*symmetry*

MDPI

*Article*

# A Moment-Targeting Normality Transformation Based on Simultaneous Optimization of Tukey *g–h* Distribution Parameters

**Zeynel Cebeci [1,*], Figen Ceritoglu [2], Melis Celik Guney [1] and Adnan Unalan [3]**

[1] Division of Biometry & Genetics, Faculty of Agriculture, University of Cukurova, Adana 01380, Türkiye; celikm@cu.edu.tr
[2] Division of Biometry & Genetics, Faculty of Agriculture, University of Siirt, Siirt 56100, Türkiye; figenyildiz@siirt.edu.tr
[3] Department of Biostatistics & Medical Informatics, Faculty of Medicine, University of Nigde Omer Halisdemir, Nigde 51240, Türkiye; aunalan@ohu.edu.tr
* Correspondence: cebeciz@gmail.com

**Abstract**

This study proposes Optimized Skewness and Kurtosis Transformation (OSKT), a novel moment-targeting normality transformation that corrects asymmetry and peakedness in non-normal data. OSKT employs a transformation function derived from the Tukey *g–h* distribution, incorporating skewness and kurtosis parameters, and is optimized by minimizing a single objective function based on the Anderson–Darling test statistic. The optimization process uses L-BFGS-B to tune the transformation parameters to find the best fit for the standard normal distribution. OSKT ensures a balance between symmetry and tail behavior by minimizing deviations from theoretical normality. It has highly competitive performance compared to the alternative, Box–Cox, Yeo–Johnson transformations, including their robust variants and moment-matching Lambert W method, for normalizing complex distributions. According to our analysis, OSKT also achieves superior normalization for highly non-Gaussian data, successfully transforming highly resistant distributions, including approximately symmetric bimodal datasets, where other methods fail.

**Keywords:** normality transformation; moment-targeting transformation; Tukey *g–h* distribution; distributional asymmetry; skewness; kurtosis

Check for updates

Academic Editors: Sergei Odintsov and Quanxin Zhu

Received: 26 January 2026
Revised: 1 March 2026
Accepted: 3 March 2026
Published: 6 March 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

## 1. Introduction

The assumption of normality is an essential step for the valid application of many parametric statistical methods including *t*-tests, analysis of variance, linear regression, and many likelihood-based inference procedures. Large deviations from normality, especially in the form of skewness (distributional asymmetry) and kurtosis (tail thickness or peakedness), can create biased parameter estimates, unreliable confidence intervals, and inflated Type I or Type II error rates, all of which can lead to misleading conclusions. These risks are particularly critical in biomedical research, environmental science, finance, and quality control, where decisions strongly depend on accurate uncertainty quantification. Therefore, data normalization clearly remains a key preprocessing strategy in such research areas and data science.

In order to mitigate non-normality, various power transformations were developed in the past. Classical transformation methods such as the square-root, logarithmic, and arcsine

variants remain popular due to their simplicity, but they require strictly positive inputs and offer only limited improvements when the data contain heavy tails or many zeros. More robust single-parameter approaches are available, such as the Box–Cox transformation (for strictly positive values) and the Yeo–Johnson transformation (for positive and negative values) which deal with the skewness by optimizing a transform exponent, λ [1,2]. However, they often underperform in the face of significant kurtosis or complex departures from normality. On the other hand, as a recent solution the Lambert W × F transformation [3] offers improved flexibility by explicitly modeling the heavy tails. Yet, difficulties still remain for non-linear distributional shapes or strongly asymmetric data.

This study proposes a normalization method that utilizes the Tukey *g*–*h* distribution [4], which models skewness and thick tails in data. In this context, it is important to examine some studies and developments that have aimed to use this distribution. Since the Tukey *g*–*h* distribution is defined through a quantile transformation of the normal distribution, its probability function cannot be handled in closed form, and exact maximum likelihood estimation is not possible. Despite the lack of a closed-form expression, it has been widely used for modeling skewed and heavy-tailed features in many studies.

One of the early studies on estimating the *g* and *h* parameters is Hoaglin's [5] letter-value-based estimation method. Hoaglin used robust letter-value summaries to estimate these parameters. Because the log-transformation of the ratios of upper and lower letter-value spacings shows an approximately linear relationship with the *g* parameter, and the log-transformation of the spacing magnitudes shows an approximately linear relationship with the *h* parameter, regression-like estimates for *g* and *h* are produced using ratios obtained from different letter-value levels. By combining these estimates, separate or joint estimators for g and h can be obtained. An implementation of this method is available in Zhan's [6] R package *TukeyGH77*.

In later years, *g* and *h* parameter estimation has generally focused on maximum likelihood-based methods. Some studies have also employed a few biologically inspired optimization algorithms for parameter estimation. More recently, computationally more demanding Bayesian estimation methods have gained attention.

Kuo and Headrick [7] derived closed-form solutions for parameters associated with the family of Tukey distributions using the method of percentiles (MOP). They defined a univariate MOP procedure and compared it with the method of moments (MOM) in the context of distribution fitting and estimation of skewness and kurtosis functions. Bee and Trapin [8] proposed a new approach to approximately compute maximum likelihood estimates of the *g*–*h* distribution, based on a frequentist reinterpretation of approximate Bayesian computation. Xu and Genton [9] employed the Tukey *g*–*h* distribution to model non-Gaussian spatial data, and the method called maximum approximate likelihood estimation was used to estimate *g* and *h* parameters. Zhang et al. [10] proposed an approach based on the Tukey *g*–*h* distribution for degradation reliability and used particle swarm optimization for parameter estimation. In their study, Möstel et al. [11] compared various classical and maximum likelihood methods for estimating the *g* and *h* parameters of Tukey *g*–*h* distributions through simulation and found the maximum approximate likelihood estimation method proposed by Xu and Genton [9] to be successful. Zhang et al. [12], in their study on improving the accuracy of lifetime predictions of mechanical components in reliability design, performed parameter estimation using the whale optimization algorithm. Al-Saadony et al. [13] proposed to determine the Tukey *g* and *h* parameters using several MCMC-based Bayesian methods. More recently, Guillaumin and Efremova [14] trained a neural network within a regression framework to estimate the parameters of the Tukey *g*–*h* distribution by minimizing a negative log-likelihood function.

This study introduces the Optimized Skewness and Kurtosis Transformation (OSKT) to address the limitations that normality transformation methods have so far struggled with. OSKT reduces overall non-Gaussianity directly instead of isolated moment distortions and provides independent control of skewness and kurtosis by integrating the Tukey *g*–*h* modeling framework. The *g* and *h* parameters are estimated by minimizing the Anderson–Darling test statistic ($A^2$) which is a goodness-of-fit measure known for its heightened sensitivity both at the center and within the tails of a distribution. Due to the simultaneous minimization of the skewness and kurtosis, OSKT achieves superior normalizations, especially for cases with severely heavy tails or multimodal or complex skewed data.

The contributions of this study can be summarized as follows:

- Unlike the traditional single-parameter adjustments (or moment-based transformations), OSKT optimizes skewness and kurtosis in a data-driven style utilizing a two-parameter approach based on the Tukey *g*–*h* distribution,
- An objective function that minimizes deviations from normality, assessed by the Anderson–Darling test statistic with Stephens' correction [15], which contributes to a more accurate and robust transformation, particularly in cases with heavy-tailed, non-unimodal but partially symmetric data,
- The proposed algorithm is effective for a broad variety of distribution shapes, including not just continuous data but also discrete data.

These contributions verify the potential of OSKT as a tool for statistical modeling and parametric data analysis, machine learning, and even other fields that require Gaussian-like data transformations.

## 2. Materials and Methods

### *2.1. Normality Transformation Methods*

There are many terms in data science and machine learning that are used interchangeably, even though they refer to distinct concepts. For example, the term clipping is sometimes considered a form of normalization, even though it only bounds extreme values within predefined limits rather than allowing them to take arbitrarily large or small magnitudes. Clipping is used to reduce the effect of outliers, but it does not change the overall distribution. However, normalization is for scaling numerical values to a specific range (e.g., $[-1, 1]$ or $[0, 1]$) to enable feature comparability. On the other hand, normality transformation modifies the distributional shape to better meet the normality assumptions required by several statistical models. Standardization is another distinct procedure in which data are transformed to have a zero mean and unit variance. Although this transformation is often incorrectly referred to as normalization, only data that are already normally distributed will result in a standard normal distribution after standardization. In summary, transformation is a general term that encompasses multiple approaches including scaling, normalization, and normality adjustments to improve analytical suitability [16]. In this study, the focus is placed on normality transformation methods.

#### 2.1.1. Overview of Normality Transformation Methods

Numerous methods have been developed for normality transformations, which can be broadly categorized into four groups: (i) traditional methods, (ii) power-based methods, (iii) moment-targeting methods, and (iv) rank-based methods. Traditional data transformation methods, some of the most widely known of which are summarized in Table 1, rely on simple monotonic functions and therefore have low computational cost, while providing several approaches to address non-normality by adjusting skewness and stabilizing variance. For example, the reciprocal transformation is useful for reducing extreme positive

skewness, although it is highly sensitive to zero or near-zero values. Transformations such as square root and logarithmic are beneficial for data with mild to moderate skewness and can stabilize variance.

**Table 1.** Traditional transformation methods.

| Abbreviation | Method | Transformation Function |
|---|---|---|
| REC | Reciprocal transformation | $T(x) = 1/x$ |
| SQR | Square root transformation | $T(x) = \sqrt{x}$ |
| LOG | Logarithmic transformation | $T(x) = ln(x)$ |
| ASN | Inverse hyperbolic sine transformation | $T(x) = ln(x + \sqrt{x^2 + 1})$ |
| EXP | Exponential transformation | $T(x) = e^x$ |

Transformation methods such as REC, SQR, and LOG are originally defined only for positive input values. However, they can also be applied to zero or negative data by adding a constant to the input to ensure positivity, using the adjustment with $x + max(0, -min(x) + \epsilon)$, where $\varepsilon$ is a small positive constant (i.e., $1 \times 10^{-6}$) to prevent zero input.

Traditional methods are straightforward and serve as adequate solutions for some cases but they address skewness or variance individually. This may become a limitation when more complex distributional deviations are present. On the other hand, power-based transformation methods offer a flexible framework due to the application of parametric transformations that modify the distributional shape. These methods improve normality and suitability for parametric modeling. The Box–Cox and Yeo–Johnson methods are among the most prominent power-based transformations. Both of them incorporate a tunable parameter that helps control the degree and direction of the transformation.

The Box–Cox transformation (BC) [1] is a well-known power transformation method that addresses substantial skewness in data by stabilizing variance and improving adherence to the normality assumption. As shown in Equation (1), BC requires strictly positive inputs and depends on a tunable shape parameter, $\lambda$, to determine the optimal transformation [17].

$$T(x;\lambda) = \begin{cases} log(x), & \lambda = 0,\ x > 0 \\ \frac{x^\lambda - 1}{\lambda}, & \lambda \neq 0,\ x > 0 \end{cases} \quad (1)$$

The Yeo–Johnson transformation (YJ) was proposed by Yeo and Johnson [2] to tackle BC's limitation of handling only positive values. As shown in Equation (2), YJ therefore generalizes the BC framework to handle zero and negative values as well, while preserving $\lambda$-parameter flexibility. Consequently, it is particularly suitable for datasets exhibiting mixed-sign observations or more complex deviations from normality:

$$T(x;\lambda) = \begin{cases} \frac{(x+1)^\lambda - 1}{\lambda}, & \lambda \neq 0,\ x \geq 0 \\ log(x+1), & \lambda = 0,\ x \geq 0 \\ -\frac{(-x+1)^{2-\lambda} - 1}{2-\lambda}, & \lambda \neq 2,\ x < 0 \\ -log(-x+1), & \lambda = 2,\ x < 0 \end{cases} \quad (2)$$

Since the transformation methods BC and YJ may not work well on complex distributions, in subsequent years, several extensions have been proposed to improve the estimation of transformation parameters and enhance the robustness of BC and YJ transformations [18–24].

Yu et al. [22] proposed an adaptive Box–Cox (ABC) transformation as a more efficient method to improve data normality. In the univariate case, where a single variable $x$ is considered, ABC identifies the optimal transformation parameter $\lambda^*$ by directly evaluating

the normality of the transformed data. For a candidate $\lambda$ in a predefined range $[a, b]$, the variable is transformed using the classical BC formula in Equation (1). The normality of the transformed variable is then evaluated using Shapiro–Wilk, a univariate normality test, producing a $p$-value $p(\lambda)$. The optimal $\lambda^*$ is chosen to maximize the log-transformed $p$-value, thus achieving the best approximation to normality:

$$\lambda^* = \arg\max_{\lambda \epsilon [a,b]} \log(p(\lambda)) \tag{3}$$

Finally, the original variable is transformed as $T(x; \lambda^*)$ using the optimal $\lambda^*$. Yu et al. [22] evaluated ABC through Monte Carlo simulations and real data, using the proportion of variables passing normality tests ($p > 0.05$) as a performance measure. Their results indicate that ABC consistently achieves higher normality rates compared to the classical BC transformation for both simulated and empirical datasets.

Raymaekers and Rousseeuw [24] proposed a reweighted maximum likelihood (RewML) approach to obtain robust and accurate transformation parameters for the BC and YJ methods. In the original versions of these algorithms, maximum likelihood (ML) estimators are applied, which are sensitive to outliers and heavy-tailed or skewed distributions. RewML addresses this issue by combining a robust initialization with an iterative reweighting scheme.

A preliminary transformation parameter $\hat{\lambda}_0$ is obtained by minimizing a robust criterion that compares the transformed values $T(x_i; \lambda)$ to the theoretical quantiles of a standard normal distribution, following the approach of Raymaekers and Rousseeuw [24], as given in Equation (4):

$$\hat{\lambda}_0 = arg\min_{\lambda} \sum_{i=1}^{n} \rho\left(\frac{(T(x_i;\ \lambda) - \hat{\mu}_M)}{\hat{\sigma}_M} - z_i\right) \tag{4}$$

In Equation (4), consider an ordered sample of univariate observations $X = (x_{(1)},\ \ldots,\ x_{(n)})$. $\rho(.)$ is the Tukey bisquare function (bounded and continuously differentiable), $\hat{\mu}_M$ and $\hat{\sigma}_M$ denote the Huber M-estimates of location and scale of $(x_i;\ \lambda)$, respectively, and $z_i = \Phi^{-1}[(i - 0.5)/n]$ are standard normal quantiles. This ensures that the initial transformation is less biased by outliers and highly skewed points.

Using $\hat{\lambda}_0$ as a starting value, a weighted log-likelihood is computed with Equation (5), which is defined in the RewML method [24] as:

$$\updownarrow_{\mathrm{w}}(\lambda) = -\frac{1}{2}\sum_{\mathrm{i}=1}^{\mathrm{n}} w_i \log[(T(x_i,\ \lambda) - \hat{\mu}_{\mathrm{w}})^2] + (\lambda - 1)\sum_{\mathrm{i}=1}^{\mathrm{n}} w_i \mathrm{g}(x_i) \tag{5}$$

where $\hat{\mu}_w$ is the weighted mean of the transformed data. The weights $w_i$ are derived from a hard rejection rule based on preliminary standardized residuals. Observations identified as potential outliers are down-weighted (typically $w_i = 0$), ensuring robustness. In the weighted log-likelihood, the term $g(x_i)$ depends on the transformation used. For BC transformation, $g(x_i) = log(x_i),\ x_i > 0$. For YJ transformation, $g(x_i) = log(x_i + 1), x_i \geq 0$ and $g(x_i) = -log(-x_i + 1)$, for negative values. This term is calculated from the derivative of the transformation with respect to $\lambda$ and ensures that the likelihood correctly accounts for the effect of the transformation. This reweighting step is iterated twice, which has been shown sufficient to achieve stable and accurate estimates.

The RewML estimator improves central normality while maintaining robustness to outliers and heavy tails and achieves smoother transformations than traditional ML estimators. It avoids masking extreme observations and is especially suitable for preprocessing before robust statistical modeling and applied data analysis. In this study, the implementation of the RewML approach to BC is referred to as the Reweighted Box–Cox (RBC)

transformation, and its application to YJ is referred to as the Reweighted Yeo–Johnson (RYJ) transformation.

The Lambert W function (LMB), denoted as $W(.)$, is a special function defined as the multivalued inverse of the function $f(w) = we^w$ [3]. More formally, for a given real or complex number $z$, LMB satisfies the identity in Equation (6):

$$z = W_k(z)e^{W_k(z)} \tag{6}$$

where $W_k(z)$ denotes the $k$-th branch of the function, as the inverse may yield multiple solutions depending on the value of $z$ and the branch index $k$. The principal branch is represented by $W_0(z) \geq -1$, while the lower branch, typically relevant for $z \in [-1/e, 0)$, is denoted $W_{-1}(z) \leq -1$. For $z \in [0, \infty)$, only one real-valued solution exists, $W_0(z)$.

In the context of data transformation, LMB has been utilized to address skewed and heavy-tailed distributions. The LMB transformation applies this function to Gaussianize non-normal data by applying an invertible transformation. For an observed variable $x$ generated from an underlying latent Gaussian variable $U \sim \mathcal{N}(0, 1)$, the skewed and heavy-tailed version can be modeled as:

$$x = T^{-1}(U; \delta, \gamma) = Ue^{\delta U}(1 + \gamma U^2) \tag{7}$$

The method assumes that the observed data are a distorted version of an underlying normal distribution, where deviations from normality arise due to skewness or excess kurtosis. The parameter $\delta$ controls skewness in the distribution: when $\delta > 0$, it reduces right skewness, and when $\delta < 0$, it reduces left skewness. The parameter $\gamma$ adjusts heavy-tailedness: when $\gamma > 0$, it reduces leptokurtosis (heavy tails), and when $\gamma < 0$, it reduces platykurtosis (light tails). The inverse transformation, which normalizes the observed data, is given by:

$$T(x; \delta, \gamma) = U = \frac{1}{\delta} W(\delta x / (1 + \gamma x^2) \tag{8}$$

The parameters $\delta$ and $\gamma$ in Equation (8) can be estimated using either moment matching or maximum likelihood estimation to minimize skewness and excess kurtosis. This two-parameter system allows the LMB transformation to simultaneously correct asymmetry and tail behaviour as well as keeping it invertible. The transformation and its inverse are based on the generalized skewness and heavy-tails (GSH) model of the Lambert W $\times$ F transformation [3] which is a widely recognized framework also offered in R packages. A major advantage of this transformation is its flexibility due to allowing simultaneous adjustment of skewness and kurtosis and supporting reliable parameter estimation for statistical models and hypothesis tests [25]. Because it is invertible, the original data can be recovered without information loss. These properties make LMB a useful and theoretically sound tool for improving the suitability of data for parametric statistical procedures and machine learning algorithms with the assumption of normality.

Parametric transformations like BC, YJ, and LMB can directly modify observed values based on their magnitudes and distributional characteristics but there are alternative rank-based normalization techniques that offer a non-parametric strategy for achieving normality. The inverse normal transformation (INT), also known as the rank-based inverse normal or quantile normalization transform, maps ordered observations to the quantiles of a standard normal distribution. INT is frequently discussed and used in many research topics, especially in genetics and genomics. Beasley et al. [26] highlighted that INTs do not necessarily maintain proper Type I error control and can reduce statistical power under some circumstances. Further studies, like that of McCaw et al. [27], have formally defined

and systematically compared direct and indirect INT-based association tests that indicate the method's evolution over time.

Ordered quantile normalization (ORQ) [28] follows the same principle as the INT and incorporates an improved quantile estimation that provides a better tail behavior. Since both of them entirely depend on the rank structure of the data rather than its absolute values, they remove meaningful information regarding scale and relative distances between observations. Although they can be perfect in satisfying normality assumptions, their loss of interpretability makes them incompatible for some analytical objectives. INT and ORQ will always produce the same transformed values for datasets that share the same rank order, independent of the actual observed values. For example, the sequences {1, 2, 3, . . .} and {10, 20, 30, . . .} will return identical transformed values because they have the same rank structure. By mapping the ranks of data points to the quantiles of a standard normal distribution, the transformed values reflect just the relative position of each observation within the dataset rather than the original scale or the distance between values. INT and QRQ can be really effective when the main objective is to normalize data to meet the assumptions of certain statistical tests or machine learning algorithms that require normality. However, they are useless when the goal is to compare or interpret the means or effect sizes of the original variables. As a result, neither INT nor ORQ is further considered in the subsequent analyses in this study.

