International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

2026 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper; all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 7th Mar 2026. Link

https://ijiemr.org/downloads.php?vol=Volume-15& issue=issue03

DOI: 10.48047/IJIEMR/V15/ISSUE 03/40

Title Improved Estimation and Testing Procedures for Linear Statistical Models under Model Violations

Volume 15, ISSUE 03, Pages: 317 – 336 Paper Authors

Sreenivasulu Arigela, M. Venkataramanaiah, B .Sarojamma





USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper as Per UGC Guidelines We Are Providing A Electronic Bar code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Improved Estimation and Testing Procedures for Linear Statistical Models under Model Violations

Sreenivasulu Arigela[1], M. Venkataramanaiah[2], B .Sarojamma[3]

[1]Research Scholar, Department of Statistics, S.V University, Tirupati [2]Professor(Rtd), Department of Statistics, S.V University, Tirupati.

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

Corresponding Author: saroja14397@gmail.com

Abstract

The ordinary least squares (OLS) estimator of the linear model is the best linear unbiased estimator when the errors are homoscedastic, serially uncorrelated and the design matrix has full column rank. These conditions rarely hold together in observational data. Heteroscedasticity and autocorrelation leave the OLS coefficients unbiased but invalidate the conventional covariance matrix, so that t-tests, F-tests and confidence intervals lose their nominal properties. Outlying and high-leverage observations bias the coefficient estimates themselves, and near-collinear designs inflate their variances. The established remedies treat these problems one at a time: heteroscedasticity-consistent and autocorrelation-consistent covariance estimators repair the variance but are computed from OLS residuals that outliers distort; robust M- and MM-estimators repair the coefficients but are usually reported with standard errors that presuppose spherical errors; feasible generalized least squares requires a correctly specified variance model; and the residual bootstrap is invalid under heteroscedasticity. This paper develops an integrated procedure, Robust Adaptive Linear Inference (RALI), which combines a high-breakdown MM-estimator of the coefficients with a leverage-adjusted sandwich covariance matrix built from the robust score contributions, and calibrates Wald-type tests of the general linear hypothesis by a restricted wild bootstrap. The estimator and its covariance are derived, regularity conditions for consistency and asymptotic normality are stated, and a seven-scenario Monte Carlo framework covering single and simultaneous violations is specified. Properties inherited from the component methods are distinguished from those that must be verified computationally. The framework offers a coherent basis for inference when several classical assumptions fail at once.

Keywords: Linear statistical models; Gauss–Markov theorem; Robust estimation; Model violations; Hypothesis testing; Heteroscedasticity; Autocorrelation; Statistical inference

Introduction

The linear regression model occupies a central position in applied statistics because it is simple to fit, easy to interpret and supported by a complete finite-sample theory under the classical assumptions. Writing the model as $Y = X\beta + \varepsilon$, the Gauss–Markov theorem states

that the OLS estimator is the best linear unbiased estimator whenever the errors have zero conditional mean and a scalar covariance matrix. When, in addition, the errors are normally distributed, the OLS coefficients are exactly normal and the usual t- and F-statistics have exact reference distributions. This combination of optimality and exactness explains the ubiquity of OLS in economics, the environmental sciences, epidemiology and engineering.

The same theory also identifies the weakness of OLS: every optimality property is conditional on an assumption that empirical data routinely violate. Cross-sectional data on households, firms or regions typically display error variances that change with the level of one or more regressors. Time-ordered observations exhibit serial correlation. Measurement errors, recording mistakes and genuinely atypical units produce outliers in the response and high-leverage points in the design. Macroeconomic and environmental regressors are often nearly collinear. The consequences of these departures differ. Heteroscedasticity and autocorrelation leave the OLS point estimates unbiased and consistent but make the conventional covariance estimator $\hat{\sigma}^2(X'X)^{-1}$ inconsistent, so that inference is misleading even in large samples. Outliers and influential observations bias the coefficients themselves, because the squared-error criterion allows a single observation to exert unbounded influence. Multicollinearity does not bias the estimates but inflates their variances, making individual coefficients unstable.

A large methodological literature addresses each of these difficulties. White (1980) introduced a covariance estimator that is consistent under heteroscedasticity of unknown form, and MacKinnon and White (1985) and Cribari-Neto (2004) proposed finite-sample refinements that adjust for leverage. Newey and West (1987) and Andrews (1991) extended the sandwich idea to serially correlated errors. Huber (1964, 1973) initiated the theory of M-estimation, and Rousseeuw (1984) and Yohai (1987) developed estimators with high breakdown point that resist a substantial proportion of contaminated observations. Efron (1979), Wu (1986) and Liu (1988) established resampling methods for regression, and Davidson and Flachaire (2008) showed how the wild bootstrap can be tuned to give reliable heteroscedasticity-robust tests in small samples. Hoerl and Kennard (1970) proposed ridge regression for near-collinear designs.

Each of these tools is well understood in isolation. Their joint behaviour is not. A sandwich covariance estimator computed from OLS residuals inherits the bias of OLS under contamination; a robust estimator reported with a covariance formula that assumes independent, identically distributed errors yields incorrect standard errors under heteroscedasticity or autocorrelation; and a bootstrap applied to OLS reproduces its

sensitivity to outliers. Applied analysts who face several violations at once are therefore left to choose among procedures that each address only part of the problem.

