Permutation-Based Inference for Benford's Law under Skewed Generalized Error Distributions

Gianfranco Piscopo[1], Valerio Ficcadenti*[2], Massimiliano Giacalone[3], and Maria Longobardi[1]

[1]Department of Mathematics and Applications Renato Caccioppoli, University of Naples Federico II, Monte S. Angelo Complex, 80126 Naples, Italy, Emails: gianfranco.piscopo@unina.it; maria.longobardi@unina.it

[2]Business School, London South Bank University, 103 Borough Road, London SE1 0AA, United Kingdom, Email: ficcadv2@lsbu.ac.uk

[3]Department of Economics, University of Campania Luigi Vanvitelli, Corso Gran Priorato di Malta 1, 81043 Capua (Caserta), Italy , Email: massimiliano.giacalone@unicampania.it

Abstract

Benford's law is widely used as a diagnostic tool in applied contexts, relying on the distribution of leading digits to assess conformity with expected logarithmic frequencies. While its theoretical properties are well understood for scale-invariant mechanisms and exponential-type models, less attention has been devoted to the combined role of distributional asymmetry and tail behaviour in data-generating processes.

This paper studies Benford compliance under the Skewed Generalized Error Distribution (SGED). We assess conformity using classical goodness-of-fit measures

*Corresponding author

(chi-square and MAD) and a multivariate inferential strategy based on the Non-Parametric Combination (NPC) methodology.

Monte Carlo evidence illustrates how skewness and tails interact in shaping digit frequencies and highlights systematic differences between univariate diagnostics and NPC-based global inference. The results support a model-aware interpretation of digit laws and provide practical guidance for rigorous Benford assessment in skewed and heavy-tailed environments.

Keywords: Benford's law; Skewed generalized error distribution; Non-parametric combination methodology (NPC); Monte Carlo design.

1 Introduction

Benford's law describes the logarithmic distribution of leading digits observed in many naturally occurring numerical datasets. Since the seminal contributions of Benford (1938); Newcomb (1881), this law has attracted sustained interest both as a theoretical regularity and as a practical diagnostic tool. In applied settings, deviations from Benford frequencies are often interpreted as potential indicators of data manipulation, recording errors, or structural anomalies, and the law has been widely employed in auditing, accounting, forensic analysis, and financial applications (Nigrini, 1996; Berger and Hill, 2011; Ausloos et al., 2021; Abrantes-Metz et al., 2011).

Despite its claimed universality, compliance with Benford's law depends on a number of elements underlying the phenomenon being studied. Its validity depends on the underlying data-generating mechanism, and substantial departures may arise even in the absence of data quality issues (for example the amount of data Ausloos et al., 2021). Understanding how distributional features such as skewness, tail behaviour, and parameterisation affect digit frequencies is therefore essential for a correct statistical evaluation of Benford compliance (Hill, 1995; Formann, 2010).

Theoretical results show that Benford compliance emerges only under specific conditions, such as scale invariance or suitable distributional mixtures (Hill, 1995), and systematic deviations may reflect legitimate distributional properties rather than data irregularities. In modern applications involving financial, economic, and scientific data, where skewness and heavy tails are pervasive, relying on Benford diagnostics without accounting for the underlying distributional structure may lead to misleading conclusions (Robledo, 2011).

To address these issues, this paper focuses on the Skewed Generalized Error Distribution (SGED), a flexible class of models that allows separate control of tail thickness and skewness, hence incorporating some of the distribution characteristics that may affect compliance. SGED-type distributions generalise the Gaussian case and encompass a wide range of shapes, making them particularly suitable for modelling asymmetric and heavy-tailed data observed in practice (Theodossiou, 1998, 2015).
We adopt a two–piece construction in the spirit of the Fernández–Steel transformation (Fernandez and Steel, 1998), which introduces skewness through asymmetric scaling of a standardised symmetric distribution. This approach preserves the interpretability of parameters and provides a coherent framework for simulation and inference. The Gaussian distribution emerges as a special and limiting case, thereby offering a natural reference point for assessing the impact of skewness and tail behaviour on Benford compliance.

The three components of the paper are therefore directly connected. Benford's law provides the digit benchmark, the SGED family supplies a controlled way to vary skewness and tail thickness, and the non-parametric combination (NPC) framework gives a joint inferential device for assessing several dependent digit Benford's laws simultaneously. This link is important because a dataset may appear approximately Benford under one marginal digit rule while failing once the dependence among first, second, and joint first-two digits is evaluated.

Therefore, the main contributions of the paper are as follows:

1. We provide a unified analysis of Benford's law under SGED models constructed via two–piece transformations, with separate control of tail thickness and skewness.

2. We adopt classical descriptive diagnostics (mean absolute deviation, MAD) and compare them with an inferential permutation-based global procedure built from synchronised chi-square tests across digit rules, combined via Fisher's method.

3. We implement a simulation framework based on standardised two–piece constructions, ensuring clear interpretation of parameters and reliable Monte Carlo assessment of digit distributions.

4. We provide a systematic comparison and ranking of SGED scenarios, including sign-split (positive and negative tail) diagnostics, showing how skewness and tail behavior jointly determine Benford compliance.

Relative to classical goodness-of-fit applications, the novelty lies in treating deviations from Benford's law as consequences of an explicit shape-controlled data-generating model rather than only as empirical discrepancies. Relative to recent model-based approaches, the paper isolates the joint roles of tail thickness and skewness through a standardised SGED design. Relative to existing permutation approaches, it applies synchronised re-sampling and Fisher-NPC aggregation to dependent digit rules, so that the inferential conclusion concerns global Benford conformity rather than separate marginal tests.

The remainder of the paper is organised as follows. Section 2 contains an overview of the literature; Section 3 reviews Benford's law and standard diagnostics; Section 4 introduces the SGED framework; Section 5 details the NPC methodology, and Section 6 describes the Monte Carlo design. The results and conclusions are reported in Section 7 and Section 8. The former reports the numerical results and implications, while the latter contains the conclusions and contribution statement.

2 Literature Review

The literature on Benford's law is vast and interdisciplinary, spanning mathematics, statistics, economics, finance, and forensic sciences. Early theoretical contributions established the mathematical foundations of the logarithmic law and its connection with scale invariance, base invariance, and multiplicative processes. In particular, Hill (1995) provides a probabilistic derivation showing how Benford-type behaviour may arise from suitable stochastic constructions, while comprehensive treatments such as those discussed in Berger and Hill (2015) emphasise the role of invariance principles and mixtures of distributions in generating Benford-conforming data.

On the applied side, Benford's law has been widely used as a diagnostic tool for data quality assessment (Diekmann, 2007), anomaly detection (Riccioni and Cerqueti, 2018), and electoral studies (Pericchi and Torres, 2011). In these contexts, conformity to the logarithmic benchmark is often interpreted as an indicator of regularity, whereas systematic deviations may signal manipulation or structural irregularities (Nigrini, 1996, 2012; Durtschi et al., 2004). However, it is well recognised that deviations from Benford's law may also reflect intrinsic characteristics of the underlying data-generating process rather than data contamination (Burgos and Santos, 2021).

From a statistical perspective, empirical assessments of Benford compliance are typically based on univariate goodness-of-fit measures, such as the chi-square statistic, the Kolmogorov–Smirnov distance, or the mean absolute deviation (MAD). While these tools are widely used, their inferential interpretation may be problematic, especially in finite samples or when digit frequencies exhibit dependence across positions. Moreover, classical approaches focus on marginal distributions and may fail to capture more complex patterns of deviation. Recent contributions have therefore emphasised the need for a more rigorous inferential framework, grounded in explicit probabilistic modelling of the Benford hypothesis (Barabesi et al., 2022, 2023).

A parallel strand of the literature highlights that Benford's law is not universal, but depends on specific structural properties of the generating distribution. In particular, support, tail behaviour, and dispersion play a crucial role. For example, Formann (2010) shows that distributions that are sufficiently spread over several orders of magnitude and exhibit right-skewness tend to display closer agreement with the Benford distribution, whereas bounded or highly concentrated distributions often lead to systematic departures.

Despite these advances, comparatively less attention has been devoted to flexible distributional families that allow tail thickness and asymmetry to be modelled jointly. In this context, the Generalized Error Distribution (GED) provides a tractable benchmark (McDonald and Newey, 1988; Nelson, 1991).

The Skewed Generalized Error Distribution extends the GED by introducing an explicit skewness parameter. Applications in financial econometrics include Theodossiou (1998), Lee et al. (2008) and Cerqueti et al. (2020). From a methodological standpoint, the SGED allows disentangling the effects of skewness and tail thickness on digit distributions.

A further issue concerns inference. When multiple digit rules are considered jointly, permutation-based approaches and nonparametric combination methods provide a coherent inferential framework (Pesarin, 2001; Pesarin and Salmaso, 2010; Giacalone, 2022; Bonnini et al., 2024b, 2025).

Recent contributions have also clarified the practical scope and theoretical properties of permutation testing, including broad methodological reviews (Bonnini et al., 2024a), applications to experimental data (Holt and Sullivan, 2023), and minimax optimality results (Kim et al., 2022).

3 Benford's law and Goodness-of-Fit Metrics

3.1 The Newcomb–Benford Law

Let D be the first significant digit of a non-zero random variable X. The Newcomb–Benford Law (NBL) states that

$$\Pr(D = d) = \log_{10}\left(1 + \frac{1}{d}\right), \qquad d = 1, \ldots, 9.$$

Originally observed by Newcomb (1881) and later rediscovered and formalised by Benford (1938), this logarithmic distribution describes a regularity in the leading digits of many empirical datasets arising from natural, social, and economic phenomena. Rather than being uniformly distributed, smaller digits occur with substantially higher frequency.

From a theoretical perspective, Benford's law is closely related to scale invariance and base invariance properties of distributions. If a random variable is scale invariant, meaning that its distribution is unchanged under multiplication by positive constants, then its leading digits necessarily follow the Benford distribution (Pinkham, 1961; Hill, 1995). Benford's law can also emerge as a limit under mixtures and aggregation of distributions (Hill, 1995), and under multiplicative processes.

Benford's law is not universal. Its validity depends critically on the underlying data-generating mechanism, the support of the distribution, and the presence of truncation, rounding, or structural constraints. Datasets confined to narrow ranges or characterised by strong scale anchors often deviate substantially from Benford behavior (Durtschi et al., 2004; Rauch et al., 2011). Consequently, understanding *why* Benford's law holds (or fails) requires careful consideration of distributional shape, tail behavior, and skewness.

3.2 Digit extraction rules and multiple targets

Digit laws can be formulated for several targets beyond the first digit. This paper focuses on three commonly used rules:

- First digit $D_1 \in \{1, \dots, 9\}$;
- Second digit $D_2 \in \{0, \dots, 9\}$;
- First-two digits $D_{12} \in \{10, \dots, 99\}$ (joint first and second digits).

