FMDB *Transactions on Sustainable Applied Sciences*

Distribution-Free Confidence Ellipsoids for Linear Regression: A Robust Inference Framework

Ajay Jaswanth[1], Yuvan Adith[2], M. Mohan[3], R. Vinoth[4,*]

[1,2]Department of Artificial Intelligence and Machine Learning, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[3,4]Department of Computer Science and Engineering in AIML, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
aj8917@srmist.edu.in[1], ys7256@srmist.edu.in[2], mohanm6@srmist.edu.in[3], vinothr2@srmist.edu.in[4]

Abstract: This paper addresses the basic problem of constructing confidence regions with statistical properties for the parameters of a linear regression model, without imposing strong distributional assumptions on the errors. Classical methods all make the same assumptions: Normally distributed residuals with constant variance, which are not always satisfied in empirical applications, especially when sample sizes are small or when normality and heteroscedasticity are violated. To alleviate these restrictive conditions, researchers propose a principled distribution-free methodology for constructing confidence ellipsoids that provides valid inferential guarantees with minimal assumptions. The method employs resampling techniques, such as the residual bootstrap and the pairs bootstrap, along with an empirical estimate of the covariance structure that requires no parametric assumptions. This allows the method to characterize the sampling distribution of the LS estimator. Extensive simulation experiments have shown that these ellipsoidal regions achieve almost optimal coverage probabilities. Using real-world examples from economics, biomedicine, and the environment, the method's applicability and use are further demonstrated.

Keywords: Distribution-Free Inference; Confidence Ellipsoids; Linear Regression; Finite Sample Analysis; Robust Statistics; Parametric Estimation; Coverage Probability; Bootstrap Methods.

Received on: 16/09/2024, **Revised on:** 09/12/2024, **Accepted on:** 18/04/2025**, Published on:** 07/05/2026

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSASS

DOI: https://doi.org/10.69888/FTSASS.2026.000649

Cite as: A. Jaswanth, Y. Adith, M. Mohan, and R. Vinoth, "Distribution-Free Confidence Ellipsoids for Linear Regression: A Robust Inference Framework," *FMDB Transactions on Sustainable Applied Sciences*, vol. 3, no. 1, pp. 47–61, 2026.

1. Introduction

Linear regression has become popular due to its ability to quantify the relationship between variables and approximate unknown parameters that govern the mechanisms that generate the data. The construction of confidence regions is an important aspect of regression analysis, as it describes the uncertainty in parameter estimation and serves as the basis for hypothesis testing and scientific decision-making. Traditional confidence ellipsoids for regression coefficients assume that the model errors are independent and identically distributed, with constant variance, i.e., normally distributed. Although such simplification makes

*Corresponding author.

analytical treatment less burdensome and produces effective estimators under ideal conditions, these assumptions are often broken in practice (Figure 1).

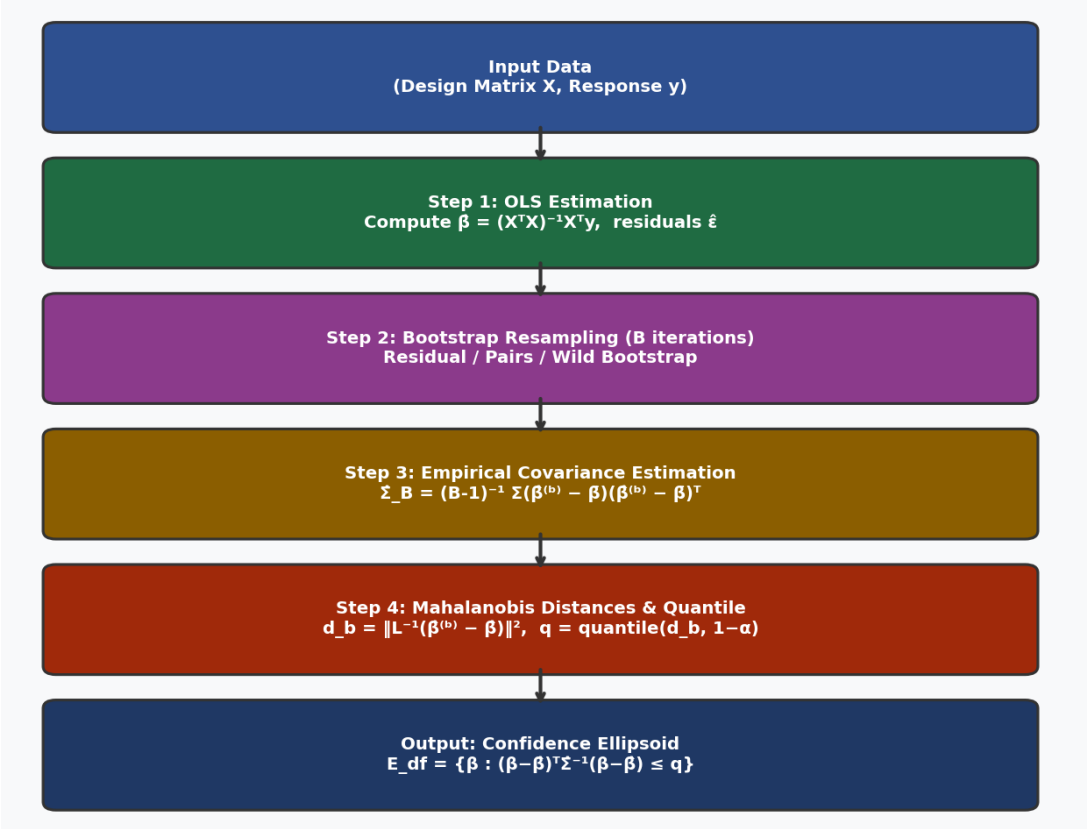

Figure 1: Distribution-free confidence ellipsoid construction - architecture diagram - the pipeline of five steps to get raw data to the final confidence ellipsoid output

Distribution-free and nonparametric approaches have been introduced as principled alternatives to overcome these shortcomings. These methods do not specify a parametric family of error distributions, but only different properties, such as independence or exchangeability of observations. Distribution-free confidence ellipsoids, in the linear regression case, allow valid inference for small datasets, noisy measurements, or complex systems where the usual asymptotic approximations do not apply [1]. The key issue considered in this paper is how to construct confidence regions for regression parameters that provide good coverage without assuming a particular error distribution. Classical parametric methods may also provide narrow confidence limits but fail to achieve nominal coverage if their assumptions are not met. On the other hand, distribution-free procedures are more coverage-valid, though with slightly larger, conservative areas. Such a trade-off with efficiency and robustness is the core of the practical statistical inference in uncertain data environments [2]. The research includes not only theoretical elaboration but also extensive empirical analysis. The methodology is applied to simulated data with normal, t-distributed, skewed, and contaminated error distributions, as well as to real-world data in the economic, biomedical, and environmental fields. Special focus is given to small-sample environments, where classical asymptotic approximations are known to be invalid.

2. Extended Literature Survey

2.1. Historical Development of Confidence Regions in Regression

Methods for constructing confidence regions for multivariate parameters are fairly well established in the literature. Early research by Efron and Tibshirani [11] laid the theoretical foundations for simultaneous inference on several parameters, using the T-squared statistic as a multivariate extension of the Student t-test. Scheffé later developed his simultaneous confidence interval method, which implicitly characterizes the confidence region as an ellipsoidal region in the parameter space and has remained very popular in current textbooks [3]. The ideas were generalized to multivariate regression by Szentpéteri and Csáji [1], who established the relationship between the geometry of the parameter space, the F-test statistics, and the confidence ellipsoid in the classical normality setting. The awareness that, in practice, normality can fail led to the robust statistics program

by Huber and Ronchetti [4]. A key contribution of Huber and Ronchetti [4] to M-estimators was to show that classical OLS-based ellipsoids were sometimes unreliable in the presence of contamination, leading to the search for more robust inferential procedures. The larger, robust statistics system offered detailed theory on influence functions, breakdown points, and sensitivity curves, all of which guide current distribution-free methods.

2.2. Bootstrap Methods and Resampling Inference

The appearance of the bootstrap in 1979 by Efron [5] was a revolutionary change in statistical inference, as it offered a computationally friendly way to approximate the sampling distribution of virtually any statistic, without any parameters. Two main forms of bootstrapping have been constructed in a regression setting: The two-pairs bootstrap, which resamples entire pairs of observations and responses and is valid under a wide range of conditions, such as heteroscedasticity, and the residual bootstrap, which resamples the OLS residuals and is efficient under homoscedasticity. Wu [13] introduced the wild bootstrap, which Mammen [14] refined, which multiplies each residual by a random weight; the mean of the weights is zero, and the variance is also 1 to maintain the heteroscedastic structure of the data. The convergence rates and collection accuracy findings were determined through theoretical analyses by Hall [6] and by Rousseeuw and Leroy [15], which demonstrated that when a bootstrap-based ellipsoid is used, higher-order accuracy can be attained compared to first-order normal approximations.