#### 2.1.2. Proposed Normality Transformation Method

The novel normality transformation method is designed to transform non-normal datasets into distributions that more closely follow the Gaussian assumption. It optimizes skewness and kurtosis based on the empirical characteristics of the observed dataset and is therefore called the Optimized Skewness and Kurtosis Transformation (OSKT).

##### Conceptual Distinction from Existing Approaches

Before going further into the details of the algorithm's structure, it would be useful to further clarify OSKT's position in comparison to the existing Tukey *g*–*h* parameter estimation procedures and transformation-based Gaussianization methods. For example, consider Hoaglin's letter-value estimation, the method of moments, and maximum likelihood estimation, which are standard techniques that use a parameter-recovery framework. These techniques assume that the collected data are generated by a Tukey *g*–*h* distribution and aim to estimate its structural parameters. When the empirical distribution deviates, the estimators end up approaching pseudo-true values that minimize internal model discrepancy without a guarantee of Gaussianity after transformation. However, recovering parameters without specifying a generative model and enforcing normality directly are two distinct concepts. Similarly, the Lambert W×F framework is provided in a structured parametric representation, where the transformation parameters are interpreted as components of a generative skewness–kurtosis correction mechanism. The objective remains as model-consistent parameter estimation rather than distributional realignment.

In contrast, OSKT reframes the task as a functional optimization problem targeting empirical Gaussianization. Unlike the methods above, it does not assume that the data is generated by a Tukey *g*–*h* distribution. Instead, the Tukey *g*–*h* function is employed as a flexible non-linear transformation kernel whose parameters are selected specifically to minimize a discrepancy measure between the transformed empirical distribution and the Gaussian law. Therefore, the resulting parameters serve as transformation controls rather than structural descriptors of a data-generating mechanism. OSKT's goal is not inference on latent shape parameters but optimal alignment with normality under a well-defined statistical criterion.

This distinction is amplified by the choice of objective function. OSKT minimizes the Anderson–Darling statistic ($A^2$) by incorporating Stephens' correction. The Anderson–Darling functional can be interpreted as a weighted Cramér–von Mises distance, where the weight function ($\psi(u) = [u\ (1-u]^{-1}$) imposes an increasing emphasis on discrepancies near the boundaries of the unit interval. Unlike likelihood-based objectives that tend to be dominated by high-density central regions, the Anderson–Darling statistic puts significant penalties on deviations in the tails. Such a tail-sensitive structure ensures that optimization simultaneously controls central distributional fit and extreme quantile behavior, which is an essential property for maintaining valid Type I error rates and reliable $p$-value calibration in parametric inference.

The incorporation of Stephens' correction reinforces the finite-sample properties of the objective. The adjusted statistic ($A^2_{adj} = A^2\ [1 + 0.75/n + 2.25/n^2]$) reduces the small-sample upward bias of the raw $A^2$ while keeping asymptotic consistency as the multiplicative factor that converges to unity. Although the correction does not eliminate sample-size dependence, it improves calibration in moderate and small samples and prevents excessive penalization in the case of limited data.

Asymptotically, OSKT inherits the statistical properties of EDF-based minimum distance estimators. The EDF satisfies a functional central limit theorem, yielding root-n convergence under standard regularity conditions. Consequently, weighted quadratic EDF functionals such as the Anderson–Darling statistic exhibit $O(1/n)$ variance decay as sample size increases. This behavior reflects the stable convergence structure of empirical process theory and mostly holds under finite-variance conditions. Hence, the optimization benefits from a discrepancy measure whose sampling variability decreases at a controlled and predictable rate.

However, objectives that are constructed directly from higher-order sample moments, especially skewness and kurtosis, can show pronounced finite-sample instability in heavy-tailed settings. Since empirical third and fourth moments are highly sensitive to extreme observations, their variance may inflate substantially as tail thickness increases. Although moment-based estimators remain consistent when the required moments exist, their sampling volatility can lead to irregular optimization and reduced robustness. On the other hand, OSKT framework leverages a more globally informative and statistically stable discrepancy structure by operating on the full empirical distribution rather than isolated moment conditions.

To conclude, OSKT is not an alternative estimator of Tukey $g$–$h$ parameters within a generative modeling framework. Rather, it amounts to a distribution-shaping methodology that employs a flexible transformation kernel, which is optimized under a tail-aware minimum-distance criterion with well-defined finite-sample and asymptotic properties. This distinction clarifies both its theoretical motivation and methodological contribution among the rest of the distributional normalization techniques.

### Description of the Algorithm

The OSKT adapts the classical Tukey $g$–$h$ distribution [4] for automatic parameter estimation by using a fully data-driven optimization strategy. In classical implementations of the Tukey distribution, $g$ and $h$ are obtained from heuristic or subjective procedures. These include visually inspecting distributional shapes, roughly searching parameter grids, or making simple assumptions, such as fixing one parameter while tuning the other. Such implementations lack a proper normalization criterion, cannot ensure convergence to an optimal Gaussian-like form, and usually perform sub-optimally. OSKT aims to be a solution for these limitations by treating parameter selection as a constrained optimization problem, determining $g$ and $h$ values that minimize the deviation of the transformed dataset from

a standard normal distribution. It aims to achieve $g = 0$ and $h = 0$ for symmetry and normal-like tail behaviour, respectively. To describe the procedure, let $x$ be the observed data and standardize it to obtain $X_s$. Then, the Tukey $g$–$h$ transformation used in OSKT can be given as:

$$T(X_s) = \begin{cases} \frac{e^{(g \cdot X_s)} - 1}{g} \cdot e^{\left(\frac{h \cdot X_s^2}{2}\right)}, & g \neq 0 \\ X_s \cdot e^{\left(\frac{h \cdot X_s^2}{2}\right)}, & g = 0 \end{cases} \quad (9)$$

In Equation (9), when $g > 0$, the transformation increases right (positive) skewness, stretching the right tail of the distribution. Conversely, when $g < 0$, it increases left (negative) skewness by elongating the left tail. When $g = 0$, the skewness component becomes inactive, and the transformation does not modify asymmetry. The parameter h, on the other hand, controls the kurtosis (i.e., the heaviness of the tails) through the second exponential term in the formula in Equation (9). For values of $h > 0$, the transformation amplifies tail thickness, resulting in a distribution with heavier tails (leptokurtic behavior). When $h = 0$, the distribution maintains a kurtosis close to that of a standard normal distribution. Importantly, the kurtosis component is symmetric around the origin and influences only tail thickness, without affecting skewness. Together, the $g$ and $h$ parameters offer a flexible mechanism to simultaneously adjust both skewness and kurtosis, enabling the transformation to handle a wide variety of non-normal distributional shapes.

According to Equation (9), if $g > 0$, the transformation increases in right (positive) skewness, stretching the right tail of the distribution. If $g < 0$, it increases in left (negative) skewness this time, by stretching the left tail of the distribution. Under the condition $g = 0$, the skewness component becomes inactive, and the transformation does not need to modify asymmetry. The parameter h controls the kurtosis (i.e., the heaviness of the tails) via the second exponential term, which means that, if $h > 0$, the transformation enlarges tail thickness. As a result, distribution will show heavier tails (leptokurtic behavior). When $h = 0$, the distribution maintains a kurtosis close to that of a standard normal distribution. It is important to note that the kurtosis component is symmetric around the origin and influences only tail thickness, without affecting skewness. In co-operation, the $g$ and $h$ parameters offer a flexible mechanism to simultaneously adjust both skewness and kurtosis, allowing the transformation to handle a wide variety of non-normal distributional shapes.

The optimal parameters $(g^*, h^*)$ are estimated by minimizing an objective function ($J$) based on the Anderson–Darling test statistic with Stephens' correction, which quantifies the deviation of the transformed data from a standard normal distribution:

$$(g^*, h^*) = \underset{\{g,h\}}{\operatorname{argmin}}\, J(g,h); J(g,h) = \left[-n - \frac{1}{n} \cdot \sum_{i=1}^{n} (2i-1)(log(P_{(i)}) + log(1 - P_{(n-i+1)}))\right]\left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right) \quad (10)$$

where $P_{(i)} = \Phi(Y_{(i)})$ stands for the standard normal cumulative distribution (CDF) calculated at the $i$-th order statistic of the transformed data. The original Anderson–Darling test [29] uses the empirical cumulative distribution function. In Equation (10), the last term, Stephens' correction factor is used to stabilize the original test statistic for small samples.

Another motivation for using the Anderson–Darling test statistic in the objective of the OSKT optimization scheme was that it assesses deviations from normality across the entire distribution with special emphasis on the tails. It also provides a continuous and differentiable measure of overall fit to a standard normal distribution which makes it suitable for gradient-based optimization algorithms such as L-BFGS-B. This optimization ensures that the transformed dataset closely approximates a standard normal distribution by reducing asymmetry and controlling tail behavior. The appropriate $g$ and $h$ parameters are not known a priori, so OSKT identifies them by solving the bounded optimization problem described above. The resulting transformation produces data with skewness and kurtosis

values that closely match those of a Gaussian distribution. The following pseudocode outlines the step-by-step OSKT procedure, including data standardization, application of the *g*–*h* transformation, and optimization of the parameters through minimization of the objective function.

OSKT's optimization strategy ensures that the transformed dataset can closely approximate a standard normal distribution by reducing asymmetry and controlling tail behavior. The optimal $g$ and $h$ are not known beforehand so OSKT finds them by solving the bounded optimization problem described earlier. The transformation results in data with skewness and kurtosis that resemble those of a Gaussian distribution. The step-by-step OSKT procedure is explained by the following pseudocode.

Pseudocode for OSKT, the proposed transformation method

1: INPUT:
   Data vector $X = [x_1, x_2, \ldots, x_n]$
   Initial guess for parameters $g_i, h_i$
   Bounds for parameters $g_b, h_b$
2: Rescale raw data via standardization:
   $X_s \leftarrow (X - \overline{X})/S_x$
3: Define *g*–*h* transformation function
   IF g $\neq 0$ THEN
      $T(X, g, h) = \frac{e^{(g \cdot X)} - 1}{g} \cdot e^{(\frac{h \cdot X^2}{2})}$
   ELSE // Tukey *h*-transformation type part (kurtosis)
      $T(X, g, h) = X \cdot e^{(\frac{h \cdot X^2}{2})}$
   ENDIF
4: Define objective function *J(g, h)*:
   $Y \leftarrow T(X_s, g, h)$ // Apply Tukey *g*–*h* transformation
   $Y_s \leftarrow (Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(N)})$ // Sort the transformed data
   $P_i \leftarrow \Phi(Y_s)$ // Calculate the Standard Normal CDF
   $J(g,h) \leftarrow A^2 = -n - \frac{1}{n} \cdot \sum_{i=1}^{n} [(2i-1) \cdot (\log(P_i) + \log(1 - P_{n-i+1}))] \left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right)$
5: Set initial guess and bounds (defaults):
   $g_i \leftarrow -1, h_i \leftarrow 0$
   $g_b \leftarrow [-1, 1], h_b \leftarrow [0, 0.5]$
6: Find optimal parameters subject to bounds using L-BFGS-B optimization:
   $(g^*, h^*) \leftarrow \underset{\{g,h\}}{argmin}\ J(g,h)$
7: Compute final transformed data *Z*:
   $Z \leftarrow T(X_s, g^*, h^*)$
8: OUTPUT: Transformed data *Z*

As shown in the pseudocode above, the OSKT method begins to work by standardizing the input data to zero mean and unit variance (see Step 2). Then, it applies the Tukey *g*–*h* transformation where parameter $g$ modulates skewness and parameter $h$ adjusts tail heaviness (Step 3). To estimate $g$ and $h$, OSKT forms an objective function based on the Anderson–Darling test statistic ($A^2$), which measures the deviation of the transformed distribution from standard normality with an emphasis on tail behavior (see Step 4).

Objectives formed upon moment-based error measures can be sensitive to sampling variability in skewness and kurtosis. Therefore, it is appealing to use the Anderson–Darling test statistic which offers a more stable optimization scheme by evaluating the whole empirical distribution against the standard normal reference with greater focus on tail fit. The initial values for parameters are selected near the identity mapping $f(x) = x$ (see Step 5)

so that the transformation begins without strong distortions and consequently improves the convergence. The bounded search space, $g \in [-1, 1]$ and $h \in [0, 0.5]$, restricts the parameter estimates to remain inside realistic adjustment levels and also prevents numerical instability that may arise from the exponential components (recall Equation (9)). In addition, they contribute to the monotonicity and interpretability of the transformation result.

Having defined its limits, OSKT searches the optimal parameter pair for $g$ and $h$ (see Step 6). The search is formulated as a bounded non-linear optimization problem, where the objective function minimizes $g$ and $h$. To solve the optimization problem, gradient-based quasi-Newton algorithms such as the L-BFGS-B can be employed due to their efficiency [30,31] and ability to handle constraints. There are alternatives of course, including Nelder–Mead, BFGS without bounds, or global optimization approaches such as differential evolution or genetic algorithms [32,33] to consider especially when the objective landscape is highly non-convex. However, L-BFGS-B is the chosen solver for OSKT because it balances computational efficiency with robustness and converges to the optimal $g^*$ and $h^*$ in a reliable way while respecting the constraints, as well. By updating $g$ and $h$ iteratively, the L-BFGS-B reduces the goodness-of-fit discrepancy from the standard normal distribution; thus, producing transformed data that more closely adheres to normality.

Choosing a suitable optimization algorithm depends on the structure of the objective function and the computational power at hand. Even though L-BFGS-B provides a satisfactory balance between stability and efficiency for the OSKT objective, other bounded optimization methods can be chosen when the search landscape becomes highly non-convex. After settling on the optimal parameters, the OSKT is applied to the standardized data (Step 7), yielding a distribution with reduced skewness and controlled tail heaviness.

To recover the original data $X$ from the OSKT-transformed values $Z$, a numerical inversion procedure is employed. Let $X_s$ denote the standardized form of $X$ (i.e., after centering and scaling by the sample mean and standard deviation). The back transformation requires finding $X_s$ that satisfies the transformation equation used in OSKT:

$$Z = \begin{cases} \frac{e^{(g^* \cdot X_s)} - 1}{g^*} \cdot e^{\left(\frac{h^* \cdot X_s^2}{2}\right)}, & \text{if } g^* \neq 0 \\ X_s \cdot e^{\left(\frac{h^* \cdot X_s^2}{2}\right)}, & \text{if } g^* = 0 \end{cases} \tag{11}$$

where $g^*$ and $h^*$ are parameters optimized during the forward transformation. As the resulting equations are transcendental, a closed-form solution for $X_s$ does not exist. Therefore, various root-finding techniques can be utilized. In this context, the Newton–Raphson (NR) method and the bisection algorithms are widely preferred options. The NR method utilizes derivatives and is a faster option because the convergence is possible in just a few iterations; thus, it is a very undemanding option in terms of computational resources, especially when large datasets are used. However, due to the strong non-linear nature of the Tukey $g$–$h$ transform, NR is prone to becoming unstable for outliers. Unless the speed is a concern, the uniroot method (UR) can be considered as an alternative to achieve reliable results. UR guarantees convergence when there is a sign change in the search interval. Therefore, it is a safer but slower solution. With the UR, a tolerance of $10^{-10}$ can be applied for numerical accuracy, but machine-level precision can also be achieved with tolerances of $10^{-10}$ to $10^{-16}$. Stability is enhanced by the restricted parameter ranges ($g^* \in [-1, 1], h^* \in [0, 0.5]$) to prevent bias resulting from excessive exponential terms. Even for large datasets (e.g., $10^7$ observations), the computational time required for back transformation is negligible compared to forward transformation and parameter optimization. After calculating $X_s$, the original data $X$ is obtained by inverting the standardization:

$$X = X_s \cdot S_X + \overline{X} \tag{12}$$

In summary, OSKT offers more flexibility than single-parameter transformations such as the BC, YJ, and their variants because it simultaneously controls skewness and tail behavior. This feature brands OSKT as a promising alternative for datasets that contain both asymmetry and heavy tails, where simpler transformations are unable to secure an effective normalization. The proposed method is also adaptable as it performs distributional adjustments and offers improved normality approximation which makes it particularly useful for applications where deviations from Gaussian assumptions hinder the analysis.

### *2.2. Datasets and Experiments*

To evaluate the comparative performance of the proposed method, OSKT, against existing transformation techniques, both simulated and real-world datasets exhibiting diverse distributional characteristics were utilized. This dual design enables a controlled examination under known conditions and a practical validation using empirical data.

#### 2.2.1. Simulated Datasets

In this study, a scenario-oriented simulation design was used to obtain a rich spectrum of skewness–kurtosis combinations that are representative of real-world data. This strategy emphasizes diversity of data shapes rather than sampling variability, allowing direct comparison of transformation performance across distinct distributional conditions.

For this purpose, ten univariate variables (v1–v10) were generated with a sample size of $n = 500$, representing positively and negatively skewed forms, as well as discrete and bounded data structures commonly encountered in applied research. In the simulation of data, the degree of departure from normality has been categorized using absolute skewness thresholds. In the scenario, values of $|g_3|$ less than 0.5 indicate slight skewness, those between 0.5 and 1.0 suggest moderate skewness, and values equal to or greater than 1.0 reflect high skewness. Such strict cutoffs, as proposed by Bulmer [34] and Doane and Seward [35], create a more nuanced assessment of transformation performance, even though the more lenient $|g_3| < 2$ rule is often used as a general guideline for approximate normality [36]. In a similar way, excess kurtosis values can be interpreted as platykurtic ($g_4$ between $-2$ and $-1$), indicating light tails; mesokurtic ($g_4 = 0$), representing normal tail weight; and leptokurtic ($g_4$ between 1 and 5), indicating heavy tails [37]. In this study, continuous variables were simulated using the Tukey *g*–*h* transformation, allowing explicit control over skewness and tail heaviness through the values of $g$ and $h$ parameters in Table 2. In addition to the continuous distributions, the simulation design also incorporated count (discrete) and bounded (proportion-type) variables to evaluate the adaptability of the transformation methods beyond the continuous domain. A complete summary of the simulated variables, distributional specifications, and empirical moment statistics is provided in Table 2.

Previous studies have shown that, at smaller sample sizes (e.g., $n = 20$ or 30), many normality tests differ in error rates, often deviating from the nominal significance level of 0.05 [37]. When the sample size is increased, the tests stabilize and show error rates closer to the nominal significance level. Additionally, small samples tend to result in relatively low test power, particularly under mild skewness. The effectiveness of normality tests is also influenced by significance level and the similarity between the underlying non-normal distribution and the best-fitting normal model [38]. Based on these findings, in this study, a relatively large sample size of $n = 500$ was selected to ensure reliable evaluation of normality transformations and to minimize variability in test performance.

**Table 2.** Summary of simulated variables used in the experiments ($n$ = 500).

| Var. | Distribution | Parameters | $g_3$ | $g_4$ | Characteristics |
|---|---|---|---|---|---|
| **v1** | Tukey $g$–$h$ | $g = -0.15$, $h = 0$ | −0.52 | 0.63 | Moderately left-skewed, near-normal |
| **v2** | Tukey $g$–$h$ | $g = 0.36$, $h = 0$ | 1.00 | 1.36 | Highly right-skewed, light-leptokuric |
| **v3** | Tukey $g$–$h$ | $g = 0.40$, $h = 0$ | 1.51 | 4.71 | Highly right-skewed, heavy-tailed |
| **v4** | Tukey $g$–$h$ | $g = -0.48$, $h = 0$ | −1.52 | 3.16 | Highly left-skewed, long left tail |
| **v5** | Tukey $g$–$h$ | $g = -0.30$, $h = 0.1$ | −1.07 | 4.00 | Highly left-skewed, light-tailed |
| **v6** | Gaussian mixture | $N$~(0, 1) + $N$~(3.05, 1) | 0.01 | −1.04 | Bimodal, platykurtic, near-symmetric |
| **v7** | Poisson | $\lambda = 3.5$ | 0.40 | −0.01 | Slightly right-skewed, discrete |
| **v8** | Beta | shape1 = 5, shape2 = 2 | −0.69 | 0.26 | Moderately left-skewed, bounded [0, 1] |
| **v9** | Beta | shape1 = 0.6, shape2 = 0.6 | 0.02 | −1.45 | U-shaped, heavy tails near 0 and 1 |
| **v10** | Binomial | size = 50, $p = 0.2$ | 0.33 | 0.37 | Slightly right-skewed, discrete |

Var.: Variables, $g_3$: Skewness, $g_4$: Kurtosis.

