimiser of

$$L(\theta) = \sum_{i=1}^n \rho\left(\frac{r_i(\theta)}{\hat{\sigma}}\right), \quad (7)$$

where ρ is a bounded, even loss that is nondecreasing on $[0, \infty)$. We admit Huber's function, with derivative $\psi(u) = \max\{-c, \min(u, c)\}$, and Tukey's biweight,

$$\rho_c(u) = \begin{cases} \frac{c^2}{6} [1 - (1 - (u/c)^2)^3], & |u| \leq c, \\ \frac{c^2}{6}, & |u| > c, \end{cases} \quad \psi_c(u) = u(1 - (u/c)^2)^2 \mathbf{1}\{|u| \leq c\}. \quad (8)$$

Following Yohai (1987) and its nonlinear extension by Fasano et al. (2012), estimation proceeds in MM fashion: a high-breakdown S-estimator supplies an initial $\tilde{\theta}$ and an M-scale $\hat{\sigma}$ solving $n^{-1} \sum_{i=1}^n \rho_0(r_i(\tilde{\theta})/\hat{\sigma}) = \delta$ for a bounded ρ_0 and $\delta = E_\Phi[\rho_0]$, after which (7) is minimised from $\tilde{\theta}$ with a redescending ψ tuned for efficiency while $\hat{\sigma}$ is held fixed. For the biweight, taking $\rho_0(r) = \rho_1(r/k_0)$ with $k_0 = 1.56$ makes the S-scale, and hence the MM-estimator, attain the asymptotic breakdown point 0.5; because breakdown depends only on k_0 and normal efficiency only on the M-step constant, the two can be tuned separately. The S-start protects the breakdown point; the M-step recovers efficiency. A median absolute deviation (MAD) scale may replace the S-scale when a fast, if less efficient, start is preferred.

Component 2: Adaptive weights. The robust normal equations attach to each observation a weight combining four factors,

$$w_i = w_i^r w_i^x w_i^v w_i^d, \quad (9)$$

where the residual weight $w_i^r = \psi(r_i/\hat{\sigma})/(r_i/\hat{\sigma})$ implements the redescending downweighting of vertical outliers; the leverage weight w_i^x downweights high-leverage Jacobian rows using a robust Mahalanobis distance d_i of F_i computed from a robust location–scatter estimate of the

10.48047/jocaaa.2026.35.05.26

Jacobian rows (for instance the MCD; Rousseeuw and Van Driessen, 1999), in Mallows form $w_i^x = \min\{1, (b/d_i)^\gamma\}$ or in Schweppe form (Hampel et al., 1986); the variance weight $w_i^v = 1/\hat{v}(x_i)$ accommodates heteroscedasticity through a working variance function (Carroll and Ruppert, 1988); and w_i^d is a dependence adjustment induced by the HAC construction in Component 3. The estimating equations become

$$\sum_{i=1}^n w_i \psi\left(\frac{r_i(\hat{\theta})}{\hat{\sigma}}\right) F_i(\hat{\theta}) = \mathbf{0}. \quad (10)$$

Component 3: Robust sandwich covariance. Writing the estimating function as $g_i(\theta) = w_i \psi(r_i/\hat{\sigma}) F_i$, a Taylor expansion of $\sum_{i=1}^n g_i(\hat{\theta}) = \mathbf{0}$ about θ_0 gives the sandwich form

$$\hat{V} = \hat{A}^{-1} \hat{B} \hat{A}^{-1}, \quad (11)$$

with the “bread”

$$\hat{A} = \frac{1}{n\hat{\sigma}^2} \sum_{i=1}^n w_i \psi'\left(\frac{r_i}{\hat{\sigma}}\right) F_i F_i', \quad (12)$$

and the “meat”, under independence,

$$\hat{B} = \frac{1}{n} \sum_{i=1}^n w_i^2 \psi\left(\frac{r_i}{\hat{\sigma}}\right)^2 F_i F_i'. \quad (13)$$

Equation (13) is the nonlinear, robust analogue of White’s (1980) HC0 meat; leverage-based finite-sample corrections in the spirit of MacKinnon and White (1985) and Cribari-Neto (2004) can be applied to the summands. When autocorrelation is present, $\hat{B}$ is replaced by the heteroscedasticity- and autocorrelation-consistent version (Newey and West, 1987; Andrews, 1991)

$$\hat{B}_{\text{HAC}} = \hat{I}_0 + \sum_{\ell=1}^L \kappa\left(\frac{\ell}{L+1}\right) (\hat{I}_\ell + \hat{I}_\ell'), \quad \hat{I}_\ell = \frac{1}{n} \sum_{i=\ell+1}^n s_i s_{i-\ell}', \quad (14)$$

where $s_i = w_i \psi(r_i/\hat{\sigma}) F_i$, κ is the Bartlett kernel, and L is the bandwidth.