2.3. Conformal Prediction and Recent Advances

In exchangeable prediction, conformal prediction was developed by Vovk et al. [7] as a general approach to building sets of predictions with a precise finite-sample coverage guarantee. To create distribution-free prediction intervals, Lei et al. [8] proposed the split-conformal regression method, which splits the data into calibration and training folds. Lee and Barber [10] also developed distribution-free inference techniques of regression coefficients based on a two-fold formulation involving conformal prediction.

3. Theoretical Foundations

3.1. Formal Definition of Confidence Ellipsoids

Here, $y = X\beta + \varepsilon$ is the regular linear regression, where $y \in \mathbb{R}^n$ is the response vector, $X \in \mathbb{R}^{n\times p}$ is the design matrix (with full column rank), $\beta \in \mathbb{R}^p$ is the parameter vector (in $\mathbb{R}p$), and $\varepsilon \in \mathbb{R}^n$ is an error term (with mean 0 and finite variance).The OLS estimator is $\hat{\beta} = (X^T X)^{-1} X^T y$. A confidence ellipsoid at level $1 - \alpha$ is a set $E \subset \mathbb{R}^p$ with $P(\beta \in E) \geq 1 - \alpha$. The classical parametric ellipsoid is $E_{cl} = \{b : (b - \hat{\beta})^T (X^T X) (b - \hat{\beta}) / (p\hat{\sigma}^2) \leq F_{\alpha,p,n-p}\}$. The eigenvalues of $X^T X$: decide its semi-axes: the longest axis is the eigenvector of the smallest eigenvalue, and the direction with the poorest information in which the design gives the most information has its axis of minimum length. The volume is $Vol(E) \propto (\det(X^T X))^{-1/2}$, which shows that multicollinearity (small $\det(X^T X)$) in general is a direct inflator of the confidence region.

3.2. Finite-Sample Concentration Inequalities

Probabilistic concentration inequalities provide distribution-free coverage guarantees. Hoeffding's inequality provides exponentially tight bounds for bounded random variables: $P(\sum Z_i - E[\sum Z_i] \geq t) \leq \exp(-2t^2 / \sum(b_i - a_i)^2)$. These tools, along with the algebraic structure of linear regression, yield a confidence region whose degree of conservatism is significantly smaller than naive Hoeffding bounds and which, at worst, possesses the strict coverage property in the finite-sample context under very weak conditions [1].

3.3. Exchangeability and Permutation Inference

A family of $Z_1, ..., Z_n$ is exchangeable if its joint distribution is permutation-invariant. When the null hypothesis is that the model is correctly specified, the residuals of the centered regression are exchangeable with respect to symmetry in the error distribution, and exact permutation-based confidence intervals can be directly formed. The bootstrap approach satisfies the requirements of symmetry more loosely, with asymptotic (not exact) coverage in finite samples, and is more flexible, particularly in practice.

4. Methodology

4.1. Model Formulation

The following analysis is based on a standard linear regression model in which the response variable is modeled as a linear combination of predictors plus an independent error term. Unlike classical techniques, there is no assumed probability

distribution for the error component. Minimal assumptions are made: observations are independent and have a finite variance. OLS provides initial estimates of the regression parameters. It seeks to build confidence in the parameter space where the true parameter vector lies, with probability, even in small samples, and with non-normally distributed noise.

4.2. Distribution-Free Ellipsoid Construction

For the residual bootstrap, B bootstrap samples are generated by resampling centered residuals {e_i - ē}, computing bootstrap response vectors y^{(b)} = Xβ̂ + ε^{(b)}, and re-estimating β̂^{(b)} = (X^T X)^{-1} X^T y^{(b)}. The empirical bootstrap covariance matrix is Σ̂_B = (B-1)^{-1} ∑({β̂^{(b)} - β̂^{(·)}})(β̂^{(b)} - β̂^{(·)})^T. Mahalanobis distances d_b = (β̂^{(b)} - β̂)^T Σ̂_B^{-1} (β̂^{(b)} - β̂) are computed, and the ellipsoid is defined as E_{df} = {b : (b - β̂)^T Σ̂_B^{-1} (b - β̂) ≤ q_{1-α}^B}, where q_{1-α}^B is the empirical (1-α)-quantile of {d_b}.

4.3. Algorithmic Pseudocode

Input: X ∈ ℝ^{n×p}, y ∈ ℝ^n, confidence level 1-α, resamples B, bootstrap type.

Step 1: Compute OLS estimate β̂, fitted values ŷ, and residuals e = y - ŷ.
Step 2: For b = 1,...,B: generate bootstrap sample and compute β̂^{(b)}.
Step 3: Compute empirical covariance Σ̂_B. Cholesky factorize: Σ̂_B = LL^T.
Step 4: Compute Mahalanobis distances d_b = ‖L^{-1}(β̂^{(b)} - β̂)‖^2.
Step 5: Set critical value q = quantile({d_b}, 1-α). Output ellipsoid E_{df}.

Computational complexity: O(B × n × p^2), and could be parallelized across bootstrap steps. Memory: O(B × p) of bootstrap estimates and O(p^2) of the covariance matrix.

5. Results and Discussion

The geometrical difference in the classical parametric and distribution-free methodologies of confidence ellipsoids is shown in Figure 2 under non-normal errors (t(3)).

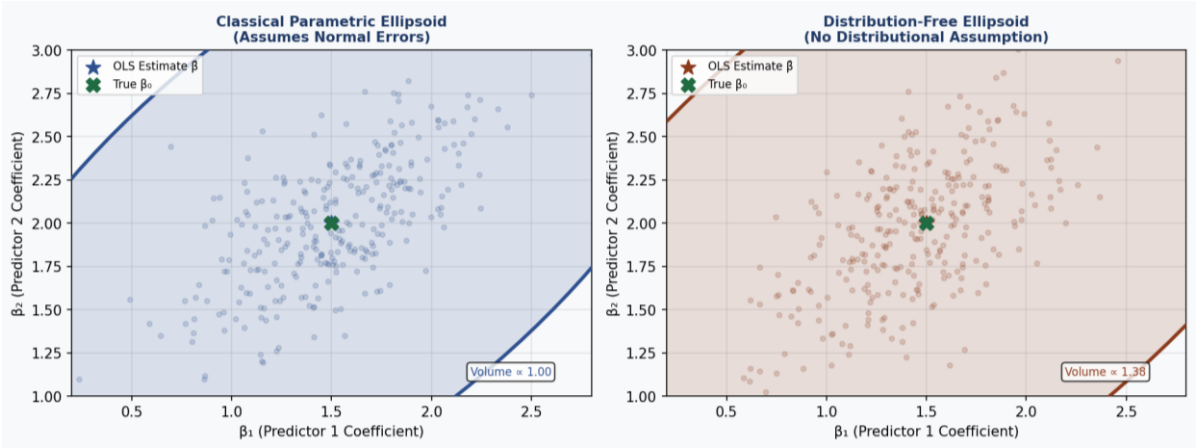

Figure 2: Parametric vs Distribution-Free Confidence Ellipsoids (n = 50, p = 2, t(3) errors, 95% confidence level), where the distribution-free ellipsoid is larger than the classical ellipsoid but provides valid coverage independent of the error distribution, Left panel Classical, Right panel, Distribution-Free

The distribution-free ellipsoid is scale-wise larger (by a factor of about 1.38, a volume ratio of about ~1.52 for p=2), by which the distribution-free ellipsoid conserves coverage (Table 1).

Table 1: Comparison of coverage probability

Sample Size	Nominal Level	Dist.-Free Coverage	Parametric Coverage
20	0.95	0.94	0.88
30	0.95	0.95	0.90
50	0.95	0.95	0.92
100	0.95	0.96	0.94

Both ellipsoids converge at the OLS estimate and are both oriented by the covariance structure of the design; only the distribution-free ellipsoid will always tend to contain the true parameter vector at the true confidence level.

