Iraqi Journal of Statistical Sciences

http://stats.uomosul.edu.iq

Modeling Income Distributions Using Gamma and Lognormal Models: A Case Study from the University of Zakho

Sameer Jalal Hamo[1*] Haithem Taha Mohammed Ali[2]

[1]Department of Mathematics, University of Zakho, Kurdistan Region, Iraq,[2] Department of Economic Sciences, University of Zakho, Kurdistan Region, Iraq.

Article information

Article history:
Received: August 28,2025
Revised: September 10,2025
Accepted: November 3,2025
Available online: June 1,2026

Keywords:
Income Distribution; Gamma Distribution; Lognormal Distribution; Goodness-of-fit tests; Numerical Algorithms.

Correspondence:
Sameer Jalal Hamo*
[1]samirjala93@gmail.com
Haithem Taha Mohammed Ali
[2]haithem.taha@uoz.edu.krd

Abstract

Income distribution modeling plays a critical role in understanding wage structures and inequality within public institutions, particularly in developing economies such as the Kurdistan Region of Iraq. This study examines the suitability of the Gamma and Lognormal distributions for modeling the income structure of university affiliates at the University of Zakho. A systematic sample of 250 employees was selected from the university's total workforce. Two estimation techniques – Maximum Likelihood Estimation and the Method of Moments – were employed to estimate the parameters of both distributions. For the Gamma model, three numerical algorithms (Newton–Raphson, Broyden, and Bisection) were applied to solve the nonlinear likelihood equations, producing stable and consistent parameter estimates. Model adequacy was evaluated using multiple goodness-of-fit tests, including Chi-square, Kolmogorov–Smirnov, Anderson–Darling, and Cramér–von Mises statistics. The results revealed that both distributions adequately describe the total salary data, with MLE providing the best overall fit (e.g., $p > 0.10$ in K–S and A–D tests) and lower variability compared with MEM. Conversely, both models performed poorly for basic salarie, suggesting the need for more flexible or mixed-distribution approaches in such structured data. Overall, the findings confirm the empirical relevance of distributional choice and estimation strategy in modeling income data. The study provides quantitative evidence supporting the use of MLE-based Gamma and Lognormal models as effective tools for assessing income dispersion and inequality in higher-education institutions.

DOI: 10.33899/iqjoss.v23i1.62337 , ©Authors, 2026, College of Computer Science and Mathematics, University of Mosul.
This is an open access article under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Economic inequality analysis, policy design, and wage structure modeling fundamentally rely on the accurate modeling of income distributions. In practice, empirical salary data are often highly right-skewed, especially within institutional contexts such as universities. This skewness typically arises from allowances, bonuses, and hierarchical rank differences, making the assumption of normality inadequate for such datasets.

Two classical models from the exponential family, the Lognormal Distribution (LND) and the Gamma Distribution (GD), have long played a pivotal role in income distribution modeling. The LND, grounded in Gibrat's law of proportionate growth, describes multiplicative income generation processes and remains popular due to its ability to model positive, skewed data [1,2]. Its empirical performance continues to attract attention in contemporary research. For instance, Jabbari Khamnei et al. (2023) examined income data from Iran and compared Lognormal, Pareto, and Gibbs–Boltzmann distributions, concluding that the LND performs moderately but tends to be outperformed by alternative heavy-tailed models in household income applications [3]. Conversely, the GD has traditionally been favored for its flexibility and closed-form expressions of shape and density [4]. Early empirical studies such as Salem and Mount (1974) [5] and McDonald and Ransom (1979) [6] indicated that the Gamma model might better capture middle-income ranges compared to the Lognormal. However, recent findings remain divided. For example, Hlasny (2021) and others reported weak fits for both LND and GD in favor of log-t or Log-Pearson families [7]. More recent developments, such as the Flexible

Interpretable Gamma model (2024), suggest that GD variants can still outperform other models in specific economic contexts [8].

In recent years, both GD and LND have continued to appear in income modeling applications. For example, Jabbari Khamnei et al. (2023) demonstrated their applicability for middle-income populations within an econophysics framework [3], while Paul (2022) confirmed their use within kinetic exchange models involving saving behavior [9].