This paper develops an estimation and testing procedure designed for the simultaneousviolation setting. The procedure, which we call Robust Adaptive Linear Inference (RALI), is assembled from three components that are mutually consistent rather than merely juxtaposed: an MMestimator of the coefficients, a leverageadjusted heteroscedasticity- and autocorrelationconsistent covariance estimator built from the same robust estimating equations, and a restricted wild bootstrap for the general linear hypothesis. The contributions of the paper are as follows.

1. A coefficient estimator that is resistant to outliers and high-leverage points and is paired with a covariance estimator that remains valid under heteroscedasticity and weak serial dependence.
2. A leverage-adjusted robust sandwich estimator that carries the leverage discount of Cribari-Neto (2004) over from OLS residuals to robust score contributions.
3. A robust Wald test of H_0: $C\beta = r$ calibrated by a null-imposed wild bootstrap in the sense of Davidson and Flachaire (2008).
4. A statement of the theoretical properties of the procedure, separating those that follow from established results on the component methods from those that require computational verification.
5. A systematic Monte Carlo framework covering seven violation scenarios, including a simultaneous-violation scenario, together with the performance measures needed to compare RALI with OLS, weighted and generalized least squares, heteroscedasticity-consistent and autocorrelation-consistent inference, robust regression and bootstrap inference.

The remainder of the paper is organised as follows. Section 2 reviews the literature. Section 3 sets out the linear model, the Gauss–Markov framework and the consequences of assumption failure, and states the research gap. Section 4 develops the proposed estimator, its covariance matrix and the associated tests, and states their theoretical properties. Section 5 describes the Monte Carlo framework. Section 6 concludes.

2. Review of Literature

2.1 Least squares and the Gauss–Markov framework

The optimality of least squares among linear unbiased estimators under spherical errors is

classical. Aitken (1935) showed that when the error covariance matrix is $\sigma^2\Omega$ with Ω known and positive definite, the generalized least squares estimator $(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$ is the best linear unbiased estimator, and that OLS is inefficient in this case. Weighted least squares is the special case in which Ω is diagonal. In practice Ω is unknown; feasible generalized least squares replaces it by an estimate obtained from a parametric model of the variance or correlation structure. The efficiency gains of feasible GLS depend on the adequacy of that model, and the estimator is no more resistant to contamination than OLS.

2.2 Diagnostics for assumption failure

Formal tests for the classical assumptions were developed largely in the 1950s to 1980s. Durbin and Watson (1950, 1951) proposed the bounds test for first-order serial correlation in least squares residuals; Breusch (1978) and Godfrey (1978) developed Lagrange multiplier tests for higher-order autocorrelation. Breusch and Pagan (1979) derived a Lagrange multiplier test for heteroscedasticity, and White (1980) proposed a general test that does not require specifying the form of the variance function. Cook (1977) introduced a distance measure for the influence of individual observations, and Belsley, Kuh and Welsch (1980) systematised leverage, studentized residuals, variance inflation factors and condition indices as diagnostic tools. These diagnostics identify which assumptions fail but do not by themselves supply a corrected inference.

2.3 Heteroscedasticity-consistent covariance estimation

White (1980) proposed estimating the covariance matrix of OLS by the sandwich form $(X'X)^{-1}(\Sigma_i\ \hat{u}_i^2 x_i x_i')(X'X)^{-1}$, which is consistent under heteroscedasticity of unknown form. This estimator, later labelled HC0, is biased downward in finite samples. MacKinnon and White (1985) introduced three modifications: HC1 rescales by $n/(n-p)$, HC2 divides each squared residual by $1-h_i$, where h_i is the leverage, and HC3 divides by $(1-h_i)^2$, an approximation to the jackknife. Long and Ervin (2000) compared these estimators by simulation and recommended HC3 whenever the sample size is below about 250. Cribari-Neto and Zarkos (2001) found that the presence of highleverage points affects the finite-sample behaviour of these estimators more than the strength of heteroscedasticity itself, and Cribari-Neto (2004) proposed HC4, in which the exponent applied to $1-h_i$ varies with the leverage of the observation. Further leveragesensitive variants appear in CribariNeto, Souza and Vasconcellos (2007) and CribariNeto and da Silva (2011). Hausman and Palmer (2012) and Imbens and Kolesár (2016) revisited small-sample behaviour and recommended degrees-of-freedom corrections. MacKinnon (2013) surveys the first three decades of this literature. Cattaneo, Jansson and Newey

(2018) showed that the standard sandwich estimators are inconsistent when the number of regressors grows with the sample size and proposed a correction. Zeileis (2004, 2006) provided a unified software implementation.

2.4 Autocorrelation-consistent covariance estimation

Newey and West (1987) proposed a heteroscedasticity- and autocorrelationconsistent estimator that weights sample autocovariances of the score contributions by a Bartlett kernel, guaranteeing a positive semidefinite result. Andrews (1991) placed kernel estimators in a general framework, derived the meansquared-error-optimal bandwidth and showed that the quadratic spectral kernel is optimal in that sense; Andrews and Monahan (1992) added a prewhitening step. Kiefer and Vogelsang (2005) developed an alternative asymptotic theory in which the bandwidth is a fixed fraction of the sample size, giving reference distributions that better approximate finitesample behaviour. Block resampling methods for dependent data, including the movingblock bootstrap of Künsch (1989) and the stationary bootstrap of Politis and Romano (1994), offer a resampling alternative to kernel estimation.