For a nonzero real-valued observation X, let $Y = |X|$ and define

$$m(Y) = \lfloor \log_{10} Y \rfloor, \qquad U(Y) = \frac{Y}{10^{m(Y)}} \in [1, 10).$$

Then the first digit is $D_1 = \lfloor U(Y) \rfloor$ and the second digit is $D_2 = \lfloor 10\, U(Y) \rfloor - 10D_1$. The first-two digits can be computed as $D_{12} = 10D_1 + D_2$.

Benford probabilities for D_2 and D_{12} are:

$$\Pr(D_{12} = k) = \log_{10}\left(1 + \frac{1}{k}\right), \quad k = 10, \dots, 99,$$

and

$$\Pr(D_2 = d) = \sum_{k=1}^{9} \log_{10}\left(1 + \frac{1}{10k + d}\right), \quad d = 0, \dots, 9.$$

The multivariate character of digit laws becomes apparent when multiple rules are analysed jointly. For example, in a given sample the digit counts for D_1 and D_2 are not independent because both derive from the same underlying observations. This dependence is one of the reasons why a global multivariate inferential framework (NPC) is helpful.

3.3 Goodness-of-fit diagnostics

Let $\mathbf{p} = (p_1, \dots, p_9)$ denote the empirical distribution of the first digit and $\mathbf{p}^{\mathrm{B}}$ the Benford probabilities. Compliance can be framed as measuring distance between $\mathbf{p}$ and $\mathbf{p}^{\mathrm{B}}$. Distance-based diagnostics provide quantitative summaries and support model ranking and screening (Nigrini, 2012; Durtschi et al., 2004).

However, the behaviour of distance measures depends on sample size and on the multinomial dependence structure of digit counts. This motivates comparing descriptive indices with inferential tests.

The Pearson chi-square statistic, given sample size n and observed first digit counts

O_d, is

$$\chi^2 = \sum_{d=1}^{9} \frac{(O_d - E_d)^2}{E_d},$$

$$E_d = n \log_{10}\left(1 + \frac{1}{d}\right).$$

Under exact Benford conformity, χ^2 is asymptotically χ^2_8, for the first digit. The test is widely used in auditing and forensic contexts (Nigrini, 1996, 2012).

Yet, it is highly sensitive to sample size and ignores ordinal structure and joint behaviour across digit rules. In large samples, even tiny deviations lead to rejection; in small samples, power may be low (Goodman, 1954; Leemis et al., 2000). This motivates complementing χ^2 with descriptive distances and multivariate permutation inference.

The mean absolute deviation (MAD), for the first digit, is defined as

$$\text{MAD} = \frac{1}{9} \sum_{d=1}^{9} \left| \hat{p}_d - \log_{10}\left(1 + \frac{1}{d}\right) \right|,$$

$$\hat{p}_d = \frac{O_d}{n}.$$

MAD provides a stable descriptive assessment across different sample sizes and is widely used in practice (Nigrini, 1999, 2012; Cerqueti and Maggi, 2025). However, it lacks a natural sampling distribution under H_0 and does not support coherent joint inference across multiple digit rules or datasets.

In practice, digit laws are routinely applied to: (i) multiple digit rules; (ii) multiple datasets; or (iii) multiple subgroups (e.g., country-by-country or firm-by-firm auditing). In such settings, analysts need a global decision while preserving dependence and avoiding multiple-testing inflation. This is precisely the domain where permutation-based multivariate methods (NPC) provide a natural solution (Pesarin, 2001; Pesarin and Salmaso, 2010).

4 SGED: Definition, Properties, and Benford Implications

The Skewed Generalized Error Distribution provides a flexible family for modelling departures from Gaussianity through separate control of tail thickness and skewness.

Intuitively, the tail parameter determines how widely observations spread across orders of magnitude, which is central for Benford behavior because leading digits are governed by logarithmic scale coverage. The skewness parameter determines whether this spread is balanced or concentrated more strongly on one side of the distribution, thereby creating direction-specific departures in positive and negative tails.

4.1 The Generalized Error Distribution

The Generalized Error Distribution (GED), also known as the exponential power distribution, provides a flexible parametric family for modelling symmetric continuous data with varying tail thickness (Subbotin, 1923; Box, 1953; Zeckhauser and Thompson, 1970; Taguchi, 1974; West, 1987; Mineo, 1989).

Key special cases illustrate the family's flexibility:

- $\nu = 2$: reduces to the Gaussian (normal) distribution
- $\nu = 1$: yields the Laplace (double exponential) distribution
- $\nu < 2$: produces heavier-than-Gaussian tails, capturing extreme values
- $\nu > 2$: produces lighter-than-Gaussian tails with sharper central peaks

In its location–scale–shape parameterization, the GED density is defined as

$$f_{\mathrm{GED}}(x; \mu, \sigma, \nu) = \frac{\nu}{2\,\sigma\,\Gamma(1/\nu)} \exp\left\{ - \left| \frac{x-\mu}{\sigma} \right|^{\nu} \right\}, \qquad x \in \mathbb{R}, \tag{1}$$

where $\mu \in \mathbb{R}$ is a location parameter, $\sigma > 0$ is a scale parameter, $\nu > 0$ controls tail thickness, and $\Gamma(\cdot)$ denotes the Gamma function.

This nesting property makes the GED a natural benchmark family for studying departures from normality in a controlled and interpretable way (Nelson, 1991; Fernandez and Steel, 1998).

For inferential and simulation purposes, it is often convenient to work with a standardized GED having zero mean and unit variance. Let $Z \sim \text{GED}(0, 1, \nu)$. The variance of Z is given by

$$\text{Var}(Z) = \frac{\Gamma(3/\nu)}{\Gamma(1/\nu)}. \tag{2}$$

A standardized GED variable with unit variance can therefore be obtained as

$$Z^* = \frac{Z}{\sqrt{\Gamma(3/\nu)/\Gamma(1/\nu)}}.$$

This explicit normalization is crucial when comparing digit distributions across different values of ν, as it isolates the effect of tail thickness from scale effects (Cerqueti et al., 2020; Giacalone, 2022).

From the perspective of Benford's law, the GED provides an analytically tractable yet highly flexible family for investigating how tail behaviour influences leading-digit frequencies. Scale invariance plays a dominant role in Benford compliance, deviations from Gaussianity induced by heavy or light tails can generate systematic compliance with the law (Hill, 1995; Leemis et al., 2000; Berger and Hill, 2015).

Figure 1 illustrates the shape of standardised GED densities for selected values of ν, highlighting the progressive transition from heavy-tailed to light-tailed regimes and the Gaussian benchmark at $\nu = 2$.

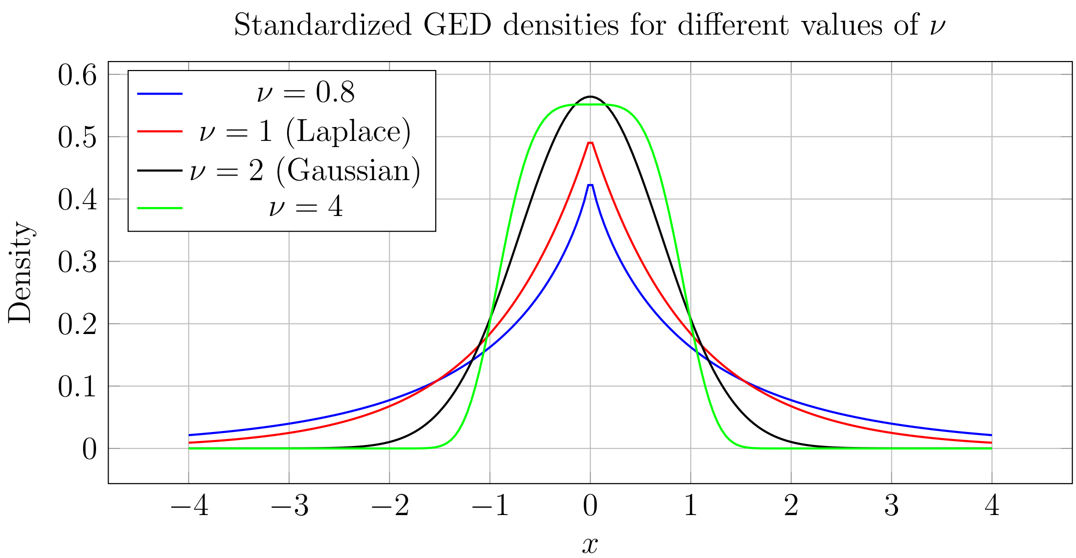

Figure 1: Shapes of standardized Generalized Error Distributions (eq. 1) for selected values of the tail parameter $\nu \in 0.8, 1, 2, 4$. Smaller values of ν produce heavier tails, while larger values lead to lighter-than-Gaussian behavior.

4.2 Two–piece skewing and the SGED

We introduce skewness via the Fernández–Steel two–piece construction (Fernandez and Steel, 1998). Let $f(\cdot \mid \vartheta)$ be a standardised symmetric base density (Normal, GED with parameter ν, or Student-t). Let $\xi > 0$ be the skewness parameter, with $\xi = 1$ yielding symmetry. Then

$$f(z \mid \xi, \vartheta) = \frac{2}{\xi + \xi^{-1}} \left[f(\xi z \mid \vartheta)\, H(-z) + f\left(\frac{z}{\xi} \mid \vartheta\right)\, H(z) \right], \tag{3}$$

where $H(z) = (1 + \operatorname{sign}(z))/2$.

The construction rescales the left and right sides of the mode differently while preserving continuity at 0 and keeping the tail functional form determined by $f(\cdot \mid \vartheta)$. When f is GED with shape ν, the resulting skewed density is the SGED used throughout.

A critical point for meaningful Benford comparisons is to avoid confounding skewness with unintended shifts in mean and variance. Under eq. (3), mean and variance depend on ξ; one therefore re-standardise.

Let

$$M_r = 2 \int_0^{\infty} x^r f(x \mid \vartheta)\, dx$$

be the absolute moment of the positive half. For standardised bases, $M_2 = 1$. Then (Würtz et al., 2006)

$$\mu_\xi = M_1 \left(\xi - \xi^{-1}\right), \qquad \sigma_\xi^2 = \left(M_2 - M_1^2\right)\left(\xi^2 + \xi^{-2}\right) + 2M_1^2 - M_2. \tag{4}$$

Standardised innovations being $Z^\star = (Z - \mu_\xi)/\sigma_\xi$ so that $\mathbb{E}(Z^\star) = 0$ and $\mathbb{V}(Z^\star) = 1$ for every ξ.

The Gaussian model arises as the central benchmark when $\nu = 2$ and $\xi = 1$. Holding $\xi = 1$ and varying ν isolates tail effects; holding $\nu = 2$ and varying ξ isolates skewness. This "Gaussian bridge" is essential for interpreting Benford compliance, because fixed-scale Gaussians are generally not Benford, while suitable scale mixtures can be approximately Benford, as emphasised in the classical literature discussed in the Introduction.