Despite their extensive use, previous studies have not reached consensus on which distribution better characterizes institutional salary data—a category that differs from national or household income due to its structured and semi-regulated nature. Moreover, few empirical studies have systematically compared estimation techniques for these models in such contexts. This gap motivates the present study, which applies both GD and LND to model the salary data of 250 affiliates at the University of Zakho (UoZ) in the Kurdistan Region (KRG) of Iraq, and evaluates their adequacy under two estimation frameworks: Maximum Likelihood Estimation (MLE) and the Moment Estimation Method (MEM). The MLE was selected for its statistical efficiency and invariance properties, ensuring asymptotically unbiased and consistent estimates. The MEM, on the other hand, was included as a comparative approach due to its computational simplicity and frequent use in applied settings where analytical tractability is required. Model adequacy was further validated through multiple goodness-of-fit (GOF) tests, including the Chi-square, Kolmogorov–Smirnov, Anderson–Darling, and Cramér–von Mises statistics, supported by the use of three numerical optimization algorithms; Newton–Raphson method (NRM, Broyden method (BRM), and Bisection method (BSM) to ensure convergence and solution stability of MLE.

Accordingly, this study contributes to the existing literature by providing an empirical comparison of two exponential-family models − GD and LND − under distinct estimation techniques and by identifying the distributional form that best fits institutional salary data. The results aim to offer practical insights for income modeling, wage policy formulation, and inequality measurement within higher education institutions and similar organizational settings.

Besides this introduction, the study is organized into four main sections before the conclusions. Section 2 presents the mathematical and analytical characteristics of GD and LND, emphasizing their applicability to positive and skewed data. Section 3 explains the estimation methods, with particular focus on MLE and MEM and the numerical algorithms used to obtain optimal solutions. Section 4 provides the empirical analysis, including descriptive statistics, summary indicators, GOF evaluations, and a comparative assessment of the fitted models and estimation methods in terms of efficiency and accuracy. The final section discusses the main findings and concludes the study.

2. Mathematical Properties of Gamma and Lognormal Distributions

Understanding the mathematical characteristics of income distribution models is essential for accurately interpreting statistical findings in socio-economic research. Among these models, members of the exponential family such as the GD and the LND are among the most widely used continuous forms in income modeling. Their mathematical tractability and parametric flexibility, together with their ability to capture the inherent skewness of income data, have made these distributions of particular interest in various applied fields. They have been extensively employed in labor economics, actuarial science, and social policy analysis due to their capability to represent both moderate and extreme income variations effectively [10,11].

2.1 Gamma Distribution

The GD is parameterized by a shape parameter (α) and a rate parameter (β), and it is defined only for positive values. In this study, the alternative parameterization is adopted, where the probability density function (PDF) is given by:

$$f_X(x;\alpha,\beta)=\begin{cases}\dfrac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} & x>0;\ \alpha,\beta>0\\ 0 & otherwise\end{cases} \tag{1}$$

Here, the rate parameter β is the reciprocal of the scale, which appears as a multiplier in the exponential term. This formulation is widely used in statistical software due to its convenience and flexibility in modeling right-skewed data, a common feature of income distributions. The mean (first moment about the origin) and variance (second moment about the mean) of the GD are given respectively by:

$$\acute{\mu}_1 = E[X] = \alpha/\beta, \quad \mu_2 = Var[X] = \alpha/\beta^2$$

The coefficient of variation (CV), which expresses dispersion relative to the mean, is thus:

$$CV = 1/\sqrt{\alpha}$$

As the shape parameter α increases, the distribution becomes more symmetric; in the limit $\alpha \to \infty$, it approaches the Normal distribution. The skewness and kurtosis coefficients, which describe the asymmetry and peakedness of the distribution, are given by:

$$\gamma_1 = 2/\sqrt{\alpha}, \qquad \gamma_2 = 6/\alpha$$

These relationships demonstrate that larger values of α lead to a more symmetric and less heavy-tailed shape, confirming the sensitivity of GD to changes in the shape parameter. The mode exists for $\alpha > 1$ and is expressed as:

$$Mode = (\alpha - 1) / \beta$$

An analytical approximation of the median can also be obtained as:

$$Median \approx (\alpha - 1/3) / \beta$$

Finally, the most widely used measure of income inequality − the Gini Coefficient (GC) − can be approximated for the GD as:

$$GC \approx (2\,\Gamma(\alpha + 0.5))/(\sqrt{\alpha}\,\Gamma(\alpha)) - 1$$

This relationship implies that as the shape parameter α increases, income inequality tends to decrease. Hence, the GD provides an analytically tractable and economically interpretable framework for modeling income distributions, particularly when assessing the dispersion and inequality characteristics of wage data [13].