2.5 Robust regression Huber (1964) introduced M-estimation of location by replacing the squared-error loss with a function that grows linearly for large residuals, and Huber (1973) extended the approach to regression and derived its asymptotic theory. M-estimators bound the influence of large residuals but not of highleverage points, so their breakdown point in regression is zero. Hampel, Ronchetti, Rousseeuw and Stahel (1986) developed the influence-function approach to robustness and the class of bounded-influence estimators. Rousseeuw (1984) introduced least median of squares and least trimmed squares, which attain a breakdown point of one half but are inefficient under normal errors; Rousseeuw and Yohai (1984) introduced S-estimators, which achieve high breakdown by minimising a robust scale of the residuals. Yohai (1987) resolved the tension between breakdown and efficiency with the MM-estimator, which starts from a high-breakdown S-estimate, fixes the resulting scale and then solves a redescending M-estimating equation tuned for high efficiency. Rousseeuw and Leroy (1987) and Maronna, Martin and Yohai (2006) give book-length treatments. Koenker and Bassett (1978) introduced regression quantiles, of which least absolute deviations is a special case. Salibian-Barrera and Zamar (2002) developed a fast bootstrap for MMestimators. Atkinson, Riani and Torti (2016) combined robust estimation with heteroscedastic regression using the forward search.

2.6 Resampling and permutation inference

Efron (1979) introduced the bootstrap. Freedman and Lane (1983) proposed a

permutation procedure for regression that permutes residuals from the reduced model and is close to exact for testing a subset of coefficients. Wu (1986) studied jackknife and bootstrap methods for regression and showed that resampling residuals is inconsistent under heteroscedasticity, proposing instead a scheme that multiplies each residual by an external random variable; Liu (1988) refined this scheme by matching the third moment of the errors, and Mammen (1993) established its asymptotic validity, naming it the wild bootstrap. Flachaire (2005) compared the wild bootstrap with the pairs bootstrap under heteroscedasticity, and Godfrey (2006) studied tests for regression models with heteroscedasticity of unknown form. Davidson and Flachaire (2008) demonstrated that the two-point Rademacher wild bootstrap, applied with the null hypothesis imposed on the restricted residuals, gives the smallest error in rejection probability for heteroscedasticity-robust t-tests. These results establish the wild bootstrap as the resampling method of choice under heteroscedasticity but do not address contamination.

2.7 Regularised estimation Hoerl and Kennard (1970) proposed ridge regression, which adds a multiple of the identity matrix to X′X and thereby trades a small bias for a substantial reduction in variance when the design is ill-conditioned. The ridge estimator is relevant to the present work only as a stabilising device for the collinearity scenario and is not itself a robust or heteroscedasticityconsistent procedure.

2.8 Summary of the literature The literature supplies a valid covariance estimator for non-spherical errors, a valid coefficient estimator for contaminated data and a valid resampling scheme for heteroscedastic errors. It does not supply a single procedure in which the coefficient estimator is robust, the covariance estimator is computed from the robust estimating equations and remains consistent under heteroscedasticity and serial dependence, and the reference distribution is obtained by a resampling scheme that respects both structures. This is the gap addressed below.

3. The Linear Model, Model Violations and the Research Gap

3.1 Model and classical assumptions

Consider the linear model

$$Y = X\beta + , \quad (1)$$

where Y is the n × 1 vector of responses, X is the n × p design matrix with first column equal to a vector of ones, β is the p × 1 vector of unknown coefficients and ε is the n × 1 vector of unobserved errors. The classical assumptions are:

- **(A1) Correct specification:** $E(\varepsilon \mid X) = 0$.
- **(A2) Spherical errors:** $\mathrm{Var}(\varepsilon \mid X) = \sigma^2 I_n$, that is, homoscedasticity and absence of serial correlation.
- **(A3) Full rank:** $\mathrm{rank}(X) = p < n$.

- **(A4) Normality:** $\varepsilon \mid X \sim N(0, \sigma^2 I_n)$, required only for exact finite-sample inference.

3.2 The OLS estimator and the Gauss–Markov theorem

The OLS estimator minimises the residual sum of squares $S(\beta) = (Y - X\beta)'(Y - X\beta)$. Differentiating,

$$\frac{\partial S}{\partial \beta} = -2X'(Y - X\beta) = 0 \implies X'X\beta = X'Y \implies \beta_{OLS} = (X'X)^{-1}X'Y. \quad (2)$$

The Hessian $2X'X$ is positive definite under (A3), so the solution is the unique minimiser. Substituting (1) gives $\hat{\beta}_{OLS} = \beta + (X'X)^{-1}X'\varepsilon$, so that under (A1) $E(\hat{\beta}_{OLS} \mid X) = \beta$ and under (A2)

$$Var(\beta_{OLS} \mid X) = \sigma^2(X'X)^{-1}, \quad (3)$$

estimated by $\hat{\sigma}^2(X'X)^{-1}$ with $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n-p}$.

The Gauss–Markov theorem states that for any other linear unbiased estimator $\tilde{\beta} = AY$, with $AX = I$, the matrix $Var(\tilde{\beta}) - Var(\beta_{OLS})$ is positive semidefinite; the proof writes $A = (X'X)^{-1}X' + D$ with $DX = 0$ and observes that $Var(\tilde{\beta}) = \sigma^2(X'X)^{-1} + \sigma^2 DD'$. Under (A4), $\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)}$ is exactly normal and the statistics $\hat{\beta}_{OLS}$ and the F-ratio for a linear restriction have exact t and F distributions with $n - p$ denominator degrees of freedom.