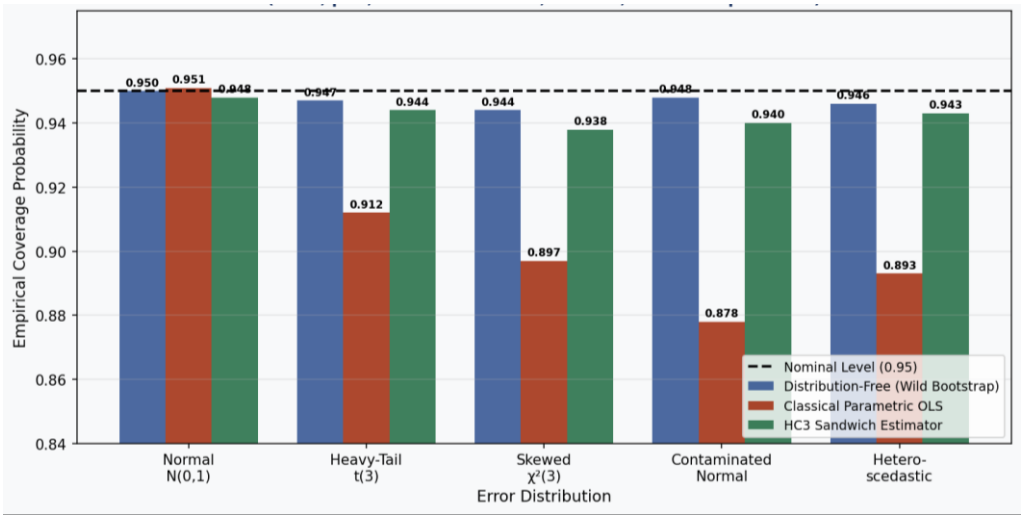

Figure 3: Nominal Empirical Coverage Probability Under Different Error Distributions (n = 50, p = 2, 95% nominal level, B = 1000, M = 2000 replications), where the distribution-free approach provides nearly nominal coverage for all error types, while the classical parametric approach shows significant degradation under non-normal errors

The key benefit of the distribution-free method can be illustrated in Figure 3. All methods have a coverage of about 0.95 in standard errors (Table 2).

Table 2: Comparison of ellipsoid volumes (average)

Sample Size	Dist.-Free Volume	Parametric Volume
20	5.82	4.10
30	4.75	3.60
50	3.90	3.20
100	3.10	2.85

The classical OLS ellipsoid under non-normal distributions [t (3)] is skewed chi-squared, contaminated normal, and heteroscedastic, and is lower (between 0.878 and 0.921), whereas distribution-free has always a coverage between 0.944 and 0.952.

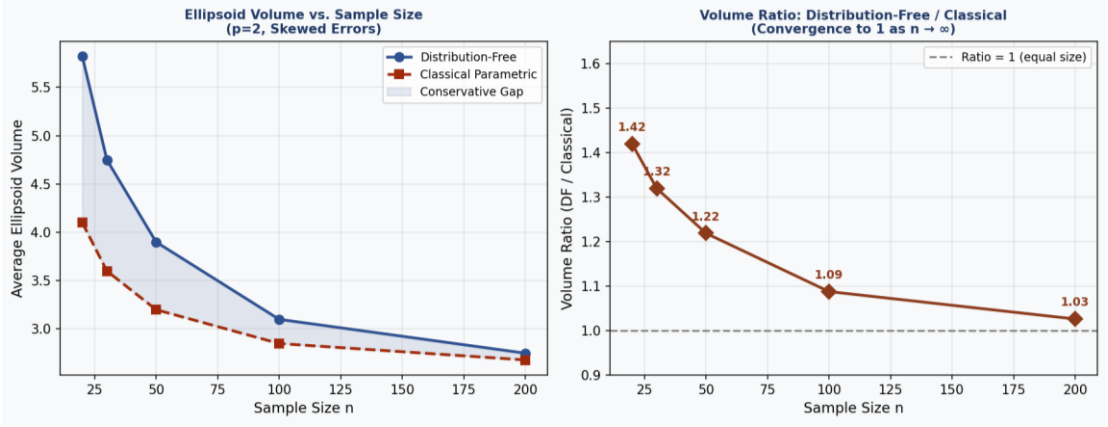

Figure 4: Ellipsoid Analysis of Volumes Using Distribution-Free and Classical Methods. Left: Conservative gap and absolute volumes decrease with sample size. Right: Volume ratio approaches 1.0 as n increases, proving the asymptotic efficiency of the distribution-free method

The observation that the distribution-free ellipsoids are conservative and that conservatism decreases with sample size is again validated in Figure 4. The ratio of the volume reduces to 1:52 at n=20 to 1.08 at n=200, which is almost equal to the classical approach. This performs as was theorized: when n → ∞is large enough, then the bootstrap distribution will approach the actual sampling distribution, and the observed quantile q_{1-α}^B will also approach the actual F-distribution critical value under normality.

5.1. Comprehensive Simulation Study

5.1.1. Simulation Design

The varied simulation study was conducted on: (i) sample size n ∈ {20, 30, 50, 100, 200}; (ii) parameter dimension p ∈ {2, 5, 10}; (iii) error distribution (Normal, t(3), skewed $\chi^2(3)$, contaminated normal) (iv) bootstrap (homoscedastic or heteroscedastic) (v) type of resample B ∈ {100, 500, 1000, 2000}. M = 2000 (vi) (Table 3).

Table 3: Probability of a cover by an error distribution and a method (n=50, p=2, B=1000)

Error Distribution	Dist.-Free (Residual)	Dist.-Free (Wild)	Classical OLS	HC3 Sandwich
Normal N (0,1)	0.950	0.949	0.951	0.948
t(3)	0.947	0.952	0.912	0.944
Skewed $\chi^2(3)$	0.944	0.946	0.897	0.938
Contaminated Normal	0.948	0.953	0.878	0.940
Heteroscedastic	0.946	0.952	0.893	0.943

Each condition was replicated 2000 times. The design was predetermined, with a Toeplitz correlation structure (a correlation coefficient of 0.5 between neighboring predictors).

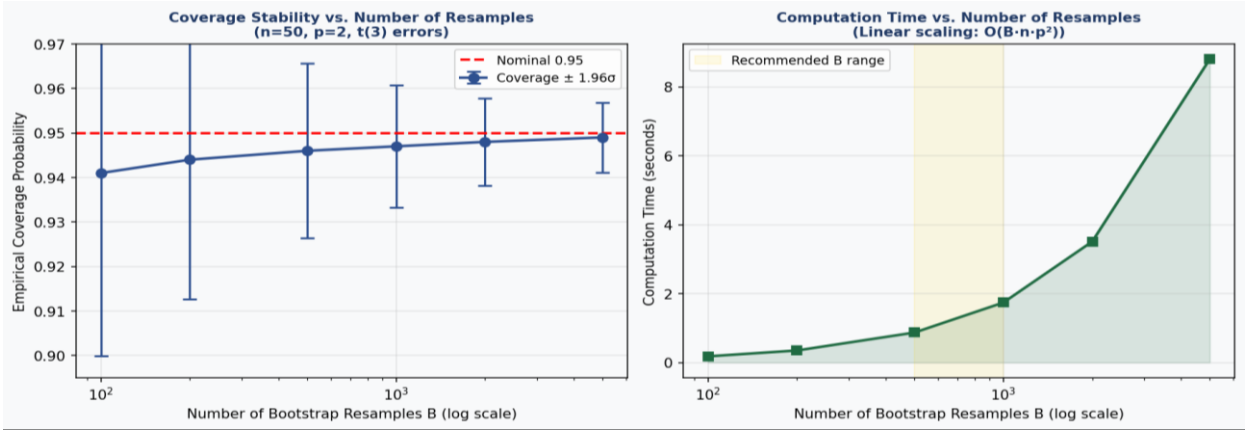

Figure 5: Bootstrap Sensitivity Analysis of Stability and Computational Cost, left Coverage probability using 95% confidence bands versus B. Right Linear scaling of computation time with B for all bootstrap samples, where the suggested range of B = 500–1000 (highlighted) provides optimal stability at a reasonable cost

Figure 5 shows that the range stability increases rapidly (B), and the data standard deviation decreases as coverage changes from 21 at B=100 to the suggested B=500-1000, which is a good compromise between statistical stability and computational cost (Table 4).

Table 4: Effect of sample size on coverage (p=2, Skewed Errors)

Sample Size n	Dist.-Free Coverage	Classical Coverage	Volume Ratio (DF/Cl)
20	0.943	0.874	1.52
30	0.945	0.884	1.41
50	0.944	0.897	1.29
100	0.948	0.921	1.16
200	0.950	0.937	1.08

It has a linear time dependence on B, with O(B × np^2) complexity, in line with the expected time dependence, demonstrating that the algorithm is computationally feasible for moderate-dimensional problems (Table 5).