2.2 Lognormal Distribution

Another important model in income modeling is the LND [10]. A random variable X is said to follow LND if its logarithm is normally distributed, i.e. $Y = ln(X) \sim N(\mu, \sigma^2)$. The PDF of X is given by [10]:

$$f_X(x;\mu,\sigma^2) = \begin{cases} \dfrac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(lnx-\mu)^2}{2\sigma^2}} & x > 0;\ \sigma^2 > 0;\ -\infty < \mu < \infty \\ 0 & otherwise \end{cases} \tag{2}$$

This property makes the LND particularly suitable for modeling variables that exhibit multiplicative growth processes, which are frequently observed in economic and income data [12]. The mean and variance of the LND are expressed as [12]:

$$\acute{\mu}_1 = E[X] = exp\left(\mu + \frac{\sigma^2}{2}\right), \qquad \mu_2 = Var[X] = (exp(\sigma^2) - 1)\, exp(2\mu + \sigma^2).$$

These expressions show how the expectation and dispersion depend on the scale parameter σ. The CV is given by [15]:

$$CV = \sqrt{(e\hat{}(\sigma^2) - 1)}$$

This indicates that income variability increases with larger values of σ. Similarly, the skewness and kurtosis coefficients, which measure asymmetry and peakedness, are given by [15]:

$$\gamma_1 = (e\hat{}(\sigma^2) + 2)\sqrt{(e\hat{}(\sigma^2) - 1)}, \qquad \gamma_2 = e\hat{}(4\sigma^2) + 2e\hat{}(3\sigma^2) + 3e\hat{}(2\sigma^2) - 6$$

The median and mode of the distribution are expressed respectively as [15]:

$$Median = exp(\mu), \qquad Mode = exp(\mu - \sigma^2)$$

The GC, can be computed as [15]:

$$GC = 2\Phi\left(\frac{\sigma_z}{\sqrt{2}}\right) - 1$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. This measure of inequality increases monotonically with the dispersion parameter σ, which explains why the LND is particularly relevant in empirical studies of income data [15].

Overall, both the GD and LND provide highly adaptable and effective models for representing income data [15]. While the GD is analytically tractable and sensitive to shape-related variations, the LND captures multiplicative growth mechanisms and exhibits longer upper tails. Together, their moment measures and GC offer a comprehensive understanding of wage variability and income inequality.

3. Parameter Estimation via MLE and MEM

Since the MLE and MEM are commonly applied to exponential family distributions such as the GD and LND, both methods are considered. Each has its own advantages, depending on the sample characteristics and the complexity of the distribution being analyzed [17].

3.1 Parameter Estimation of GD

For the PDF of GD according to Eq. 1, the log-likelihood function for a random sample $x_1, x_2, \ldots, x_n$ is:

$$ln\, f_X(\alpha, \beta; x_1, \ldots, x_n) = n\alpha\, ln\beta - n\, ln\Gamma(\alpha) + (\alpha - 1)\sum_{i=1}^{n} Lnx_i - \beta\sum_{i=1}^{n} x_i \quad (3)$$

The MLE estimates are obtained by solving the partial derivatives of Eq. 3 with respect to α and β. Because these equations are nonlinear, iterative numerical methods such as Newton-Raphson are typically required [18]. Alternatively, the MEM equates sample and theoretical moments. The mean and variance of GD are $E[X] = \alpha/\beta$ and $Var[X] = \alpha/\beta^2$ respectively. Given a sample mean $\bar{x}$ and variance S^2, the moment estimators are $\hat{\alpha} = (\bar{x})^2/S^2$ and $\hat{\beta} = S^2/\bar{x}$. This method is computationally simpler but may yield less efficient estimators, particularly in small samples [19].

3.2 Parameter Estimation of LND

For the PDF of LND according to Eq. 2, the log-likelihood function for a random sample $x_1, x_2, \ldots, x_n$ is:

$$ln\, f_X(\mu,\, \sigma^2; x_1, \ldots, x_n) = -n\, ln(\sigma\sqrt{(2\pi)}) - \sum_{i=1}^{n} Lnx_i - (1/(2\sigma^2))\sum_{i=1}^{n}(ln\, x_i - \mu)^2 \quad (4)$$

Maximizing Eq. 4 by solving the partial derivatives with respect to μ and σ^2 yields closed-form MLE estimates $\hat{\mu} = \sum_{i=1}^{n} lnx_i/n$ and $\widehat{\sigma^2} = \sum_{i=1}^{n}(lnx_i - \hat{\mu})^2/n$. In MEM, based on the population and sample moments, the moment estimators of LND are $\hat{\mu} = ln\left(\sqrt{n}\bar{x}^2/\sqrt{\sum_{i=1}^{n} x_i^2}\right)$ and $\widehat{\sigma^2} = \ln\left(\sum_{i=1}^{n} x_i^2/n\bar{x}^2\right)$. Though straightforward, this approach is less efficient than MLE when the data exhibit high skewness [16].