In Table 2, the variables v1–v5 were generated with the related functions in the R packages *TukeyGH77* [6] and *groupcompare* [39]. The variable v6 was generated using the base R function "rnorm", v7 using "rpois", v8–v9 using "rbeta", and v10 using "rbinom", with the parameters specified in Table 2. Following data generation, all variables were rescaled in magnitude to approximate a standard normal distribution. A shifting procedure was applied to allow for the transformation methods (LOG, SQR, and BC) that require strictly positive values. After this procedure, descriptive statistics were calculated using the "describe" function from the *psych* package [40] in R. The computed skewness ($g_3$) and kurtosis ($g_4$) statistics are reported in Table 2, along with descriptions of the distributional characteristics. Figure 1 visualizes simulated variables in histograms, violin plots, and Q-Q plots, offering a glance at their skewness, kurtosis, and conformity (or deviation) from normal distribution assumptions.

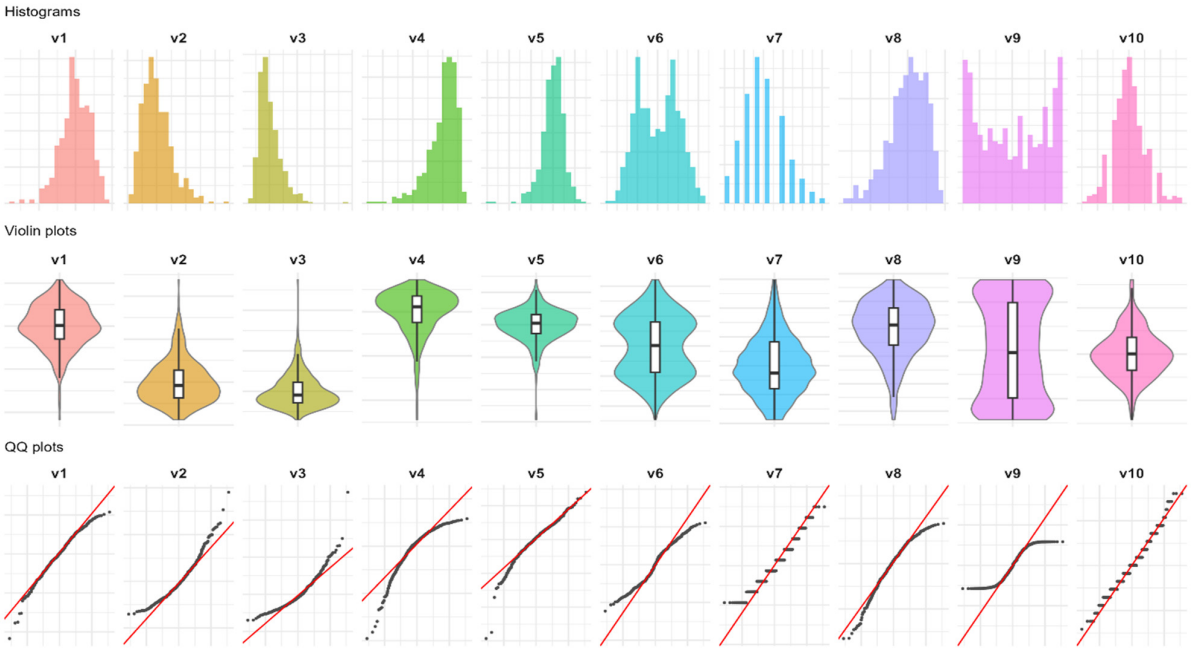



**Figure 1.** Distributional visualizations of simulated data.

### 2.2.2. Real-World Datasets

To evaluate the performance of different normalization methods on real-world data, three of R's built-in datasets (*airquality*, *co2*, and *USJudgeRatings*) as well as the *Wine Quality Red* dataset from the UCI Machine Learning Repository [41] were utilized. The analysis of

these datasets was restricted to variables exhibiting a consistent lack of normality across the full range of statistical tests employed in this study.

Additionally, three traits (leaf circularity, seed area, and seed width) from a doctoral study [42], which examined 967 specimens representing 52 species of the genus *Onobrychis* (sainfoin) based on various morphological and molecular genetic characteristics, were incorporated. A random selection of traits was performed, including only those previously confirmed to exhibit a consistent lack of normality across all employed statistical tests. As an example of a medical dataset, we also used the *Liver Disorders* dataset [43] from the UCI Machine Learning Repository, which contains 345 observations with several blood test measurements related to alcohol-induced liver dysfunction. For the normality transformation in this dataset, two non-normal variables, alkaline phosphatase and alanine aminotransferase, were selected. The datasets, variables, and their characteristics used in the experiments are summarized in Table 3 and visualized in Figure 2.

**Table 3.** Summary of real data variables used in the experiments.

| Var. | Dataset | Variable | $n$ | $g_3$ | $g_4$ |
|---|---|---|---|---|---|
| v1 | *USJudgeRatings* | Phys | 43 | −1.504 | 2.159 |
| v2 | *Wine Quality Red* | Fixed acidity | 1599 | 0.981 | 1.120 |
| v3 | *Wine Quality Red* | Citric acid | 1599 | 0.318 | −0.793 |
| v4 | *co2* | CO2 | 468 | 0.242 | −1.223 |
| v5 | *airquality* | Ozone | 153 | 1.210 | 1.112 |
| v6 | *Onobrychis* | Leaf circularity | 967 | −0.170 | −0.091 |
| v7 | *Onobrychis* | Seed solidity | 967 | −1.463 | 2.912 |
| v8 | *Onobrychis* | Seed width | 967 | 0.964 | 0.435 |
| v9 | *Liver Disorders* | Alkaline phosphatase | 345 | 0.750 | 0.700 |
| v10 | *Liver Disorders* | Alanine aminotransferase | 345 | 3.04 | 13.50 |

Var.: Variables, $g_3$: Skewness, $g_4$: Kurtosis, $n$: Sample size.

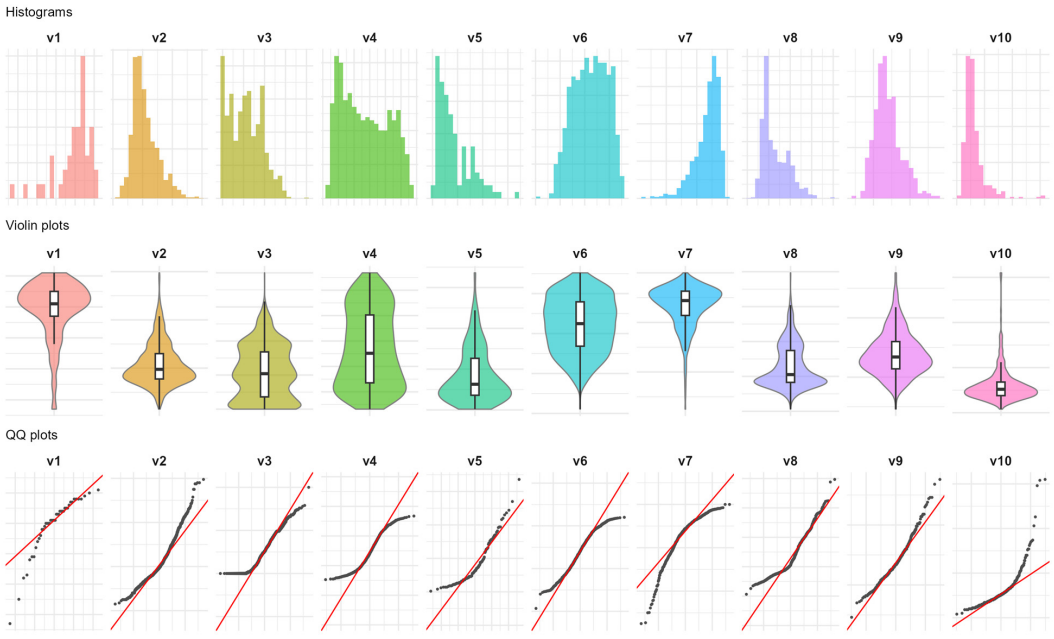



**Figure 2.** Distributional visualizations of the real-world data.

Based on the simulated and real datasets, the comparative performance of OSKT and a set of commonly applied transformation methods was evaluated through moment-based metrics, the Pearson $p$-statistic, and eight normality tests, as described in the following section.

2.2.3. Performance Evaluation of Normality Transformation Methods

Normality tests are statistical tools developed to assess the level of conformity of a dataset to a normal distribution. In data preprocessing workflows, normality tests play a critical dual role in which they guide the selection of appropriate normalization techniques for the problem and provide quantitative evidence of the performance of the applied transformation. In the literature, a wide range of statistical procedures has been developed to test normality. For example, Arnastauskaitė et al. [44] evaluated forty different normality tests under varying sample sizes and distributional conditions in a recent comprehensive benchmark study.

Normality tests can be categorized into three or four principal methodological groups, each designed to capture different forms of deviation from the Gaussian distribution [37,38,45]. Firstly, the frequency-based methods such as the Pearson chi-square test [46] assess deviations by comparing observed and expected bin frequencies derived from a histogram representation of the data. Among these, the most commonly used metric is the Pearson P/df (PPM), which has also been employed as the main performance criterion in the *bestNormalize* package [47] in R. PPM is not a *p*-value but a measure of deviation from normality. It is calculated by dividing the Pearson chi-square statistic by its degrees of freedom ($df = k - 3$, where $k$ is the number of bins). The number of bins is generally determined as $k = \left\lfloor 2\,n^{2/5} \right\rfloor$, following the Terrell–Scot rule. Smaller values of PPM indicate that the data more closely approximates a normal distribution. By scaling the chi-square statistic in this way, the metric provides a comparable normality score across different transformations and sample sizes, avoiding the inflation of the raw chi-square statistic due to larger $n$ or more bins.

As another group of normality tests, the empirical distribution function (EDF)-based tests examine discrepancies between the empirical and theoretical cumulative distribution functions, providing sensitivity across the full range of observations, particularly in the tails. Regression or correlation-based procedures evaluate linearity under normal probability plots and mostly rely on correlation coefficients as test statistics. Finally, the moment-based tests assess skewness and kurtosis to quantify asymmetry and tail extremity.

Mathematical formulations and statistical properties of normality tests have been comprehensively studied in the literature [37,44]. These are details that are outside the scope of the present study; thus, an exhaustive methodological review is not provided here. Instead, several commonly used and recently developed univariate normality tests are summarized in Table 4, highlighting their primary approaches and sensitivity characteristics. Performance of normality tests is highly influenced by sample size, skewness, and kurtosis. The "sensitivity" column in Table 4 reflects each normality test's ability to detect deviations from a normal distribution, pointing to the statistical power of the test. Tests with high sensitivity can detect even small or moderate departures from normality, whereas those with medium sensitivity can detect some deviations but might miss minor or specific irregularities in the distribution. Tests with low sensitivity detect only prominent deviations and perform poorly at detecting minor departures from normality.

**Table 4.** Types and comparative features of univariate normality test methods.

| Test | Type | Approach | Sensitivity | Description |
|---|---|---|---|---|
| PPM | FB | Compares observed versus expected frequencies in histogram bins | Medium | Captures general deviations; simple to compute |
| SW | RCB | Correlation between ordered sample and normal scores | High | Powerful for small and medium samples; captures skewness and kurtosis |
| SF | RCB | Variant of SW optimized for larger samples | Medium | Similar to SW but computationally easier for large samples |
| JB | MB | Uses skewness and kurtosis standardized moments | Medium | Skewness and kurtosis-focused; less sensitive to tails |
| RJB | MB | Robust version of JB using median-based scale | Medium–High | Less sensitive to outliers; improves testing under contaminated data |
| DAG | MB | Combines skewness and kurtosis with transformation | Low–Medium | Tolerates outliers; less sensitive to tail deviations |
| AD | EDF | Weighted CVM, gives more weight to tails | High | Sensitive to tails; good for extreme values |
| KS | EDF | Maximum difference between ECDF and CDF | Medium | Sensitive to both central and some tail deviations |
| LIL | EDF | Modification of KS; accounts estimated mean and variance | Medium | Robust for small or medium samples |
| CVM | EDF | Integral of squared differences between ECDF and CDF | High | Considers all points of ECDF; detects center and tail deviations |
| ZA | EDF | Log-transformed ECDF | High | Sensitive to both center and tail deviations; uses $p$-value via Monte Carlo simulations |
| ZC | EDF | Weighted ECDF | High | Enhances sensitivity to both center and tail; uses $p$-value via Monte Carlo simulations |

CDF: Cumulative distribution function; CVM: Cramér–von Mises; ECDF: Empirical cumulative distribution function; EDF: Empirical distribution function; FB: Frequency-based; MB: Moment-based; RCB: Regression/correlation-based, PPM: Pearson P/df statistics, SW: Shapiro–Wilk; SF: Shapiro–Francia; JB: Jarque–Bera; RJB: Robust Jarque–Bera; DAG: D'Agostino omnibus; AD: Anderson–Darling; KS: Kolmogorov–Smirnov; LIL: Liliefors;; ZA: Zhang–Wu 1; ZC: Zhang–Wu 2.

Based on the classification and sensitivity of the normality tests presented in Table 4, commonly used normality tests, such as Shapiro–Wilk (SW) [48], robust Jarque–Bera (RJB) [49], D'Agostino omnibus (DAG) [50], Anderson–Darling (AD) [29], Kolmogorov–Smirnov (KS) [51,52], and Liliefors (LIL) [53] including Pearson P/df statistics (PPM) [46,47], were selected and applied in this study to evaluate the performance of the examined transformation methods. We also included Zhang–Wu 1 (ZA) [54] and Zhang–Wu 2 (ZC) [55] tests since Uhm and Yi [56] suggested the ZC test when data is small in size and the ZA test when data is of moderate size for asymmetric distributions. This selection covers a wide range of methodological categories and varying levels of sensitivity, comprehensively assessing the performance of transformation methods in achieving approximate normality.

2.2.4. Computational Tools and Setting

The analyses were conducted with R version 4.5.1 [57] on a laptop computer that has an Intel i9 CPU and 64 GB of RAM. A custom R script was developed for the implementation of the normality transformations. In this script, the transformations were carried out using the functions "sqrt_x", "log_x", "arcsinh_x", "boxcox", "yeojohnson", and "lambert" from the *bestNormalize* package [46]. The proposed method, OSKT, and its corresponding back transformation were implemented in R and Matlab (version 2025b), named "oskt" and "backoskt", respectively (Appendices A and B). For the function calls, the initial values for the parameters g and h were set to (0.1, 0.1), with lower and upper bounds assigned as $(-1, 0)$ and $(1, 0.5)$, respectively.

To assess the performance of the proposed method, OSKT, in normalizing data, we applied eight broadly used statistical tests for evaluating normality of the transformed datasets. These tests consider different aspects of deviation from the normal distribution, including skewness, kurtosis, cumulative distribution differences, and overall goodness of fit. SW and KS tests were performed using the base R functions "shapiro.test" and "ks.test". The PPM, LIL, and AD tests were respectively carried out using the "pearson.test", "lillie.test", and "ad.test" functions from the *nortest* package [58]. The D'Agostino omnibus test DAG was executed using the "dagoTest" function from the *fBasics* package [59]. Custom R functions were developed for the ZA, ZC, and robust RJB tests. For the ZA and ZC tests, the *p*-values were computed via 10,000 Monte Carlo runs.

As an overall performance criterion, the number of tests confirming normality (Nt) at a significance level of $\alpha = 0.05$ was used, corresponding to the null hypothesis "$H_0$: the transformed data follow a normal distribution". In all corresponding computations, we used excess kurtosis because the kurtosis for a standard normal distribution is three. To visually inspect the normality results, the histograms, violin plots and density plots were plotted using the *ggplot2* package [60].

To assess of the scalability and computational demands of the tested normalization methods, execution times were measured on simulated data of increasing sizes from $10^3$ to $10^7$ observations. The data were generated using the *g*–*h* distribution with the parameters of $A = 0$, $B = 1$, $g = 0.36$, $h = 0$ and standardized before transformation. For each sample size, the elapsed time (μs) to apply each normalization method was recorded using the system clock. The logarithms of the recorded computation times were used for visualization and comparison in order to ensure simplicity and clarity in the presentation.

For the back transformation analyses, the simulated variables above were generated with $10^7$ observations and subjected to the OSKT. The transformed values obtained for each variable were then recovered to their original values using a custom R function coded for back transformation. Root mean squares error (RMSE) was calculated from the differences between the original values and the recovered values by both the Newton–Raphson (NR) and uniroot (UR) methods, and the computation times (s) for the back transformation were also calculated.

## 3. Results and Discussion

### *3.1. Analyses for Simulated Data*

Normality test results for the examined transformations of the simulated variable v1 are presented in Table 5. As shown in Figure 3, the original data (ORG) exhibited mild deviations from normality, with a moderate negative skewness ($g_3 = -0.524$) and slightly elevated kurtosis ($g_4 = 0.635$). Although the KS test did not reject normality ($p = 0.054$), other tests strongly rejected it ($p < 0.05$), indicating that the distribution deviated from normality mainly because of the tail behavior.

**Table 5.** Results of normality tests for the transformed data of simulated variable 1 (v1).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | −0.524 | 0.635 | 1.523 | 0.054 | <0.001 | <0.001 | 0.006 | <0.001 | 0.046 | 0.000 | <0.001 | 0 |
| **SQR** | −1.653 | 7.975 | 3.086 | 0.007 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LOG** | −12.380 | 218.006 | 14.436 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | −2.178 | 10.830 | 4.214 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **BC** | −0.210 | −0.133 | 1.368 | 0.783 | 0.030 | 0.146 | 0.170 | 0.141 | 0.371 | 0.062 | 0.174 | 7 |
| **ABC** | −0.039 | −0.356 | 1.441 | 0.652 | 0.137 | 0.385 | 0.136 | 0.192 | 0.210 | 0.345 | 0.432 | 8 |
| **RBC** | −0.179 | −0.183 | 1.241 | 0.760 | 0.050 | 0.215 | 0.187 | 0.200 | 0.340 | 0.098 | 0.204 | 7 |
| **YJ** | −0.035 | −0.322 | 1.518 | 0.657 | 0.164 | 0.457 | 0.132 | 0.285 | 0.215 | 0.391 | 0.477 | 8 |
| **RYJ** | −0.088 | −0.266 | 1.318 | 0.696 | 0.140 | 0.439 | 0.182 | 0.350 | 0.257 | 0.276 | 0.454 | 8 |
| **LMB** | **0.000** | −0.119 | **1.027** | 0.674 | 0.271 | 0.934 | 0.115 | 0.933 | 0.233 | 0.679 | 0.718 | 8 |
| **OSKT** | −0.183 | **0.090** | 1.032 | 0.834 | 0.051 | 0.229 | 0.253 | 0.199 | 0.454 | 0.147 | 0.179 | 8 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero, and the lowest Pearson P statistic value.

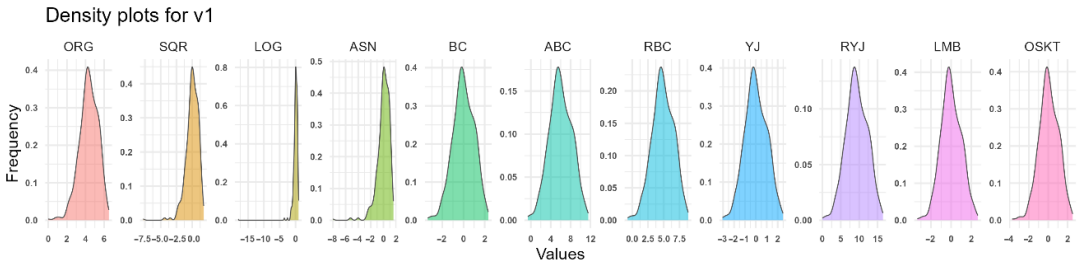



**Figure 3.** Density plots of the transformed data of simulated variable 1 (v1).

The basic transformation methods, such as SQR, LOG, and ASN, further distorted the distribution. These transformations aggravated the asymmetry and peakedness, resulting in extreme skewness and kurtosis (e.g., LOG: $g_3 = -12.380, g_4 = 218.006$) and completely rejecting normality in all tests ($Nt = 0$). Thus, simple monotonic transformations tended to overcorrect the data, moving it further from the Gaussian form.