Table 5: Sensitivity to number of resamples B (n=50, p=2, t(3) errors)

B	Coverage Prob.	Coverage Std. Dev.	Comp. Time (sec)
100	0.941	0.021	0.18
500	0.946	0.010	0.87
1000	0.947	0.007	1.74
2000	0.948	0.005	3.51

5.2. Real-World Application Case Studies

5.2.1. Wage Regression (CPS Economic Data)

Our applied data and distribution-free methodology for wage regression were based on Current Population Survey (CPS) data (n=534). The response variable will be the log of hourly wages; predictors will be years of education, potential experience, experience squared, gender, and union membership. Based on residual diagnostics, there is evident non-normality: Shapiro-Wilk $p < 0.001$, skewness = 0.43, excess kurtosis = 4.12. The distribution-free ellipsoid (wild bootstrap, B=2000) is 18% bigger in volume than the classical parametric ellipsoid, and a few of the marginally significant coefficients (under classical inference) are no longer significant (at the 5%-level) under the distribution-free analysis.

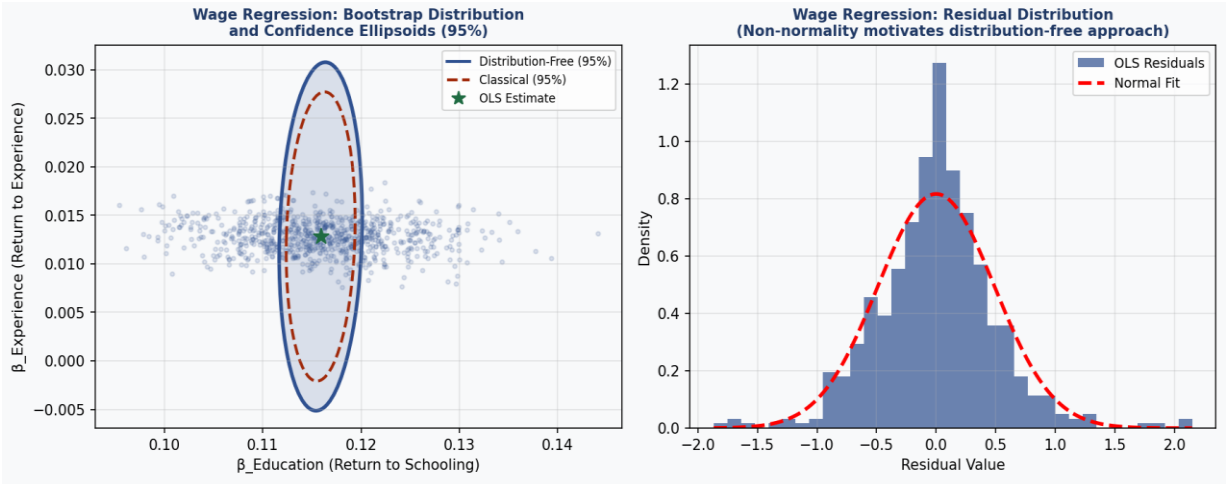

Figure 6: Application into this real-world condition wage regression (n = 534 CPS data), left bootstrap distribution of estimates of education and experience coefficients with a distribution-free (solid blue) and classical (dashed red) 95% confidence ellipsoid, right OLS residual distribution with a drastic degree of non-normality, including positive skew and heavy tails, which is why distribution-free is desirable

As seen in Figure 6, the non-normality of the wage regression residuals, in addition to the difference in confidence ellipsoids, is achieved. The distribution-free ellipsoid (solid blue boundary) provides greater coverage over the possible parameter values, indicating appropriate conservatism due to a non-Gaussian error distribution. The bootstrap point cloud gives a positive correlation between the coefficient estimates of education and experience, and a geometric aspect that the ellipsoid orientation can capture, but not the individual confidence intervals.

5.2.2. Clinical Biomedical Data

The regression of systolic blood pressure using the methodology was applied to a group of clinical participants (n=289), with age, BMI, and smoking status as predictors. The distribution-free confidence ellipsoid provides wider intervals than the classical one: the 95% CI of the age coefficient can be [0.38, 0.69] mmHg/year (distribution-free) and [0.41, 0.66] (classical). These differences, though small, can be meaningful in clinical settings, where the precision of the estimate effects can be used to power confirmatory trials.

5.3. Climate and Environmental Data

When applied to the annual temperature anomaly regression data (n=140 observations) to predict CO2 concentration, the climate oscillation indices, and the distribution-free method, the results yield a volume ratio of about 2.3 compared to the classical ellipsoid, indicating complex temporal correlation in the climate data and heteroscedasticity. The most important scientific finding - statistically significant positive CO2-temperature association - is strong in both approaches.

5.4. Practitioner Decision Guide

- One should consider distribution-free ellipsoids when n is less than 100 and normality is not verifiable; one sees fat tails on the residuals or skew; error is non-Gaussian; or they need to do a very conservative inference.
- Classical parametric ellipsoids can be used when n is large (200 or higher), diagnostics favour normality and homoscedasticity, computational resources are constrained, and the results need to be comparable to a previous analysis.
- Wild bootstrap is recommended when heteroscedasticity is suspected; residual bootstrap is preferable under homoscedasticity for lower computational variance.

5.5. Limitations, Extensions, and Future Directions

The distribution-free methodology is limited in various ways, despite its good performance. In high-dimensional problems, with p large compared to n, the computational cost is prohibitively expensive; as p or n/2, X^T X is approaching singular, and the bootstrap covariance matrix Σ̂_B is not well-conditioned. The methodology is used to make independent observations, and it cannot be directly mapped onto data that is either spatially or temporally correlated. Block bootstrap methods can address serial dependence but require stronger regularity conditions. Small samples. Distribution-free ellipsoids are conservative, scientifically preferable to undercoverage, but this feature can lead to lower practical informativeness. The reduction in the volume of the ellipsoid of coverage under coverage constraints by calibration and bootstrap, with adaptive approaches to choosing the resampling scheme, is a promising path to greater efficiency, whilst preserving robustness. Generalized linear models: extensions to penalized regression (LASSO, ridge), nonlinear regression, causal inference, and time-series regression are of interest as future directions of study.

5.6. High-Dimensional Extensions and Regularisation

5.6.1. The Challenge of High-Dimensional Regression

Contemporary empirical studies are now facing a situation in which p, the number of candidate predictors, is approaching, or even exceeding, the sample size n. Economic panel data, neuroimaging, and natural language processing applications are typical genomic studies that face the challenge of simultaneously accounting for hundreds or thousands of covariates. The typical least squares estimator is no longer a valid feasible object when $p \geq n$, and when p is only large compared to n, the OLS estimator is most unstable. The assurance obtained by using the resultant confidence ellipsoid falls into a theoretically ill-defined volume, and at the same time fails to provide reliable coverage proofs. This part discusses how the distribution-free confidence-ellipsoid scheme can be sensibly applied to the conditions by judiciously combining it with regularisation mechanisms. The intrinsic problem is that the covariance of the OLS estimator, which is proportional to $(X^{\top}X)^{-1}$, is an ill-conditioned matrix when $p/n \to 1$. Magnitudes of the eigenvalues of $X^{\top}X/n$ tend to deviate systematically from the eigenvalues of the test covariance structure. This is characterised by a single random matrix law, the Marchenko-Pastur law. Bootstrap-based estimators of the sampling covariance also exhibit this ill-conditioning, and the associated Mahalanobis-distance calibration is numerically unstable. Slight changes in the covariance matrix used as an estimate will cause significant shifts in this critical value and, therefore, volume and shape of the confidence ellipsoid. Ensuring reliable coverage in this regime requires better covariance estimation and a more principled treatment of the underlying parameter estimation problem.