The simulation proceeds in four steps:

1. Start from a standardised symmetric base Z (Normal, standardised GED, or stan-

dardised Student-t).

2. Apply two–piece skewing with parameter ξ via left/right scaling.

3. Re-standardize to $Z^\star$ using eq. (4).

4. Build location–scale observations via $X = \mu + \sigma Z^\star$ – with (μ, σ) fixed in Monte Carlo so that digit effects reflect only shape.

This construction matches widely used implementations in computational finance and ensures that comparisons across (ξ, ν) are not confounded by accidental changes in mean/variance, see Figure 2.

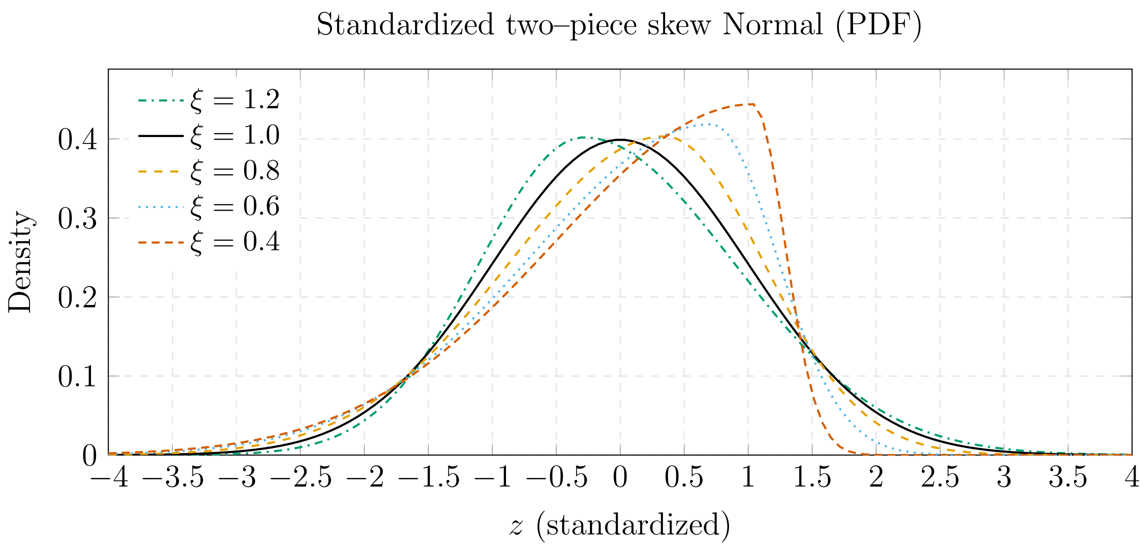

Figure 2: Standardized two–piece (Fernández–Steel) skew Normal (GED with $\nu = 2$): PDF for $\xi \in \{1.2, 1, 0.8, 0.6, 0.4\}$ using eq. (3). The two–piece skewing is applied to a standardized symmetric base and then re-standardized to mean 0 and variance 1.

5 Permutation-Based Inference and the NPC Framework

Classical assessments of Benford conformity are typically based on univariate goodness-of-fit statistics and asymptotic reference distributions. While effective in simple settings, such approaches may be inadequate when multiple digit-based diagnostics are considered jointly or when deviations from Benford's law are weak but structurally coherent. We therefore adopt a permutation-based inferential framework grounded in the Nonparametric Combination (NPC) methodology.

Let $\mathcal{T}_1, \ldots, \mathcal{T}_J$ denote a collection of digit-based diagnostics (e.g., χ^2 and MAD for first digit, second digit, and first-two digits). Each statistic $\mathcal{T}_j$ induces a partial hypothesis

$$H_{0j}: \quad \text{no deviation from Benford's law according to } \mathcal{T}_j.$$

The global null is $H_0 = \cap_{j=1}^J H_{0j}$.

In Benford testing, the data induce multinomial digit counts. Under H_0, the distribution of digit counts is fully specified by Benford probabilities. This permits two equivalent resampling strategies: (i) *Permutation/exchangeability* constructions based on label permutations in suitable contrast representations; (ii) *Parametric bootstrap under* H_0 (multinomial resampling from Benford probabilities), which yields the same reference distribution for digit-count-based statistics.

In our implementation, for each configuration, we generate B Monte Carlo replicates under H_0 to obtain the reference distribution of each partial statistic. The same replicate index is shared across all partial statistics so that dependence among partial tests is preserved.

Given partial p-values $p_1, \ldots, p_J$, we adopt Fisher's combination:

$$T_{\text{NPC}} = -2 \sum_{j=1}^{J} \log(p_j),$$

and evaluate its null distribution using the same resampling replicates, consistent with the NPC prescription (Pesarin, 2001; Pesarin and Salmaso, 2010, 2012). The procedure is advantageous because it delivers a single global decision while maintaining finite-sample validity; it preserves the dependence structure across digit rules because resampling is joint; it supports ranking of models/configurations by global evidence against Benford.

6 Monte Carlo Design

The empirical results presented below are obtained from a carefully parameterised Monte Carlo simulation pipeline, which we describe below.

6.1 Parameter grid and replication scheme

Throughout the simulation we fix $(\mu, \sigma) = (0, 1)$ and generate SGED samples with the `fGarch::rsged()` implementation in R. The evaluated grid is

$$(\xi, \nu) \in \{0.25, 0.50, \ldots, 5.00\} \times \{0.25, 0.50, \ldots, 5.00\},$$

which yields 400 parameter configurations. For each grid cell we simulate $R = 1000$ independent samples of size $n = 10{,}000$ and, for each inferential calculation, we use $B = 500$ synchronised permutation resamples under Benford's law compliance. The resulting archive therefore contains $400,000$ Monte Carlo replications. The Gaussian benchmark remains the symmetric case $(\xi, \nu) = (1, 2)$, but the expanded design allows us to study a much broader continuum of tail regimes and asymmetry levels than the earlier pilot grids. The sample size $n = 10{,}000$ was chosen to make digit-frequency discrepancies visible without relying only on asymptotic approximations, while $R = 1000$ replications per cell stabilise the reported means, medians, and rejection rates over the 400-configuration grid. The choice $B = 500$ provides a fine enough permutation resolution for ranking and rejection summaries while keeping the full Monte Carlo experiment computationally feasible. A summary of the parameters used is presented in Table 1

Table 1: SGED Monte Carlo design.

Component	Specification
Shape grid ν	$\{0.25, 0.50, \ldots, 5.00\}$
Skewness grid ξ	$\{0.25, 0.50, \ldots, 5.00\}$
Configurations	$20 \times 20 = 400$
Sample size per replication	$n = 10{,}000$
Monte Carlo replications per cell	$R = 1000$
Permutation resamples per test	$B = 500$
Digit rules	D_1, D_2, D_{12}
Contexts	combined, positive-only, negative-only
Location/scale	$(\mu, \sigma) = (0, 1)$

6.2 Digit rules, sign contexts, and descriptive diagnostics

For each simulated sample we extract digits from absolute magnitudes, and we repeat the same workflow on the positive subsample $X_+ = \{X_i : X_i > 0\}$ and on the magnitudes of the negative subsample $|X_-| = \{|X_i| : X_i < 0\}$. For each context and each digit rule we compute Pearson's chi-square statistic, its asymptotic reference p-value, a permutation p-value calibrated under the Benford multinomial null, and MAD. This design supports three complementary views of conformity: classical goodness-of-fit, finite-sample permutation calibration, and descriptive distance.

6.3 Implemented NPC statistic

The generic NPC framework described in Section 5 is specialised in the current implementation to the three dependence-linked chi-square tests associated with D_1, D_2, and D_{12}. If p_1, p_2, and p_{12} denote the synchronized permutation p-values for these three digit rules, the implemented global statistic is

$$T_{\mathrm{NPC}} = -2\left\{\log(p_1) + \log(p_2) + \log(p_{12})\right\}.$$

The permutation reference set is generated from a joint Benford multinomial table on $\{10, \ldots, 99\}$ and then marginalised to the first- and second-digit counts, so the natural dependence between digit rules is preserved by construction. MAD is reported throughout the manuscript as a complementary descriptive statistic and as a separate ranking criterion, but it is not used as a partial input to the implemented NPC statistic in this study.

Across the $R = 1000$ replications of each cell, chi-square and MAD are summarised by their means, while permutation and NPC p-values are summarised by medians and rejection rates.

Consider K configurations (models/scenarios/datasets), indexed by $k = 1, \ldots, K$. For each configuration k we observe a sample $\{X_i^{(k)}\}_{i=1}^{n_k}$. Let $m \in \mathcal{M}$ denote the digit position or digit rule under investigation (e.g., $m = 1$ for first digit, $m = 2$ for second digit, $m = 12$ for first-two digits).

Let $D_{i,m}^{(k)} \in \mathcal{D}_m$ be the extracted digit category from $X_i^{(k)}$ at rule m, with $\mathcal{D}_1 = \{1, \ldots, 9\}$ for the first digit and $\mathcal{D}_2 = \{0, \ldots, 9\}$ for the second digit. Denote by $O_{d,m}^{(k)}$ the observed counts and by

$$\hat{p}_{d,m}^{(k)} = \frac{O_{d,m}^{(k)}}{n_k}$$

the empirical digit probabilities.

Let $p_{d,m}^{\mathrm{B}}$ denote the Benford probabilities for digit rule m. For the first digit, $p_{d,1}^{\mathrm{B}} =$

$\log_{10}(1 + 1/d)$ for $d = 1, \ldots, 9$.

For each configuration k and each digit rule $m \in \mathcal{M}$, define a partial test statistic $T_m^{(k)}$ measuring deviation from Benford's law. In this paper we applied it to the chi-square statistic.

$$T_{m,\chi^2}^{(k)} = \sum_{d \in \mathcal{D}_m} \frac{\left(O_{d,m}^{(k)} - n_k p_{d,m}^{\mathrm{B}}\right)^2}{n_k p_{d,m}^{\mathrm{B}}}.$$

In general, the NPC procedure can accommodate a vector of J partial statistics $\{T_j^{(k)}\}_{j=1}^{J}$, where j may index both the digit rule and the discrepancy measure.

For each configuration k we generate a permutation distribution under Benford conformity. Let π_b denote the b-th random permutation, $b = 1, \ldots, B$. The permutation acts on the digit labels by reallocating observed digits to Benford categories in a way consistent with the null. Operationally, one can implement this by resampling digits from the Benford multinomial model (parametric bootstrap under H_0), or by permuting a suitable set of contrast signs when inference is built on symmetry-based transformations. Therefore in our implementation, the permutation framework is equivalent to a parametric bootstrap under the Benford multinomial null. In all cases, the same permutation index b is applied jointly across all partial statistics within configuration k, thereby preserving dependence across digit positions.