4. Real Data Application

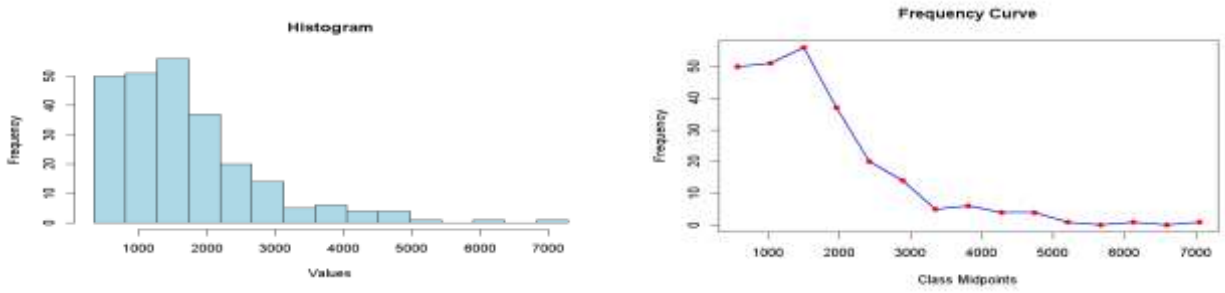
This section presents an empirical evaluation of the suitability of the GD and LND for modelling income data, focusing on the total salary (TS) and basic salary (BS) of employees at the UoZ in KRG of Iraq for December 2024.

4.1 Data Description and Descriptive Statistics

This section provides an empirical assessment of the suitability of the GD and LND distributions for modeling income data. The analysis focuses on the TS and BS of employees at UoZ in the KRG of Iraq for December 2024. A sample of 250 employees was selected systematically from a population of 973 staff members. It includes workers, administrative staff, and academic faculty across different ranks, ensuring full representation of the university's employment structure.

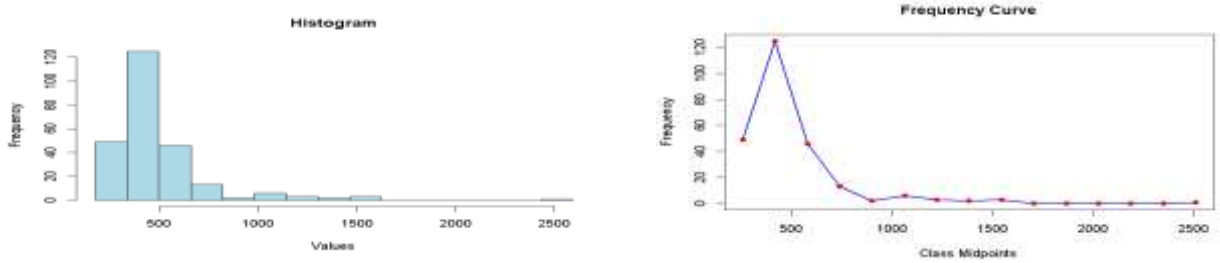
The analysis presents descriptive statistics and applies two estimation methods − MLE and MEM − to estimate key statistical indicators for GD and LND. All computations were performed using the R software environment. Furthermore, numerical optimization algorithms, including NRM, BRM, and BSM, were used to solve parameter estimation problems for the GD [17,18].

Figures 1 and 2 present the histograms and frequency polygons for the TS and BS, visually illustrating the right-skewed nature of both distributions.



a **b**

Figure 1: (a) Histogram and (b) frequency polygon of TS for UoZ staff in the KRG of Iraq (1,000 IQD) for December 2024



a **b**

Figure 2: (a) Histogram and (b) frequency polygon of BS for UoZ staff in the KRG of Iraq (1,000 IQD) for December 2024

Table 1 summarizes the descriptive statistics for TS and BS, including measures of mean, variance, standard deviation (SD), CV, median (M_d), mode ((M_o), skewness (γ_1), kurtosis (γ_2) and GC.

Table 1: Summary of Descriptive Statistics for TS and BS Data of UoZ Staff in the KRG of Iraq (1,000 IQD)

Descriptive Statistics	TS	BS
Mean	1686.1582	484.0480
Variance	1161488.6867	68548.2146
SD	1077.7238	261.8171
CV	63.9159	54.0891
M_d	1469.0000	429.0000
M_o	440.1000, 500.1000, 4513.2000	368.0000
γ_1	1.6723	3.4568
γ_2	7.0601	21.3594
G	0.3293	0.2336

The inequality and variation indicators in the Table 1 − such as G, CV, and γ_1 − demonstrate a higher degree of income disparity in TS ($G \approx 0.33,\ CV \approx 63.92\%,\ \gamma_1 \approx 1.67$) compared to BS ($G \approx 0.23,\ CV \approx 54.09\%,\ \gamma_1 \approx 3.46$). These results highlight the considerable impact of non-basic income components (e.g., allowances and bonuses) on overall salary distribution.

4.2 Goodness-of-Fit Evaluation

After applying the estimation techniques – MLE and MEM – to both GD and LND, the fitted distributions were evaluated for their suitability in representing the income data.