3.3 Consequences of assumption failure

If (A2) fails because $Var(\varepsilon \mid X) = \sigma^2\Omega$ with $\Omega \neq I_n$, the estimator (2) remains unbiased but its covariance becomes $\sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$, which (3) does not estimate consistently; t- and F-statistics no longer have their nominal distributions, and GLS is more efficient. If (A1) fails because some observations are drawn from a contaminating distribution, the OLS estimator is biased: the influence of an observation on $\hat{\beta}_{OLS}$ is proportional to its residual and to its leverage, and neither is bounded. If (A4) fails but (A1)–(A3) hold, OLS retains its Gauss–Markov optimality among linear estimators, but nonlinear estimators can be substantially more efficient when the error distribution has heavy tails, and finite-sample inference is only approximate. If (A3) nearly fails, the smallest eigenvalue of $X'X$ is close to zero, the diagonal elements of $(X'X)^{-1}$ are large and the coefficients are estimated with low precision.

3.4 Violation scenarios

The procedure developed below is evaluated under seven scenarios that isolate each departure and then combine them.

- **Scenario A:** normal, homoscedastic, independent errors (all assumptions hold).
- **Scenario B:** heteroscedastic errors, $Var(\varepsilon_i) = \sigma^2 \exp(\gamma z_i)$ for a regressor or index z_i.
- **Scenario C:** first-order autoregressive errors, $\varepsilon_t = \varphi\varepsilon_{t-1} + \eta_t$, $|\varphi| < 1$.

- **Scenario D:** heavy-tailed or skewed errors, for example Student t with few degrees of freedom.
- **Scenario E:** a proportion π of observations replaced by vertical outliers or bad leverage points.
- **Scenario F:** near-collinear design with high variance inflation factors.
- **Scenario G:** heteroscedasticity, autocorrelation, heavy tails and contamination occurring together.

3.5 Research gap

Conventional OLS inference is valid only in Scenario A. Heteroscedasticity-consistent and autocorrelation-consistent covariance estimators restore valid inference in Scenarios B and C provided the point estimator is sound, but their sandwich is computed from OLS residuals and so inherits the bias of OLS in Scenario E. Robust M- and MM-estimators restore a sound point estimator in Scenarios D and E, but the covariance formulae with which they are conventionally reported assume spherical errors and are therefore incorrect in Scenarios B and C. Feasible GLS requires a correctly specified covariance model and is not robust. The residual bootstrap is inconsistent under heteroscedasticity, and the pairs and wild bootstraps, when applied to OLS, reproduce its sensitivity to contamination. No widely used procedure therefore remains valid in Scenario G. The specific problem addressed here is to provide Wald-type inference on Cβ that is simultaneously robust to contamination, consistent under heteroscedasticity and weak dependence, and well calibrated in finite samples, without requiring the analyst to model the variance structure.

4. Methodology: Robust Adaptive Linear Inference

The procedure has three stages: a robust point estimator, a robust sandwich covariance estimator computed from the same estimating equations, and a resampling calibration of the test statistic. The stages are not independent devices placed side by side; the second is the covariance implied by the first, and the third resamples from the residuals and weights of the first.

4.1 Stage 1: MM-estimation of the coefficients

Let ρ_c denote Tukey's biweight function,

$$\rho_c(u) = 1 - \left(1 - \left(\frac{u}{c}\right)^2\right)^3 \quad for\ |u| \le c, \quad \rho_c(u) = 1 \quad for\ |u| > c, \qquad (4)$$

with derivative ψ_c = ρ_c′. Following Yohai (1987), the estimator is constructed in three steps.

Step 1. Obtain an S-estimate β_S and residual scale $\hat{s}$ by minimising $\hat{s}$ subject to $\frac{1}{n}\sum_{i=1}^{n} \rho_{c_0}\left(\frac{y_i - x_i'\beta}{\hat{s}}\right) = \delta$, where $\delta = E_\Phi[\rho_{c_0}]$ and $c_0 = 1.547$, so that the breakdown point of $(\beta_S, \hat{s})$ is one half.

Step 2. Fix the scale ŝ.

Step 3. Define β̂ _MM as the solution of the estimating equation

$$\sum_{i=1}^{n} \psi_{c_1}\left(\frac{y_i - x_i'\beta}{\hat{s}}\right) x_i = 0 \tag{5}$$

with $c_1 = 4.685$, obtained by iteration from β_S. Y ohai (1987) p roved t hat t his c hoice retains the breakdown point of the initial estimate while attaining 95 per cent asymptotic efficiency under normal errors.

Writing $r_i = \frac{y_i - x_i'\beta_{MM}}{\hat{s}}$ and $w_i = \frac{\psi_{c_1}(r_i)}{r_i}$, normal equations $X'W(Y - X\beta_{MM}) = 0$ $W = diag(w_1, \ldots, w_n)$, so that equation (5) is equivalent to the weighted

with

$$\beta_{MM} = (X'WX)^{-1}X'WY \tag{6}$$

where W depends on β̂ _MM and the fixed point is found by iteratively reweighted least squares. Observations with $|r_i| > c_1$ receive weight zero and have no influence on the estimate. When the design is ill-conditioned, indicated by a maximum variance inflation factor above ten or a large condition index, X′WX in (6) is replaced by X′WX + λI with λ chosen by cross-validation on the robust residuals; this ridge-augmented variant is used only in Scenario F and is not part of the default procedure.