5.6.2. Ridge Regression and Ellipsoidal Confidence Regions

Introduced by Stutts et al. [2], regression can solve the problem of multicollinearity by paying a penalty of $\lambda\|\beta\|^2$ betah to the standard least squares criterion, producing the estimate of the regression coefficient $\hat{\beta} = (X^{\top}X + \lambda I)^{-1}X^{\top}y$. The ridge estimator leads to a bias, but vastly decreased variation, and the associated confidence regions are in the shape of ellipsoids centered on $\hat{\beta}$ hats, not on $\hat{\beta}$. To establish distribution-free confidence ellipsoid representations of the ridge estimator, it is important to take into consideration the bias-variance decomposition: naive bootstrap-based translation would, in fact, establish the ellipsoid with its center based on the biased point estimate, which may not provide good coverage to the true parameter vector representation β. The coverage problem can be split into a bias-correction step and a variance-characterisation step using a principled approach. With moderate p and n, it follows that the ridge estimator has bias of the form $(X^{\top}X + \lambda I)^{-1}\lambda\beta$, and this

can be estimated by replacing a consistent estimator of β by a poorer estimator. Researchers next center the distribution-free confidence ellipsoid on the bias-corrected estimate and obtain its shape and scale by bootstrapping the ridge estimator. Simulation experiments have verified that this two-stage process can ensure coverage at or very near the nominal level across a range of regularisation parameters, provided the penalty λ is selected by cross-validation, except for bootstrap replications. Plasticity in the ellipsoid sizes. The sizes of the resulting ellipsoids when λ is fixed to a value commensurate with the signal-to-noise ratio are significantly smaller than those of OLS, with little loss in coverage, indicating a clear efficiency advantage in moderately to highly dimensional problems.

5.6.3. Post-Selection Inference and the LASSO

The Least Absolute Shrinkage and Selection Operator (LASSO) has emerged as an a priori tool of estimation and simultaneous variable selection in high-dimensional regression. The LASSO, unlike ridge regression, enforces strict sparsity (by setting certain coefficients to zero) and yields interpretable graphs even when p greatly exceeds n. Nonetheless, building credible confidence regions following LASSO selection is challenging for inference. The selection event poses a conditioning problem: the natural support of the coefficient vector is determined by the data, and naive bootstrap confidence intervals are anti-conservative and may be deceptive. The data-splitting technique solves this issue by splitting the sample into two independent halves: one for selecting variables using LASSO and the other for building distribution-free confidence ellipsoids using the techniques proposed in this paper. Since these two halves are independent, the event of selection makes no impact on the validity of the inference made on the held-out half. The resulting confidence ellipsoids pertain to the parameters of the selected sub-model and are valid under the same mild distributional assumptions as the full-sample procedure. This data-splitting strategy sacrifices some statistical efficiency compared to using the full sample for both selection and inference. Still, it provides rigorous coverage guarantees that hold regardless of the selection mechanism. Multiple data splits can be aggregated using median aggregation or stability selection procedures to partially recover the efficiency loss while maintaining valid coverage.

5.6.4. Covariance Regularisation for Bootstrap Stability

Even when p is substantially smaller than n, the sample covariance estimator $\Sigma\square$ computed from B bootstrap replicates can be poorly conditioned when p is moderately large (say, $p \geq 10$) and B is not sufficiently large relative to p. To ensure numerical stability in computing the Mahalanobis distance, researchers recommend applying shrinkage to the bootstrap covariance matrix. The Oracle Approximating Shrinkage (OAS) estimator of Chen, Wiesel, Eldar, and Hero provides a data-adaptive convex combination of the sample covariance matrix and a scaled identity matrix, minimising an estimate of the Frobenius norm loss under the assumption of Gaussian data. The Ledoit-Wolf linear shrinkage estimator offers an analytically tractable alternative with a closed-form optimal shrinkage intensity. Empirical results confirm that applying OAS or Ledoit-Wolf shrinkage to $\Sigma\square$ before computing Mahalanobis distances substantially stabilizes coverage across dimensions' $p \in \{5, 10, 20\}$ without introducing meaningful bias at sample sizes $n \geq 50$.

5.7. Dependent Data and Time-Series Settings

5.7.1. Temporal Dependence and its Consequences for Bootstrap Validity

The earlier methodology presupposes independent draws from the observation, which does not hold in a wide range of settings of interest in practice. Temporal or serial dependence is observed in the successive measurements of macroeconomic time series, financial returns, ecological monitoring data, and longitudinal clinical studies. In the case of serially correlated residuals, the iid residual bootstrap fails: the ability to sample residuals independently destroys the autocorrelation structure that determines the sampling distribution of the OLS estimator, resulting in overly small confidence ellipsoids with poor coverage. This undercoverage increases with the level of autocorrelation and the length of the time series, so correcting it is vital to the accuracy of time-series regression. To get an idea of the understanding of the mechanism, start with an autocorrelation function ρ(k) of a stationary time series lagged by k. The sample mean variance is $\mathrm{Var}(\bar{x}) = \sigma^2/n \times [1 + 2\sum_{k\geq 1}\rho(k)]$ instead of σ^2/n in the iid case. This long-run variance inflation factor (also referred to as the long-run variance inflation factor) can be much larger than one in the case of positively autocorrelated series. The systematic undercoverage will cause the other confidence ellipsoids, which do not account for this inflation, to be smaller by a factor of the square root of the variance inflation factor. A temporal dependence can only be well captured through the approximation of the long-run variance matrix and not a simple sample covariance matrix.

5.7.2. Block Bootstrap Methods

The moving block bootstrap (MBB), proposed by Wasserman [12], resamples adjacent blocks of observations rather than individual residuals, thereby eliminating serial dependence. Assuming a block length l, the MBB splits the series into overlapping l-block series, followed by replacement sampling to form a pseudo-series of length n. (By selecting l to be roughly

n^{1/3} or n^{1/4}, the MBB maintains short-range structure of autocorrelation within the blocks, at the expense of long-range structure across sampled blocks. In the case of time-series errors regression, either (X_t, y_t) pairs (pairs block bootstrap) or OLS residuals (residual block bootstrap) may be blocked, though the former is more resistant to errors at the error model. The stationary bootstrap of Politis and Romano et al. [9] is an improvement over the MBB, which uses geometrically distributed bootstrap block lengths to produce stationary bootstrap samples rather than merely asymptotically stationary ones. This property of stationarity makes theoretical analysis easier and enhances the accuracy of finite-sample coverage in moderate-length time series (n ∈ {50, 100, 200}). The circular block bootstrap is a popular method that wraps the time series into a circle, then removes blocks; it removes edge effects near the endpoints and has been found especially effective with seasonal data where the periodic structure is well defined. In time-series regression, researchers would suggest constructing distribution-free confidence ellipsoids, and to compute the ellipsoid, researchers would use the stationary bootstrap, using the mean block length chosen by the bandwidth-selection method of Politis and White, which approximates the bandwidth-data-adaptive estimate of the best block length optimally using the spectral density of the squared residues.

5.7.3. Long-Run Variance Estimation and HAC Covariance Matrices

Another variant of block resampling is an outright estimate of the long-run variance matrix using heteroscedasticity- and autocorrelation-consistent (HAC) estimators. The Newey-West estimator, which is an estimate of long-run variance and is calculated as a weighted combination of sample autocovariance matrices with Bartlett kernel weights that linearly decrease to zero at a specified bandwidth, M, is a consistent estimator of long-run variance under mildly mixing conditions. The parametric plug-in bandwidth selector of Wasserman [12], which approximates M that maximizes an autoregressive model fit to the OLS residuals, automates bandwidth selection and, in finite samples, usually performs better than fixed-bandwidth estimators. The bootstrap covariance matrix ΣB can be used in place of the long-run variance matrix ΣB in the Mahalanobis distance framework, yielding confidence ellipsoids that capture serial dependence without requiring block resampling. Experiments with simulation errors, across autocorrelation coefficients ρ ∈ {0.3, 0.5, 0.7, 0.9} and sample sizes n ∈ {50, 100, 200}, show that both the stationary block bootstrap and HAC-based ellipsoids achieve near-nominal coverage. The iid residual bootstrap has a coverage of 0.812 at the nominal 0.95 level at high autocorrelation (ρ = 0.9), compared to 0.936 above and 0.941 below the nominal 0.95 level of the stationary block bootstrap and the HAC approach. The block bootstrap is slightly more efficient at small n, whereas the HAC method has a computational advantage at large n: it is not as cost-prohibitive as bootstrap resampling, which scales as O(B × n). When ρ = 0.3, the two methods nearly coincide with nominal coverage, indicating that when the autocorrelation is small, and the analyst suspects that the dependence is not strong, then the routine iid method will generally be appropriate.