Let $T_{j,b}^{(k)}$ denote the partial statistic computed under permutation π_b. The one-sided partial permutation p-value is

$$p_j^{(k)} = \frac{1 + \sum_{b=1}^{B} \mathbb{I}\left(T_{j,b}^{(k)} \geq T_{j,\mathrm{obs}}^{(k)}\right)}{B + 1}.$$

The associated permutation empirical CDF is

$$\widehat{F}_j^{(k)}(t) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(T_{j,b}^{(k)} \leq t\right).$$

Let $\mathbf{p}^{(k)} = (p_1^{(k)}, \ldots, p_J^{(k)})$ be the vector of partial p-values. The global NPC statistic for configuration k is computed via a combining function $\phi(\cdot)$ applied to $\mathbf{p}^{(k)}$. In this paper we use Fisher, as mentioned earlier.

$$T_{\mathrm{F}}^{(k)} = -2 \sum_{j=1}^{J} \log\Bigl(p_j^{(k)}\Bigr) .$$

To obtain the permutation distribution of the global statistic, we compute the same combination on the permutation partial p-values $\{p_{j,b}^{(k)}\}$ generated under π_b:

$$T_{\mathrm{NPC},b}^{(k)} = \phi\Bigl(p_{1,b}^{(k)}, \ldots, p_{J,b}^{(k)}\Bigr) ,$$

where

$$p_{j,b}^{(k)} = \frac{1 + \sum_{b'=1}^{B} \mathbb{I}\Bigl(T_{j,b'}^{(k)} \geq T_{j,b}^{(k)}\Bigr)}{B+1}.$$

The global NPC p-value is then

$$p_{\mathrm{NPC}}^{(k)} = \frac{1 + \sum_{b=1}^{B} \mathbb{I}\Bigl(T_{\mathrm{NPC},b}^{(k)} \geq T_{\mathrm{NPC,obs}}^{(k)}\Bigr)}{B+1}.$$

To compare configurations, we define a ranking score that increases with deviation from Benford. A natural choice is either the global statistic $T_{\mathrm{NPC,obs}}^{(k)}$ or its global p-value $p_{\mathrm{NPC}}^{(k)}$ (smaller p indicates stronger deviation). Accordingly, the NPC-based conformity ranking can be defined as

$$k_1 \prec k_2 \quad \Longleftrightarrow \quad p_{\mathrm{NPC}}^{(k_1)} > p_{\mathrm{NPC}}^{(k_2)},$$

so that configurations with larger NPC p-values are deemed closer to Benford conformity. Panel 1 presents a summary of the steps taken to implement the NPC part.

7 Results and discussion

Two regularities emerge from the 20×20 SGED experiment. Benford conformity is concentrated in a narrow neighborhood of symmetry and very heavy tails; and the second-digit rule is materially more tolerant than either the first digit or the joint first-two-digit rule, whereas the Fisher-NPC aggregation is more selective than any marginal diagnostic.

The intuition is that very heavy tails spread observations over several logarithmic

Algorithm 1 NPC test and ranking for multi-configuration Benford conformity

Require: Configurations $k = 1, \dots, K$ with samples $\{X_i^{(k)}\}_{i=1}^{n_k}$; digit rules $\mathcal{M}$; partial statistics $\{T_j\}_{j=1}^{J}$; permutations B; combining function ϕ

Ensure: Global NPC p-values $\{p_{\mathrm{NPC}}^{(k)}\}$ and ranking of configurations

```
for k = 1, ..., K do
    Extract digits D_{i,m}^{(k)} for all m ∈ M and compute observed partial statistics {T_{j,obs}^{(k)}}_{j=1}^{J}
    for b = 1, ..., B do
        Generate permutation π_b under H_0 (Benford conformity) and compute partial statistics {T_{j,b}^{(k)}}_{j=1}^{J}
    end for
    Compute partial permutation p-values p_j^{(k)} from {T_{j,b}^{(k)}}_{b=1}^{B}
    Compute global statistic T_{NPC,obs}^{(k)} = φ(p_1^{(k)}, ..., p_J^{(k)})
    for b = 1, ..., B do
        Compute p_{j,b}^{(k)} using the permutation reference set and set T_{NPC,b}^{(k)} = φ(p_{1,b}^{(k)}, ..., p_{J,b}^{(k)})
    end for
    Compute global p-value p_NPC^{(k)} = (1 + Σ_{b=1}^{B} 𝕀(T_{NPC,b}^{(k)} ≥ T_{NPC,obs}^{(k)})) / (B+1)
end for
Rank configurations by decreasing p_NPC^{(k)} (higher ⇒ closer to Benford)
```

orders of magnitude, making leading digits less tied to a single local scale. Skewness disrupts this mechanism asymmetrically by stretching one tail more than the other, so the positive and negative components can show different Benford behavior. The NPC region is narrower because it requires simultaneous adequacy of first-digit, second-digit, and first-two-digit evidence, rather than accepting conformity under whichever marginal rule is most permissive.

7.1 Metric-specific rankings

Tables 2, 3, and 4 compare the top five configurations under the chi-square and MAD criteria for each digit rule. Across all three digit rules, the ranking surfaces are dominated by symmetric or near-symmetric configurations with small values of ν. For the first digit, both criteria are led by $(\xi, \nu) = (1.00, 0.25)$ and $(1.00, 0.50)$. For the second digit, the same neighborhood remains dominant, but mild departures from symmetry remain competitive. For the joint rule, the concentration around symmetry reappears, with admissible configurations essentially restricted to the heavy-tailed left end of the grid. The ranking comparison tables therefore point to a common geometric picture: Benford conformity is best when the SGED is both highly heavy-tailed and close to symmetric, and it deteriorates sharply once either ν increases or $|\xi - 1|$ becomes substantial.

Table 2: Top 5 parameter combinations for **first**: Chi-Square (left) versus MAD ranking (right). Lower values indicate better Benford compliance. Configurations ranked independently.

Chi-Square Ranking		**MAD Ranking**	
(ξ, ν)	χ^2_{mean}	(ξ, ν)	MAD_{mean}
$(1.00, 0.25)$	8.25	$(1.00, 0.25)$	0.00239
$(1.00, 0.50)$	8.64	$(1.00, 0.50)$	0.00246
$(1.00, 0.75)$	20.82	$(1.00, 0.75)$	0.00403
$(1.25, 0.75)$	43.73	$(1.25, 1.00)$	0.00633
$(1.25, 1.00)$	50.76	$(1.25, 0.75)$	0.00652

Table 3: Top 5 parameter combinations for **second**: Chi-Square (left) versus MAD ranking (right). Lower values indicate better Benford compliance. Configurations ranked independently.

Chi-Square Ranking		**MAD Ranking**	
(ξ, ν)	χ^2_{mean}	(ξ, ν)	MAD_{mean}
$(1.00, 0.50)$	8.96	$(1.00, 0.50)$	0.00239
$(1.00, 0.25)$	9.13	$(1.00, 0.75)$	0.00241
$(1.00, 0.75)$	9.13	$(1.00, 0.25)$	0.00242
$(0.75, 1.00)$	9.35	$(0.75, 1.00)$	0.00243
$(0.75, 0.75)$	9.59	$(0.75, 0.75)$	0.00246

Table 4: Top 5 parameter combinations for **joint**: Chi-Square (left) versus MAD ranking (right). Lower values indicate better Benford compliance. Configurations ranked independently.

Chi-Square Ranking		**MAD Ranking**	
(ξ, ν)	χ^2_{mean}	(ξ, ν)	MAD_{mean}
$(1.00, 0.25)$	89.10	$(1.00, 0.25)$	0.00079
$(1.00, 0.50)$	89.63	$(1.00, 0.50)$	0.00080
$(1.00, 0.75)$	104.17	$(1.00, 0.75)$	0.00087
$(1.25, 0.75)$	136.44	$(1.25, 0.75)$	0.00105
$(1.25, 1.00)$	145.54	$(1.25, 1.00)$	0.00105

7.2 Global NPC ranking

The global Fisher-NPC ranking is even more concentrated, see Table 5. The top two configurations, $(\xi, \nu) = (1.00, 0.25)$ and $(1.00, 0.50)$, achieve median global p-values 0.4840 and 0.4601, with empirical rejection rates of only 6.1% and 7.2%, respectively. The third-best configuration, $(1.00, 0.75)$, already falls to a median NPC p-value of 0.0339 and a rejection rate of 56.9%. Beyond this point the ranking collapses to the permutation floor. In other words, once the three digit rules are assessed jointly, the empirical Benford region becomes extremely localized around symmetry and the most heavy-tailed part of the explored SGED family.

Table 5: Top 10 parameter combinations by NPC global p-value (median across repetitions). Higher p-value = better Benford compliance. NPC combines first/second/joint digit tests via Fisher's method with dependence-preserving permutations. Parameters: ξ (skewness), ν (shape).

Rank	ξ	ν	pNPC	Reject Rate
1	1.00	0.25	0.4840	0.061
2	1.00	0.50	0.4601	0.072
3	1.00	0.75	0.0339	0.569
4	0.25	0.25	0.0020	1.000
5	0.50	0.25	0.0020	1.000
6	0.75	0.25	0.0020	1.000
7	1.25	0.25	0.0020	1.000
8	1.50	0.25	0.0020	1.000
9	1.75	0.25	0.0020	1.000
10	2.00	0.25	0.0020	1.000

This table is especially informative when compared with the metric-specific rankings above. Several configurations that remain competitive under isolated second-digit or MAD criteria fail the global test once the first-digit and joint first-two-digit evidence is incorporated. The practical implication is that marginal Benford conformity is materially easier to obtain than joint conformity across digit rules.

The digit-specific summary tables and heatmaps show where this tightening occurs. The summary tables 6, 7 and 8 rank configurations by mean chi-square; the heatmaps 3, 4 and 5 complement them by reporting the full surface of MAD values, cell-wise chi-square annotations, and NPC significance markers.

For the first digit, the best configuration $(\xi, \nu) = (1.00, 0.25)$ has mean $\chi^2 = 8.25$, median permutation p-value 0.4910, and MAD = 0.00239. The neighbouring symmetric heavy-tailed cell $(1.00, 0.50)$ remains close, with a mean $\chi^2 = 8.64$ and median permutation p-value 0.4401. By contrast, already at $(1.00, 0.75)$ the median permutation p-value drops to 0.0120, after which the surface deteriorates rapidly. The first-digit heatmap therefore exhibits a small low-deviation basin centred at symmetry and very small ν, surrounded by a broad region of systematic rejection.

The second digit is markedly more permissive. The top three configurations are

Table 6: Top 10 parameter combinations for first digit (ranked by Chi-Square). Parameters: ξ (skewness), ν (shape).