10.48047/jocaaa.2026.35.05.26

Component 4: Robust tests and intervals. For the linear hypothesis $H_0: C\theta = q$ against $H_1: C\theta \neq q$, with C of full row rank s , the robust Wald statistic is

$$W = n(C\hat{\theta} - q)'(C\hat{V}C')^{-1}(C\hat{\theta} - q) \xrightarrow{d} \chi_s^2 \quad \text{under } H_0. \quad (15)$$

Individual coefficients use $t_j = \hat{\theta}_j / \sqrt{[\hat{V}/n]_{jj}}$, with confidence intervals $\hat{\theta}_j \pm z_{1-\alpha/2} \sqrt{[\hat{V}/n]_{jj}}$. A robust likelihood-ratio-type statistic based on the ρ -objective,

$$\Lambda = \frac{2}{K} [L(\hat{\theta}_{H_0}) - L(\hat{\theta})], \quad K = \frac{E[\psi^2]}{E[\psi']}, \quad (16)$$

with $\hat{\theta}_{H_0}$ the constrained minimiser, provides a complementary test whose null distribution is asymptotically χ_s^2 .

Component 5: Wild bootstrap calibration. Because (15) and (16) rely on first-order asymptotics that are unreliable in small samples and in strongly curved problems, URNI calibrates them by a wild bootstrap. For $b = 1, \dots, B$, generate

$$y_i^{*(b)} = f(x_i, \hat{\theta}) + t_i^{(b)} \hat{\sigma} r_i^\dagger, \quad (17)$$

where $r_i^\dagger$ is a scaled robust residual and the $t_i^{(b)}$ are independent mean-zero, unit-variance weights—Rademacher or the two-point weights of Mammen (1993)—then recompute the studentised statistic and read critical values from the bootstrap distribution (Wu, 1986; Liu, 1988; Mammen, 1993; Davidson and Flachaire, 2008). The wild bootstrap is chosen precisely because it preserves each observation's conditional variance and is therefore valid under heteroscedasticity, unlike the residual bootstrap of Freedman (1981). To avoid recomputing the MM-estimator on every replicate, the fast robust bootstrap of Salibián-Barrera and Zamar (2002) resamples the linearised reweighted representation of $\hat{\theta}$, solving only a linear system per replicate; a block or wild-block variant preserves dependence. Studentising the resampled statistic follows the pivoting guidance of Hall and Wilson (1991).

3.4 Algorithm 1: Univariate URNI

Step 1. Compute an initial high-breakdown S-estimator $\tilde{\theta}$ from a number of elemental subsets, retaining the fit with the smallest robust scale.

Step 2. Compute the M-scale $\hat{\sigma}$ from the residuals $r_i(\tilde{\theta})$ by solving $n^{-1} \sum_{i=1}^n \rho_0(r_i/\hat{\sigma}) = \delta$.

Step 3. Set $\theta^{(0)} = \tilde{\theta}$ and $t = 0$.

Step 4. Evaluate residuals $r_i^{(t)}$ and Jacobian rows $F_i^{(t)}$.

Step 5. Form residual weights w_i^r from $\psi(r_i^{(t)}/\hat{\sigma})$.

Step 6. Form leverage weights w_i^x from robust Mahalanobis distances of $\{F_i^{(t)}\}$.

Step 7. Form variance weights w_i^y from the working variance function; set $w_i = w_i^r w_i^x w_i^y$.

Step 8. Solve the damped weighted Gauss–Newton step

$$\theta^{(t+1)} = \theta^{(t)} + (F^{(t)'} W F^{(t)} + \lambda I)^{-1} F^{(t)'} W \psi^{(t)} \hat{\sigma}, \quad (18)$$

with $W = \text{diag}(w_1, \dots, w_n)$, $\psi^{(t)} = (\psi(r_1^{(t)}/\hat{\sigma}), \dots, \psi(r_n^{(t)}/\hat{\sigma}))'$, and Levenberg–Marquardt damping $\lambda \geq 0$.