For the GD, Table 2 presents the fitting results of TS and BS data under four approaches: MEM and three numerical optimization methods – NRM, BRM, and BSM – for MLE. GOF was assessed using several standard tests, including the Chi-square test (CST), Kolmogorov-Smirnov test (KST), Anderson-Darling test (ADT), and Cramér-von Mises test CvMT tests.

Table 2: Test statistics and p-values from GOF tests for TS and BS data using GD estimated by four approaches: MEM and MLE via NRM, BRM, and BSM

Test statistics & p-value for GOD Tests	TS				BS			
	Estimation Methods				Estimation Methods			
	MLE			MEM	MLE			MEM
	NRM	BRM	BSM		NRM	BRM	BSM	
CST (χ^2)	8.3096	8.3096	8.6723	6.3131	108.4469	108.4469	108.5072	101.0000
p-value	0.2163	0.2163	0.1929	0.3890	0.0000	0.0000	0.0000	0.0000
KST (D)	0.0589	0.0589	0.0611	0.0643	0.1713	0.1713	0.1712	0.6948
p-value	0.3520	0.3520	0.3081	0.2531	0.0000	0.0000	0.0000	0.0000
ADT (A^2)	0.9700	0.9700	1.0069	1.2189	9.1354	9.1354	9.1372	9.0000
p-value	0.3734	0.3734	0.3536	0.2604	0.00003	0.0000	0.0000	0.0000
CvMT (w^2)	0.1122	0.1122	0.1189	0.1219	1.6607	1.6607	1.6608	56.7030
p-value	0.5285	0.5285	0.5002	0.4883	0.0000	0.0000	0.0000	0.0000

As shown in the Table 2, the GD provided a satisfactory fit to the TS data across all MLE-based methods, with p-values exceeding 0.20. In contrast, BS data failed to satisfy the GOF conditions under both MLE and MEM approaches, with p-values near 0.00, suggesting that the GD is not appropriate for modelling BS data.

Similarly, Table 3 summarizes the fitting outcomes of TS and BS data for LND under the same estimation methods.

Table 3: Test statistics and p-values from GOF tests for TS and BS data using LND estimated by MLE and MEM

Test Statistics & p-value for GOD Tests	TS		BS	
	Estimation Methods		Estimation methods	
	MLE	MEM	MLE	MEM
CST (χ^2)	9.9502	13.2841	79.3109	122.9025
p-value	0.1268	0.0387	0.0000	0.0000
KST (D)	0.0518	0.0541	0.1409	0.1527
p-value	0.5138	0.4572	0.0001	0.00002
ADT (A^2)	0.7688	1.1067	5.2103	9.6893
p-value	0.5039	0.3058	0.0023	0.00001
CvMT (w^2)	0.1166	0.1401	0.9343	1.6584
p-value	0.5098	0.4215	0.0035	0.00007

The outcomes listed in the Table 3 indicate that the LND fits the TS data reasonably well when using MLE (e.g., Chi-square p-value = 0.1268). For BS data, both estimation methods result in poor model fit, consistent with the behavior observed under the GD, Table 2.

In conclusion, both GD and LND proved effective in modelling TS at UoZ, with the Lognormal model exhibiting a slight advantage in capturing the distributional tails. This is reflected in higher empirical kurtosis (approximately 10.8) and a more flexible structure for modeling extreme values.

These findings support the broader literature that highlights the applicability of GD and LND in income distribution studies [14,19], and they underscore the importance of employing rigorous estimation techniques and multiple GOF diagnostics in applied statistical modeling.

4.3 Parameter Estimation and Comparative Diagnostics

Table 4 presents the estimated indicators of the GD for the target sample TS under three MLE-based estimation approaches: NRM, BRM, and BSM. The results reveal a consistent structural pattern among these approaches, where all statistical indicators exhibit nearly identical values with minimal numerical variation. This strong internal consistency is further supported by the Hessian matrix analysis, in which the second derivatives were -3.77×10^{2} and -9.20×10^{8} with a positive determinant ($\approx 1.85 \times 10^{11}$), confirming that the log-likelihood surface is concave and that the obtained MLE solutions correspond to a unique global maximum. These numerical properties validate the stability and reliability of the MLE-based estimates for the GD model applied to the target sample.