ξ	ν	χ^2	pPerm	MAD
1.00	0.25	8.25	0.4910	0.00239
1.00	0.50	8.64	0.4401	0.00246
1.00	0.75	20.82	0.0120	0.00403
1.25	0.75	43.73	0.0020	0.00652
1.25	1.00	50.76	0.0020	0.00633
1.00	1.00	64.10	0.0020	0.00775
0.75	1.00	108.69	0.0020	0.01003
1.00	1.25	143.46	0.0020	0.01207
1.25	1.25	143.53	0.0020	0.01117
0.75	0.75	159.93	0.0020	0.01215

$(1.00, 0.50)$, $(1.00, 0.25)$, and $(1.00, 0.75)$, with mean chi-square values 8.96, 9.13, and 9.13, and median permutation p-values between 0.4940 and 0.5080. Moreover, configurations such as $(0.75, 1.00)$ and $(1.25, 1.00)$ still remain statistically competitive. The second-digit heatmap thus displays a wider conformity plateau around symmetry, indicating that this digit rule is less sensitive to moderate perturbations in skewness and tail thickness than the first-digit or joint rules.

Table 7: Top 10 parameter combinations for second digit (ranked by Chi-Square). Parameters: ξ (skewness), ν (shape).

ξ	ν	χ^2	pPerm	MAD
1.00	0.50	8.96	0.5080	0.00239
1.00	0.25	9.13	0.4940	0.00242
1.00	0.75	9.13	0.5030	0.00241
0.75	1.00	9.35	0.4661	0.00243
0.75	0.75	9.59	0.4501	0.00246
1.25	1.00	9.73	0.4471	0.00249
1.25	0.75	10.21	0.3932	0.00254
1.50	1.00	10.34	0.3613	0.00257
1.00	1.00	10.68	0.3623	0.00261
1.75	1.00	10.98	0.3313	0.00265

The joint rule is again restrictive, but not uniformly so. The best configuration $(1.00, 0.25)$ attains mean $\chi^2 = 89.10$, median permutation p-value 0.5050, and MAD $=$ 0.00079; $(1.00, 0.50)$ and $(1.00, 0.75)$ remain acceptable or borderline with median permutation p-values 0.4631 and 0.1507. Once the grid moves toward lighter tails or appreciable asymmetry, however, the joint-digit p-values collapse to the permutation floor. Relative to the first digit, the joint rule preserves a small buffer at $\nu = 0.75$, but it still rules out most of the parameter space once global evidence is aggregated.

Table 8: Top 10 parameter combinations for joint digit (ranked by Chi-Square). Parameters: ξ (skewness), ν (shape).

ξ	ν	χ^2	pPerm	MAD
1.00	0.25	89.10	0.5050	0.00079
1.00	0.50	89.63	0.4631	0.00080
1.00	0.75	104.17	0.1507	0.00087
1.25	0.75	136.44	0.0020	0.00105
1.25	1.00	145.54	0.0020	0.00105
1.00	1.00	155.60	0.0020	0.00110
0.75	1.00	212.17	0.0020	0.00131
1.00	1.25	245.85	0.0020	0.00144
1.25	1.25	254.67	0.0020	0.00146
0.75	0.75	257.35	0.0020	0.00149

Taken together, the digit-specific surfaces yield a clear hierarchy. The second digit offers the broadest near-Benford region, the first digit is materially stricter, and the joint rule becomes highly selective once the model leaves the symmetric heavy-tailed corner. The global NPC ranking should therefore be interpreted as the intersection of these three marginal geometries rather than as a stand-alone black-box score.

7.3 Sign-split evidence under skewness

The sign-split results reveal how skewness acts through tail asymmetry rather than through a uniform shift in conformity; this can be appreciated from Table 9. In the symmetric cases, positive and negative subsamples behave similarly, of course. For example, the top two global NPC configurations, $(1.00, 0.25)$ and $(1.00, 0.50)$, exhibit close agreement across combined, positive-only, and negative-only analyses. Once ξ moves away from 1, however, one tail can remain close to Benford while the other collapses. This aligns with the literature connecting Benford's law to rank-size laws.

This effect is stark in the extreme left-skew case $(\xi, \nu) = (0.25, 0.25)$: the combined NPC median is at the permutation floor, the positive-only NPC median is also at the floor, but the negative-only NPC median rises to 0.5100 with only 5.7% rejections.

The same pattern is visible in the aggregated chi-square/MAD comparison, Table 10. The best combined configurations remain centered at $(1.00, 0.25)$ and $(1.00, 0.50)$, but asymmetric cells can display sharp positive/negative divergence. For instance, $(1.25, 0.75)$ has a combined average chi-square of 63.46, yet its positive and negative averages split into

Table 9: Top 10 SGED parameter combinations ranked by global NPC p-value (combined context). NPC combines first-digit, second-digit, and joint first-two-digit Benford tests via Fisher's method with dependence-preserving permutations. Higher p-values indicate better compliance with Benford's Law. Columns show NPC p-values (median across repetitions) and rejection rates (proportion of repetitions with $p < 0.05$, shown as percentages) for combined (all values), positive-only, and negative-only subsamples. Parameters: ξ (skewness), ν (shape).

Rank	ξ	ν	pNPC (Comb)	pNPC (Pos)	pNPC (Neg)	Rej% (Comb)	Rej% (Pos)	Rej% (Neg)
1	1.00	0.25	0.4840	0.4850	0.4691	6.1	4.2	6.9
2	1.00	0.50	0.4601	0.5030	0.4870	7.2	5.8	6.0
3	1.00	0.75	0.0339	0.1527	0.1407	56.9	28.5	28.7
4	0.25	0.25	0.0020	0.0020	0.5100	100.0	100.0	5.7
5	0.25	0.50	0.0020	0.0020	0.4182	100.0	100.0	9.5
6	0.25	0.75	0.0020	0.0020	0.0978	100.0	100.0	35.0
7	0.25	1.00	0.0020	0.0020	0.0080	100.0	100.0	86.3
8	0.25	1.25	0.0020	0.0020	0.0020	100.0	100.0	99.5
9	0.25	1.50	0.0020	0.0020	0.0020	100.0	100.0	100.0
10	0.25	1.75	0.0020	0.0020	0.0020	100.0	100.0	100.0

40.75 and 107.72, respectively. Sign-specific ranking tables produced by the simulation reinforce this mirror geometry: right-skewed configurations dominate the positive-only top lists, whereas left-skewed configurations dominate the negative-only ones.

Table 10: Top 10 SGED parameter combinations ranked by chi-square statistic (combined context, averaged across first-digit, second-digit, and joint digit tests). Lower χ^2 and MAD values indicate better compliance with Benford's Law. Columns show mean χ^2 statistics and Mean Absolute Deviation (MAD) values across repetitions for combined (all values), positive-only, and negative-only subsamples. Parameters: ξ (skewness), ν (shape).

Rank	ξ	ν	χ^2 (Comb)	χ^2 (Pos)	χ^2 (Neg)	MAD (Comb)	MAD (Pos)	MAD (Neg)
1	1.00	0.25	35.49	35.48	35.64	0.00187	0.00261	0.00263
2	1.00	0.50	35.75	35.51	35.27	0.00188	0.00263	0.00260
3	1.00	0.75	44.71	40.35	40.10	0.00244	0.00307	0.00306
4	1.25	0.75	63.46	40.75	107.72	0.00337	0.00361	0.00615
5	1.25	1.00	68.68	53.27	87.24	0.00329	0.00460	0.00550
6	1.00	1.00	76.79	56.23	56.24	0.00382	0.00429	0.00430
7	0.75	1.00	110.07	172.95	52.23	0.00459	0.00804	0.00462
8	1.00	1.25	133.98	84.88	85.14	0.00545	0.00585	0.00587
9	1.25	1.25	136.94	73.14	134.28	0.00516	0.00577	0.00649
10	0.75	0.75	142.29	251.51	40.61	0.00537	0.00909	0.00373

These direction-specific differences are also visible in the sign-specific heatmaps. Figures 6, 7, and 8 report the positive-only conformity surfaces. In these figures, the horizontal axis reports ν and the vertical axis reports ξ, whereas parameter pairs in the text are written as (ξ, ν). They confirm that right-skewed configurations can retain comparatively good Benford compliance on the positive tail even when the pooled analysis deteriorates.

Figures 9, 10, and 11 provide the complementary negative-only surfaces. Read together with the positive-only figures, they make the mirror-image effect of skewness transparent: left-skewed configurations improve conformity on the negative tail, whereas right-skewed configurations improve conformity on the positive tail.

From the sign-split analysis, one understands that skewness should be interpreted as a direction-specific perturbation of Benford behaviour. For asymmetric SGED models, reporting only pooled diagnostics can obscure where the departure actually arises.

The 20×20 SGED experiment yields four main lessons for applied Benford analysis.

The global conformity region is not diffuse: only the symmetric and very heavy-tailed configurations $(1.00, 0.25)$ and $(1.00, 0.50)$ remain clearly non-rejected under the Fisher-NPC criterion. Moving from $\nu = 0.50$ to $\nu = 0.75$ at symmetry already reduces the median global p-value from 0.4601 to 0.0339, indicating that Benford conformity under SGED is a sharply localised property rather than a generic consequence of non-Gaussianity.

The choice of digit rule materially affects the inferential picture. The second digit admits a much broader near-Benford region than either the first digit or the joint first-two-digit rule, so a researcher or a practitioner who relies only on second-digit evidence can easily classify a configuration as approximately Benford even when the first digit or the joint rule rejects decisively. Multi-rule reporting is therefore a minimum standard.

Skewness, moreover, operates through tail-specific distortions rather than a uniform shift in conformity. The positive-only and negative-only tables show a mirror pattern: under SGED's constraints, right-skew improves the positive tail at the expense of the negative tail, and vice versa. For auditing or forensic applications, sign-stratified Benford diagnostics can be informative whenever the underlying mechanism is asymmetric.

Finally, the Fisher-NPC test is stricter than marginal rankings because it requires simultaneous adequacy across three dependent digit rules. By preserving the joint dependence structure through synchronised multinomial resampling, the global test avoids the false comfort that can arise when one rule fits well and another fails. In practice, marginal diagnostics should be treated as descriptive whereas the global NPC statistic supplies the formal inferential decision.

8 Final Remarks

The large-scale SGED experiment, covering a 20×20 grid in (ξ, ν) with 1000 replications per cell, yields a clear empirical message. Benford conformity is strongest in a narrow symmetric and very heavy-tailed region, weakens rapidly as tails lighten and becomes tail-specific as skewness increases. The second digit remains the most permissive marginal diagnostic, whereas the first digit and the joint first-two-digit rule are materially more selective. These results are obtained under a controlled Monte Carlo design with fixed location and scale parameters and should not be interpreted as implying universal Benford conformity for SGED processes in empirical applications.