Step 9. If the objective (7) does not decrease, increase λ (step-halving) and repeat Step 8.

Step 10. Check convergence: stop if $\|\theta^{(t+1)} - \theta^{(t)}\|/\|\theta^{(t)}\| < 10^{-8}$ or if the relative change in (7) is below 10^{-10} ; otherwise set $t \leftarrow t + 1$ and return to Step 4. The iteration count is capped at 200; failure to converge within this cap is reported rather than silently accepted.

Step 11. At convergence compute $\hat{A}$ and $\hat{B}$ (or $\hat{B}_{\text{HAC}}$) and form $\hat{V}$.

Step 12. Compute Wald, t -type, and ρ -based statistics and asymptotic confidence intervals.

Step 13. Calibrate by the fast robust wild bootstrap and report bootstrap critical values and intervals together with the point estimates.

4. Main Research

4.1 Multivariate extension

For model (3), let $\hat{\Sigma}$ be a robust estimate of the error scatter and define the squared robust Mahalanobis residual

$$D_i^2(\theta) = (\mathbf{Y}_i - \mathbf{f}(x_i, \theta))' \hat{\Sigma}^{-1} (\mathbf{Y}_i - \mathbf{f}(x_i, \theta)). \quad (19)$$

The multivariate objective is

$$L_m(\theta) = \sum_{i=1}^n \rho \left(\sqrt{D_i^2(\theta)} \right), \quad (20)$$

with $\hat{\Sigma}$ obtained by an S- or MCD-type step (Lopuhaä, 1989; Rousseeuw and Van Driessen, 1999; Bilodeau and Duchesne, 2000). Joint estimation differs from separate univariate fits in four respects: it exploits the cross-response correlation encoded in $\hat{\Sigma}$, and so gains efficiency exactly as SUR improves on equation-by-equation least squares (Zellner, 1962); it yields a covariance for the full parameter vector that supports joint tests across responses; it permits parameters shared across response equations to be estimated coherently; and it downweights a unit as a whole when its multivariate residual is aberrant, rather than trimming one coordinate at a time. Differentiating (20) gives the stacked estimating equations

$$\sum_{i=1}^n w_i \frac{\psi(D_i)}{D_i} J_i' \hat{\Sigma}^{-1} (\mathbf{Y}_i - \mathbf{f}(x_i, \hat{\theta})) = \mathbf{0}, \quad (21)$$

where $J_i = \partial \mathbf{f}(x_i, \theta) / \partial \theta'$ is the $m \times P$ Jacobian of the stacked mean function. The multivariate sandwich covariance retains the form (11) with

$$\hat{A} = \frac{1}{n} \sum_{i=1}^n w_i \frac{\psi(D_i)}{D_i} J_i' \hat{\Sigma}^{-1} J_i, \quad \hat{B} = \frac{1}{n} \sum_{i=1}^n w_i^2 \frac{\psi(D_i)^2}{D_i^2} J_i' \hat{\Sigma}^{-1} \mathbf{u}_i \mathbf{u}_i' \hat{\Sigma}^{-1} J_i, \quad (22)$$

where $\mathbf{u}_i = \mathbf{Y}_i - \mathbf{f}(x_i, \hat{\theta})$, and a HAC version paralleling (14) applies when units are ordered in time. The joint hypothesis $H_0: C \text{vec}(\theta) = q$ is tested by the Wald statistic (15) with $\hat{\theta}$ and its multivariate $\hat{V}$, and is calibrated by a multivariate wild bootstrap that multiplies the whole residual vector $\mathbf{u}_i$ by a scalar weight $t_i^{(b)}$, preserving the within-unit covariance (Salibián-Barrera et al., 2008).

4.2 Theoretical properties

We record the main properties as propositions with assumptions and proof outlines. Throughout, $\psi = \rho'$ is bounded and continuous, θ_0 is identified and interior to a compact θ , f is twice continuously differentiable, the weights w_i are bounded, and the relevant limiting matrices exist and are nonsingular.

Proposition 1 (Consistency). Under the stated assumptions and the nonlinear analogue of Jennrich's uniform convergence condition, the URNI estimator defined by (10) is strongly consistent for θ_0 .