Compared to the basic methods, flexible power-based transformations showed significant improvements in symmetry and tail behavior. For example, the BC transformation reduced both skewness ($g_3 = -0.210$) and kurtosis ($g_4 = -0.133$), achieving acceptance of normality in seven out of eight tests ($Nt = 7$). Meanwhile, the transformation methods ABC and RBC yielded slightly better results, with ABC achieving full acceptance ($Nt = 8$), having moment statistics closer to zero ($g_3 = -0.039$, $g_4 = -0.356$). Similarly, YJ and its robust variant, RYJ, both achieved eight accepted tests ($Nt = 8$), hence confirming that reweighted estimation and flexible parameterization improve the distributional fit even for moderately skewed data.

Within the compared methods, LMB and OSKT gave the most satisfactory results by having moment statistics closest to zero and the most consistent no-rejection in all normality tests ($Nt = 8$). In particular, LMB showed a near-zero skewness ($g_3 = 0.000$), while OSKT yielded the lowest kurtosis value ($g_4 = 0.09$). Therefore, both LMB and OSKT can be considered marginally superior to YJ in achieving distributions that are almost symmetric and mesokurtic. As shown in Figure 3, these transformations generated bell-shaped densities closely matching the theoretical normal curve, while basic transformations such as LOG or ASN led to severely distorted, heavy-tailed shapes.

Normality test results for the transformed data of the simulated variable v2 can be viewed in Table 6, and the corresponding density plots are presented in Figure 4. The original v2 data revealed a distinctly right-skewed and leptokurtic distribution, with $g_3 = 1.000$ and $g_4 = 1.361$. In addition, normality was observed to be rejected in all

statistical tests ($Nt = 0$), indicating a huge deviation from a Gaussian distribution. Continuing with the classical methods, SQR reduced both asymmetry and excess kurtosis ($g_3 = 0.079$, $g_4 = 0.178$), with eight tests confirming normality ($Nt = 8$). ASN showed moderate improvement ($g_3 = -0.062$, $g_4 = -0.359$), and normality was confirmed by eight tests. In contrast to others, LOG overcorrected the right tail, resulting in an extreme negative skewness ($g_3 = -2.699$) and a highly inflated kurtosis ($g_4 = 19.373$), and therefore failed to normalize the data ($Nt = 0$).

**Table 6.** Results of normality tests for the transformed data of simulated variable 2 (v2).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | 1.000 | 1.361 | 3.809 | 0.005 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **SQR** | 0.079 | 0.178 | 1.405 | 0.981 | 0.669 | 0.532 | 0.421 | 0.478 | 0.861 | 0.907 | 0.979 | 8 |
| **LOG** | −2.699 | 19.373 | 4.073 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | −0.062 | −0.359 | 1.182 | 0.781 | 0.270 | 0.311 | 0.473 | 0.168 | 0.368 | 0.494 | 0.459 | 8 |
| **BC** | **0.000** | 0.235 | **0.950** | 0.982 | 0.690 | 0.568 | 0.536 | 0.492 | 0.864 | 0.884 | 0.975 | 8 |
| **ABC** | 0.033 | 0.208 | 1.182 | 0.984 | 0.712 | 0.601 | 0.509 | 0.528 | 0.872 | 0.914 | 0.982 | 8 |
| **RBC** | −0.015 | 0.250 | 1.036 | 0.982 | 0.664 | 0.527 | 0.546 | 0.455 | 0.860 | 0.871 | 0.945 | 8 |
| **YJ** | **0.000** | −0.219 | 1.632 | 0.916 | 0.719 | 0.710 | 0.854 | 0.657 | 0.634 | 0.895 | 0.831 | 8 |
| **RYJ** | −0.002 | −0.219 | 1.550 | 0.924 | 0.719 | 0.710 | 0.854 | 0.657 | 0.636 | 0.895 | 0.831 | 8 |
| **LMB** | **0.000** | −0.128 | 1.259 | 0.970 | 0.838 | 0.913 | 0.877 | 0.917 | 0.809 | 0.971 | 0.935 | 8 |
| **OSKT** | 0.077 | **−0.026** | 1.577 | 0.897 | 0.575 | 0.772 | 0.845 | 0.771 | 0.595 | 0.653 | 0.712 | 8 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

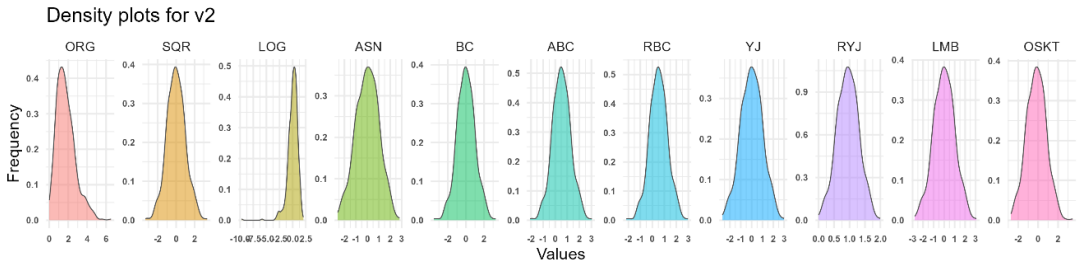



**Figure 4.** Density plots of the transformed data of simulated variable 2 (v2).

Scrolling through the simulated variable v2 results, it is seen that power- and moment-targeting transformations performed better. For example, BC and YJ achieved near-zero skewness ($g_3 \approx 0$) and were accepted by all normality tests ($Nt = 8$). Among the extended versions of BC/YJ, ABC produced a small skewness ($g_3 = 0.033$) and a moderate kurtosis ($g_4 = 0.208$) with full test acceptance ($Nt = 8$). RBC exhibited a low negative skewness ($g_3 = -0.015$) and a positive kurtosis ($g_4 = 0.250$) and was also accepted by eight tests ($Nt = 8$). RYJ resulted in a very small negative skewness ($g_3 = -0.002$) and a modest negative kurtosis ($g_4 = -0.219$) and was accepted by all tests, too ($Nt = 8$).

The LMB method produced perfect skewness ($g_3 = 0$) and slightly negative kurtosis ($g_4 = -0.128$) and was also accepted by all tests ($Nt = 8$). The OSKT also yielded a small skewness ($g_3 = 0.077$) and the smallest kurtosis ($g_4 = -0.026$), showing full acceptance by all tests ($Nt = 8$).

Onwards with the transformed data of the simulated variable v3, the corresponding normality test results are presented in Table 7, and density plots are shown in Figure 5. The original v3 distribution displayed a prominent right skewness ($g_3 = 1.510$) and leptokurtosis ($g_4 = 4.713$), indicating a substantial deviation from normality ($Nt = 0$). The density plot confirms this with a long right tail and reflects the heavy skew and peakedness of the raw data.

**Table 7.** Results of normality tests for the transformed data of simulated variable 3 (v3).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | 1.510 | 4.713 | 5.036 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **SQR** | 0.405 | 0.806 | 1.554 | 0.119 | <0.001 | <0.001 | <0.001 | <0.001 | 0.002 | 0.020 | 0.004 | 1 |
| **LOG** | −3.169 | 28.134 | 2.223 | 0.027 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | 0.194 | −0.214 | 1.114 | 0.672 | 0.098 | 0.146 | 0.044 | 0.137 | 0.230 | 0.472 | 0.519 | 7 |
| **BC** | 0.111 | 1.004 | 1.145 | 0.345 | <0.001 | <0.001 | 0.010 | 0.002 | 0.036 | 0.040 | 0.004 | 1 |
| **ABC** | 0.187 | 0.895 | 1.209 | 0.271 | <0.001 | 0.001 | 0.004 | 0.002 | 0.020 | 0.042 | 0.005 | 1 |
| **RBC** | −0.095 | 1.505 | 1.046 | 0.534 | <0.001 | <0.001 | 0.040 | <0.001 | 0.116 | 0.023 | <0.001 | 2 |
| **YJ** | 0.003 | −0.082 | 0.832 | 0.758 | 0.606 | 0.902 | 0.468 | 0.984 | 0.335 | 0.834 | 0.904 | 8 |
| **RYJ** | −0.041 | **−0.070** | 0.827 | 0.714 | 0.568 | 0.855 | 0.546 | 0.924 | 0.269 | 0.773 | 0.832 | 8 |
| **LMB** | **0.000** | 0.501 | 0.932 | 0.769 | 0.048 | 0.199 | 0.259 | 0.111 | 0.351 | 0.284 | 0.148 | 7 |
| **OSKT** | 0.226 | 1.384 | **0.768** | 0.807 | <0.001 | <0.001 | 0.579 | <0.001 | 0.403 | 0.065 | <0.001 | 4 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

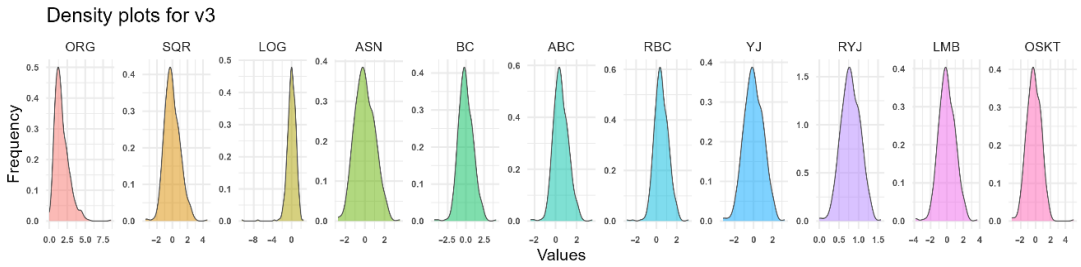



**Figure 5.** Density plots of the transformed data of simulated variable 3 (v3).

Classical transformations showed varying performances for the simulated variable v3. For example, SQR had partially reduced skewness ($g_3 = 0.405$) and kurtosis ($g_4 = 0.806$), with one normality test accepting the transformed data ($Nt = 1$), which was somewhat an improvement but only a limited one. LOG overcorrected the distribution again, producing an extreme negative skewness ($g_3 = -3.169$) and a highly inflated kurtosis ($g_4 = 28.134$), and failed to normalize the data ($Nt = 0$). It was interesting how ASN achieved a much more balanced adjustment, producing nearly symmetric and slightly platykurtic data ($g_3 = 0.194$, $g_4 = -0.214$) that passed seven normality tests ($Nt = 7$), making it the best-performing classical method.

Like their classical counterparts, power-based transformations also showed a variety of performances. BC was modestly symmetric ($g_3 = 0.111$) and still showed high kurtosis ($g_4 = 1.004$) and only passed one test ($Nt = 1$). Its extension, ABC, slightly improved kurtosis ($g_4 = 0.895$) but received low acceptance ($Nt = 1$). RBC revealed a small negative skew ($g_3 = -0.095$) but a high kurtosis ($g_4 = 1.505$), passing only two tests ($Nt = 2$). However, YJ and RYJ transformations were much more successful this time. YJ produced an almost perfect normality ($g_3 = 0.003$, $g_4 = -0.082$) and was confirmed by all eight tests ($Nt = 8$), while RYJ achieved similar moment statistics ($g_3 = -0.041$, $g_4 = -0.070$) with full test acceptance ($Nt = 8$).

Moment-targeting methods demonstrated generally solid but varying performances. LMB reduced skewness to zero ($g_3 = 0$) and moderately decreased kurtosis ($g_4 = 0.501$), passing seven tests ($Nt = 7$). OSKT had a modest symmetry ($g_3 = 0.226$), retained its relatively high kurtosis ($g_4 = 1.384$), but was accepted by four tests ($Nt = 4$). However, OSKT was distinguished by achieving the lowest Pearson P-statistic ($PPM = 0.768$), suggesting a good overall approximation to normality if the medium test acceptance was ignored.

The density plots in Figure 5 illustrate the numerical results in Table 7, confirming the enhanced symmetry and improved tail behavior achieved through the more effective transformations. As can be seen clearly, the YJ and LMB transformations produce approximately symmetric and mesokurtic distributions that closely resemble the theoretical normal curve. The OSKT-transformed data did not pass the DAG, SW, RJB, and ZC tests, possibly due to a slightly thin right tail, but produced the lowest PPM value among all methods. These tests show that, despite a favorable overall alignment of frequencies as in the case of PPM, small discrepancies in shape can still matter. Thus, it can be concluded that, although PPM can serve as a valuable diagnostic indicator, it should be used with caution as a criterion because minor tail deviations may be overlooked.

Table 8 presents the analyses with simulated variable v4. Overall, the results showed a notable deviation from normality. The original distribution was clearly left-skewed ($g_3 = -1.524$) and leptokurtic ($g_4 = 3.156$). All eight normality tests rejected the null hypothesis of normality ($Nt = 0$), signaling a significant departure from the Gaussian shape. This is also visually verified in Figure 6, where the density plot shows the existence of a left tail.

**Table 8.** Results of normality tests for the transformed data of simulated variable 4 (v4).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | −1.524 | 3.156 | 5.995 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **SQR** | −2.796 | 13.425 | 9.536 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LOG** | −12.240 | 205.634 | 32.154 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | −3.449 | 18.351 | 13.832 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **BC** | −0.684 | **−0.021** | 3.054 | 0.005 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ABC** | −0.241 | −0.718 | 1.859 | 0.363 | <0.001 | 0.003 | <0.001 | <0.001 | 0.041 | <0.001 | 0.001 | 1 |
| **RBC** | −0.336 | −0.621 | 2.136 | 0.217 | <0.001 | 0.001 | <0.001 | <0.001 | 0.010 | <0.001 | <0.001 | 1 |
| **YJ** | −0.256 | −0.680 | 1.895 | 0.384 | <0.001 | 0.004 | <0.001 | <0.001 | 0.048 | 0.000 | 0.001 | 1 |
| **RYJ** | −0.395 | −0.519 | 2.209 | 0.135 | <0.001 | 0.000 | <0.001 | <0.001 | 0.003 | <0.001 | 0.004 | 1 |
| **LMB** | **0.000** | 0.113 | **0.536** | 0.894 | 0.907 | 0.784 | 0.923 | 0.777 | 0.581 | 0.964 | 0.939 | 8 |
| **OSKT** | −0.172 | 0.054 | 0.596 | 0.916 | 0.110 | 0.294 | 0.712 | 0.257 | 0.634 | 0.218 | 0.241 | 8 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

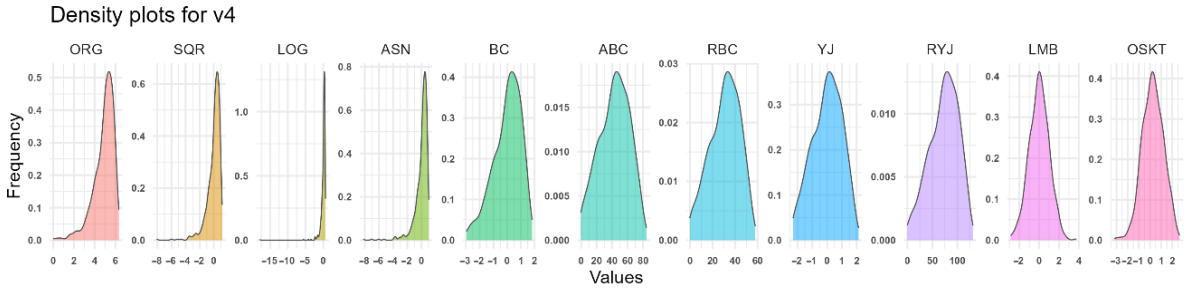



**Figure 6.** Density plots of the transformed data of simulated variable 4 (v4).

The classical transformations SQR, LOG, and ASN were not efficient in correcting negative skewness. SQR increased the asymmetry, yielding a skewness of $g_3 = -2.796$ and a kurtosis of $g_4 = 13.425$. LOG overcorrected the data, resulting in an extreme negative skewness ($g_3 = -12.240$) and an ultimately high kurtosis ($g_4 = 205.634$). Similarly, ASN performed poorly, producing a highly distorted distribution ($g_3 = -3.449$; $g_4 = 18.351$). As a result, none of the classical transformations could pass the normality tests ($Nt = 0$).

In comparison to the classical methods, power transformations were somewhat more efficient. For example, the BC method improved the left skewness ($g_3 = -0.684$) and brought kurtosis close to zero ($g_4 = -0.021$). Despite these values, BC's normality was

rejected by all tests ($Nt = 0$). ABC and RBC had reductions in skewness to some extent (ABC: $g_3 = -0.241$; RBC: $g_3 = -0.336$) but both were accepted by only one test, $Nt = 1$. YJ's results improved both in terms of symmetry and tail shape ($g_3 = -0.256; g_4 = -0.680$), but its normality tests were not as impressive ($Nt = 1$). RYJ had close results to the YJ ($g_3 = -0.395; g_4 = -0.519$) but also had low acceptance ($Nt = 1$).

On the other hand, the moment-targeting transformations LMB and OSKT were quite effective. LMB had a perfectly symmetric distribution ($g_3 = 0.000$) with near-mesokurtic characteristics ($g_4 = 0.113$) and was confirmed by all normality tests ($Nt = 8$). The OSKT also performed very well, with a highly symmetric ($g_3 = -0.172$) and mesokurtic ($g_4 = 0.054$) distribution, achieving $Nt = 8$ in the normality tests. Both LMB and OSKT showed the lowest PPM statistics, which indicates their superiority in restoring normality.

Overall, the results showed the inefficiency of classical monotonic transformations in normalizing highly left-skewed or leptokurtic data. For example, even when the data were shifted towards positive values to ensure their advantage, SQR, LOG, ASN, and BC transformations still failed frequently. In comparison, power transformation methods BC, ABC, RBC, and YJ offered some improvements but also focused on right-skew correction which, in the end, made their performance limited when dealing with left-skewed data.

As for moment-based transformations LMB and OSKT, the explicit adjustment of skewness and kurtosis allows them to maintain symmetry and mesokurtosis regardless of direction. Therefore, they achieve almost perfect results from the normality tests. These results verify the superiority of moment-targeting methods in handling highly asymmetric or heavy-tailed data over the classical methods.

With the next simulated variable, v5, it was aimed to assess the effectiveness of the compared transformation methods on negatively skewed and leptokurtic data. Here, it was questioned whether moderate reductions in skewness and kurtosis could enhance the performance of flexible parametric transformations, such as YJ and its variants. It was also investigated whether moment-targeting approaches, LMB and OSKT, would maintain their superior normalizations under less extreme conditions.

As shown in Table 9, the original v5 data exhibited a distinctly left-skewed ($g_3 = -1.065$) and leptokurtic ($g_4 = 3.997$) distribution. All statistical tests rejected the assumption of normality ($Nt = 0$), confirming a substantial deviation from the Gaussian form. Consistent with the results obtained for v4, the traditional transformations SQR, LOG, and ASN were ineffective for correcting negative skewness. Instead, they exacerbated the distributional asymmetry and sharply increased kurtosis. For example, the LOG transformation produced an extreme skewness of $g_3 = -15.516$ and a kurtosis of $g_4 = 293.438$, indicating a severe overcorrection. Accordingly, none of these monotonic transformations achieved normality ($Nt = 0$).

The performance of power-based and moment-targeting transformations showed a more favorable trend. The BC transformation moderately improved the data structure ($g_3 = -0.221$, $g_4 = 0.824$), resulting in normality acceptance by four tests ($Nt = 4$). Its adjusted variants, ABC and RBC, achieved further enhancement in both skewness and kurtosis, with ABC ($g_3 = 0.050$, $g_4 = 0.593$) accepted by seven tests ($Nt = 7$), and RBC ($g_3 = -0.027$, $g_4 = 0.624$) accepted by six tests ($Nt = 6$). These outcomes suggest that ABC can substantially mitigate negative skewness when appropriately parameterized.