From a methodological standpoint, the study demonstrates the value of coupling descriptive and inferential diagnostics. Mean chi-square and MAD surfaces identify where deviation grows, but the Fisher combination of synchronised digit-rule permutation p-values provides the sharpest global statement. Only two configurations remain clearly non-rejected under this global criterion, illustrating how demanding joint Benford conformity becomes once dependence across digit rules is respected.

The analysis emphasises that Benford conformity is inherently a multivariate property and that a joint evaluation of first and second digits provides a more reliable assessment than separate univariate diagnostics.

For practitioners, key lessons are that heavy tails by themselves do not guarantee Benford conformity, and that near-symmetry is crucial and must be taken into account. Pooled diagnostics can mask strong sign asymmetries, requiring a sharpened scrupulousness in compliance evaluation. So any rigorous assessment should distinguish between marginal conformity on individual digit rules and global conformity across all rules jointly.

Statements and Declarations

Funding

No funding was received for conducting this study.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Author contributions

All authors contributed to the conception and design of the study, methodological development, and interpretation of results. The first draft of the manuscript was prepared by the authors, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability

This work relies on simulation experiments. All numerical steps needed to replicate the findings can be carried out using the algorithms and parameter grids described in the manuscript. The code and random seeds are publicly available at `https://github.com/mevalerio/Piscopo_et_al_2026_NPC_Benford.git`.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

References

Abrantes-Metz, R. M., Villas-Boas, S. B., and Judge, G. (2011). Tracking the libor rate. *Applied Economics Letters*, 18(10):893–899.

Ausloos, M., Ficcadenti, V., Dhesi, G., and Shakeel, M. (2021). Benford's laws tests on S&P500 daily closing values and the corresponding daily log-returns both point to huge non-conformity. *Physica A: Statistical Mechanics and its Applications*, 574:125969.

Barabesi, L., Cerasa, A., Cerioli, A., and Perrotta, D. (2022). On characterizations and tests of Benford's law. *Journal of the American Statistical Association*, 117(540):1887–1903.

Barabesi, L., Cerioli, A., and Di Marzio, M. (2023). Statistical models and the Benford hypothesis: A unified framework. *TEST*, 32(4):1479–1507.

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78(4):551–572.

Berger, A. and Hill, T. P. (2011). Benford's law strikes back: no simple explanation in sight for mathematical gem. *The Mathematical Intelligencer*, 33(1):85–91.

Berger, A. and Hill, T. P. (2015). *An Introduction to Benford's Law*. Princeton University Press.

Bonnini, S., Assegie, G. M., and Trzcinska, K. (2024a). Review about the permutation approach in hypothesis testing. *Mathematics*, 12:2617.

Bonnini, S., Borghesi, M., and Giacalone, M. (2024b). Advances on multisample permutation tests for "V-shaped" and "U-shaped" alternatives with application to circular economy. *Annals of Operations Research*, 342(3):1655–1670.

Bonnini, S., Borghesi, M., Piscopo, G., and Giacalone, M. (2025). A non-parametric test for a two-way analysis of variance. *Mathematics*, 13(7):1131.

Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika*, 40:318–335.

Burgos, A. and Santos, A. (2021). The Newcomb–Benford law: Scale invariance and a simple Markov process based on it. *American Journal of Physics*, 89(9):851–861.

Cerqueti, R., Giacalone, M., and Mattera, R. (2020). Skewed non-Gaussian GARCH models for cryptocurrencies volatility modelling. *Information Sciences*, 527:1–26.

Cerqueti, R. and Maggi, M. (2025). Classes of probability measures built on the properties of Benford's law. *AStA Advances in Statistical Analysis*, 109(1):197–216.

Diekmann, A. (2007). Not the first digit! Using benford's law to detect fraudulent scientific data. *Journal of Applied Statistics*, 34(3):321–329.

Durtschi, C., Hillison, W., and Pacini, C. (2004). The effective use of Benford's law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting*, 5(1):17–34.

Fernandez, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371.

Formann, A. K. (2010). The Newcomb–Benford law in its relation to some common distributions. *PLOS ONE*, 5(5):e10541.

Giacalone, M. (2022). Optimal forecasting accuracy using l_p-norm combination. *Metron*, 80(2):187–230.

Goodman, L. A. (1954). Kolmogorov–smirnov tests for psychological research. *Psychological Bulletin*, 51(2):160–168.

Hill, T. P. (1995). A statistical derivation of the significant-digit law. *Statistical Science*, 10(4):354–363.

Holt, C. A. and Sullivan, S. P. (2023). Permutation tests for experimental data. *Experimental Economics*, 26(4):775–812.

Kim, I., Balakrishnan, S., and Wasserman, L. (2022). Minimax optimality of permutation tests. *The Annals of Statistics*, 50(1):225–251.

Lee, M.-C., Su, J.-B., and Liu, H.-C. (2008). Value-at-risk in us stock indices with skewed generalized error distribution. *Applied Financial Economics Letters*, 4(6):425–431.

Leemis, L. M., Schmeiser, B. W., and Evans, D. L. (2000). Survival distributions satisfying Benford's law. *The American Statistician*, 54(4):236–241.

McDonald, J. B. and Newey, W. K. (1988). Partially adaptive estimation of regression models via the generalized t distribution. *Econometric theory*, 4(3):428–457.

Mineo, A. (1989). The norm-p estimation of location, scale and simple linear regression parameters. In *Statistical Modelling: Proceedings of GLIM 89 and the 4th International Workshop on Statistical Modelling*, New York, NY. Springer New York.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the econometric society*, pages 347–370.

Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*, 4(1):39–40.

Nigrini, M. J. (1996). A taxpayer compliance application of Benford's law. *Journal of the American Taxation Association*, 18(1):72–91.

Nigrini, M. J. (1999). I've got your number: how a mathematical phenomenon can help uncover fraud. *Journal of Accountancy*, 187(5):79–83.

Nigrini, M. J. (2012). *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. John Wiley & Sons.

Pericchi, L. and Torres, D. (2011). Quick anomaly detection by the Newcomb—Benford Law, with applications to electoral processes data from the USA, Puerto Rico and Venezuela. *Statistical Science*, pages 502–516.

Pesarin, F. (2001). *Multivariate Permutation Tests: With Applications in Biostatistics*. Wiley & Sons.

Pesarin, F. and Salmaso, L. (2010). *Permutation tests for complex data: theory, applications and software*. John Wiley & Sons.

Pesarin, F. and Salmaso, L. (2012). A review and some new results on permutation testing for multivariate problems. *Statistics and Computing*, 22(2):639–646.

Pinkham, R. S. (1961). On the distribution of first significant digits. *Annals of Mathematical Statistics*, 32(4):1223–1230.

Rauch, B., Göttsche, M., Brähler, G., and Engel, S. (2011). Fact and fiction in EU-governmental economic data. *German Economic Review*, 12(3):243–255.

Riccioni, J. and Cerqueti, R. (2018). Regular paths in financial markets: Investigating the Benford's law. *Chaos, Solitons & Fractals*, 107:186–194.

Robledo, A. (2011). Laws of Zipf and Benford, intermittency, and critical fluctuations. *Chinese Science Bulletin*, 56(34):3643–3648.

Subbotin, M. T. (1923). On the law of frequency of errors. *Matematicheskii Sbornik*, 31(2):296–301.

Taguchi, T. (1974). On Fechner's thesis and statistics with norm p. *Annals of the Institute of Statistical Mathematics*, 26(1):175–193.

Theodossiou, P. (1998). Financial data and the skewed generalized t distribution. *Management Science*, 44:1650–1661.

Theodossiou, P. (2015). Skewed generalized error distribution of financial assets and option pricing. *Multinational Finance Journal*, 19(4):223–266.

West, M. (1987). On scale mixtures of normal distributions. *Biometrika*, 74(3):646–648.

Würtz, D., Chalabi, Y., and Luksan, L. (2006). Parameter estimation of ARMA models with GARCH/APARCH errors. *Journal of Statistical Software*, 17(1):1–56.

Zeckhauser, R. and Thompson, M. (1970). Linear regression with non-normal error terms. *The Review of Economics and Statistics*, pages 280–286.

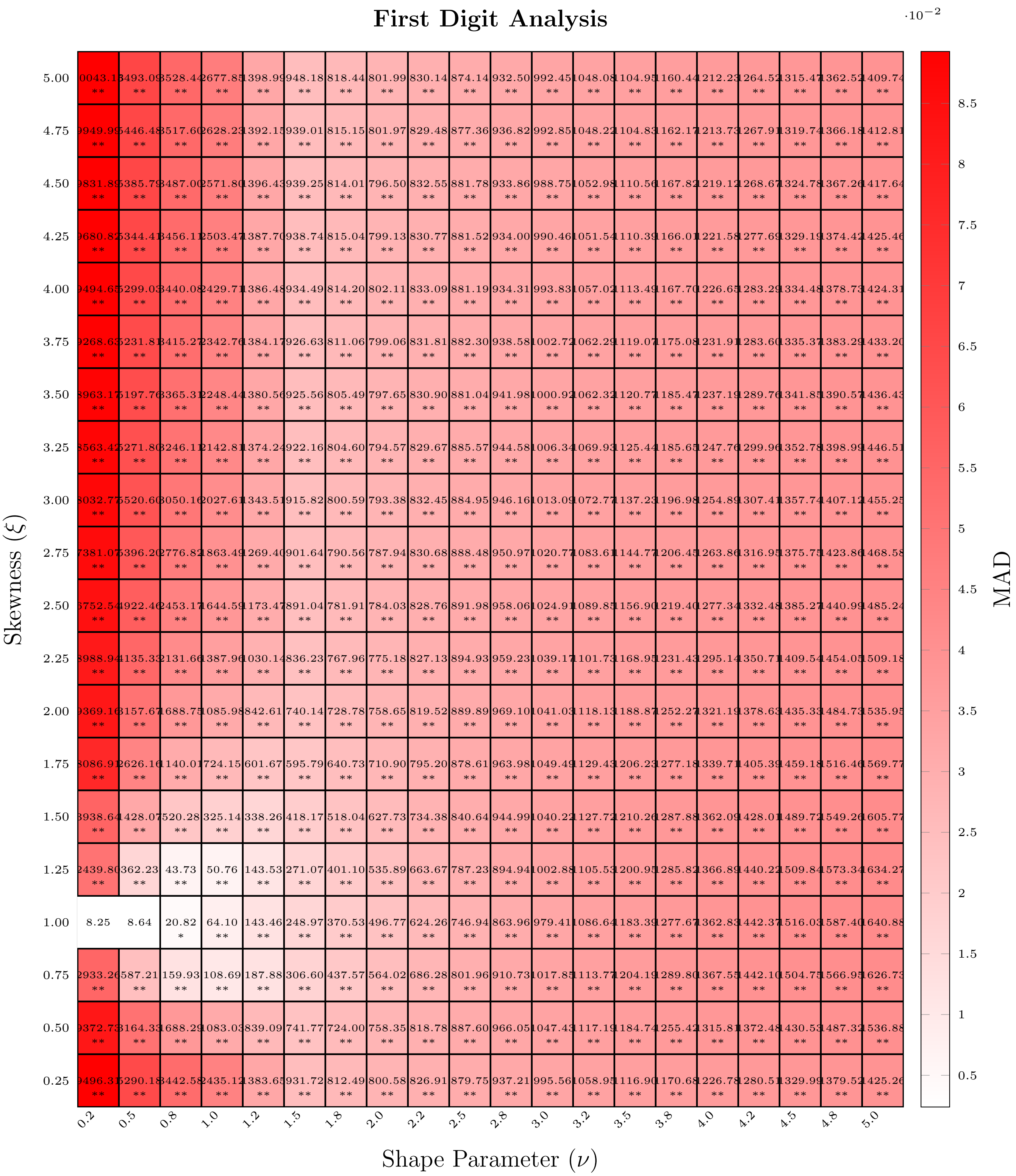

Figure 3: **First-Digit Combined Heatmap.** Color scale represents Mean Absolute Deviation (MAD) across the (ξ, ν) grid. Each cell shows the χ^2 statistic with NPC significance asterisks below: (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$.