4.2 Stage 2: leverage-adjusted robust sandwich covariance

Equation (5) defines an M-estimator with score function $g(\beta) = \Sigma_i \psi(r_i)x_i$, where $\psi = \psi_{c_1}$ and ŝ is treated as fixed. A first-order expansion of g about the true β gives the sandwich representation

$$Var(\beta_{MM}) \approx B^{-1}MB^{-1}, \quad B = \frac{1}{n}\sum_{i=1}^{n} \psi'(r_i)x_i x_i', \quad M = \frac{1}{n}Var(\sum_{i=1}^{n} \psi(r_i)x_i). \tag{7}$$

The bread B is estimated by its sample analogue $\hat{B} = n^{-1}\Sigma_i \psi'(\hat{r}_i)x_i x_i'$ with $\hat{r}_i$ the standardised MM-residuals. The meat M is where the treatment of non-spherical errors enters. Define the robust score contributions $\hat{g}_i = \hat{s}\ \psi(\hat{r}_i)\ x_i$ and the leverage $h_i = x_i'(X'WX)^{-1}x_i$ from the weighted hat matrix. Under heteroscedasticity alone, M is estimated by

$$M_{HC} = \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{g}_i \hat{g}_i'}{(1 - h_i)^{\delta_i}}, \quad \delta_i = min\left\{4, \frac{nh_i}{p}\right\}, \tag{8}$$

which transfers the leverage-dependent discount of Cribari-Neto (2004) from OLS residuals to robust scores. The discount counteracts the downward bias that highleverage observations induce in the squared score contributions, a bias that Cribari-Neto and Zarkos (2001) identified as the dominant source of finite-sample distortion. Because ψ is bounded, $\hat{g}_i$ is bounded in the residual direction, and because the discount grows with h_i, the

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

contribution of a bad leverage point is limited in the design direction as well. When serial dependence is present, M is estimated by a Bartlett-kernel autocorrelation-consistent form in the sense of Newey and West (1987),

$$M_{HAC} = \Gamma_0 + \sum_{\ell=1}^{L}\left(1 - \frac{\ell}{L+1}\right)(\Gamma_\ell + \Gamma_\ell'), \quad \Gamma_\ell = \frac{1}{n}\sum_{t=\ell+1}^{n} \hat{g}_t \hat{g}_{t-\ell}', \tag{9}$$

with the truncation lag L chosen by the rule of Andrews (1991). The RALI covariance estimator is

$$V_{RALI} = \frac{1}{n} B^{-1} M B^{-1}, \quad M \in \{M_{HC}, M_{HAC}\}, \tag{10}$$

with M̂ _HC used by default and M̂ _HAC when the Durbin–Watson or Breusch–Godfrey diagnostics computed on the robust residuals indicate serial correlation. The estimator is p × p; B̂ and M̂ are both p × p, so the product is well defined.

Comparison with existing covariance estimators clarifies the role of each component. The OLS estimator $\hat{\sigma}^2(X'X)^{-1}$ is valid only under (A2). The White estimator and its HC1–HC4 refinements are consistent under heteroscedasticity but use OLS residuals and the unweighted hat matrix, so both the bread and the meat are contaminated in Scenario E. The GLS covariance $\sigma^2(X'\Omega^{-1}X)^{-1}$ is valid only when Ω is correctly specified. The Newey–West estimator is consistent under heteroscedasticity and weak dependence but again relies on OLS residuals. The classical MM covariance, proportional to $(X'X)^{-1}$ scaled by the ratio $E[\psi^2]/(E[\psi'])^2$, is valid only under spherical errors. Estimator (10) combines the robust bread of the MMestimator with a meat that is both robust and heteroscedasticity- or autocorrelationconsistent.

4.3 Stage 3: hypothesis testing

Consider the general linear hypothesis H_0: $C\beta = r$ against H_1: $C\beta \neq r$, where C is a q × p matrix of rank q and r is q × 1. Tests of a single coefficient (q = 1) and joint tests of several coefficients are special cases. The robust Wald statistic is

$$W_{RALI} = (C\beta_{MM} - r)'[CV_{RALI}C']^{-1}(C\beta_{MM} - r). \tag{11}$$

Under H_0 and the regularity conditions of Section 4.4, W_{RALI} converges in distribution to a chi-squared variable with q degrees of freedom. For q = 1 the corresponding t-type statistic is $t_{RALI} = \frac{c'\beta_{MM} - r}{(c'V_{RALI}c)^{1/2}}$, asymptotically standard normal; in small samples a t reference distribution with n − p degrees of freedom, or with the adjusted degrees of freedom of Imbens and Kolesár (2016), is preferable. An F-type version divides (11) by q and refers it to the F distribution with q and n − p degrees of freedom.