YJ and RYJ transformations were similarly successful in normalization. YJ yielded nearly symmetric and moderately leptokurtic data ($g_3 = 0.049$, $g_4 = 0.662$) and was accepted by five tests ($Nt = 5$). RYJ produced very close results ($g_3 = 0.134$, $g_4 = 0.653$), with five tests confirming normality ($Nt = 5$). They both outperformed the classical BC transformation, though their performances remained slightly below those of ABC and RBC. Among the moment-based methods, LMB achieved perfect symmetry ($g_3 = 0.000$)

but retained excess kurtosis ($g_4 = 1.588$), with normality acceptance by only one test ($Nt = 1$). In contrast, the OSKT performed outstandingly, achieving a balanced and nearly ideal distribution ($g_3 = 0.111$, $g_4 = -0.095$), with all eight normality tests confirming its effectiveness ($Nt = 8$). To sum up, OSKT once again stood out as the most useful approach, achieving full normalization with minimal residual deviation in moment statistics. The density plots in Figure 7 confirmed these results by showing that OSKT resembled the theoretical normal curve for v5.

The next simulated variable, v6, represents one of the most complex cases examined in this study as summarized in Table 10. The original data exhibited near-perfect symmetry ($g_3 = 0.008$) but pronounced platykurtosis ($g_4 = -1.035$), indicating a bimodal distribution pattern, as clearly illustrated in Figure 8. This structural deviation reflects a severe departure from the mesokurtic form of a normal curve, characterized by an underrepresentation of central density and overextended tails.

**Table 9.** Results of normality tests for the transformed data of simulated variable 5 (v5).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | −1.065 | 3.997 | 2.104 | 0.032 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **SQR** | −3.248 | 25.490 | 4.304 | 0.000 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LOG** | −15.516 | 293.438 | 27.936 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | −4.777 | 43.814 | 7.482 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **BC** | −0.221 | 0.824 | 1.132 | 0.588 | 0.004 | <0.001 | 0.007 | 0.002 | 0.155 | 0.176 | 0.112 | 4 |
| **ABC** | 0.050 | 0.593 | 0.823 | 0.727 | 0.072 | 0.021 | 0.084 | 0.054 | 0.295 | 0.458 | 0.527 | 7 |
| **RBC** | −0.027 | 0.624 | 0.846 | 0.707 | 0.057 | 0.014 | 0.053 | 0.047 | 0.269 | 0.425 | 0.538 | 6 |
| **YJ** | 0.049 | 0.662 | 0.904 | 0.707 | 0.050 | 0.009 | 0.068 | 0.034 | 0.270 | 0.387 | 0.436 | 5 |
| **RYJ** | 0.134 | 0.653 | 0.859 | 0.723 | 0.038 | 0.005 | 0.081 | 0.018 | 0.289 | 0.358 | 0.308 | 5 |
| **LMB** | **0.000** | 1.588 | 1.232 | 0.381 | <0.001 | <0.001 | 0.001 | <0.001 | 0.047 | 0.020 | 0.002 | 1 |
| **OSKT** | 0.111 | **−0.095** | **0.777** | 0.971 | 0.117 | 0.561 | 0.617 | 0.574 | 0.805 | 0.092 | 0.303 | 8 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

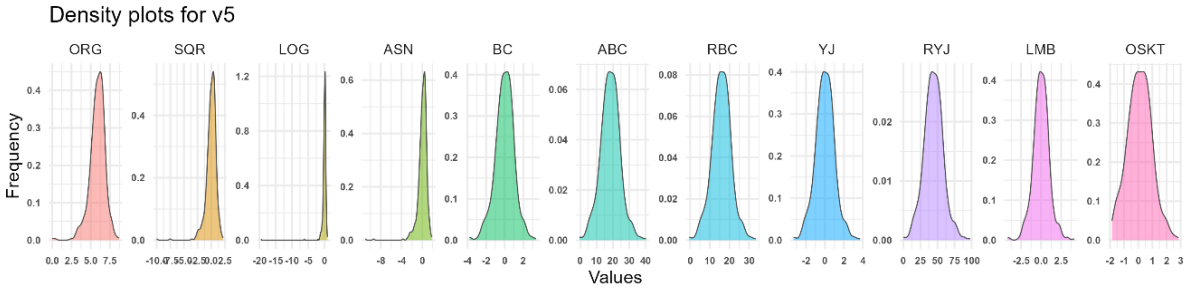



**Figure 7.** Density plots of the transformed data of simulated variable 5 (v5).

The classical monotonic transformations SQR, LOG, and ASN were unable to correct the structural irregularity of v6. SQR and ASN slightly altered the shape but failed to meaningfully improve kurtosis, while LOG introduced excessive negative skewness ($g_3 = -4.888$) and extreme kurtosis ($g_4 = 57.460$), further distorting the data. As a result, all three traditional transformations were rejected by all normality tests ($Nt = 0$).

The power-based transformations were limited in performance. Despite minor numerical adjustments to skewness and kurtosis, none could restore a mesokurtic structure. Their skewness values remained slightly negative (e.g., BC: $g_3 = -0.183$; YJ: $g_3 = -0.082$), and kurtosis stayed well below zero ($g_4 \approx -1.0$). They also failed in all normality tests ($Nt = 0$).

**Table 10.** Results of normality tests for the transformed data of simulated variable 6 (v6).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | 0.008 | −1.035 | 4.523 | 0.003 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **SQR** | −0.463 | −0.160 | 4.677 | <0.001 | <0.001 | 0.000 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LOG** | −4.888 | 57.460 | 8.668 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | −0.581 | **−0.322** | 6.136 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **BC** | −0.183 | −0.873 | 4.196 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ABC** | −0.131 | −0.939 | 4.332 | 0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.003 | 0 |
| **RBC** | −0.163 | −0.901 | 4.177 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.003 | 0 |
| **YJ** | −0.082 | −1.003 | 4.523 | 0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.002 | 0 |
| **RYJ** | −0.087 | −1.001 | 4.514 | 0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LMB** | **0.000** | −1.036 | 4.318 | 0.002 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.002 | 0 |
| **OSKT** | −0.041 | 0.665 | **2.054** | 0.412 | 0.012 | 0.223 | 0.021 | 0.034 | 0.058 | 0.255 | 0.089 | 5 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

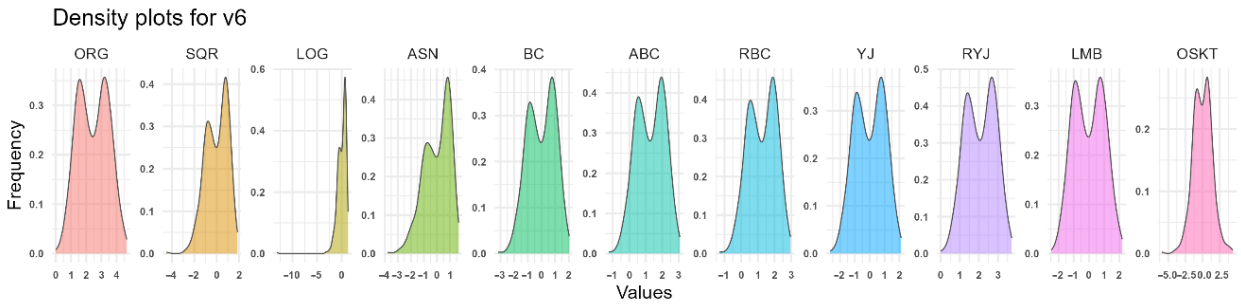



**Figure 8.** Density plots of the transformed data of simulated variable 6 (v6).

The moment-based method LMB maintained perfect symmetry ($g_3 = 0.000$) but did not overcome the flat-tailed nature of the data ($g_4 = -1.036$) and completely rejected normality ($Nt = 0$). However, the OSKT method demonstrated a striking performance. It simultaneously addressed both skewness and kurtosis, achieving a well-balanced distribution ($g_3 = -0.041$; $g_4 = 0.665$). This reflected as a substantial statistical improvement, with five out of eight normality tests accepting the transformed data ($Nt = 5$), which was the highest performance among all methods tested.

All of the results above are visually represented by the density plots in Figure 8. It can be seen that OSKT could reshape the original flattened, bimodal pattern into a smooth, symmetric, and moderately peaked distribution that nearly resembles the theoretical curve. The analyses with variable v6 demonstrated OSKT's strength against more complex structures such as platykurtosis or bimodality, owing to direct optimization of high-order moments and an adaptive mechanism. OSKT simultaneously corrects both skewness and kurtosis, which is better for overall shape balance compared to single-parameter optimization used by its competitors, YJ and BC. The other moment-based method, LMB, is constrained in handling unimodal frameworks and cannot capture multimodal features.

Table 11 presents the results for the simulated variable v7, revealing only moderate deviations from normality in the original data ($g_3 = 0.401; g_4 = -0.008$), but all statistical tests rejected the null hypothesis of normality ($Nt = 0$). Therefore, even small departures in skewness and kurtosis can lead to statistically significant non-normality when sample sizes are large or when the data deviate from the assumptions of continuous normality.

**Table 11.** Results of normality tests for the transformed data of simulated variable 7 (v7).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | 0.401 | −0.008 | 66.536 | <0.001 | <0.001 | <0.001 | <0.001 | 0.002 | <0.001 | <0.001 | <0.001 | 0 |
| **SQR** | −0.904 | 1.471 | **65.414** | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LOG** | −3.768 | 14.214 | 68.186 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | −0.610 | 0.095 | **65.414** | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **BC** | −0.735 | 1.038 | **65.414** | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ABC** | 0.067 | **−0.085** | 66.536 | <0.001 | <0.001 | 0.750 | <0.001 | 0.810 | <0.001 | <0.001 | 0.002 | 2 |
| **RBC** | −0.448 | 0.446 | **65.414** | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **YJ** | −0.017 | −0.206 | 66.536 | <0.001 | <0.001 | 0.975 | <0.001 | 0.693 | <0.001 | <0.001 | <0.001 | 2 |
| **RYJ** | 0.006 | −0.207 | 66.536 | <0.001 | <0.001 | 0.984 | <0.001 | 0.697 | <0.001 | <0.001 | <0.001 | 2 |
| **LMB** | **0.000** | −0.263 | 66.536 | <0.001 | <0.001 | 0.907 | <0.001 | 0.501 | <0.001 | <0.001 | 0.001 | 2 |
| **OSKT** | 0.087 | −0.135 | 66.536 | <0.001 | <0.001 | 0.697 | <0.001 | 0.651 | <0.001 | <0.001 | <0.001 | 2 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

The classical methods SQR, LOG, and ASN failed to improve normality and even intensified the distortions in some cases. For instance, SQR transformation increased skewness and kurtosis (SQR: $g_3 = -0.904$, $g_4 = 1.471$), while LOG resulted in an over-correction, producing extreme negative skewness ($g_3 = -3.768$) and inflated kurtosis ($g_4 = 14.214$). None of these methods achieved acceptance by any normality test ($Nt = 0$), reaffirming that classical approaches are not effective for modest or structurally induced asymmetry. Similarly, BC also increased skewness and kurtosis, rejecting all normality tests ($g_3 = -0.735$, $g_4 = 1.038$).

On the other hand, ABC, YJ, RYJ, LMB, and OSKT methods brought more balanced results. ABC was almost symmetrically perfect ($g_3 = 0.067; g_4 = -0.085$) and returned two tests confirming normality ($Nt = 2$). RYJ reduced skewness ($g_3 \approx 0$) and improved kurtosis ($g_4 \approx -0.21$), also resulting in $Nt = 2$. The moment-targeting LMB and OSKT performed closely with each other, both producing near-zero skewness (LMB: $g_3 = 0.000$; OSKT: $g_3 = 0.087$) and modestly negative kurtosis ($g_4 = -0.135$) and $Nt = 2$. Although these improvements were limited by the statistical test results, the distributions displayed better symmetry and smoother tails, as also demonstrated in Figure 9.

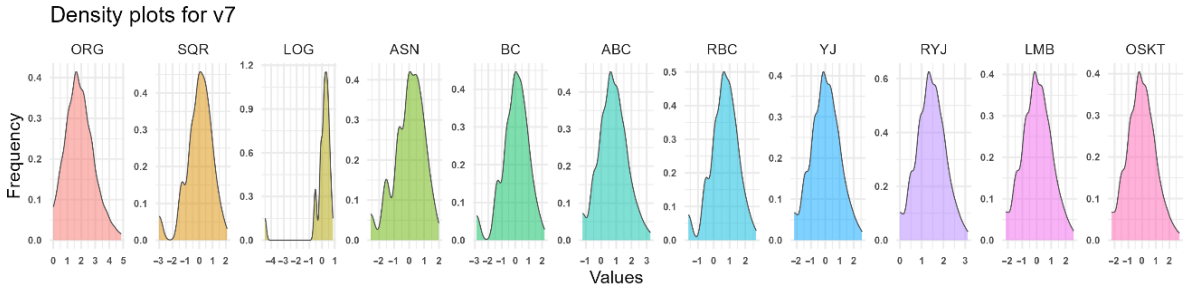



**Figure 9.** Density plots of the transformed data of simulated variable 7 (v7).

To summarize some implications of the results, it is worth mentioning the origin of the simulated variable v7. It was generated from a Poisson process, which produces discrete values with integer values and variances that depend on the means. Such structural features break the assumptions of continuous normality, making full Gaussian restorations infeasible. Hence, even flexible transformations such as BC and YJ may not completely correct the discrete shape, which appears to be the case in this experiment. However, LMB and OSKT yielded smoother, more symmetric distributions despite the discrete nature of v7.

As shown in Table 12, the analysis of the simulated variable v8 indicates differences in performance between the compared methods. The original data exhibited noticeable deviation from normality, with moderate left skewness ($g_3 = -0.694$) and slightly elevated kurtosis ($g_4 = 0.262$). Normality was rejected by almost all tests ($Nt = 1$), confirming that the original distribution was distinctly non-Gaussian. For variable v8, the methods SQR, LOG, and ASN worsened both skewness and kurtosis. For example, SQR increased asymmetry ($g_3 = -1.624$) and kurtosis ($g_4 = 4.372$), while LOG and ASN generated extreme skewness (LOG: $g_3 = -7.148$) and inflated kurtosis (LOG: $g_4 = 79.669$; ASN: $g_4 = 4.918$). None of these methods passed any normality test ($Nt = 0$), showing that classical transformations are inadequate for left-skewed or moderately kurtotic data.

**Table 12.** Results of normality tests for the transformed data of simulated variable 8 (v8).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | −0.694 | 0.262 | 2.727 | 0.069 | <0.001 | <0.001 | <0.001 | <0.001 | 0.000 | <0.001 | <0.001 | 1 |
| **SQR** | −1.624 | 4.372 | 4.859 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LOG** | −7.148 | 79.669 | 17.914 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | −1.882 | 4.918 | 8.195 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **BC** | −0.330 | −0.471 | 1.823 | 0.485 | <0.001 | 0.003 | <0.001 | <0.001 | 0.091 | <0.001 | 0.002 | 2 |
| **ABC** | −0.100 | −0.715 | 1.209 | 0.431 | <0.001 | 0.013 | 0.006 | <0.001 | 0.066 | <0.001 | 0.007 | 2 |
| **RBC** | −0.238 | −0.587 | 1.446 | 0.523 | <0.001 | 0.008 | 0.002 | 0.000 | 0.114 | <0.001 | 0.007 | 2 |
| **YJ** | −0.136 | −0.648 | 1.159 | 0.473 | 0.001 | 0.019 | 0.012 | 0.000 | 0.085 | 0.001 | 0.016 | 2 |
| **RYJ** | −0.117 | −0.664 | 1.250 | 0.461 | 0.001 | 0.019 | 0.013 | <0.001 | 0.080 | 0.002 | 0.008 | 2 |
| **LMB** | **0.000** | −0.460 | **1.086** | 0.702 | 0.103 | 0.212 | 0.281 | 0.040 | 0.264 | 0.162 | 0.217 | 7 |
| **OSKT** | −0.126 | **−0.188** | 1.123 | 0.821 | 0.140 | 0.423 | 0.431 | 0.385 | 0.435 | 0.241 | 0.357 | 8 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

Among the power-based transformation methods, YJ reduced skewness and kurtosis towards zero ($g_3 = -0.136$; $g_4 = -0.648$). Each method passed two normality tests ($Nt = 2$), indicating partial correction of asymmetry and tail behavior. ABC and RBC slightly outperformed BC, showing that they can provide additional benefits for moderately skewed data.

Moment-targeting transformations showed a clear advantage. The LMB method achieved perfect symmetry ($g_3 = 0.000$) and slightly negative kurtosis ($g_4 = -0.460$), resulting in acceptance by seven of the eight normality tests ($Nt = 7$). The OSKT performed even better, producing nearly symmetric and slightly mesokurtic data ($g_3 = -0.126$; $g_4 = -0.188$) and achieving the highest number of accepted tests ($Nt = 8$). OSKT not only minimized deviation in moment statistics but also provided the strongest statistical and visual alignment with a normal distribution confirmed by a small PPM value. The density plots in Figure 10 confirm these results.

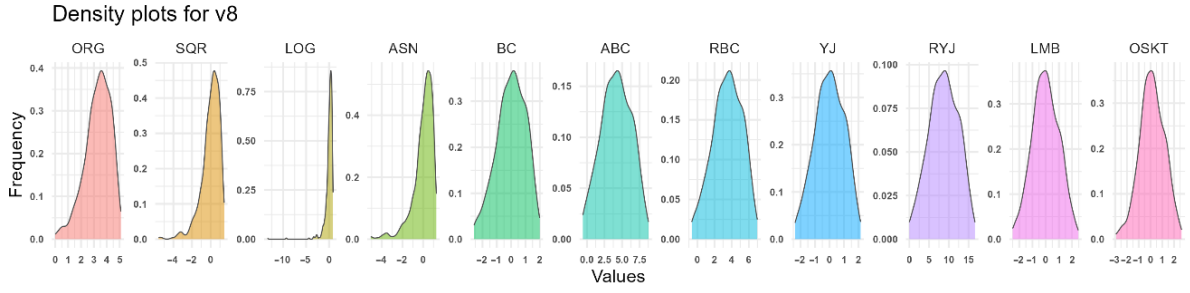



**Figure 10.** Density plots of the transformed data of simulated variable 8 (v8).

The results for the simulated variable v9 are summarized in Table 13. As shown in the table, normalizing this dataset was a particular challenge. The original data displayed a U-shaped distribution, characterized by slight skewness ($g_3 = 0.024$) but pronounced kurtosis ($g_4 = -1.454$). All normality tests rejected the original distribution ($Nt = 0$), indicating significant deviation from the Gaussian form.

**Table 13.** Results of normality tests for the transformed data of simulated variable 9 (v9).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | 0.024 | −1.454 | 12.364 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **SQR** | −0.493 | −1.026 | 9.900 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LOG** | −1.835 | 3.518 | 39.654 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | −0.331 | −1.350 | 18.159 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **BC** | −0.452 | −1.085 | 9.959 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ABC** | −0.276 | −1.288 | 14.373 | 0.000 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **RBC** | −0.383 | −1.174 | 11.218 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **YJ** | −0.129 | −1.422 | 18.391 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **RYJ** | −0.124 | −1.424 | 18.414 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LMB** | **0.000** | −1.454 | 12.123 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **OSKT** | 0.012 | **−0.817** | **4.891** | 0.211 | <0.001 | 0.042 | <0.001 | <0.001 | 0.009 | <0.001 | <0.001 | 1 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

Classical methods SQR, LOG, and ASN failed to improve normality. Both skewness and kurtosis remained far from zero, and all normality tests were rejected. Similarly, power-based methods, BC, YJ, and their extensions ABC, RBC, and RYJ, did not correct the distribution well and were unable to normalize the data.

Despite its moment-targeting approach, the LMB transformation produced a shape very similar to that of the original U-shaped distribution. Skewness was corrected to zero ($g_3 = 0$), but kurtosis remained extreme ($g_4 = -1.454$), resulting in rejection by all tests. It suggests that LMB's optimization, while effective for unimodal or moderately skewed distributions, struggles with distributions that are strongly bimodal or U-shaped. The OSKT, on the other hand, showed the most promising adjustment. It reduced both skewness ($g_3 = 0.012$) and kurtosis ($g_4 = -0.817$) and achieved partial success, with the Kolmogorov–Smirnov test accepting normality ($Nt = 1$). The Pearson P-statistic decreased significantly ($PPM = 4.891$), indicating an increased alignment with the normal distribution. This is confirmed by Figure 11, which shows that OSKT partially flattened the U-shaped structure and shifted the density closer to a bell-shaped form.