10.48047/jocaaa.2026.35.05.26

Outline. The scaled objective $n^{-1}L(\theta)$ converges uniformly on θ to a limit uniquely minimised at θ_0 ; boundedness and continuity of ρ give the uniform law of large numbers, and identifiability gives the unique minimum. The argument follows Huber (1967) and Jennrich (1969), extended to nonlinear S- and MM-functionals by Fasano et al. (2012). *Status: proved, given the cited regularity.*

Proposition 2 (Asymptotic normality). Under the assumptions of Proposition 1 together with nonsingularity of A ,

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(\mathbf{0}, A^{-1}BA^{-1}), \quad (23)$$

where A and B are the population limits of (12) and (13).

Outline. A mean-value expansion of the estimating equation (10) about θ_0 , the central limit theorem applied to $n^{-1/2} \sum_{i=1}^n g_i(\theta_0)$, and consistency of the Jacobian of the estimating function give (23); see Fasano et al. (2012) for the nonlinear case and Huber (1967) for the general sandwich. *Status: proved, given the cited regularity.*

Proposition 3 (Influence function and bounded influence). The influence function of the URNI functional T at the model distribution G is

$$\text{IF}(y, x; T, G) = A^{-1} \psi\left(\frac{r}{\sigma}\right) w(x) \frac{\partial f(x, \theta_0)}{\partial \theta}, \quad (24)$$

which is bounded in the response direction whenever ψ is bounded and bounded in the design direction whenever $w(x) \partial f(x, \theta_0) / \partial \theta$ is bounded.

Outline. This is standard for M-functionals (Hampel, 1974; Hampel et al., 1986); the bounded, redescending ψ of (8) caps the vertical influence, and the leverage weight w^x caps the design influence, so the gross-error sensitivity is finite. *Status: proved.*

Proposition 4 (Breakdown behaviour). With a high-breakdown S-start, the MM-type nonlinear estimator inherits a high (up to 50%) breakdown point, in the sense appropriate to nonlinear regression.

Outline. Breakdown in nonlinear models must be defined through the behaviour of fitted values rather than parameter norms (Stromberg and Ruppert, 1992; Sakata and White, 2001); the S-scale bounds the effect of contamination, and the M-step, initialised at the S-fit, cannot reduce the breakdown point (Yohai, 1987). Because a closed-form breakdown expression is unavailable in the nonlinear case, we state this as an inherited property under the conditions of Fasano et

10.48047/jocaaa.2026.35.05.26

al. (2012) rather than as a sharp value. *Status: partly proved; the numerical breakdown value is model-dependent and left open.*

Proposition 5 (Efficiency at the Gaussian model). The tuning constant governs efficiency at the normal model: Huber's ψ with $c = 1.345$ and the biweight with $c = 4.685$ each yield approximately 95% asymptotic efficiency relative to least squares under Gaussian errors, without sacrificing the breakdown point set by the S-start (Yohai, 1987; Maronna et al., 2006).

Outline. Efficiency at the model follows from Proposition 2 by comparing $A^{-1}BA^{-1}$ with σ^2Q^{-1} when the errors are Gaussian; the constants are the standard ones. *Status: established.*

Proposition 6 (Covariance consistency). The sandwich estimator $\hat{V}$ in (11) is consistent for the asymptotic covariance under heteroscedasticity, and $\hat{V}$ formed with $\hat{B}_{HAC}$ is consistent under additional weak dependence, provided the bandwidth satisfies $L \rightarrow \infty$ with $L/n \rightarrow 0$.

Outline. Consistency of $\hat{A}$ and $\hat{B}$ follows from uniform laws of large numbers; the HAC case uses the kernel arguments of Newey and West (1987) and Andrews (1991). *Status: proved, given standard mixing and moment conditions.*

Proposition 7 (Bootstrap validity). The wild bootstrap distribution of the studentised statistic converges to the same limit as the statistic itself, so that bootstrap critical values are asymptotically valid; the fast robust reweighting reproduces this limit to first order.