Asymptotic critical values under-reject or over-reject in small samples, and the direction of the error depends on the leverage structure. The statistic is therefore calibrated by a restricted wild bootstrap, following the principle established by Davidson and

Flachaire (2008) that imposing the null hypothesis on the resampling scheme yields the most accurate rejection probabilities. The steps are:

1. Compute the restricted estimate $\tilde{\beta} = \beta_{MM} - VC'(CVC')^{-1}(C\beta_{MM} - r)$, the projection of β_{MM} onto the null set, and the restricted residuals $\tilde{u}_i = y_i - x_i'\tilde{\beta}$. Transform the residuals as $\tilde{u}_i^{\dagger} = \frac{\tilde{u}_i}{(1-h_i)^{1/2}}$.
2. For b = 1, …, B, draw independent Rademacher variables $v_i^{(b)}$ taking the values −1 and +1 with equal probability, and form the bootstrap responses $y_i^{(b)} = x_i'\tilde{\beta} + v_i^{(b)}\,\tilde{u}_i^{\dagger}$.
3. Refit the unrestricted MM-estimator and the covariance (10) on the bootstrap sample and compute $W^{(b)}$ for the hypothesis $C\beta = C\tilde{\beta}$.
4. Compute the bootstrap p-value $p^* = \frac{1 + \#\{b : W^{(b)} \geq W_{RALI}\}}{1+B}$ and reject H_0 at level α if $p^* \leq \alpha$.

The Rademacher distribution preserves the second moment of each residual and, because the null is imposed, the bootstrap datagenerating process satisfies H_0 exactly. Multiplying each residual by an independent sign preserves the observed heteroscedasticity, and refitting the MMestimator on each bootstrap sample preserves the down-weighting of outlying observations. A confidence set for Cβ at level $1-\alpha$ is obtained by inverting the test, and a confidence interval for a single coefficient is $c'\hat{\beta}_{MM} \pm q^*_{1-\alpha}(c'\hat{V}_{RALI}\,c)^{1/2}$, where q^* is the bootstrap quantile of $|t_{RALI}|$.

4.4 Theoretical properties The following properties are inherited from established results on the component methods and hold under the stated conditions; they are not reproved here.

Breakdown and efficiency. Because Stage 2 and Stage 3 do not alter the point estimator, $\hat{\beta}_{MM}$ retains the finite-sample breakdown point of the initial S-estimate, which approaches one half for $c_0 = 1.547$, and the asymptotic Gaussian efficiency of 95 per cent for $c_1 = 4.685$ (Yohai, 1987).

Consistency and asymptotic normality. Assume (R1) $E[\psi(\varepsilon/s_0)] = 0$ at the true coefficient and the limiting scale s_0, which holds when the error distribution is symmetric about zero; (R2) $n^{-1}X'X$ converges to a positive definite matrix; (R3) the maximum leverage $\max h_i$ tends to zero; (R4) ψ is bounded and continuously differentiable; and (R5) the errors are either independent or strongly mixing with summable mixing coefficients. Under (R1)–(R5), $\sqrt{n}(\hat{\beta}_{MM} - \beta)$ converges in distribution to $N(0, B^{-1}MB^{-1})$ (Yohai, 1987; Maronna, Martin and Yohai, 2006). Under (R1)–(R5) with independent errors, $\hat{M}_{HC}$ is a consistent estimator of M by the

argument of White (1980); the leverage discount in (8) is a finite-sample correction that vanishes asymptotically under (R3). Under (R5) with $L \to \infty$ and $L/n \to 0$, $\hat{M}$ _HAC is consistent by the results of Newey and West (1987) and Andrews (1991). Consequently W_RALI converges to a chisquared distribution with q degrees of freedom under H_0.

Bootstrap validity. For an asymptotically pivotal statistic such as W_RALI, the wild bootstrap is asymptotically valid under heteroscedasticity of unknown form (Mammen, 1993), and imposing the null provides an asymptotic refinement (Davidson and Flachaire, 2008). SalibianBarrera and Zamar (2002) established the validity of bootstrapping MM-estimators.

Influence. The influence function of β_{MM} at a point (x, y) is proportional to $\psi\left(\frac{y - x'\beta}{s}\right) B^{-1}x$. S ince ψ is bounded and redescending, the influence is bounded in the residual direction and vanishes for gross residuals; the discount in (8) further limits the contribution of high-leverage observations to the covariance estimate.

Properties requiring computational verification. Finite-sample size control of the bootstrap-calibrated test, coverage of the resulting confidence intervals under Scenarios B to G, the magnitude of the efficiency loss relative to OLS under Scenario A, and the power of the test under H_1 are not established by the results above and must be examined by simulation. The framework for doing so is specified in Section 5. The present paper makes no numerical claims on these points.

5. Main Research: Monte Carlo Evaluation Framework

This section specifies the design under which the procedure is to be compared with its

5.1 Design

Table 1. Simulation Design

competitors. It is presented as a framework: the design, the data-generating processes, the benchmark methods and the performance measures are fixed so that any investigator can execute the comparison and obtain results that are directly comparable across studies.