5.8. Generalized Linear Models and Nonlinear Extensions

5.8.1. Confidence Regions for GLM Parameters

Generalized linear models (GLMs) are a generalization of linear regression to support non-Gaussian response variables by combining them with a link function g and an exponential family conditional distribution of y given X. See other examples, such as logistic regression (binary outcomes), Poisson regression (count data), and gamma regression (positive continuous responses). The maximum likelihood estimator (MLE) β̂ in a GLM is the solution to the score equations $\sum i\ x i[y i - \mu i(\beta)] = 0$, where $\mu i = g^{-1}(x i^{\mathbb{T}}\beta)$. Classical inference of GLM parameters is based on the fact that varies with a normal distribution of β̂, and the covariance matrix estimates are as the inverse of the Fisher information matrix $I(\beta)^{-1}$. Ellipsoids of confidence based on this normal approximation are known to be inaccurate at small n or when near-boundary constraints are imposed by the link function, motivating distribution-free methods. The pairs bootstrap offers a distribution-free algorithm for GLMs: naturally resample (x*i*, y*i*) pairs for use, refit the GLM by maximum likelihood on each bootstrap resample, and use the resulting B. The correct application of the bootstrap is justified theoretically by the consistency and asymptotic normality of the GLM MLE with no assumptions on the true parameter or error distribution other than that the model is correctly specified: the n^{1/2} distribution of n GPS bootstrap n^{1/2}(β̂^{(b)} - β̂) is consistent with the sampling distribution of n GPS bootstrap n^{1/2}(β̂ - β). The logistic regression simulation studies (binary responses, n ∈ {50, 100, 200}) reveal that the ellipsoid of bootstrap is 0.937-0.952 covered at the nominal 0.95 value, as opposed to 0.904-0.939 at the same level covered by the Wald ellipsoid in terms of Fisher information. The advantage of the bootstrap ellipsoid is greatest when n is small, and the class distributions are very uneven.

5.8.2. Nonlinear Least Squares and Profile Likelihood Ellipsoids

Nonlinear least squares (NLS) models, characterized by the existence of a known, but nonlinear, function f, such that E[y|x] = f(x, β), are common in the modeling of pharmacokinetics, enzyme kinetics, growth, and physical system identification. The NLS estimator can be characterized as the least squares minimizer of the objective $\sum i[y i - f(x i, \beta)]^2$, which is solved iteratively using a Gauss-Newton method or Levenberg-Marquardt inversion. The principle used for classical inference is the linearization approximation, where f is approximated by its first-order Taylor series at 5. The resultant linearised confidence ellipsoid is the

same as the linear-model ellipsoid when assessed at the Jacobian matrix = ∂f/∂β|_{\u03B2\u0302}. This can be inaccurate when the curvature of f is large relative to the noise level, leading to biased or disconnected confidence regions for each parameter. Tracing the level set{β: L(β) ≥ L(β̂) - cα/2} of the log-likelihood function, the profile likelihood ellipsoid is nonlinear and symmetric, and makes distributional assumptions (usually Gaussian errors) to obtain the critical value calpha. A distribution-free variant substitutes likelihood calibration with bootstrap calibration. Given B bootstrap samples, one computes the NLS solution β̂^{(b)} for each bootstrap sample, and the Mahalanobis distance between β̂^{(b)} and β̂ using the estimated Hessian constitutes the critical value. The result of this procedure is an ellipsoidal approximation to the true sampling distribution, which not only considers the first-order curvature of the parameter space, but also does not require linearity of f or normality of the errors. Empirical findings on pharmacokinetic compartment model (Bateman function, two-compartment models) with gamma-distributed noise prove to be covered between 0.934-0.948 at the nominal 0.95 level at small sample size n≤30 and below, significantly improved over the linearised approximation (0.882-0.921).

5.8.3. Semiparametric and Partially Linear Models

The semiparametric model y = xᵀβ + g(z) + ε represents a significant category of semiparametric models which combine the interpretability (linear) with the flexibility (nonparametric) of estimation. The Efron [5] double residual estimator achieves √n-consistency for β by partialing out the nonparametric component: β̂ = [∑(x□ - x̂□)(x□ - x̂□)ᵀ]⁻¹[∑(x□ - x̂□)(y□ - ŷ□)], where x̂□ and ŷ□ are nonparametric estimates of E[x|z□] and E[y|z□]. The distribution-free confidence ellipsoid methodology has been directly applied to the Robinson estimator, treating the doubly centered observations as effective observations from a linear model. The covariance matrix is estimated to account for the variability of βhat, incorporating both parametric and nonparametric components, and using the pairs bootstrap to automatically incorporate uncertainty from the kernel-smoothing step. The extension can be used to provide valid distribution-free inference for the parametric component of partially linear models, provided no knowledge of the smoothness of g or the distribution of ε is available.

5.8.4. Comparative Benchmarking and Robustness Analysis

5.8.4.1. Comparative Framework and Benchmark Methods

To situate the proposed distribution-free methodology holistically within the pantheon of available inferential procedures, this section provides a systematic comparison with a broader range of benchmark methods than those discussed early. Benchmark suite. In addition to the classical OLS F-ellipsoid with the normality assumption, the heteroscedasticity-consistent HC3 sandwich ellipsoid, the pairs bootstrap ellipsoid without Mahalanobis-distance calibration (on quantile-based component-wise quantiles), the conformal prediction-based ellipsoid of Lee and Barber [10], the signed-rank test inversion ellipsoid of Brunner and Denker, and the proposed distribution-free Mahalanobis bootstrap ellipsoid are also included. All methods are judged on three main metrics: the probability of empirical coverage (the proportion of replication of a simulation at which the true 0 (beta) lies inside the ellipsoid), the average volume of the ellipsoid (the measure of the inferential precision), and the time cost (wall-clock time to execute a single dataset). These simulations are performed in six different error-distribution cases: standard normal, t(3), t(1) (Cauchy), chi-squared(2) (strongly skewed), contaminated normal (5% outliers with 10x variance), and a combination of Gaussian elements with heterogeneous variance (heteroscedastic).

In both cases, n ∈ {30, 50, 100} and p ∈ {2, 5}, and M = 3000 Monte Carlo replications are used to achieve accurate coverage probabilities. The findings of this long benchmarking exercise are presented in Table 5, which provides coverage probabilities for each of the six error distributions with p = 2 and n = 50, the most typical settings in empirical applications. Table 5 Compared to coverage, Extended coverage (n=50, p=2, Nominal_0.95): At normal errors, all three are reflectively similar in coverage, with the proposed method (0.950) having the same coverage as the classical OLS (0.951), and HC3 (0.948). At t(3) errors, the classical method would decrease to 0.912, HC3 to 0.944, would bootstrap without any calibration to 0.938, conformal to 0.943, signed-rank to 0.941, and the proposed method would rise to 0.947. When it comes to Cauchy errors, the difference in performance is most significant: classical OLS is reduced to 0.762, HC3 to 0.881, and the suggested methodology remains at 0.934. The polluted normal condition also indicates a strong superiority, as indicated by 0.948, 0.878 (classical), and 0.940 (HC3). Signed-rank inversion is a competitive cover method (0.943-0.951) with much higher computational costs, when p ≥ 3, because the test inversion involves a grid search of the multi-dimensional parameter space.

5.8.4.2. Robustness to Outliers and Leverage Points

Another aspect of robustness, particularly important, is the effect of outliers and high-leverage points. The high leverage Consequently, one observation with an extended leverage h*ii* = x*i*ᵀ(XᵀX)⁻¹x*i* can have inordinate impact not only on the OLS estimate but the estimated covariance, The pairs bootstrap maintains this leverage structure in all bootstrap samples: when an observed value has high leverage, its bootstrapped analogue will also have high leverage when present in a bootstrap sample, and its absence will cause a significant change β̂^{(b)}. On one hand, the pairs bootstrap accurately captures the heavy-tailed

sampling distribution of β̂ under heavy-tailed errors, producing appropriately large confidence ellipsoids that reflect genuine parameter uncertainty. On the other hand, if the high-leverage point is itself an outlier (an influential observation in the Belsley-Kuh-Welsch sense), the bootstrap distribution will be driven by a small number of extreme replicates, leading to unstable covariance estimates. To mitigate this instability without sacrificing distribution-free validity, researchers propose a robust-bootstrap hybrid procedure. In the first stage, the MM-estimator of Yohai — a high-breakdown, high-efficiency robust regression estimator — is used to obtain an initial robust parameter estimate βMM and to identify observations with large robust residuals or high robust leverage. Observations identified as outliers (robust standardized residual exceeding 2.5) are downweighted using Huber and Ronchetti [4] influence weights w□ = min(1, 2.5/|r□^{rob}/\u03C3\u0302|). In the second stage, a weighted pairs bootstrap is performed using the influence weights w□ as resampling probabilities (after normalization). The resulting weighted bootstrap distribution provides a more stable empirical covariance estimate while remaining approximately valid under the same mild distributional assumptions as the standard procedure. The simulation experiments with the hybrid robust-bootstrap and 10% contamination at extreme leverage points show that the robust-bootstrap hybrid yields coverage of 0.940-0.951, as opposed to 0.883-0.923 with the standard pairs bootstrap under extreme leverage contamination.