Outline. Validity of the wild bootstrap under heteroscedasticity follows Mammen (1993) and Davidson and Flachaire (2008); the linearised reweighting is justified by Salibián-Barrera and Zamar (2002). *Status: first-order validity proved; higher-order refinement in the nonlinear case is conjectured and left for empirical study.*

4.3 Computational complexity

The per-iteration cost of NLS is dominated by forming and solving the $p \times p$ normal system, $O(np^2 + p^3)$, which is $O(np^2)$ for $n \gg p$. Weighted M-estimation adds only the $O(n)$ reweighting, so it shares the same order per iteration. The MM procedure adds the one-time cost of the S-start, which is dominated by subsampling: evaluating N elemental subsets costs $O(Np^3)$ plus $O(Nn)$ for scale evaluation, incurred once before the iterations. The sandwich covariance (11) costs $O(np^2)$, and the HAC meat (14) adds $O(Lnp^2)$. The bootstrap multiplies the estimation cost by B ; the fast robust bootstrap reduces each replicate to the solution of a single linear system, $O(p^3)$ after an initial $O(np^2)$ set-up, so its total cost is $O(np^2 + Bp^3)$ rather than B full re-estimations. In the multivariate case the corresponding orders scale with the stacked dimension P and include the $O(m^3)$ cost of factorising $\hat{\Sigma}$.

5. Conclusion

This paper has assembled a single framework, URNI, for inference in univariate and multivariate nonlinear models that is simultaneously resistant to outliers and leverage, valid under heteroscedasticity and weak dependence, and calibrated in finite samples by resampling. The construction inherits its robustness from the MM strategy of Yohai (1987) as extended to nonlinear models by Fasano et al. (2012), its covariance validity from the sandwich and HAC literature (White, 1980; Newey and West, 1987; Andrews, 1991), and its finite-sample calibration from the wild and fast robust bootstraps (Mammen, 1993; Salibián-Barrera and Zamar, 2002). The multivariate extension treats correlated responses jointly, gaining the efficiency that motivated seemingly unrelated regression (Zellner, 1962) while retaining affine-equivariant robustness (Bilodeau and Duchesne, 2000; Rousseeuw et al., 2004). The theoretical section states consistency, asymptotic normality, bounded influence, breakdown, efficiency, and covariance and bootstrap validity, and marks which of these are proved under cited conditions and which are conjectural.

The limitations are explicit. No simulation or data analysis is presented, so the finite-sample size and power of the proposed tests, and the accuracy of the bootstrap calibration in curved nonlinear problems, remain to be assessed empirically. The choice of tuning constants trades efficiency against robustness and interacts with response-surface curvature in ways that merit study. High-dimensional θ , near-collinear Jacobians, and the nonconvexity of the robust objective—which admits multiple local minima and makes the S-start consequential—all pose practical challenges. Future work should provide a Monte Carlo evaluation across representative growth and dose-response models, application to real data, an examination of breakdown and bandwidth choices specific to nonlinear designs, and a software implementation making the framework routinely available.

References

- Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3), 817–858.
- Bates, D. M., & Watts, D. G. (1980). Relative curvature measures of nonlinearity. *Journal of the Royal Statistical Society, Series B*, 42(1), 1–25.
- Bates, D. M., & Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*. New York: Wiley.
- Beaton, A. E., & Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16(2), 147–185.
- Bilodeau, M., & Duchesne, P. (2000). Robust estimation of the SUR model. *Canadian Journal of Statistics*, 28(2), 277–288.

10.48047/jocaaa.2026.35.05.26

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms. *Journal of the Institute of Mathematics and Its Applications*, 6(1), 76–90.

Carroll, R. J., & Ruppert, D. (1988). *Transformation and Weighting in Regression*. New York: Chapman and Hall.

Cribari-Neto, F. (2004). Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics & Data Analysis*, 45(2), 215–233.

Davidson, R., & Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics*, 146(1), 162–169.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.

Dennis, J. E., & Schnabel, R. B. (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, NJ: Prentice-Hall.

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3), 189–228.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1), 1–26.

Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Fasano, M. V., Maronna, R. A., Sued, M., & Yohai, V. J. (2012). Continuity and differentiability of regression M functionals. *Bernoulli*, 18(4), 1284–1309.

Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, 13(3), 317–322.

Freedman, D. A. (1981). Bootstrapping regression models. *The Annals of Statistics*, 9(6), 1218–1228.

Gallant, A. R. (1975). Nonlinear regression. *The American Statistician*, 29(2), 73–81.

Gallant, A. R. (1987). *Nonlinear Statistical Models*. New York: Wiley.

Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of Computation*, 24(109), 23–26.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer.

Hall, P., & Wilson, S. R. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics*, 47(2), 757–762.

10.48047/jocaaa.2026.35.05.26

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346), 383–393.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley.

Holland, P. W., & Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. *Communications in Statistics – Theory and Methods*, 6(9), 813–827.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1), 73–101.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 221–233. Berkeley: University of California Press.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, 1(5), 799–821.

Huber, P. J. (1981). *Robust Statistics*. New York: Wiley.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40(2), 633–643.

Kauermann, G., & Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association*, 96(456), 1387–1396.

Koenker, R., & Portnoy, S. (1990). M estimation of multivariate regressions. *Journal of the American Statistical Association*, 85(412), 1060–1068.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2), 164–168.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13–22.

Liu, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *The Annals of Statistics*, 16(4), 1696–1708.

Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54(3), 217–224.

Lopuhaä, H. P. (1989). On the relation between S-estimators and M-estimators of multivariate location and covariance. *The Annals of Statistics*, 17(4), 1662–1683.

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3), 305–325.

10.48047/jocaaa.2026.35.05.26

- Malinvaud, E. (1970). The consistency of nonlinear regressions. *The Annals of Mathematical Statistics*, 41(3), 956–969.
- Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 21(1), 255–285.
- Maronna, R. A., Martin, R. D., & Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*. Chichester: Wiley.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2), 431–441.
- Moré, J. J. (1978). The Levenberg–Marquardt algorithm: Implementation and theory. In G. A. Watson (Ed.), *Numerical Analysis* (Lecture Notes in Mathematics, Vol. 630, pp. 105–116). Berlin: Springer.
- Moré, J. J., & Sorensen, D. C. (1983). Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing*, 4(3), 553–572.
- Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708.
- Nocedal, J., & Wright, S. J. (2006). *Numerical Optimization* (2nd ed.). New York: Springer.
- Ratkowsky, D. A. (1983). *Nonlinear Regression Modeling: A Unified Practical Approach*. New York: Marcel Dekker.
- Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, 79(388), 871–880.
- Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. New York: Wiley.
- Rousseeuw, P. J., & Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3), 212–223.
- Rousseeuw, P. J., Van Aelst, S., Van Driessen, K., & Agulló, J. (2004). Robust multivariate regression. *Technometrics*, 46(3), 293–305.
- Rousseeuw, P. J., & Yohai, V. J. (1984). Robust regression by means of S-estimators. In J. Franke, W. Härdle, & R. D. Martin (Eds.), *Robust and Nonlinear Time Series Analysis* (Lecture Notes in Statistics, Vol. 26, pp. 256–272). Berlin: Springer.
- Sakata, S., & White, H. (2001). S-estimation of nonlinear regression models with dependent and heterogeneous observations. *Journal of Econometrics*, 103(1–2), 5–72.

10.48047/jocaaa.2026.35.05.26

Salibián-Barrera, M., Van Aelst, S., & Willems, G. (2008). Fast and robust bootstrap. *Statistical Methods and Applications*, 17(1), 41–71.

Salibián-Barrera, M., & Zamar, R. H. (2002). Bootstrapping robust estimates of regression. *The Annals of Statistics*, 30(2), 556–582.

Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear Regression*. New York: Wiley.

Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, 24(111), 647–656.

Stromberg, A. J. (1993). Computation of high breakdown nonlinear regression parameters. *Journal of the American Statistical Association*, 88(421), 237–244.

Stromberg, A. J., & Ruppert, D. (1992). Breakdown in nonlinear regression. *Journal of the American Statistical Association*, 87(420), 991–997.

Tabatabai, M. A., & Argyros, I. K. (1993). Robust estimation and testing for general nonlinear regression models. *Applied Mathematics and Computation*, 58(1), 85–101.

Tatsuoka, K. S., & Tyler, D. E. (2000). On the uniqueness of S-functionals and M-functionals under nonelliptical distributions. *The Annals of Statistics*, 28(4), 1219–1243.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838.

White, H. (1981). Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association*, 76(374), 419–433.

Wu, C. F. J. (1981). Asymptotic theory of nonlinear least squares estimation. *The Annals of Statistics*, 9(3), 501–513.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, 14(4), 1261–1295.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15(2), 642–656.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57(298), 348–368.