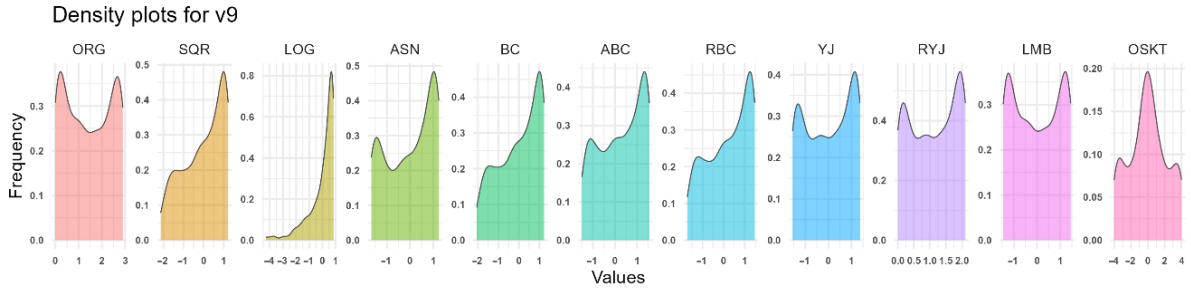



**Figure 11.** Density plots of the transformed data of simulated variable 9 (v9).

The pattern followed by the simulated variable v9 was significantly bimodal and displayed heavy tails. It was quite resistant, and no method but OSKT could fully restore normality. OSKT provided the best adjustment, reducing skewness and kurtosis and

offering partial statistical improvement. Compared with other methods, it was more robust in handling such a challenging, highly irregular distribution.

As shown in Table 14, the simulated variable v10 exhibited only mild departures from normality before transformation. The original distribution was slightly right-skewed ($g_3 = 0.331$) and moderately leptokurtic ($g_4 = 0.367$), leading most normality tests to reject the null hypothesis ($Nt = 0$). Classical transformations, SQR, LOG, and ASN, did not improve the distribution and, in some cases, even increased the asymmetry. For power-based methods, BC passed one test, RBC passed two tests, ABC and RYJ passed three tests, and YJ passed four tests. While some of these methods were fairly successful in reducing skewness, they were less effective than OSKT at correcting kurtosis. In general, YJ, LMB, and OSKT provided the most consistent results, each passing four tests ($Nt = 4$). YJ and LMB were effective at correcting skewness, whereas OSKT excelled at reducing kurtosis and tail heaviness for approximating normality.

**Table 14.** Results of normality tests for the transformed data of simulated variable 10 (v10).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | 0.331 | 0.367 | 39.114 | <0.001 | <0.001 | <0.001 | <0.001 | 0.003 | <0.001 | 0.018 | 0.011 | 0 |
| **SQR** | −0.554 | 1.841 | **38.109** | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.001 | <0.001 | 0 |
| **LOG** | −6.783 | 90.623 | 37.509 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | −0.882 | 1.953 | 38.255 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **BC** | −0.029 | 0.494 | **38.109** | 0.001 | <0.001 | 0.009 | <0.001 | 0.112 | <0.001 | 0.048 | 0.025 | 1 |
| **ABC** | 0.051 | 0.417 | **38.109** | <0.001 | <0.001 | 0.021 | <0.001 | 0.168 | <0.001 | 0.062 | 0.063 | 3 |
| **RBC** | −0.045 | 0.513 | **38.109** | 0.001 | <0.001 | 0.006 | <0.001 | 0.094 | <0.001 | 0.046 | 0.057 | 2 |
| **YJ** | 0.024 | 0.343 | **38.109** | 0.001 | <0.001 | 0.055 | <0.001 | 0.279 | <0.001 | 0.080 | 0.081 | 4 |
| **RYJ** | −0.025 | 0.369 | **38.109** | 0.001 | <0.001 | 0.042 | <0.001 | 0.241 | <0.001 | 0.068 | 0.081 | 3 |
| **LMB** | **0.000** | 0.279 | **38.109** | 0.001 | <0.001 | 0.116 | <0.001 | 0.399 | <0.001 | 0.096 | 0.094 | 4 |
| **OSKT** | −0.069 | **−0.001** | 40.427 | 0.002 | <0.001 | 0.626 | <0.001 | 0.798 | <0.001 | 0.085 | 0.098 | 4 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

Figure 12 graphically verifies the results above. Overall, LMB was favorable for skewness correction and OSKT for kurtosis adjustment. Meanwhile, ABC, RBC, and RYJ offered moderate improvements in moment adjustment.

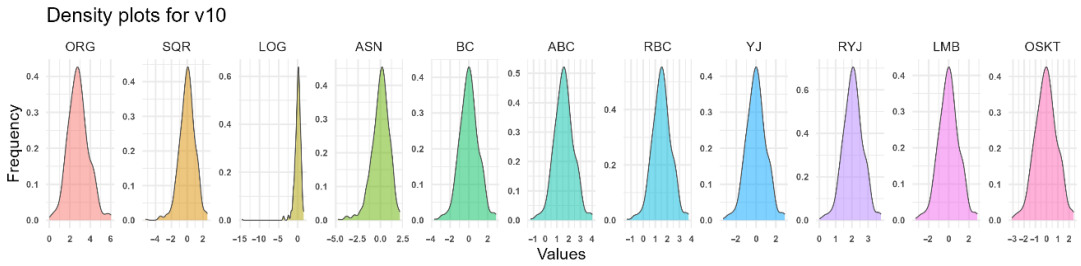



**Figure 12.** Density plots of the transformed data of simulated variable 10 (v10).

In general, LMB achieved skewness values closest to zero, standing out in asymmetry correction. OSKT showed slightly higher skewness than LMB, but only occasionally. It was better at reducing kurtosis, often yielding the lowest values. In other words, LMB was particularly good at correcting skewness while OSKT was superior in adjusting tail heaviness and peakedness. ABC, RBC, and RYJ provided decent results in reducing skewness and kurtosis to some extent but were generally less effective than LMB and OSKT. The methods ABC, RBC, and RYJ may be considered as intermediate options.

In the bigger picture, there is much potential in combining the insights from LMB and OSKT methods and considering ABC, RBC or RYJ as alternative methods to develop a comprehensive strategy for achieving normality across diverse data structures.

### *3.2. Analyses for Real Data*

In this sub-section, starting with Table 15 the analyses for real data are presented. As shown the table, the original data (real-data variable v1) contains strong asymmetry (highly left-skewed) and heavy tails. In general, the transformation methods ABC, YJ, RYJ, LMB, and OSKT substantially outperformed the classical methods SQR, LOG, and ASN, as well as the power-based BC transformation.

**Table 15.** Results of normality tests for the transformed data of real-data variable 1 (v1).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | −1.504 | 2.159 | 2.395 | 0.049 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.015 | 0.016 | 0 |
| **SQR** | −2.859 | 9.944 | 4.123 | 0.012 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LOG** | −5.537 | 31.291 | 14.954 | 0.000 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | −2.739 | 8.572 | 5.784 | 0.006 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **BC** | −1.414 | 1.792 | 2.395 | 0.069 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.019 | 0.021 | 1 |
| **ABC** | −0.117 | −0.865 | 1.665 | 0.910 | 0.371 | 0.726 | 0.503 | 0.452 | 0.593 | 0.858 | 0.748 | 8 |
| **RBC** | −0.891 | 0.137 | 1.000 | 0.196 | 0.008 | 0.008 | 0.007 | 0.030 | 0.005 | 0.173 | 0.278 | 3 |
| **YJ** | −0.366 | −0.674 | 2.395 | 0.649 | 0.270 | 0.538 | 0.276 | 0.431 | 0.188 | 0.810 | 0.728 | 8 |
| **RYJ** | −0.357 | −0.683 | 2.395 | 0.659 | 0.279 | 0.550 | 0.287 | 0.437 | 0.198 | 0.821 | 0.717 | 8 |
| **LMB** | **0.000** | **0.001** | 1.665 | 0.835 | 0.931 | 0.831 | 0.664 | 0.803 | 0.431 | 0.993 | 0.999 | 8 |
| **OSKT** | −0.016 | −0.441 | **0.668** | 0.925 | 0.977 | 0.987 | 0.888 | 0.993 | 0.631 | 1.000 | 0.997 | 8 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

For this dataset, OSKT achieved the best overall performance, producing the lowest PPM statistic and being confirmed as normal by all eight statistical tests ($Nt = 8$). LMB and ABC also achieved near-perfect normalization ($Nt = 8$), with skewness and kurtosis values close to zero, indicating effective correction of asymmetry and tail heaviness. YJ and RYJ performed well, with all eight tests confirming normality. They both had slightly high PPM values, suggesting minor deviations from the theoretical Gaussian distribution. RBC had an average performance, passing three normality tests, but was less successful than the transformations mentioned above.

Table 16 shows the results for real-data variable v2, which exhibits moderate right-skewness and heavy tails. The original data clearly deviates from normality, as all eight statistical tests rejected the null hypothesis ($Nt = 0$). The classical transformations, such as SQR and LOG, offered limited or even counterproductive improvements. SQR slightly reduced skewness and kurtosis, while LOG increased the asymmetry, producing extreme negative skewness and inflated kurtosis. ASN performed relatively better, passing two normality tests ($Nt = 2$). Among the power transformation methods, YJ achieved nearly perfect skewness ($g_3 = -0.001$), demonstrating effective asymmetry correction, but it could not address the heavy-tailed nature of the data, as most tests still rejected normality.

The rest of the compared methods, BC, ABC, RBC, RYJ, LMB, and OSKT, did not achieve statistical normality ($Nt = 0$), although some of them reduced skewness and kurtosis in different amounts. However, OSKT produced the lowest PPM statistic, suggesting the smallest overall deviation from the theoretical Gaussian distribution.

**Table 16.** Results of normality tests for the transformed data of real-data variable 2 (v2).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | 0.981 | 1.120 | 13.251 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **SQR** | 0.219 | 0.429 | 6.915 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LOG** | −2.435 | 26.131 | 5.978 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | −0.066 | **0.120** | 5.401 | 0.006 | <0.001 | 0.313 | <0.001 | 0.323 | <0.001 | 0.004 | 0.004 | 2 |
| **BC** | 0.068 | 0.622 | 7.913 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.001 | <0.001 | 0 |
| **ABC** | 0.097 | 0.573 | 7.913 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **RBC** | −0.216 | 1.440 | 4.435 | 0.000 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **YJ** | −0.001 | 0.200 | 6.905 | 0.001 | <0.001 | 0.207 | <0.001 | 0.271 | <0.001 | 0.005 | 0.004 | 2 |
| **RYJ** | −0.184 | 0.385 | 3.549 | 0.006 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LMB** | **0.000** | 0.358 | 8.159 | 0.000 | <0.001 | 0.010 | <0.001 | 0.032 | <0.001 | 0.001 | <0.001 | 0 |
| **OSKT** | −0.214 | 0.311 | **3.243** | 0.012 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

Table 17 presents the results for real-data variable v3 which shows a minor right-skewness ($g_3 = 0.318$) and a relatively flat peak ($g_4 = -0.793$). As this is a moderate deviation from normality compared with v1 and v2, all eight normality tests rejected the original data ($Nt = 0$), confirming that v3 is non-Gaussian. In addition, none of the transformation methods succeeded in achieving normality. Although some optimized transformations brought the moment statistics closer to the theoretical values (LMB achieved perfect skewness ($g_3 = 0$) the consistent rejection across all tests indicates that v3's distribution has structural features that cannot be corrected by univariate transformations. Such features may include latent multimodality as seen in Figure 2 earlier, localized outliers, or irregularities near the center of the distribution, which distort the density without greatly affecting moment statistics.

**Table 17.** Results of normality tests for the transformed data of real-data variable 3 (v3).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | 0.318 | −0.793 | 32.484 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **SQR** | −0.532 | −0.641 | 20.258 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LOG** | −2.156 | 3.770 | 64.561 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | −0.265 | −1.158 | 19.600 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **BC** | −0.866 | **−0.062** | 21.531 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ABC** | −0.040 | −0.977 | 17.075 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **RBC** | −0.610 | −0.529 | 18.423 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **YJ** | −0.066 | −1.082 | 24.022 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **RYJ** | −0.024 | −1.069 | 23.240 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LMB** | **0.000** | −1.081 | 21.692 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **OSKT** | 0.911 | 5.998 | **15.684** | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

The OSKT method also performed poorly for this variable. It overcorrected the distribution, producing pronounced positive skewness ($g_3 = 0.911$) and highly inflated kurtosis ($g_4 = 5.998$), deviating further from the Gaussian shape in terms of moments. However, the method yielded the lowest PPM statistic among all methods, suggesting that, despite its moment overcorrection, it provides a closer overall fit to the empirical distribution in terms of overall goodness of fit. In summary, v3's structural non-normality resisted the correction attempts of the compared transformation methods. This is a clear indicator of the limitations of univariate normalizations when applied to complex real-world datasets.

The next analysis is of real-data variable v4, which exhibits a pronounced platykurtic distribution as shown in Table 18. It demonstrated strong resistance to normalization for the majority of the methods. Firstly, the original data failed all eight normality tests ($Nt = 0$), indicating a substantial departure from the Gaussian shape. Classical transformations such as SQR, LOG, and ASN were completely ineffective ($Nt = 0$), and even advanced methods like BC, YJ, and LMB could not achieve normality. Notably, LMB achieved ideal skewness ($g_3 = 0$), but both LMB and YJ retained severe kurtosis deficits ($g_4 \approx -1.2$), indicating that the main source of non-normality in v4 is its excessive flatness rather than asymmetry. The OKST transformed v4 successfully and passed all normality tests ($Nt = 8$) with the smallest PPM value (0.630).

**Table 18.** Results of normality tests for the transformed data of real-data variable 4 (v4).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | 0.242 | −1.223 | 6.325 | 0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **SQR** | −0.232 | −1.029 | 4.420 | 0.015 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LOG** | −1.723 | 7.072 | 9.636 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | −0.213 | −1.233 | 6.755 | 0.003 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **BC** | −0.207 | −1.060 | 4.313 | 0.015 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ABC** | −0.207 | −1.060 | 4.313 | 0.015 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **RBC** | −0.195 | −1.073 | 4.430 | 0.016 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **YJ** | −0.055 | −1.252 | 5.993 | 0.012 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **RYJ** | −0.052 | −1.252 | 6.042 | 0.012 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LMB** | **0.000** | −1.274 | 6.335 | 0.011 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **OSKT** | 0.072 | **−0.331** | **0.630** | 0.983 | 0.289 | 0.528 | 0.625 | 0.247 | 0.872 | 0.433 | 0.463 | 8 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

The variable v5's results are presented in Table 19. In this round, classical transformations provided some partial improvements. For example, the methods SQR and ASN achieved moderate success ($Nt = 5$ and $Nt = 2$, respectively), reducing asymmetry but not fully addressing heavy tails. The power transformations ABC and RBC were decent or above average ($Nt = 6$ and $Nt = 7$), while YJ and RYJ passed only four tests ($Nt = 4$). The OSKT method achieved partial normalization ($Nt = 4$) but indicated moderate improvement in skewness and kurtosis. It performed less effectively than the methods BC and LMB, likely because its simultaneous optimization of skewness and kurtosis is more suitable for complex distributions, whereas v5's non-normality is mostly from simple right-skewness.

**Table 19.** Results of normality tests for the transformed data of real-data variable 5 (v5).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | 1.210 | 1.112 | 6.702 | 0.012 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.001 | 0 |
| **SQR** | 0.444 | −0.435 | 1.699 | 0.154 | 0.013 | 0.128 | 0.002 | 0.096 | 0.003 | 0.261 | 0.384 | 5 |
| **LOG** | −2.283 | 12.627 | **1.348** | 0.576 | <0.001 | <0.001 | 0.001 | <0.001 | 0.141 | <0.001 | <0.001 | 2 |
| **ASN** | 0.470 | −0.838 | 2.292 | 0.064 | <0.001 | 0.054 | <0.001 | 0.003 | <0.001 | 0.018 | 0.050 | 2 |
| **BC** | 0.008 | 0.001 | 1.436 | 0.428 | 0.106 | 0.967 | 0.071 | 0.913 | 0.057 | 0.486 | 0.490 | 8 |
| **ABC** | 0.097 | −0.176 | 1.370 | 0.363 | 0.119 | 0.807 | 0.046 | 0.902 | 0.036 | 0.534 | 0.616 | 6 |
| **RBC** | −0.245 | 0.727 | 1.524 | 0.614 | 0.028 | 0.400 | 0.144 | 0.120 | 0.168 | 0.283 | 0.066 | 7 |
| **YJ** | 0.087 | −0.994 | 1.897 | 0.325 | 0.027 | 0.154 | 0.016 | 0.001 | 0.026 | 0.324 | 0.300 | 4 |
| **RYJ** | 0.061 | −0.995 | 2.248 | 0.352 | 0.029 | 0.157 | 0.018 | 0.001 | 0.033 | 0.316 | 0.287 | 4 |
| **LMB** | **−0.001** | **−0.346** | 1.480 | 0.485 | 0.221 | 0.697 | 0.115 | 0.879 | 0.082 | 0.668 | 0.655 | 8 |
| **OSKT** | 0.511 | 1.089 | 1.655 | 0.666 | 0.022 | 0.039 | 0.346 | 0.007 | 0.218 | 0.263 | 0.035 | 4 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

For the real-data variable v6, the OSKT method was the only one to achieve remarkable improvement, yielding the lowest skewness ($g_3 = -0.261$) and kurtosis ($g_4 = 0.846$) values and the best overall performance ($Nt = 3$), as shown in Table 20. Although the normality tests were not fully accepted, OSKT significantly reduced both asymmetry and tail heaviness compared to other methods, showing its superiority in dealing with highly non-Gaussian distributions with complex shape deviations.

**Table 20.** Results of normality tests for the transformed data of real-data variable 6 (v6).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | −0.167 | −0.907 | 4.007 | 0.005 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **SQR** | −0.603 | **0.286** | 5.447 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LOG** | −6.052 | 97.370 | 9.884 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | −0.789 | 0.351 | 7.960 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **BC** | −0.177 | −0.896 | 3.982 | 0.005 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ABC** | −0.125 | −0.945 | 3.731 | 0.007 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **RBC** | −0.155 | −0.918 | 3.765 | 0.006 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **YJ** | −0.099 | −0.958 | 3.722 | 0.009 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **RYJ** | −0.104 | −0.955 | 3.722 | 0.009 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LMB** | **0.000** | −0.968 | 3.258 | 0.021 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **OSKT** | −0.261 | 0.846 | **0.748** | 0.871 | <0.001 | <0.001 | 0.435 | <0.001 | 0.534 | 0.001 | <0.001 | 3 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

In Table 21, for the variable v7, the classical methods SQR, LOG, and ASN greatly increased both skewness and kurtosis, amplifying distributional asymmetry and tail heaviness. They also failed to correct the severe non-normality of the original data ($Nt = 0$). The methods ABC and RBC, as well as YJ and RYJ, achieved minimal improvements, remaining far from normality. On the other hand, LMB and especially the OSKT method produced substantial normalization. OSKT achieved near-zero skewness ($g_3 = -0.084$) and a kurtosis value of 0.215, passing seven normality tests ($Nt = 7$), the highest among all compared methods. This finding further corroborates that OSKT demonstrates superior transformation performance by simultaneously capturing both symmetry and tail behavior.

**Table 21.** Results of normality tests for the transformed data of real-data variable 7 (v7).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | −1.463 | 2.911 | 9.621 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **SQR** | −2.638 | 12.407 | 14.752 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LOG** | −11.847 | 217.880 | 38.900 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | −3.363 | 18.292 | 20.519 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **BC** | −0.670 | −0.069 | 5.253 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ABC** | −0.243 | −0.682 | 3.195 | 0.005 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **RBC** | −0.243 | −0.682 | 3.195 | 0.005 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **YJ** | −0.247 | −0.654 | 3.023 | 0.005 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **RYJ** | −0.388 | −0.517 | 4.009 | 0.000 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LMB** | **−0.057** | 0.392 | 1.544 | 0.107 | 0.000 | 0.070 | 0.001 | 0.058 | 0.002 | 0.067 | 0.005 | 4 |
| **OSKT** | −0.084 | **0.215** | **1.264** | 0.883 | 0.022 | 0.368 | 0.203 | 0.211 | 0.569 | 0.197 | 0.127 | 7 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

In Table 22, it can be seen that none of the methods achieved full normalization which shows that variable v8 exhibits a particularly resistant non-normal structure. Classical methods SQR, LOG, and ASN slightly reduced skewness but failed to correct the pronounced leptokurtosis. The methods BC, RYJ, and LMB achieved marginal improvement ($Nt = 1$),

reducing asymmetry while improving tail behavior. The OSKT method achieved near-zero skewness ($g_3 = 0.043$) but retained a relatively high kurtosis ($g_4 = 3.680$), suggesting that the deviation from normality arises mostly from peakedness rather than asymmetry. It implies that, as OSKT successfully balances skewness and kurtosis in moderately distorted data, extreme leptokurtic structures like with v8 can be challenging to fully normalize.