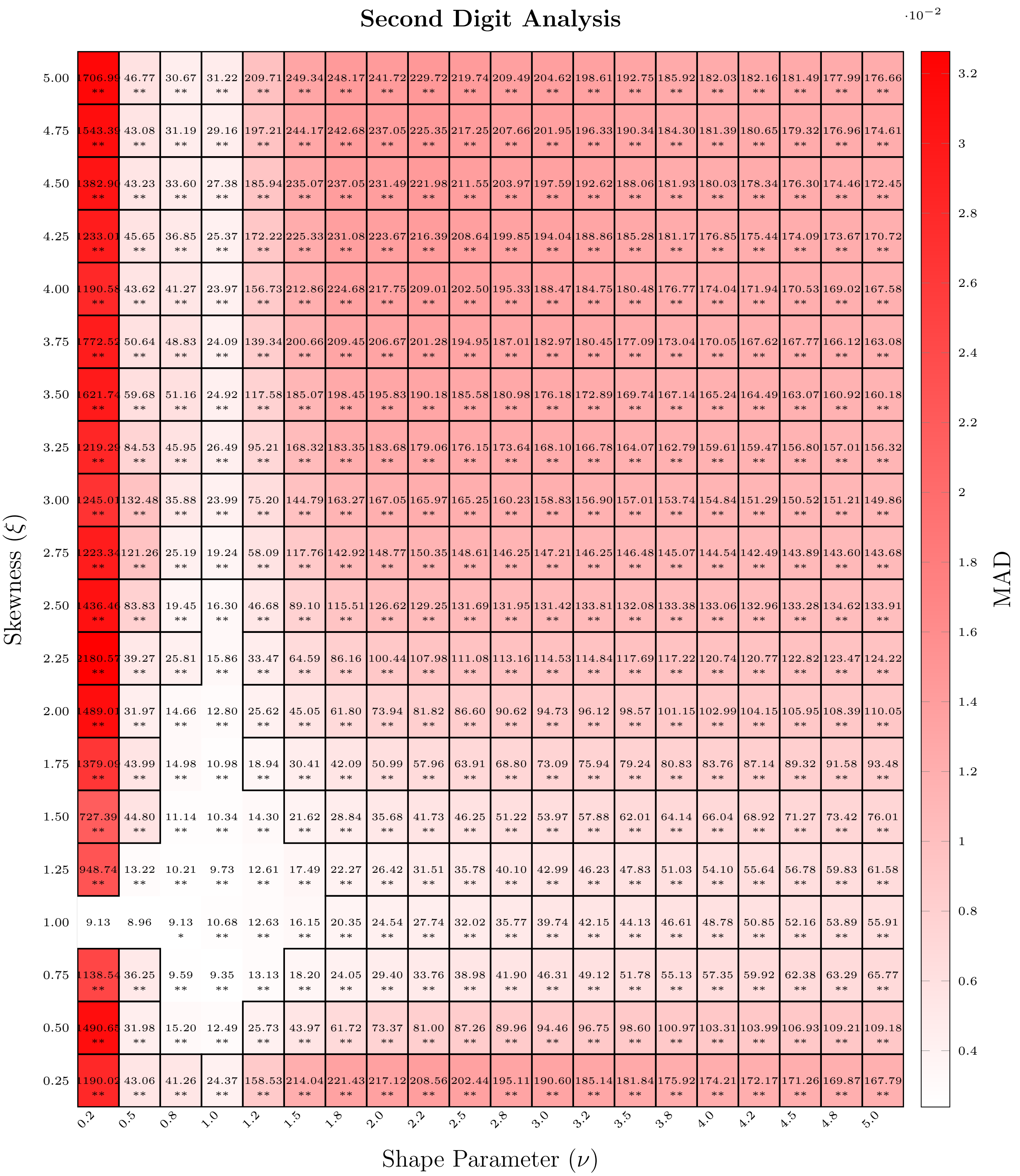

Figure 4: **Second-Digit Combined Heatmap.** Color scale represents Mean Absolute Deviation (MAD) across the (ξ, ν) grid. Each cell shows the χ^2 statistic with NPC significance asterisks below: (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$.

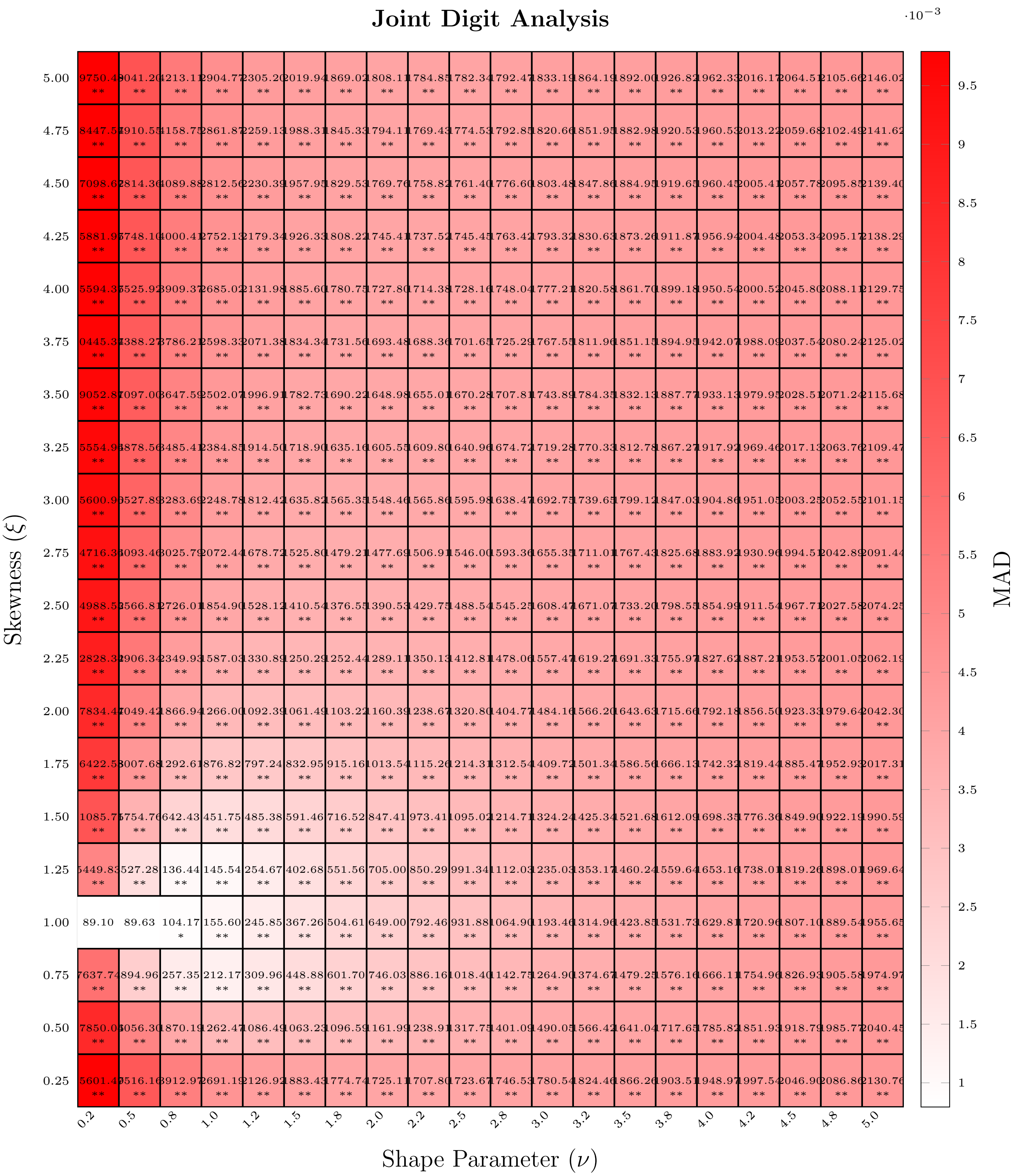

Figure 5: **Joint First-Two-Digit Combined Heatmap.** Color scale represents Mean Absolute Deviation (MAD) across the (ξ, ν) grid. Each cell shows the χ^2 statistic with NPC significance asterisks below: (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$.