Table 4: Estimated Indicators of GD for the TS (1000 ID)

Indicators	MLE-NRM	MLE-BRM	MLE-BSM
$\hat{\alpha}$	2.8998	2.8998	2.9388
$\hat{\beta}$	0.0017	0.0017	0.0017
$\hat{\mu}_1$	1686.1580	1686.1580	1686.1580
$\hat{\mu}_2$	980456.6200	980456.6200	967434.4400
$\hat{S}$	990.1801	990.1801	983.5825
$\widehat{CV}$	0.5872	0.5872	0.5833
$\widehat{M}_d$	1496.3441	1496.3441	1498.8126
$\widehat{M}_o$	1104.6845	1104.6845	1112.4075
$\hat{\gamma}_1$	1.1745	1.1745	1.1667
$\hat{\gamma}_2$	2.0691	2.0691	2.0416
$\hat{G}$	0.3174	0.3174	0.3155

It can be shown that the estimated values of the shape parameter $\hat{\alpha}$ are between $2.8998 - 2.9388$ across the three MLE-based approaches, indicating a highly consistent structure of the GD. The rate parameter $\hat{\beta}$ remains identical among these methods at 0.0017, confirming the stability of the estimation. The first moment $\hat{\mu}_1$ (mean) is approximately $1{,}686.158$ thousand ID for all methods, showing that the estimated mean is directly aligned with the empirical mean of the target sample (TS). Both $\hat{\mu}_2$ and $\hat{\sigma}$ also exhibit very close values, around $980{,}000$ and 990 respectively, indicating minimal variability among the MLE-based results. $\widehat{CV} \approx 58\%$ reflects moderate dispersion in income data, while the skewness $\hat{\gamma}_1$ and kurtosis $\hat{\gamma}_2$ estimates demonstrate a distribution moderately skewed to the right and slightly peaked, which is consistent across all MLE procedures. The GC estimate ($\hat{G} \approx 0.3174$ also remains stable across methods, suggesting similar levels of income inequality. Overall, these findings confirm that the MLE-based estimators provide consistent, stable, and statistically reliable results for fitting the GD to the target sample.

Table 5 presents the estimated indicators of the LND for the target sample (TS) under the MLE approach. The results exhibit internal consistency among the distributional indicators, demonstrating that the estimated parameters provide a coherent representation of the income data. The Hessian matrix analysis, in which the second derivatives were -2.48×10^3 and -8.48×10^5 with a positive determinant ($\approx 2.10 \times 10^9$), confirms that the log-likelihood surface is concave and that the obtained MLE solutions correspond to a unique global maximum. These numerical findings verify the stability and robustness of the MLE-based estimates in modeling the target sample using the LND.

Table 5: Estimated Indicators of LND for the TS data (in 1000 ID)

Estimation	MLE
$\hat{\mu}$	7.2480
$\widehat{\sigma^2}$	0.3704
$\hat{\mu}_1$	1691.1922
$\hat{\mu}_2$	1282277.3100
$\hat{\sigma}$	1132.3768

$\widehat{CV}$	0.6696
$\widehat{M}_d$	1405.2690
$\widehat{M}_o$	970.2698
$\hat{\gamma}_1$	2.3089
$\hat{\gamma}_2$	10.7693
$\hat{G}$	0.3331

The findings indicate that the MLE method provides a coherent and stable representation of the LND for the target sample (TS). The estimated mean $\hat{\mu}_1 = 1691.1922$ thousand ID closely aligns with the observed sample mean, confirming the adequacy of the fitted model in capturing the central tendency of the data. The estimated variance $\widehat{\sigma^2} = 0.3704$ and standard deviation $\hat{\sigma} = 1132.3768$ reflect moderate dispersion in income values, with $\widehat{CV} \approx 66.96\%$ indicating a relatively high degree of variability among the salaries. The estimated median $\widehat{M}_d = 1405.2690$ and mode $\widehat{M}_o = 970.2698$ are lower than the mean, suggesting a positively skewed distribution with most income values concentrated below the average. This is supported by the skewness ($\widehat{\gamma}_1 = 2.3089$) and kurtosis ($\hat{\gamma}_2 = 10.7693$), which indicate pronounced asymmetry and heavy tails. The estimated GC ($\hat{G} = 0.333$) further confirms the presence of noticeable income inequality within the TS. Overall, these results emphasize that the MLE-based Lognormal model effectively characterizes the distribution of salaries, capturing both the skewness and the inequality inherent in the observed income data.

5. Discussion and Conclusions

In this study, the GD and LND – both members of the exponential family – were applied to model the TS and BS of employees at the UoZ in KRG of Iraq. The empirical analysis, performed under comparable estimation settings, aimed to assess the adequacy and stability of these models using MLE, complemented by a Hessian matrix verification to ensure concavity of the likelihood surface and the validity of the estimated optima.