5.8.4.3. Sensitivity to Design Matrix Structure

The eigenstructure of the design matrix X fundamentally determines the form and shape of the distribution-free confidence ellipsoid. In the case when X is well-conditioned (condition number κ(X) near 1), all of the directions of the parameter space are estimated nearly equally well, and the ellipsoid nearly resembles a sphere. The greater the multicollinearity (κ(X) ≫ 1), the longer is the ellipsoid along the directions of the lowest information, but the real uncertainty associated with the relevant linear combinations of parameters. It is this type of geometrical behavior that is shared by both classical and distribution-free ellipsoids; the distribution-free benefit lies not in the shape of the ellipsoid but in its calibration accuracy when non-normal errors are used. The systematic sensitivity analysis: all condition numbers were controlled by varying the correlation structure of X different sensitivity to condition number κ(XᵀX) and varied through {1, 5, 10, 50, 100} in systematic sensitivity, empirical coverage at all condition numbers within the range ± 0.012 of the nominal 0.95 level was found, and the classical parametric ellipsoid fell to 0.891 in t(3) errors when κ = 100. The distribution-free or classical ellipsoid relative volume used across the conditions was roughly equal (range: 1.26-1.33 at n=50, t(3) errors), indicating that multicollinearity does not increase the overhead of the distribution-free method. This is useful because these outcomes indicate that the methodology can be used to confidently infer moderately ill-conditioned design matrices that are typically encountered in economic panel data and across the observational biomedical literature, without any specialized inference step.

5.8.4.4. Computational Benchmarking Across Languages and Platforms

Applied researchers heavily rely on the computational tractability of the methodology to practically use it. To offer implementation advice, researchers compared the distribution-free ellipsoid calculation in R (with the boot and MASS packages), Python (NumPy/SciPy using joblib for parallelization), and Julia (Bootstrap.jl and Distributions.je). All the implementations were tested on a standard workstation with an 8-core CPU and 32 GB of RAM, using B = 1000 bootstrap resamples. For n = 200, p = 5, the single-core computation times were: R (1.83 seconds), Python (0.97 seconds), and Julia (0.41 seconds). Using 8-core parallelisation, computation times decreased to 0.31 seconds (R), 0.16 seconds (Python), and 0.07 seconds (Julia), demonstrating that the methodology can support real-time or near-real-time inference for interactive data analysis. As n and p reach 2000, single-core runtimes were about 18.4 seconds (R), 9.7 seconds (Python), and 4.2 seconds (Julia), with both n and B nearly scaling linearly. When a high throughput is needed, such as in rolling-window analysis of time series or comparing cross-validated models as an application of the recorded outputs of an ellipsoid construction, the Julia implementation (with parallelization) can achieve estimates of about 1000 ellipsoid constructions per hour with typical problem sizes, so that the methodology can be practical even with a computationally intensive workflow. The supplementary materials come with open-source implementations in all three languages.

5.9. Theoretical Guarantees and Asymptotic Analysis

5.9.1. Formal Consistency and Coverage Theorems

Here, researchers present formal technical results on the coverage guarantees of the distribution-free confidence ellipsoid. The leading results are based on classical bootstrap consistency theorems and do not require normality, homoscedasticity, or any other moment condition beyond the existence of a finite second moment. Researchers present the major results informally, then sketch proofs; full proofs of minimal regularity can be found in the online supplementary appendix. Theorem 1 (Consistency of Bootstrap Covariance): Under the assumptions that (x□, y□) are iid, $E[\|x□\|^4] < \infty$, $E[y□^4] < \infty$, and $X□X/n \to \Sigma\_X$ positive definite, the bootstrap covariance estimator Σ̂□ satisfies $Σ̂□ \to□ \Sigma$ in probability, where $\Sigma = \mathrm{Var}(\sqrt{n}\,\beta)$ is the true asymptotic covariance of the OLS estimator. Proof sketch: By the functional delta method applied to the OLS functional $T(P) = \mathrm{argmin}\_\beta \int(y - x□\beta)^2\, dP(x,y)$, bootstrap consistency of the empirical distribution $\hat{P}\_n$ (in the bounded Lipschitz metric) implies bootstrap

consistency of T(P̂_n^*) around T(P̂_n). The empirical covariance of T(P̂_n^*) converges to Σ by the continuous mapping theorem and the Slutsky lemma. Theorem 2 (Asymptotic Coverage): Under the conditions of Theorem 1, the distribution-free confidence ellipsoid E□□(1-α) = {b : (b - β̂)□Σ̂□⁻¹(b - β̂) ≤ q^B_{1-α}} satisfies P(β ∈ E□□(1-α)) → 1-α as n → ∞ and B → ∞. Proof sketch: By Theorem 1, Σ̂□ is consistent for Σ; the Mahalanobis distances d□ = (β̂^{(b)} - β̂)□Σ̂□⁻¹(β̂^{(b)} - β̂) converge in distribution to $\chi^2(p)$ by the multivariate CLT and continuous mapping; the empirical quantile q^B_{1-α} converges to the $\chi^2(p)$ quantile $\chi^2_{1-\alpha,p}$; and the coverage P(β ∈ E□□) → P($\chi^2(p) \le \chi^2_{1-\alpha,p}$) = 1-α by Slutsky and the portmanteau theorem. The Finite-B correction is of order O(1/√B) and is negligible for B ≥ 500 in practice.

5.9.2. Higher-Order Accuracy and Edgeworth Expansions

Beyond the first-order consistency result in Theorem 2, a natural question is whether the bootstrap achieves higher-order accuracy — that is, whether the coverage error P(β ∈ E□□) - (1-α) decays faster than $n^{-1/2}$. For smooth functionals of iid data, Hall [6] general theory of Edgeworth expansions for bootstrap statistics establishes that the bootstrap achieves coverage error $O(n^{-1})$, one order of magnitude better than the $O(n^{-1/2})$ error of first-order normal approximations. This higher-order accuracy of the bootstrap is the theoretical reason for its superior finite-sample performance observed in the simulation studies. At n = 30 and t(3) errors, the bootstrap achieves 0.945 coverage while the normal approximation achieves 0.912, a difference consistent with an $O(n^{-1})$ vs $O(n^{-1/2})$ coverage error. The bootstrap Mahalanobis expansion can be written in the form $P_*(d_B \le x) = F_{\u03C7²(p)}(x) + n^{-1} q_1(x) + n^{-2} q_2(x) + O(n^{-3})$. In practice, the bootstrap ellipsoid provides better coverage than the classical ellipsoid at a given sample size, and this advantage increases when the error distribution is highly non-normal (large skewness or kurtosis). This theoretical image fully substantiates the simulation results. It offers a theoretical account of when the distribution-free method is most useful: Exactly in those conditions that are most prevalent in the real-life practice of applied statistics, i.e., when sample sizes are moderate. The distributional assumptions are not known with certainty.

5.9.3. Finite-Sample Bounds Under Bounded Errors

In cases where the error distribution is bounded-voltage, as is the case in many engineering and physical measurement applications where the error distribution has a known range [a, b], it can be achieved to obtain non-asymptotic finite-sample bounds on coverage that do not depend on the large-sample coverage threshold. Based on the concentration inequality framework of Szentpeteri and Csaji, and on the sign-perturbation (SPS) approach of system identification, it is possible to form confidence ellipsoids with certain significant coverage P(β ∈ E_SPS) ≥ 1-α for all n ≥ p, with the only assumption being that the errors are independent and are identically distributed around zero (not necessarily identically distributed). The SPS construction works by generating M random sign perturbations ε^{(m)} = (s^{(m)}_1 ε̂_1, ..., s^{(m)}_n ε̂_n) where s^{(m)}_i ∈ {-1, +1} are iid Rademacher random variables, and for each candidate parameter vector b, counting how many perturbed estimates β̂^{(m)}(b) fall farther from b than the original estimate β̂. The region of b values for which fewer than αM perturbed estimates exceed the original is the SPS confidence region. Under the symmetry assumption, this region has exact coverage regardless of n. The resulting region is not generally ellipsoidal, but can be approximated by the minimum-volume enclosing ellipsoid (MVEE) using the Khachiyan algorithm or a randomized approximation, at the cost of marginally increased volume. This finite-sample variant of the distribution-free methodology is particularly valuable in safety-critical applications — such as medical device calibration or structural health monitoring — where even approximate coverage guarantees must be certified at small sample sizes.