Factor	Levels
Sample size n	50, 100, 250, 500
Number of slope coefficients p	3, 5, 10
Heteroscedasticity (Scenario B)	$\gamma = 0.5$ (low), 1.0 (moderate), 2.0 (severe)
Autocorrelation (Scenario C)	$\varphi = 0.3$ (weak), 0.6 (moderate), 0.9 (strong)
Error distribution (Scenario D)	N(0,1); Student t with 3 and 5 df; skewed (chisquared, centred)
Contamination (Scenario E)	$\pi = 0.05, 0.10, 0.20$; vertical outliers and bad leverage points
Collinearity (Scenario F)	Toeplitz correlation $\rho^{\wedge}\lvert j-k\rvert$ with ρ chosen to give maximum VIF $\approx 5, 10, 30$
Simultaneous (Scenario G)	$\gamma = 1$, $\varphi = 0.6$, t_3 errors, $\pi = 0.10$ combined
Monte Carlo replications	10,000 per cell (5,000 where the bootstrap makes this infeasible)
Bootstrap replications	B = 999
Nominal level	$\alpha = 0.05$; 95 per cent confidence intervals

In every cell the true coefficient vector is fixed at $\beta = (1, 0.5, \ldots, 0.5)'$ and the regressors are

drawn from a multivariate normal distribution with the stated correlation structure. For power curves, the first slope coefficient is set to a grid of values between 0 and 1 while the remaining coefficients are held at their null values.

5.2 Data-generating processes

Scenario A draws ε from $N(0, I_n)$. Scenario B sets $\varepsilon_i = \eta_i \exp(\gamma x_\{i1\}/2)$ with η_i standard normal, so that the variance is $\exp(\gamma x_\{i1\})$. Scenario C generates ε as a stationary AR(1) process with innovation variance chosen so that the marginal variance is one. Scenario D draws ε from the stated distribution, scaled to unit variance where the variance exists. Scenario E first generates data as in Scenario A and then replaces a fraction π of the responses by $y_i + 10\sigma_y$ (vertical outliers) or, in a second variant, replaces both the response and the first regressor of a fraction π of observations by values placed far from the bulk of the data (bad leverage points). Scenario F alters only the design correlation. Scenario G applies the heteroscedastic scaling of B to an AR(1) process of C with t_3 innovations and then introduces contamination as in E.

5.3 Benchmark methods

Seven procedures are compared. (i) OLS with the classical covariance (3). (ii) Weighted least squares with weights proportional to $\exp(-\hat{\gamma}\, x_\{i1\})$, where $\hat{\gamma}$ is estimated by regressing log squared OLS residuals on $x_\{i1\}$, in Scenario B; feasible GLS with an estimated AR(1) transformation in Scenario C. (iii) OLS with HC3 and HC4 covariances. (iv) OLS with the Newey–West covariance and the Andrews bandwidth. (v) The MMestimator with its classical covariance. (vi) OLS with the pairs bootstrap and the wild bootstrap. (vii) RALI as specified in Section 4. Methods (ii) and (iv) are applied only in the scenarios for which they are designed; the others are applied throughout.

5.4 Performance measures For each method, cell and coefficient the following quantities are recorded across replications: bias, defined as the mean estimate minus the true value; relative bias, the bias divided by the true value; the Monte Carlo variance; mean squared error, the sum of squared bias and variance; root mean squared error; the empirical coverage of the nominal 95 per cent confidence interval; the average interval length; the empirical rejection frequency of a true null hypothesis on the first slope (Type I error); the rejection frequency along the power grid; the outofsample root mean squared prediction error on an independent test sample of the same size; and the mean computing time per fit. Coverage and Type I error are reported with Monte Carlo standard errors, which for a nominal 0.05 level and 10,000 replications are approximately 0.002.

5.5 Expected behaviour and its evidential status

Prior published simulations supply expectations against which populated results may be checked, and the following statements are attributed to those studies rather than to the present paper. Long and Ervin (2000) found that HC0 over-rejects for $n \leq 250$ and that HC3 controls size well for n as small as 25 under pure heteroscedasticity. Cribari-Neto (2004) found that HC4 improves on HC3 when high-leverage points are present. Davidson and Flachaire (2008) found that the null-imposed Rademacher wild bootstrap gives smaller size distortion than asymptotic HC-based tests. Yohai (1987) and Maronna, Martin and Yohai (2006) document that MM-estimators retain small bias under contamination up to the breakdown point while OLS bias grows with the contamination fraction. On the basis of Section 4.4 the following are expected of RALI, and are to be verified by executing the framework: a small efficiency loss relative to OLS in Scenario A, of the order implied by the 95 per cent tuning; bias comparable to the MM-estimator in Scenarios D and E; size and coverage close to nominal in Scenarios B and C, where the MM-estimator with its classical covariance is expected to fail; and, in Scenario G, size and coverage closer to nominal than any of the single-purpose competitors. Any departure from these expectations would indicate either a failure of a regularity condition, most plausibly (R3) when p is large relative to n, or a limitation of the bandwidth or leveragediscount rules.

5.6 Sensitivity Analysis Conclusions are assessed for stability by varying one design factor at a time from the Scenario G baseline: sample size, number of regressors, contamination fraction, error distribution, heteroscedasticity strength, autocorrelation strength and collinearity level. Two sensitivities merit particular attention. As p/n grows, the boundedleverage condition weakens and the results of Cattaneo, Jansson and Newey (2018) imply that all sandwich estimators, including (10), become biased; the framework records the ratio p/n at which coverage departs from nominal. As φ approaches 0.9, the choice of truncation lag becomes decisive, and the fixed-bandwidth reference distributions of Kiefer and Vogelsang (2005) provide an alternative that the framework accommodates as an additional benchmark.

6. Conclusion

The Gauss–Markov theorem guarantees the optimality of ordinary least squares only when the errors are spherical, the model is correctly specified and the design has full rank, and exact inference additionally requires normality. Applied data violate these conditions frequently and, more importantly, simultaneously. The methodological literature offers a valid covariance estimator for non-spherical errors, a valid coefficient

estimator for contaminated data and a valid resampling scheme for heteroscedastic errors, but these devices have been developed and evaluated separately, and each inherits the weaknesses of the estimator on which it is built.