The findings confirm that both GD and LND adequately represent the distributional structure of TS, exhibiting the characteristic right-skewed pattern typical of income data. This right-tail behavior reflects the exponential-family nature of both models, which inherently accommodate asymmetry arising from variable income components such as allowances, bonuses, and position-based differences. The estimated parameters for both models under MLE showed a high degree of internal consistency, supported by negative-definite Hessian matrices, confirming that the obtained estimates correspond to unique global maxima. In contrast, the BS data did not achieve satisfactory GOF under either model, suggesting that its more rigid structure – defined by fixed salary scales – produces limited variability inconsistent with the flexibility of continuous skewed models. Comparatively, the MLE results demonstrate that both distributions are statistically appropriate for TS, though with subtle differences in their ability to capture tail behavior. The Lognormal model exhibited slightly higher empirical kurtosis and a marginally greater capacity to accommodate extreme income values, while the Gamma model displayed superior numerical stability and robustness across different iterative optimization methods (NRM, BRM and BSM). These findings reinforce the reliability of MLE as a consistent framework for modeling positively skewed income data, where all parameter estimates, dispersion indicators, and inequality measures (such as the GC) converge toward stable and interpretable results. In the GOF comparisons, the MEM was also applied as a complementary estimation approach to provide a reference benchmark for evaluating the performance of MLE. The consistently higher p-values and lower test statistics obtained under MLE validate its superior efficiency and accuracy in fitting empirical salary distributions compared with MEM.

Overall, the results of this study lead to four main conclusions:

1. Model adequacy: Both GD and LND provide satisfactory fits for total salaries, capturing the pronounced right-skewness and income inequality typical of institutional pay structures.
2. Exponential-family suitability: The observed skewness and tail behavior are consistent with the exponential-family nature of GD and LND, confirming their theoretical relevance for income data modeling.
3. Estimation reliability: The inclusion of Hessian matrix diagnostics affirms the concavity and stability of the MLE-based solutions, demonstrating the robustness of the estimation framework.
4. Practical implications: While LND shows a slight advantage in modeling extreme incomes, GD remains a dependable and computationally stable model, particularly suitable for institutional income datasets.

In conclusion, both the Gamma and Lognormal models proved theoretically and empirically appropriate for modeling income distributions with positive skewness. Their convergence under MLE reinforces their shared exponential-family

structure and analytical tractability, offering valuable tools for understanding income inequality and wage dispersion in higher-education institutions. Future research may explore extended or three-parameter forms of these models to capture additional complexity, particularly in the modeling of basic salaries where boundedness and discreteness may be influential.

Acknowledgment

The authors would express they're thanks to University of Zakho and College of Administration and Economics, to support this report.

Conflict of interest

The author has no conflict of interest.

Ethical Approval

Ethical approval was not required for this study as it did not involve human participants, personal data.

References

1. Gibrat, R. (1931). Les inégalités economiques, (Paris, France: Librairie du Recueil Sirey).
2. Aitchison, J., & Brown, J. A. (1957). C. (1957). The Lognormal Distribution. Cambridge University. https://doi.org/10.2307/1235218.
3. Khamnei, H. J., Nikannia, S., Fathi, M., & Ghorbani, S. (2023). Modeling income distribution: An econophysics approach. Mathematical Biosciences and Engineering, 20(7), 13171-13181. https://doi.org/10.3934/mbe.2023587.
4. Mori, S., Nakata, D., & Kaneda, T. (2015). An application of the gamma distribution to the income distribution and the estimation of potential food demand functions. Modern Economy, 6(9), 1001-1017. http://dx.doi.org/10.4236/me.2015.69095.
5. Salem, A. B., & Mount, T. D. (1974). A convenient descriptive model of income distribution: the gamma density. Econometrica: journal of the Econometric Society, 1115-1127.https://doi.org/10.2307/1914221.
6. McDonald, J. B., & Ransom, M. R. (1979). Functional forms, estimation techniques and the distribution of income. Econometrica: Journal of the Econometric Society, 1513-1525. https://doi.org/10.2307/1914015.
7. Hlasny, V. (2021). Parametric representation of the top of income distributions: Options, historical evidence, and model selection. Journal of Economic Surveys, 35(4), 1217-1256. https://doi.org/10.1111/joes.12435.
8. Wagener, M., Bekker, A., Arashi, M., & Punzo, A. (2024). Uncovering a generalised gamma distribution: from shape to interpretation. Results in Applied Mathematics, 22, 100461. https://doi.org/10.1016/j.rinam.2024.100461.
9. Paul, S., Mukherjee, S., Joseph, B., Ghosh, A., & Chakrabarti, B. K. (2022). Kinetic exchange income distribution models with saving propensities: inequality indices and self-organized poverty level. Philosophical Transactions of the Royal Society A, 380(2224), 20210163.https://doi.org/10.1098/rsta.2021.0163.
10. Young, G. A., Smith, R. L., & Smith, R. L. (2005). Essentials of statistical inference (Vol. 16). Cambridge University Press. https://doi.org/10.1017/CBO9780511755392.
11. Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2008). Mathematical statistics with applications (Vol. 7). Belmont, CA: Thomson Brooks/Cole.
12. David, H. A., & Nagaraja, H. N. (2004). Order statistics. John Wiley & Sons. https://doi.org/10.1002/0471722162.
13. Chotikapanich, D., & Griffiths, W. E. (2008). Estimating income distributions using a mixture of gamma densities. In Modeling income distributions and Lorenz curves (pp. 285-302). New York, NY: Springer New York.https://doi.org/10.1007/978-0-387-72796-7_16.
14. Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: keys and clues: BioScience, 51(5), 341-352.https://doi.org/10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2.
15. Teekens, R., & Koerts, J. (1972). Some statistical implications of the log transformation of multiplicative models. Econometrica: Journal of the Econometric Society, 793-819.https://doi.org/10.2307/1912069.
16. Kleiber, C., & Kotz, S. (2003). Statistical size distributions in economics and actuarial sciences. John Wiley & Sons. https://doi.org/10.1002/0471457175.
17. Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. The computer journal, 7(4), 308-313.https://doi.org/10.1093/comjnl/8.1.27.

18. Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). Section 17.4. Second-order conservative equations. Numerical recipes: The art of scientific computing, 3rd ed., Cambridge University Press, New York, 7.
19. Gupta, A. K., & Nadarajah, S. (2004). Handbook of Gamma Distribution, Chapman and Hall.

نمذجة توزيعات الدخل باستخدام نموذجي جاما واللوغاريتم الطبيعي: دراسة حالة في جامعة زاخو

سمير جلال حمو[1] ، هيثم طه محمد علي [2]

[1]قسم الرياضيات ، جامعة زاخو ، إقليم كردستان ، العراق ، [2]قسم العلوم الاقتصادية ، جامعة زاخو ، إقليم كردستان ،العراق.

الخلاصة :يُعدّ نمذجة توزيع الدخل أداة أساسية لفهم هيكل الأجور وقياس عدم المساواة داخل المؤسسات العامة، ولا سيّما في الاقتصادات النامية مثل إقليم كوردستان العراق. تهدف هذه الدراسة إلى تقييم مدى ملاءمة توزيعي جاما (Gamma) واللوغارتمي الطبيعي (Lognormal) في توصيف هيكل الدخل للعاملين في جامعة زاخو.

اعتمدت الدراسة على عينة منتظمة مكونة من 250 موظفًا تم اختيارهم من إجمالي قوى العمل في الجامعة. واستُخدمت طريقتان لتقدير المعلمات هما :طريقة الإمكان الأعظم (Maximum Likelihood Estimation) وطريقة العزوم (Method of Moments) لتقدير معلمات كلا التوزيعين. بالنسبة لتوزيع جاما، جرى تطبيق ثلاث خوارزميات عددية هي نيوتن–رافسون (Newton–Raphson) وبرويدن (Broyden) والقطع الثنائي (Bisection) لحل معادلات الإمكان غير الخطية، وقد نتج عنها تقديرات مستقرة ومتسقة للمعلمات.

تم تقييم ملاءمة النماذج باستخدام عدة اختبارات لحسن المطابقة تضمنت اختبار (Chi–square)، وكولموغوروف–سميرنوف (Kolmogorov–Smirnov)، وأندرسون–دارلينغ (Anderson–Darling)، وكرامر –فون ميزس (Cramér–von Mises) وأظهرت النتائج أن كلا التوزيعين يقدمان تمثيلاً مناسبًا لبيانات الرواتب الكلية، في حين تفوقت طريقة الإمكان الأعظم من حيث جودة المطابقة $p > 0.10$ في اختباري كولموغوروف–سميرنوف واندرسون–دارلينغ وانخفاض التباين مقارنة بطريقة العزوم. بالمقابل، أظهرت النتائج ضعف أداء النموذجين عند تطبيقهما على الرواتب الأساسية، مما يشير إلى الحاجة إلى نماذج أكثر مرونة أو توزيعات مختلطة للتعامل مع مثل هذه البيانات المنظمة.

بوجه عام، تؤكد النتائج الأهمية التطبيقية لاختيار التوزيع المناسب وطريقة التقدير في تحليل بيانات الدخل. كما تقدم الدراسة أدلة كمية تدعم استخدام توزيعي جاما واللوغارتمي الطبيعي المعتمدين على الإمكان الأعظم كأدوات فعّالة في تقييم تشتت الدخل وقياس عدم المساواة في مؤسسات التعليم العالي.

الكلمات المفتاحية : توزيع الدخل، توزيع جاما، التوزيع اللوغاريتمي الطبيعي، اختبارات الملاءمة، الخوارزميات العددية.