**Table 22.** Results of normality tests for the transformed data of real-data variable 8 (v8).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | 0.959 | 0.435 | 17.618 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **SQR** | 0.470 | −0.284 | 10.730 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LOG** | −2.708 | 37.692 | 7.827 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | 0.297 | −0.724 | 8.508 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **BC** | 0.141 | 0.708 | 8.838 | <0.001 | <0.001 | 0.205 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 1 |
| **ABC** | 0.259 | 0.177 | 9.162 | <0.001 | <0.001 | 0.008 | <0.001 | 0.002 | <0.001 | <0.001 | <0.001 | 0 |
| **RBC** | −1.118 | 13.521 | 8.195 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **YJ** | 0.034 | −0.436 | 7.332 | <0.001 | <0.001 | 0.001 | <0.001 | 0.003 | <0.001 | <0.001 | <0.001 | 0 |
| **RYJ** | −0.168 | **−0.001** | 7.955 | <0.001 | <0.001 | 0.010 | <0.001 | 0.100 | <0.001 | <0.001 | <0.001 | 1 |
| **LMB** | 0.176 | 0.072 | 8.225 | <0.001 | <0.001 | 0.037 | <0.001 | 0.069 | <0.001 | <0.001 | <0.001 | 1 |
| **OSKT** | **0.043** | 3.680 | **5.116** | 0.016 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

According to the results presented in Table 23, the alkaline phosphatase (v9) variable in the *Liver Disorders* dataset has a skewed and moderately peaked distribution ($g_3 = 0.747$, $g_4 = 0.698$), failing all normality tests. The LOG transformation yielded the worst results for this variable, even causing a significant increase in skewness and kurtosis coefficients. While the transformation methods BC, ABC, and RBC, along with SQR, reduced skewness somewhat, they did not completely eliminate the kurtosis problem. Therefore, these methods only passed a small number of normality tests ($Nt = 3$–$4$). RJ and its robust version, RYJ, and LMB transformations produce more consistent results. These methods reduced both skewness and kurtosis to reasonable levels ($g_3 \approx 0$, $g_4 \approx 0.3$–$0.5$) and passed all normality tests ($Nt = 8$). However, OSKT achieved the best performance among the moment-targeting methods by reducing both skewness ($g_3 = -0.132$) and kurtosis ($g_4 = 0.141$). It also performed a transformation that approached a Gaussian-like structure, particularly in terms of the PPM value, with a value of 0.782.

**Table 23.** Results of normality tests for the transformed data of real-data variable 9 (v9).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | 0.747 | 0.698 | 3.419 | 0.018 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.001 | <0.001 | 0 |
| **SQR** | −0.085 | 1.318 | 1.316 | 0.461 | <0.001 | <0.001 | 0.066 | 0.002 | 0.077 | 0.165 | <0.001 | 4 |
| **LOG** | −7.088 | 92.270 | 5.043 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | −0.302 | 0.611 | 1.424 | 0.536 | 0.036 | 0.005 | 0.338 | 0.008 | 0.119 | 0.294 | 0.045 | 4 |
| **BC** | 0.057 | 0.876 | 1.674 | 0.420 | 0.007 | 0.006 | 0.036 | 0.024 | 0.059 | 0.203 | 0.018 | 3 |
| **ABC** | 0.179 | 0.616 | 1.722 | 0.304 | 0.010 | 0.019 | 0.014 | 0.041 | 0.025 | 0.205 | 0.111 | 3 |
| **RBC** | −0.115 | 1.430 | 1.458 | 0.453 | <0.001 | <0.001 | 0.071 | <0.001 | 0.071 | 0.146 | <0.001 | 4 |
| **YJ** | 0.007 | 0.347 | 0.958 | 0.636 | 0.271 | 0.374 | 0.280 | 0.368 | 0.195 | 0.710 | 0.479 | 8 |
| **RYJ** | −0.121 | 0.513 | 1.039 | 0.584 | 0.149 | 0.110 | 0.441 | 0.114 | 0.154 | 0.588 | 0.147 | 8 |
| **LMB** | **0.000** | 0.384 | 0.958 | 0.625 | 0.231 | 0.316 | 0.270 | 0.316 | 0.185 | 0.672 | 0.415 | 8 |
| **OSKT** | −0.132 | **0.141** | **0.782** | 0.760 | 0.368 | 0.501 | 0.641 | 0.442 | 0.335 | 0.648 | 0.465 | 8 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

According to the results presented in Table 24, the alanine aminotransferase (v10) variable from the *Liver Disorders* dataset exhibits a strongly right-skewed and heavily leptokurtic distribution ($g_3 = 3.037$, $g_4 = 13.501$), leading all normality tests to reject the null hypothesis. The classical transformations (SQR, LOG, ASN, BC, ABC, and RBC) proved insufficient for this variable. Although SQR, BC and ABC slightly reduced skewness, they failed to effectively correct the severe tail heaviness, resulting in persistently high kurtosis values. Moreover, the LOG transformation performed particularly poorly, generating extreme negative skewness and an excessively inflated kurtosis coefficient ($g_4 = 26.715$). Consequently, all of these classical transformations failed every normality test ($Nt = 0$). More advanced transformation families, namely YJ, RYJ, and LMB, provided noticeable improvements. The YJ transformation achieved almost perfect symmetry ($g_3 = -0.033$) and substantially reduced kurtosis ($g_4 = 0.524$), passing six normality tests ($Nt = 6$). Similarly, RYJ achieved balanced moment values ($g_3 = -0.170$, $g_4 = 0.604$), though slightly less effectively than the original YJ, and therefore passed fewer tests ($Nt = 4$). The LMB transformation also moderated the tail behavior but did not sufficiently reduce kurtosis, leading to persistent rejections by most tests ($Nt = 0$). Among all methods, OSKT delivered the best performance. It produced the most Gaussian-like distribution, with near-zero skewness ($g_3 = -0.109$) and the lowest kurtosis value among all transformations ($g_4 = 0.085$). In addition, OSKT yielded the best PPM value (2.087), indicating a distribution shape closest to the normal reference curve. Most importantly, OSKT passed the highest number of normality tests ($Nt = 8$), demonstrating superior effectiveness particularly for variables with extreme skewness and heavy tails such as v10.

**Table 24.** Results of normality tests for the transformed data of real-data variable 10 (v10).

| TM | $g_3$ | $g_4$ | PPM | KS | SW | RJB | AD | DAG | LIL | ZA | ZC | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ORG** | 3.037 | 13.501 | 11.028 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **SQR** | 1.344 | 3.878 | 4.589 | 0.000 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **LOG** | −2.745 | 26.715 | 4.569 | 0.000 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **ASN** | 0.979 | 1.601 | 3.271 | 0.005 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **BC** | 0.359 | 3.608 | 3.649 | 0.007 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.000 | <0.001 | 0 |
| **ABC** | 0.495 | 3.309 | 3.649 | 0.005 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.000 | <0.001 | 0 |
| **RBC** | −0.159 | 5.645 | 2.290 | 0.010 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0 |
| **YJ** | **−0.033** | 0.524 | 2.547 | 0.353 | 0.074 | 0.036 | 0.056 | 0.162 | 0.037 | 0.507 | 0.430 | 6 |
| **RYJ** | −0.170 | 0.604 | 2.560 | 0.343 | 0.041 | 0.008 | 0.081 | 0.047 | 0.036 | 0.365 | 0.184 | 4 |
| **LMB** | 0.506 | 2.225 | 3.358 | 0.014 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.001 | <0.001 | 0 |
| **OSKT** | −0.109 | **0.085** | **2.087** | 0.654 | 0.427 | 0.563 | 0.458 | 0.595 | 0.215 | 0.744 | 0.757 | 8 |

TM: Transformation methods; Nt: Number of tests that confirmed the normality of the transformed data; Bold values highlight the moment statistics nearest zero and the lowest Pearson P-statistic value.

### *3.3. Comparison of Computing Times*

Table 25 and Figure 13 present a detailed comparison of computing times (in log-microseconds) across all tested normalization methods for varying sample sizes ranging from $10^3$ to $10^7$ observations. The results reveal clear computational differences between simple, power-based, and optimization-based transformations.

On small to moderate sample sizes (up to $10^4$), the transformation methods SQR, LOG, and ASN were computationally efficient. However, as soon as the sample size increased beyond $10^5$, the computation times began to increase gradually, reaching 4.0–5.6 log µs at $10^7$ samples. This shows that even basic algebraic transformations cause quantifiable computational costs when applied to very large datasets.

**Table 25.** Computing times (log μs) across the normalization methods.

| $n$ | SQR | LOG | ASN | BC | ABC | RBC | YJ | RYJ | LMB | OSKT |
|---|---|---|---|---|---|---|---|---|---|---|
| $10^3$ | 0.000 | 0.000 | 0.000 | 0.000 | 5.301 | 5.568 | 0.000 | 5.672 | 4.477 | 4.903 |
| $10^4$ | 0.000 | 0.000 | 0.000 | 4.602 | 8.998 | 6.599 | 4.301 | 6.651 | 5.176 | 5.986 |
| $10^5$ | 0.000 | 0.000 | 4.301 | 5.342 | 8.999 | 7.549 | 5.301 | 7.644 | 5.935 | 7.312 |
| $10^6$ | 4.000 | 4.699 | 4.699 | 6.309 | 8.999 | 8.563 | 6.320 | 8.664 | 6.850 | 8.199 |
| $10^7$ | 4.954 | 5.556 | 5.653 | 7.256 | 8.999 | 9.586 | 7.325 | 9.717 | 7.789 | 9.119 |

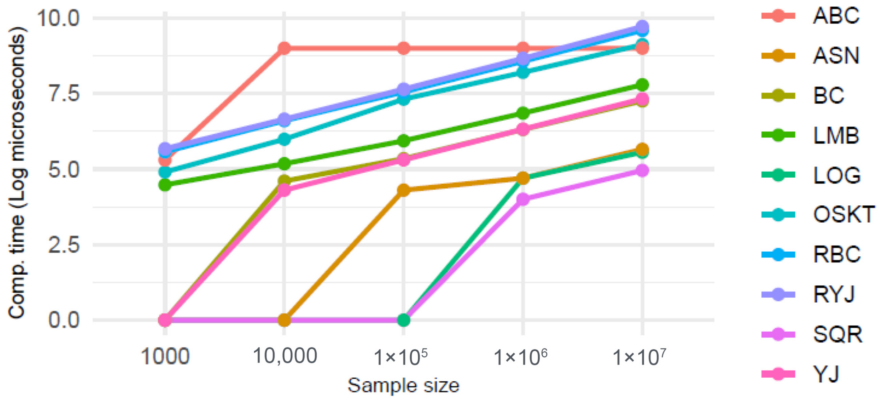



**Figure 13.** Computing times across the normalization methods.

Power-based transformations such as BC, ABC, and RBC generally required more computation time than simple transformations due to their parameter estimation processes. Their computing times increased steadily with sample size, with ABC and RBC consistently showing the highest values among this group (reaching approximately 9.6–9.7 log μs at $n = 10^7$). These results indicate that the adaptive or regularized extensions of BC introduce additional computational overhead. Among the parameter-driven and optimization-based methods, YJ, LMB, and OSKT exhibited substantially higher computing times, reflecting their iterative optimization nature. YJ and RYJ required moderate computational effort, scaling up to about 7.3–9.7 log μs at the largest sample size. LMB showed slightly lower times than OSKT, remaining below 7.8 log μs even for the largest datasets. OSKT, as expected, was the most computationally intensive method due to its dual-parameter optimization of skewness and kurtosis, reaching 9.1 log μs at $n = 10^7$.

Despite its higher computational demand, OSKT's processing time remains within a comparable range to other advanced methods such as ABC, RBC, and YJ, suggesting that its additional complexity is computationally manageable. This trade-off between computational cost and normalization accuracy highlights that OSKT's iterative optimization, while slower than simple or power-based transformations, offers a justified increase in processing time given its superior performance in restoring normality across diverse data structures. Preliminary results indicate that the speed-optimized implementations of the OSKT function, utilizing Rcpp [61] and OpenMP [62], can reduce computation times by approximately 5–6 times compared with the basic R implementation (Appendix A).

### *3.4. Results of Back Transformation Experiment*

According to the results in Table 26, both NR and UR methods provide spectacularly high numerical accuracy in the regression of OSKT-transformed values, even in the presence of very large datasets that contain $10^7$ observations. The RMSE for all tested variables consistently lies between $10^{-16}$ and $10^{-14}$. This means that the inversion is performed with machine-level precision, showing no numerical distortion. Between the two methods, NR

generally achieves slightly higher RMSE values than UR (e.g., for v6 and v9), but both provide sufficient precision for practical and methodological purposes.

**Table 26.** Root mean square errors and computing times (s) by back transformation methods.

| Vars. | NR_RMSE | UR_RMSE | NR_CT | UR_CT |
|---|---|---|---|---|
| v1 | $4.600420 \times 10^{-16}$ | $5.793887 \times 10^{-16}$ | 14.80 | 113.45 |
| v2 | $2.149265 \times 10^{-16}$ | $7.923919 \times 10^{-16}$ | 15.78 | 192.53 |
| v3 | $1.923779 \times 10^{-16}$ | $6.050859 \times 10^{-16}$ | 16.48 | 191.81 |
| v4 | $1.722727 \times 10^{-16}$ | $6.838369 \times 10^{-16}$ | 16.71 | 192.48 |
| v5 | $8.793360 \times 10^{-16}$ | $9.415690 \times 10^{-16}$ | 14.36 | 105.73 |
| v6 | $1.064861 \times 10^{-13}$ | $5.865035 \times 10^{-15}$ | 16.49 | 301.36 |
| v7 | $3.786589 \times 10^{-16}$ | $4.224095 \times 10^{-16}$ | 13.79 | 193.70 |
| v8 | $2.106872 \times 10^{-16}$ | $4.492934 \times 10^{-16}$ | 16.24 | 258.87 |
| v9 | $4.309473 \times 10^{-14}$ | $3.389857 \times 10^{-15}$ | 19.05 | 352.11 |
| v10 | $1.038500 \times 10^{-15}$ | $1.307846 \times 10^{-15}$ | 12.23 | 116.26 |

NR_RMSE: Root mean square error obtained with Newton–Raphson; UR_RMSE: Root mean square error obtained with Uniroot method; NR_CT: Computing time (s) for Newton–Raphson; UR_CT: Computing time (s) for Uniroot method.

Unlike the numerical accuracy, there was a significant difference in computational times between the two methods. The NR showed rapid convergence, completing the conversion for each variable in only 12–20 s. As the UR method relies on interval bracketing instead of analytical derivatives, it ran longer for the variables, from approximately 100 to 350 s. It is observed that, for variables such as v6 and v9, which show greater departure from normality than others, both the RMSE is higher and the running time is longer. These findings suggest that NR is preferable for machine learning modeling studies, including intensive bootstrap loops, as the default inversion technique due to its speed and consistent near-machine-precision accuracy. Furthermore, UR offers additional robustness, making it a valuable alternative for analyzing smaller data sizes.

Finally, a qualitative comparison of the main advantages and disadvantages of the normalization methods in this study is presented in Appendix C as a table. The table covers both theoretical conclusions and experimental results from our study, highlighting OSKT's flexibility and moment-targeting capabilities among the compared methods.

## 4. Conclusions

This study compared the performance of traditional, power-based, and moment-targeting transformations using ten real and simulated variables spanning a broad spectrum of distributional forms, including right- and left-skewed, leptokurtic, platykurtic, bounded, discrete, and bimodal. The results indicated peculiar strengths and limitations between the tested approaches.

Among the traditional transformations, the SQR and LOG methods handled mildly right-skewed data but generally failed when there was negative skewness, flatness, or multimodality. The ASN method was a moderate improvement for positively skewed variables but proved unsuitable for distributions outside bounded domains.

The power-based transformation methods BC, ABC, RBC, YJ, and RYJ were highly flexible and achieved near-perfect normalizations in moderately skewed or leptokurtic distributions such as v2 to v5 in the simulated dataset. BC and ABC performed particularly well for right-skewed data, whereas RBC and RYJ offered moderate improvements for variables with heavier tails or negative skewness. YJ outperformed BC in general because it can accommodate zero and negative values. However, the performance of all power-based methods declined for highly non-Gaussian or structurally complex data, where a single shape parameter could not simultaneously control both skewness and kurtosis.

The moment-targeting methods, LMB, and the proposed method OSKT displayed the most consistent performances in each simulation scenario. LMB generally returned the smallest skewness values, restoring symmetry even in extremely left-skewed distributions such as v4 and v5 in the simulated dataset. On the other hand, OSKT achieved superior control of kurtosis and tail behavior, performing exceptionally well in challenging cases such as the symmetric bimodal variable v6, where it was the only method to achieve formal normality accepted by all normality tests. Both transformations also performed very well for bounded and discrete variables, such as v7 to v10 in the simulated dataset, indicating their broad applicability beyond the continuous domain. The LMB transformation uses two parameters and is theoretically invertible. Note that the effects of the parameters are not entirely independent due to the functional structure of LMB. For example, a subtle adjustment of skewness affects tail behavior, and vice versa, which limits LMB's ability to simultaneously control both torsion and acuity in distributions that contain significant asymmetry and heavy tails. Unlike the LMB approach, the OSKT optimizes the Tukey *g*–*h* parameters using an Anderson–Darling-based objective, which enables explicit, independent corrections to both moments, enabling more accurate normalizations for even highly non-Gaussian or bimodal distributions.

According to the results, traditional and power transformations still work well for some specific skewed distributions, but moment-targeting methods like OSKT can offer a better option by providing a more general-purpose, distribution-agnostic normalization framework. By optimizing both skewness and kurtosis simultaneously, OSKT restores approximate normality even under extreme or structurally irregular conditions. In light of these findings, it is clear that OSKT would be a versatile preprocessing tool for statistical modeling and machine learning applications that require approximately normal residuals or input features.

According to the analyses with real data, no single transformation method could achieve optimal performance across the distributional scenarios. Methods such as LMB, BC, ABC, and RBC were most successful at restoring approximate normality for variables with substantial skewness and heavy tails, especially v1 and v5. In contrast with the aforementioned methods, OSKT had the advantage when the main source of non-normality involved complex kurtosis deficiencies (seen in v4) and was the only method capable of producing statistically acceptable Gaussian conformity. As for v2 and v3, there were limitations that remained even after transformation, hinting that residual deviations from normality may be caused by multimodality or other structural irregularities that cannot be adequately addressed by univariate transformations such as YJ and RYJ. On a collective level, these findings emphasize that the suitability of a transformation method heavily depends on the underlying distributional pathology. Therefore, they support the use of adaptive, data-driven transformation methods in empirical research instead of relying on a single universal approach.

The optimization algorithm used by OSKT can simultaneously target both skewness and kurtosis. It differs from traditional and power-based methods such as BC, ABC, RBC, YJ, and RYJ, which typically optimize a single parameter, $\lambda$. Consequently, OSKT has higher computational complexity. Despite the complexity leading to longer processing times, especially for very large datasets, the increase in computation was moderate and justified by the significant improvement in normalization performance. According to the comparative analysis, the computing times of OSKT were still within a comparable range to those of ABC, RBC, and YJ. However, the disadvantage of the computational cost was already outweighed by the accuracy and robustness achieved through simultaneous optimization of skewness and kurtosis. We developed an R package named *osktnorm* (version 1.0.1), which is available on GitHub (https://github.com/zcebeci/osktnorm)

(accessed on 25 January 2026), to support and facilitate the practical application of the proposed method. This package contains both normalization and reverse transformation processes, leveraging C++ implementation and OpenMP parallelization to ensure high computational performance. Building on this framework, future research will cover the key areas where OSKT can be extended. To achieve this, the first plan is to research additional transformation methods, such as the g-and-k distribution, as an alternative tail modeling analysis for diverse datasets. Second, considering the rapidly growing size of large-scale data, we aim to further optimize computational efficiency by exploring GPU-based acceleration alongside the current OpenMP implementation. Finally, we will adapt the proposed method for multivariate data and integrate it smoothly with postprocessing analysis pipelines, which will be among the priorities of future research.