This paper has proposed Robust Adaptive Linear Inference, a procedure in which an MM-estimator supplies the coefficients, a leverage-adjusted sandwich estimator computed from the same robust estimating equations supplies a covariance matrix that is consistent under heteroscedasticity and weak serial dependence, and a null-imposed wild bootstrap supplies the reference distribution for Wald-type tests of general linear hypotheses. The construction is internally consistent in that the bread and the meat of the sandwich derive from one robust criterion, and the resampling scheme preserves both the heteroscedastic structure of the residuals and the down-weighting of outlying observations. The procedure inherits proved breakdown, efficiency, consistency and asymptotic-normality properties from its components, and its finite-sample behaviour is to be assessed by the Monte Carlo framework specified in Section 5.

The procedure is not proposed as a universal replacement for least squares. When diagnostics indicate that the classical assumptions hold, OLS is more efficient and its exact finite-sample theory is preferable. When a single violation is present and its form is known, the targeted remedy—weighted least squares under a known variance function, generalized least squares under a known correlation structure, HC3 or HC4 under heteroscedasticity alone—may be adequate and is computationally cheaper. The advantage of the integrated procedure is expected precisely in the setting where several assumptions fail together and the analyst cannot, or does not wish to, model the failures parametrically. Its principal limitations are the assumption of vanishing maximal leverage, which restricts it to settings where the number of regressors is small relative to the sample size; the restriction to weak dependence in the autocorrelation-consistent component; the loss of reliability of any equivariant estimator once contamination exceeds the breakdown point; and the computational cost of the bootstrap. Extensions to clustered and panel data, to generalized linear models and to high-dimensional designs with valid postselection inference are natural directions for further work.

References

Aitken, A. C. (1935). On least squares and linear combination of observations. Proceedings of the Royal Society of Edinburgh, 55, 42–48.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica, 59(3), 817–858.

Andrews, D. W. K., & Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica, 60(4), 953–966.

Atkinson, A. C., Riani, M., & Torti, F. (2016). Robust methods for heteroskedastic regression. Computational Statistics & Data Analysis, 104, 209–222. Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: Wiley.

Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. Australian Economic Papers, 17(31), 334–355.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica, 47(5), 1287–1294.

Cattaneo, M. D., Jansson, M., & Newey, W. K. (2018). Inference in linear regression models with many covariates and heteroscedasticity. Journal of the American Statistical Association, 113(523), 1350–1361.

Cook, R. D. (1977). Detection of influential observation in linear regression. Technometrics, 19(1), 15–18.

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

Cribari-Neto, F., & da Silva, W. B. (2011). A new heteroskedasticity-consistent covariance matrix estimator for the linear regression model. AStA Advances in Statistical Analysis, 95(2), 129–146.

Cribari-Neto, F., Souza, T. C., & Vasconcellos, K. L. P. (2007). Inference under heteroskedasticity and leveraged data. Communications in Statistics – Theory and Methods, 36(10), 1877– 1888.

Cribari-Neto, F., & Zarkos, S. G. (2001). Heteroskedasticity-consistent covariance matrix estimation: White's estimator and the bootstrap. Journal of Statistical Computation and Simulation, 68(4), 391–411.

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

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression: I. Biometrika, 37(3–4), 409–428.

Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression: II. Biometrika, 38(1–2), 159–178.

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

Flachaire, E. (2005). Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap. Computational Statistics & Data Analysis, 49(2), 361–376.

Freedman, D., & Lane, D. (1983). A nonstochastic interpretation of reported significance levels. Journal

of Business & Economic Statistics, 1(4), 292–298.

Godfrey, L. G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. Econometrica, 46(6), 1293–1301.

Godfrey, L. G. (2006). Tests for regression models with heteroskedasticity of unknown form. Computational Statistics & Data Analysis, 50(10), 2715–2733.

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

Hausman, J., & Palmer, C. (2012). Heteroskedasticity-robust inference in finite samples. Economics Letters, 116(2), 232–235.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. Technometrics, 12(1), 55–67.

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

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

Imbens, G. W., & Kolesár, M. (2016). Robust standard errors in small samples: some practical advice. The Review of Economics and Statistics, 98(4), 701–712.

Kiefer, N. M., & Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. Econometric Theory, 21(6), 1130–1164.

Koenker, R., & Bassett, G. (1978). Regression quantiles. Econometrica, 46(1), 33–50.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. The Annals of Statistics, 17(3), 1217–1241.

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.

MacKinnon, J. G. (2013). Thirty years of heteroskedasticity-robust inference. In X. Chen & N. R. Swanson (Eds.), Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis (pp. 437–461). New York: Springer.

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

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.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55(3), 703–708.

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. Journal of the American Statistical Association, 89(428), 1303–1313.

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., & Yohai, V. J. (1984). Robust regression by means of Sestimators. In J. Franke, W. Härdle & R. D. Martin (Eds.), Robust and Nonlinear Time Series Analysis (pp. 256–272). New York: Springer.

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

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

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.

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. Journal of Statistical Software, 11(10), 1–17.

Zeileis, A. (2006). Object-oriented computation of sandwich estimators. Journal of Statistical Software, 16(9), 1–16.