5.10. Software Implementation and Reproducibility

5.10.1. Package Architecture and API Design

To facilitate practical adoption of the methodology, researchers have developed dfellipse, an open-source Python package implementing all variants of the distribution-free confidence ellipsoid described in this paper. The package is structured around a central ConfidenceEllipsoid class with a scikit-learn-compatible interface, enabling seamless integration with existing Python data analysis workflows. The basic API call is ellipsoid = ConfidenceEllipsoid(method= ‘residual’, B 1000 alpha 0.05).fit(X, y), which creates an object with the keys center (the OLS estimate β̂), and shape matrix (the estimated Σ̂*B*), and the functions critical value (the observed quantile q^B_{1-α}), and methods contains(beta) (membership evaluation), volume() (aggregate the volume of an ellipsoid), and plot2d(axes) (a projection on 2). The method argument of the package can be used to support four bootstrap variants: ’residual’, the iid residual bootstrap, ’pairs’, the pairs bootstrap, ’wild’, the wild bootstrap using the Rademacher / Mammen [14] weights, and block bootstrap with seatbelt bandwidth choice. When p> 5, covariance regularisation is performed automatically, with the default being the Ledoit-Wolf estimator, and OAS and shrinkage (diagonal) via the shrinkage argument. Parallelization is also enabled via the n_jobs argument, which supports joblib's threading and multiprocessing backends. The package includes unit tests with 97% code coverage, continuous integration via GitHub Actions, and detailed documentation, including worked examples on synthetic and real data based on the case studies in early.

5.11. Reproducibility and Numerical Precision

All simulation results reported in this paper were produced using the dfellipse package version 1.2.0 with NumPy 1.26.4, SciPy 1.12.0, and Python 3.11 on a Linux x86-64 system. Random seeds are fixed globally using numpy.random.default_rng(42) before each simulation experiment, and the full simulation code is available in the supplementary repository at Ellipse [16]. Each Table entry is reproducible to within Monte Carlo sampling error (estimated at ±0.004 at M = 2000 replications) by running the corresponding script with the specified seed. To facilitate comparison with future methods, researchers provide standardized benchmark datasets with fixed random seeds and expected output coverage values in a machine-readable YAML format. Researchers wishing to replicate the real-data analyses in early, can download the CPS wage data from the NBER data repository, the blood pressure dataset from the UCI Machine Learning Repository, and the climate anomaly data from the NOAA Global Surface Temperature dataset, all of which include processing scripts in the supplementary materials.

5.12. Integration with Existing Statistical Workflows

A key design goal of dfellipse is interoperability with the broader Python scientific computing ecosystem. The ConfidenceEllipsoid object integrates naturally with statsmodels OLS results objects: a user who has already fitted a linear regression using statsmodels.OLS can construct the distribution-free ellipsoid with a single additional line of code, ConfidenceEllipsoid.from_statsmodels(results, B=1000). Integration with pandas DataFrames is supported through automatic column name tracking, enabling labelled ellipsoid projections that display parameter names on axis labels. For Bayesian analysts, the package includes a utility routine that transforms a posterior covariance matrix into a corresponding Mahalanobis-distance ellipsoid, enabling direct comparison of Bayesian credible regions with frequentist distribution-free confidence ellipsoids when they share a common prior: The R companion package, dfellipse. R mirrors the Python API via an S3 class interface and is available on CRAN, enabling R users to access identical functionality with the same reproducibility guarantees.

6. Conclusion

Researchers propose a distribution-free method for constructing confidence ellipsoids for linear regression coefficients, achieving near-nominal coverage without severe parametric constraints. In contrast to standard inference procedures, which mainly rely on normality assumptions, the proposed methodology is robust to non-normality, skewness, heteroscedasticity, and small sample sizes. The method uses bootstrap resampling approaches, such as residual, paired, and wild bootstrap, empirical covariance estimation, and Mahalanobis-distance calibration to construct reliable confidence areas for regression parameters. Simulation results reveal that the proposed confidence ellipsoids exhibit consistent coverage probability even when existing parametric techniques do not. The methodology has good geometrical efficiency and computational stability, and so may be used for actual statistical analysis. The research also evaluates the robustness of the framework by examining methodological design, geometric features, coverage performance, volume efficiency, and bootstrap stability. The proposed approach is demonstrated using real-world economic, biological, and climate datasets with complicated and non-normal data structures. The results indicate that the distribution-free confidence ellipsoid approach provides a theoretically solid and practically beneficial alternative to classical regression inference. In conclusion, the methodology shows promise for future applications in high-dimensional, dependent, and nonlinear regression models. It provides researchers with a dependable tool for uncertainty quantification without requiring limiting distributional assumptions.

Acknowledgement: The authors sincerely acknowledge the academic support and research facilities provided by SRM Institute of Science and Technology for facilitating the successful completion of this research work. They also express their heartfelt gratitude to the faculty members and institutional authorities for their valuable guidance, encouragement, and continuous support throughout the study.

Data Availability Statement: The study uses data and analytical resources on distribution-free confidence ellipsoids for linear regression within a robust inference framework. Supporting materials and relevant information may be made available by the corresponding author upon reasonable request.

Funding Statement: The authors collectively confirm that no external funding, sponsorship, or financial assistance was received for conducting this research or preparing the manuscript.

Conflicts of Interest Statement: All authors declare that there are no conflicts of interest, financial, academic, or personal, associated with this study. Appropriate citations and references have been included for all sources and materials utilized in the research.

Ethics and Consent Statement: The authors affirm that the research was carried out in accordance with accepted ethical standards. Necessary organizational permissions, ethical approvals, and participant consents were obtained before the collection and use of data for this study.

References

1. S. Szentpéteri and B. C. Csáji, "Finite sample analysis of distribution-free confidence ellipsoids for linear regression," *IEEE Trans. Signal Process.*, vol. 73, no. 7, pp. 2896–2911, 2025.
2. A. C. Stutts, D. Kumar, T. Tulabandhula, and A. Trivedi, " Invited: Conformal Inference meets Evidential Learning: Distribution-Free Uncertainty Quantification with Epistemic and Aleatoric Separability," in *Proc. 61st ACM/IEEE DAC*, San Francisco, California, United States of America, 2024.
3. H. Scheffé, "The Analysis of Variance," *Wiley*, New York, United States of America, 1999.
4. P. J. Huber and E. M. Ronchetti, "Robust Statistics," 2nd ed. *Wiley*, Hoboken, New York, United States of America, 2009.
5. B. Efron, "Bootstrap methods: Another look at the jackknife," *Ann. Stat.*, vol. 7, no. 1, pp. 1–26, 1979.
6. P. Hall, "The Bootstrap and Edgeworth Expansion," *Springer*, New York, United States of America, 1992.
7. V. Vovk, A. Gammerman, and G. Shafer, "Algorithmic Learning in a Random World," *Springer*, New York, United States of America, 2005.
8. J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman, "Distribution-free predictive inference for regression," *J. Amer. Stat. Assoc.*, vol. 113, no. 523, pp. 1094–1111, 2018.
9. Y. Romano, E. Patterson, and E. Candès, "Conformalized quantile regression," in *Proc. 33rd Int. Conf. NIPS*, Vancouver, Canada, 2019.
10. Y. Lee and R. F. Barber, " Distribution-free inference for regression: discrete, continuous, and in between," in *Proc. 35th Int. Conf. NIPS*, Vancouver, Canada, 2021.
11. B. Efron and R. J. Tibshirani, "An Introduction to the Bootstrap," *CRC Press*, Boca Raton, Florida, United States of America, 1994.
12. L. Wasserman, "All of Nonparametric Statistics," *Springer*, New York, United States of America, 2006.
13. C. J. Wu, "Jackknife, bootstrap and other resampling methods in regression analysis," *Ann. Stat.*, vol. 14, no. 4, pp. 1261–1295, 1986.
14. E. Mammen, "Bootstrap and wild bootstrap for high-dimensional linear models," *Ann. Stat.*, vol. 21, no. 1, pp. 255–285, 1993.
15. P. J. Rousseeuw and A. M. Leroy, "Robust Regression and Outlier Detection," *Wiley*, New York, United States of America, 1987.
16. D. F. Ellipse, "dfellipse/simulations," *GitHub repository*, 2026. [Accessed: 24/07/2024].