**Author Contributions:** Conceptualization, Z.C.; methodology, Z.C. and M.C.G.; software, Z.C.; validation, F.C.; formal analysis, F.C. and M.C.G.; investigation, A.U.; resources, M.C.G.; data curation, M.C.G. and A.U.; writing—original draft preparation, Z.C., M.C.G., F.C. and A.U.; writing—review and editing, A.U. and M.C.G.; visualization, F.C.; supervision, Z.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** The present work received no external funding.

**Data Availability Statement:** The simulation data used in this study is described in a reproducible manner. The real-world datasets are openly accessible through R's *datasets* package and the UCI Machine Learning Repository. The implementation of the proposed method is available as the supplementary codes. Further inquiries can be directed to the corresponding author. The R and C++ implementation of the proposed method can be accessed at https://github.com/zcebeci/osktnorm (accessed on 25 January 2026).

**Acknowledgments:** We would like to thank K. Erkoc and M. Sakiroglu for providing the *Onobrychis* dataset.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| OSKT | Optimized Skewness and Kurtosis Transformation |
| REC | Reciprocal Transformation |
| SQR | Square Root Transformation |
| LOG | Logarithmic Transformation |
| ASN | Inverse Hyperbolic Sine Transformation |
| EXP | Exponential Transformation |
| BC | Box–Cox Transformation |
| YJ | Yeo–Johnson Transformation |
| ABC | Adaptive Box–Cox Transformation |
| RewML | Reweighted Maximum Likelihood |
| ML | Maximum Likelihood |
| RBC | Reweighted Box–Cox Transformation |
| RYJ | Reweighted Yeo–Johnson |
| LMB | The Lambert W Function |
| GSH | Generalized Skewness and Heavy-tails Model |
| INT | The Inverse Normal Transformation |
| ORQ | Ordered Quantile Normalization |
| CDF | Cumulative Distribution Function |
| AD | Anderson–Darling |
| $A^2$ | Anderson–Darling Test Statistic |
| GLM | Generalized Linear Model |

| | |
|---|---|
| PPM | Pearson P/df Statistics |
| EDF | Empirical Distribution Function |
| SW | Shapiro–Wilk |
| RJB | Robust Jarque–Bera |
| DAG | D'Agostino Omnibus Test |
| KS | Kolmogorov–Smirnov |
| LIL | Liliefors |
| ZA | Zhang–Wu 1 |
| ZC | Zhang–Wu 2 |
| CVM | Cramér–Von Mises |
| ECDF | Empirical Cumulative Distribution Function |
| FB | Frequency-based |
| MB | Moment-based |
| RCB | Regression/Correlation-based |
| SF | Shapiro–Francia |
| JB | Jarque–Bera |
| ORG | Original Data |
| Var. | Variable |
| TM | Transformation Method |
| Nt | Number of Tests that Confirmed the Normality of the Transformed Data |
| NR_RMSE | Root Mean Square Error Obtained with Newton–Raphson |
| UR_RMSE | Root Mean Square Error Obtained with Uniroot Method |
| NR_CT | Computing Time (s) for Newton–Raphson |
| UR_CT | Computing Time (s) for Uniroot Method |

## Appendix A

```
# R implementation of OSKT
oskt <- function(x,
      init_params = c(0.1, 0.1),
      lower_bounds = c(−1, 0),
      upper_bounds = c(1, 0.5)) {
   tryCatch({
      x <- as.numeric(scale(x))
      transformgh <- function(x, g, h) {
         if (g != 0) {
            return((exp(g * x) − 1)/g * exp(h * x^2/2))
         } else {
            return(x * exp(h * x^2/2))
         }
      }
      objfun <- function(params) {
         g <- params[1]
         h <- params[2]
         y <- transformgh(x, g, h)
         test_statistic <- nortest::ad.test(y)$statistic
         return(test_statistic)
      }
      opt <- optim(par = init_params,
                   fn = objfun,
                   method = "L-BFGS-B",
```

```
                lower = lower_bounds,
                upper = upper_bounds)
    gopt <- opt$par[1]
    hopt <- opt$par[2]
    transformed <- transformgh(x, gopt, hopt)
    return(list(transformed = transformed, g = gopt, h = hopt))
  }, error = function(e) {
    warning(paste0("Transformation failed: ", e$message, ". Returning original data."))
    return(list(transformed = x, g = NA, h = NA))
  })
}

# Back-transformation for OSKT
backoskt <- function(Z, Xmean, Xsd, gopt, hopt, tol = 1e−8, maxiter = 1e3) {
  invert_single <- function(z) {
    f <- function(xs) {
      if (gopt != 0) {
        ((exp(gopt * xs) − 1)/gopt) * exp(0.5 * hopt * xs^2) − z
      } else {
        xs * exp(0.5 * hopt * xs^2) − z
      }
    }
    lower <- −10; upper <- 10
    root <- tryCatch(
      uniroot(f, lower = lower, upper = upper,
      tol = tol, maxiter = maxiter)$root,
      error = function(e) NA
    )
    return(root)
  }
  Xs <- sapply(Z, invert_single)
  X <- Xs * Xsd + Xmean
  return(X)
}
```

## Appendix B

```
% MATLAB version 2025b implementation of OSKT
function result = oskt(x, init_params, lower_bounds, upper_bounds)
    if nargin < 2
        init_params = [0.1, 0.1];
    end
    if nargin < 3
        lower_bounds = [−1, 0];
    end
    if nargin < 4
        upper_bounds = [1, 0.5];
    end

    try
```

```
        % Standardize input
        x = (x − mean(x))/std(x);

        % Define the g–h transformation
        function y = transformgh(x, g, h)
            if g ~ = 0
                y = (exp(g * x) − 1)/g .* exp(h * x.^2/2);
            else
                y = x .* exp(h * x.^2/2);
            end
        end

        % Objective function: Anderson-Darling test statistic
        function stat = objfun(params)
            g = params(1);
            h = params(2);
            y = transformgh(x, g, h);
            % Use adtest to get statistic
            [~, ~, stat] = adtest(y);
        end

        % Optimization using fmincon
        options = optimoptions('fmincon', 'Display', 'off');
        [opt_params, ~] = fmincon(@objfun, init_params, [], [], [], [], ...
                                  lower_bounds, upper_bounds, [], options);

        % Apply optimal transformation
        gopt = opt_params(1);
        hopt = opt_params(2);
        transformed = transformgh(x, gopt, hopt);

        % Return results
        result.transformed = transformed;
        result.g = gopt;
        result.h = hopt;

    catch ME
        warning(['Transformation failed: ' ME.message '. Returning original data. ']);
        result.transformed = x;
        result.g = NaN;
        result.h = NaN;
    end
end
```

## Appendix C

**Table A1.** Overall comparison of commonly used normalization methods relative to OSKT.

| Method | Advantages | Disadvantages | Comparison to OSKT |
|---|---|---|---|
| ASN/LOG/SQR | Simple; widely used; intuitive | Often corrects only skewness or reduces variance but limited in tail adjustments | OSKT is more flexible and can simultaneously adjust skewness and kurtosis |
| BC | Simple; widely used; interpretable | Uses single parameter, cannot simultaneously correct skewness and kurtosis, and cannot handle negative values | Less flexible and may fail for strongly asymmetric or heavy-tailed distributions |
| YJ | Simple; widely used; handles negative values; interpretable | Single parameter and limited control for some tail behaviors | Less flexible and may not fully normalize skewed or heavy-tailed data |
| ABC | Can stabilize variance and handle skewed data better | Single parameter and limited tail adjustment, furthermore may be sensitive to outliers | Less flexible and cannot fully correct both skewness and kurtosis |
| RBC | Robust to outliers and improves stability over BC | Single parameter and limited simultaneous control, may reduce effectiveness for heavy tails | Less flexible and cannot fully normalize skewness and kurtosis |
| RYJ | Robust to outliers and can handle negative values | Single parameter and limited control for some tail behaviors | Less flexible and may not fully normalize asymmetric or heavy-tailed data |
| LMB | Corrects skew and kurtosis, theoretically grounded | Can be unstable for extreme tails; parameter estimation can be complex and may fail for bimodality cases | Comparable to OSKT but sometimes less robust in cases of severely non-normal or bimodal data |
| OSKT | The $g$ and $h$ parameters allow control of skewness and tail behavior, robust normalization for heavy-tailed distributions, invertible, and data-driven optimization | Requires numerical optimization and slightly higher computational cost | Provides better normalization across a wide range of challenging distributions |

## References

1. Box, G.E.; Cox, D.R. An analysis of transformations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1964**, *26*, 211–243. [CrossRef]
2. Yeo, I.K.; Johnson, R.A. A new family of power transformations to improve normality or symmetry. *Biometrika* **2000**, *87*, 954–959. [CrossRef]
3. Goerg, G.M. Lambert W random variables—A new family of generalized skewed distributions with applications to risk estimation. *Ann. Appl. Stat.* **2011**, *5*, 2197–2223. [CrossRef]
4. Tukey, J.W. *Exploratory Data Analysis*; Reading/Addison-Wesley: Boston, MA, USA, 1977.
5. Hoaglin, D.C. Summarizing shape numerically: The g-and-h distributions. In *Exploring Data Tables, Trends, and Shapes*; Hoaglin, D.C., Mosteller, F., Tukey, J.W., Eds.; Wiley: Hoboken, NJ, USA, 1985; pp. 461–513. [CrossRef]
6. Zhan, T. *R Package*, version 0.1.4; TukeyGH77: Tukey g-&-h Distribution; R Foundation for Statistical Computing: Vienna, Austria, 2025. [CrossRef]

7. Kuo, T.C.; Headrick, T.C. Simulating univariate and multivariate Tukey g-and-h distributions based on the method of percentiles. *Int. Sch. Res. Not.* **2014**, *2014*, 645823.
8. Bee, M.; Trapin, L. A simple approach to the estimation of Tukey's g–h distribution. *J. Stat. Comput. Simul.* **2016**, *86*, 3287–3302. [CrossRef]
9. Xu, G.; Genton, M.G. Tukey g-and-h random fields. *J. Am. Stat. Assoc.* **2017**, *112*, 1236–1249. [CrossRef]
10. Zhang, J.; Zhao, Y.; Liu, M.; Kong, L. A Tukey's g-and-h distribution based approach with PSO for degradation reliability modeling. *Eng. Comput.* **2019**, *36*, 1699–1715. [CrossRef]
11. Möstel, L.; Fischer, M.; Pfälzner, F.; Pfeuffer, M. Parameter estimation of Tukey-type distributions: A comparative analysis. *Commun. Stat.-Simul. Comput.* **2021**, *50*, 957–992. [CrossRef]
12. Zhang, J.; Zhang, A.; Wang, M.; Wei, Y.; Gui, P. Quantile SN curve and its application based on Tukey's g-and-h distribution and whale optimization algorithm. *J. Phys. Conf. Ser.* **2021**, *1746*, 012075. [CrossRef]
13. Al-Saadony, M.F.; Mohammed, B.K.; Melik, H.N. Bayesian estimation for the Tukey GH distribution with an application. *Period. Eng. Nat. Sci.* **2022**, *10*, 112–119. [CrossRef]
14. Guillaumin, A.P.; Efremova, N. Tukey g-and-h neural network regression for non-Gaussian data. *arXiv* **2024**, arXiv:2411.07957.
15. Stephens, M.A. EDF statistics for goodness of fit and some comparisons. *J. Am. Stat. Assoc.* **1974**, *69*, 730–737. [CrossRef]
16. Refaat, M. *Data Preparation for Data Mining Using SAS*; Elsevier: Amsterdam, The Netherlands, 2010.
17. Sugasawa, S.; Kubokawa, T. *Box-Cox Transformed Linear Mixed Models for Positive-Valued and Clustered Data*; CIRJE-F-957; The University of Tokyo, Graduate School of Economics: Tokyo, Japan, 2015. Available online: http://www.cirje.e.u-tokyo.ac.jp/research/dp/2004/2015/2015cf957.pdf (accessed on 25 January 2026).
18. Vélez, J.I.; Correa, J.C.; Marmolejo-Ramos, F. A new approach to the Box–Cox transformation. *Front. Appl. Math. Stat.* **2015**, *1*, 12. [CrossRef]
19. Atkinson, A.C.; Riani, M.; Corbellini, A. The box–cox transformation: Review and extensions. *Stat. Sci.* **2021**, *36*, 239–255. [CrossRef]
20. Blum, L.; Elgendi, M.; Menon, C. Impact of Box-Cox transformation on machine-learning algorithms. *Front. Artif. Intell.* **2022**, *5*, 877569. [CrossRef]
21. Hamasha, M.M.; Ali, H.; Hamasha, S.D.; Ahmed, A. Ultra-fine transformation of data for normality. *Heliyon* **2022**, *8*, e0937. [CrossRef]
22. Yu, H.; Sang, P.; Huan, T. Adaptive box–cox transformation: A highly flexible feature-specific data transformation to improve metabolomic data normality for better statistical analysis. *Anal. Chem.* **2022**, *94*, 8267–8276. [CrossRef]
23. Riani, M.; Atkinson, A.C.; Corbellini, A. Automatic robust Box-Cox and extended Yeo–Johnson transformations in regression. *Stat. Methods Appl.* **2023**, *32*, 75–102. [CrossRef]
24. Raymaekers, J.; Rousseeuw, P.J. Transforming variables to central normality. *Mach. Learn.* **2024**, *113*, 4953–4975. [CrossRef]
25. Lehtonen, J. The Lambert W function in ecological and evolutionary models. *Methods Ecol. Evol.* **2016**, *7*, 1110–1118. [CrossRef]
26. Beasley, T.M.; Erickson, S.; Allison, D.B. Rank-based inverse normal transformations are increasingly used, but are they merited? *Behav. Genet.* **2009**, *39*, 580–595. [CrossRef] [PubMed]
27. McCaw, Z.R.; Lane, J.M.; Saxena, R.; Redline, S.; Lin, X. Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics* **2020**, *76*, 1262–1272. [CrossRef] [PubMed]
28. Peterson, R.A.; Cavanaugh, J.E. Ordered quantile normalization: A semiparametric transformation built for the cross-validation era. *J. Appl. Stat.* **2020**, *47*, 2312–2327. [CrossRef] [PubMed]
29. Anderson, T.W.; Darling, D.A. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Ann. Math. Stat.* **1952**, *23*, 193–212. [CrossRef]
30. Mannel, F.; Aggrawal, H.O. A structured L-BFGS method with diagonal scaling and its application to image registration. *J. Math. Imaging Vis.* **2025**, *67*, 7. [CrossRef]
31. Zhang, Z.; Yuan, G.; Qin, Z.; Luo, Q. An improvement by introducing LBFGS idea into the Adam optimizer for machine learning. *Expert Syst. Appl.* **2025**, *296*, 129002. [CrossRef]
32. Chauhan, D.; Jung, D.; Yadav, A. Advancements in multimodal differential evolution: A comprehensive review and future perspectives. *arXiv* **2025**, arXiv:2504.00717. [CrossRef]
33. Taha, Z.Y.; Abdullah, A.A.; Rashid, T.A. Optimizing feature selection with genetic algorithms: A review of methods and applications. *Knowl. Inf. Syst.* **2025**, *67*, 9739–9778. [CrossRef]
34. Bulmer, M.G. Statistical inference. *Princ. Stat.* **1979**, 165–187.
35. Doane, D.P.; Seward, L.E. Measuring skewness: A forgotten statistic? *J. Stat. Educ.* **2011**, *19*, 1–18. [CrossRef]
36. George, D.; Mallery, P. *SPSS for Windows Step by Step: A Simple Guide and Reference*; 17.0 Update; Allyn & Bacon: Boston, MA, USA, 2010.
37. Kamath, A.; Poojari, S.; Varsha, K. Assessing the robustness of normality tests under varying skewness and kurtosis: A practical checklist for public health researchers. *BMC Med. Res. Methodol.* **2025**, *25*, 206. [CrossRef] [PubMed]

38. Hernandez, H. Testing for normality: What is the best method. *Forsch. Res. Rep.* **2021**, *6*, 1–38.
39. Cebeci, Z.; Ozdemir, F.; Yildiztepe, E. *R Package*, version 1.0.1; Groupcompare: Comparing Two Groups Using Various Descriptive Statistics; R Foundation for Statistical Computing: Vienna, Austria, 2025. [CrossRef]
40. Revelle, M. *R Package*, version 2.5.6; psych: Procedures for Psychological, Psychometric, and Personality Research; Northwestern University: Evanston, IL, USA, 2025. [CrossRef]
41. Cortez, P.; Cerdeira, A.; Almeida, F.; Matos, T.; Reis, J. Modeling wine preferences by data mining from physicochemical properties. *Decis. Support Syst.* **2009**, *47*, 547–553. [CrossRef]
42. Erkoc, K. Re-Evaluation of Taxa Included in Genus *Onobrychis* Using Genomic and Phenomic Approaches. Ph.D. Thesis, Department of Bioengineering, Adana A. Türkeş Science & Technology University, Adana, Turkey, 2025.
43. UCI Machine Learning Repository. *Liver Disorders [Dataset]*; UCI Machine Learning Repository: Irvine, CA, USA, 2016. [CrossRef]
44. Arnastauskaitė, J.; Ruzgas, T.; Bražėnas, M. An exhaustive power comparison of normality tests. *Mathematics* **2021**, *9*, 788. [CrossRef]
45. Henderson, A.R. Testing experimental data for univariate normality. *Clin. Chim. Acta* **2006**, *366*, 112–129. [CrossRef] [PubMed]
46. Pearson, K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1900**, *50*, 157–175. [CrossRef]
47. Peterson, R. Finding optimal normalizing transformations via bestNormalize. *R J.* **2021**, *13*, 310–329. [CrossRef]
48. Shapiro, S.S.; Wilk, M.B. An analysis of variance test for normality (complete samples). *Biometrika* **1965**, *52*, 591–611. [CrossRef]
49. Gel, Y.R.; Gastwirth, J.L. A robust modification of the Jarque-Bera test of normality. *Econ. Lett.* **2008**, *99*, 30–32. [CrossRef]
50. D'Agostino, R.B. An omnibus test of normality for moderate and large size samples. *Biometrika* **1971**, *58*, 341–348. [CrossRef]
51. Kolmogorov, A.N. Sulla determinazione empirica di una legge di distribuzione. *G. dell'Ist. Ital. Degli Attuari* **1933**, *4*, 83–91.
52. Smirnov, N. Table for estimating the goodness of fit of empirical distributions. *Ann. Math. Stat.* **1948**, *19*, 279–281. [CrossRef]
53. Lilliefors, H.W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *J. Am. Stat. Assoc.* **1967**, *62*, 399–402. [CrossRef]
54. Zhang, J. Powerful goodness-of-fit tests based on the likelihood ratio. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2002**, *64*, 281–294. [CrossRef]
55. Zhang, J.; Wu, Y. Likelihood-ratio tests for normality. *Comput. Stat. Data Anal.* **2005**, *49*, 709–721. [CrossRef]
56. Uhm, T.; Yi, S. A comparison of normality testing methods by empirical power and distribution of P-values. *Commun. Stat.-Simul. Comput.* **2023**, *52*, 4445–4458. [CrossRef]
57. R Core Team. *R: A Language and Environment for Statistical Computing*, version 4.5.1; R Foundation for Statistical Computing: Vienna, Austria, 2025. Available online: https://www.R-project.org/ (accessed on 25 January 2026).
58. Gross, J.; Ligges, U. *R Package*, version 1.0-4; nortest: Tests for Normality; R Foundation for Statistical Computing: Vienna, Austria, 2015. [CrossRef]
59. Wuertz, D.; Setz, T.; Chalabi, Y.; Boshnakov, G.N. *R Package*, version 4052.98; fBasics: Rmetrics—Markets and Basic Statistics; R Foundation for Statistical Computing: Vienna, Austria, 2024. [CrossRef]
60. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.
61. Eddelbuettel, D.; Balamuta, J.J. Extending R with C++: A brief introduction to Rcpp. *Am. Stat.* **2018**, *72*, 28–36. [CrossRef]
62. Dagum, L.; Menon, R. OpenMP: An industry standard API for shared-memory programming. *IEEE Comput. Sci. Eng.* **1998**, *5*, 46–55. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.