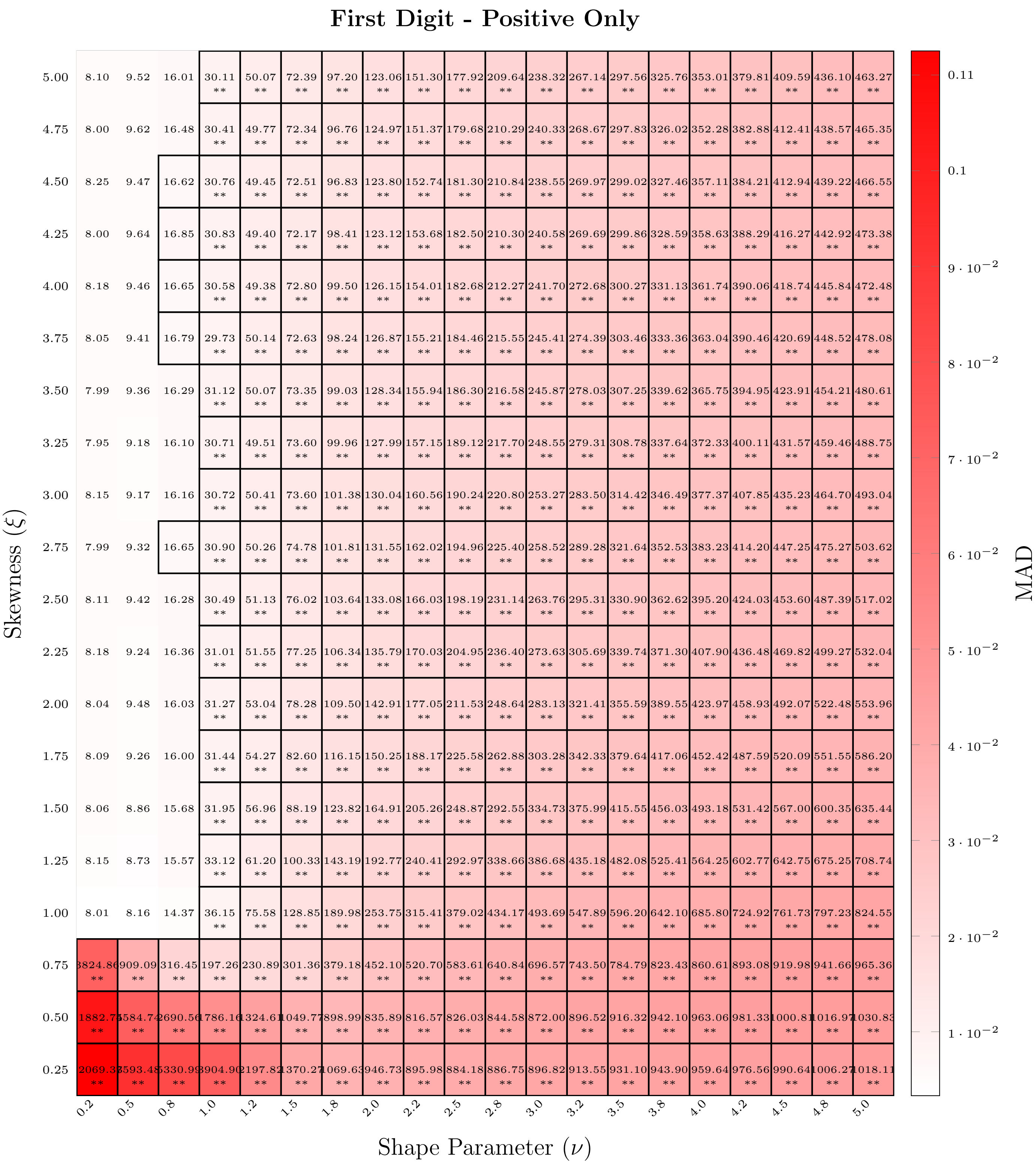

Figure 6: **First Digit Combined Heatmap - Positive Only.** Color scale represents Mean Absolute Deviation (MAD) across the (ξ, ν) grid. Each cell shows the χ^2 statistic with NPC significance asterisks below: (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$.

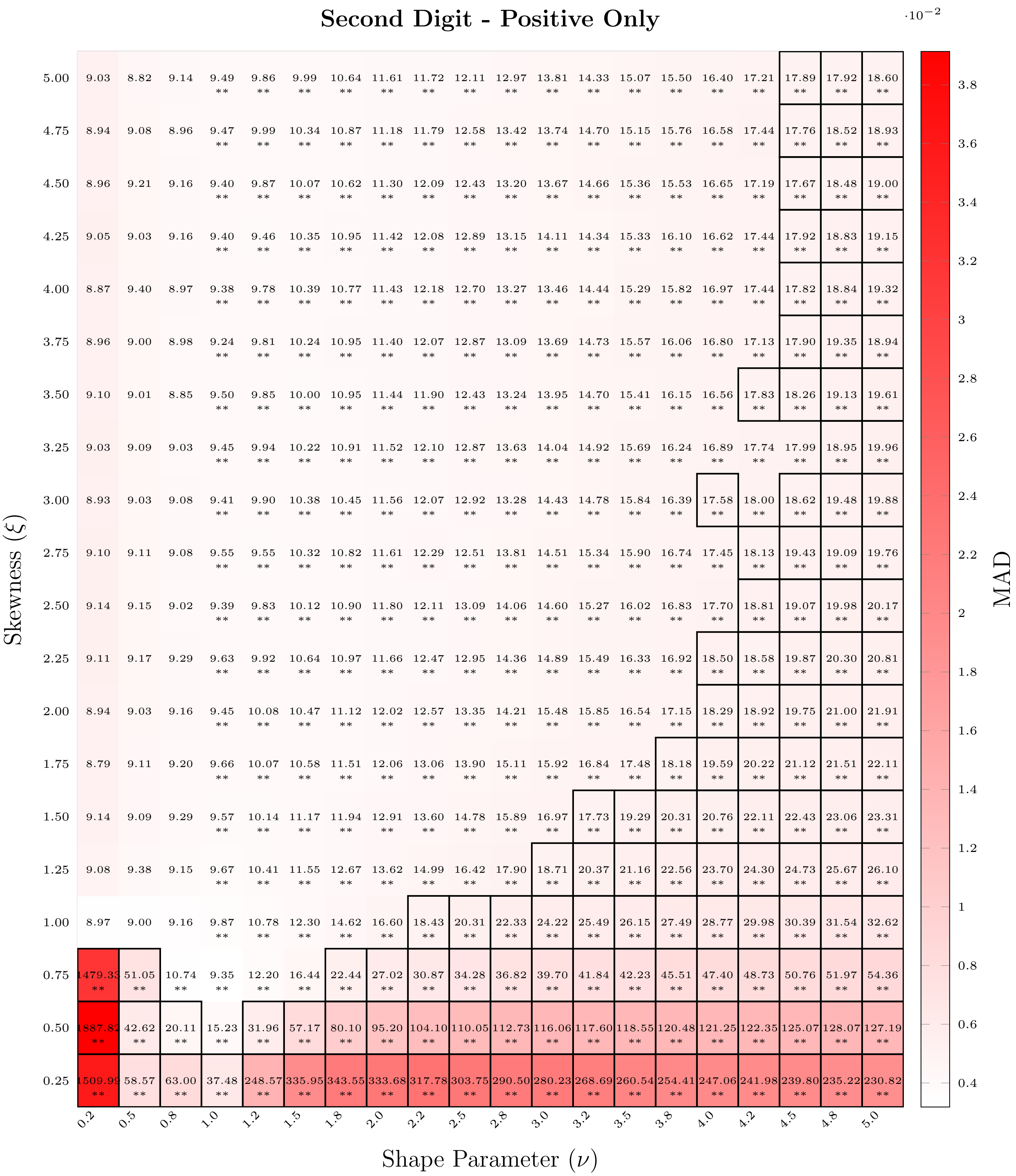

Figure 7: **Second Digit Combined Heatmap - Positive Only.** Color scale represents Mean Absolute Deviation (MAD) across the (ξ, ν) grid. Each cell shows the χ^2 statistic with NPC significance asterisks below: (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$.

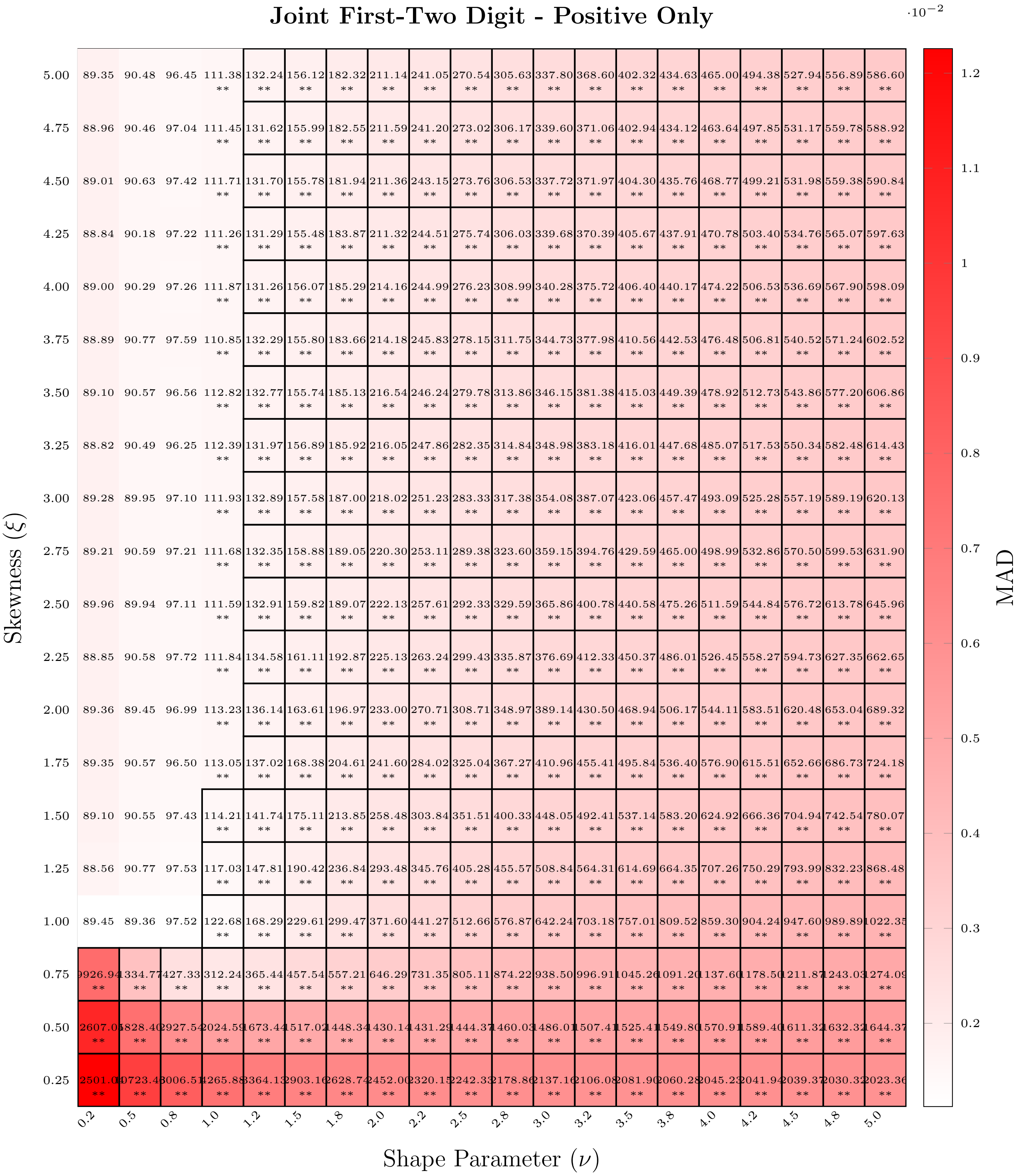

Figure 8: **Joint First-Two Digit Combined Heatmap - Positive Only.** Color scale represents Mean Absolute Deviation (MAD) across the (ξ, ν) grid. Each cell shows the χ^2 statistic with NPC significance asterisks below: (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$.

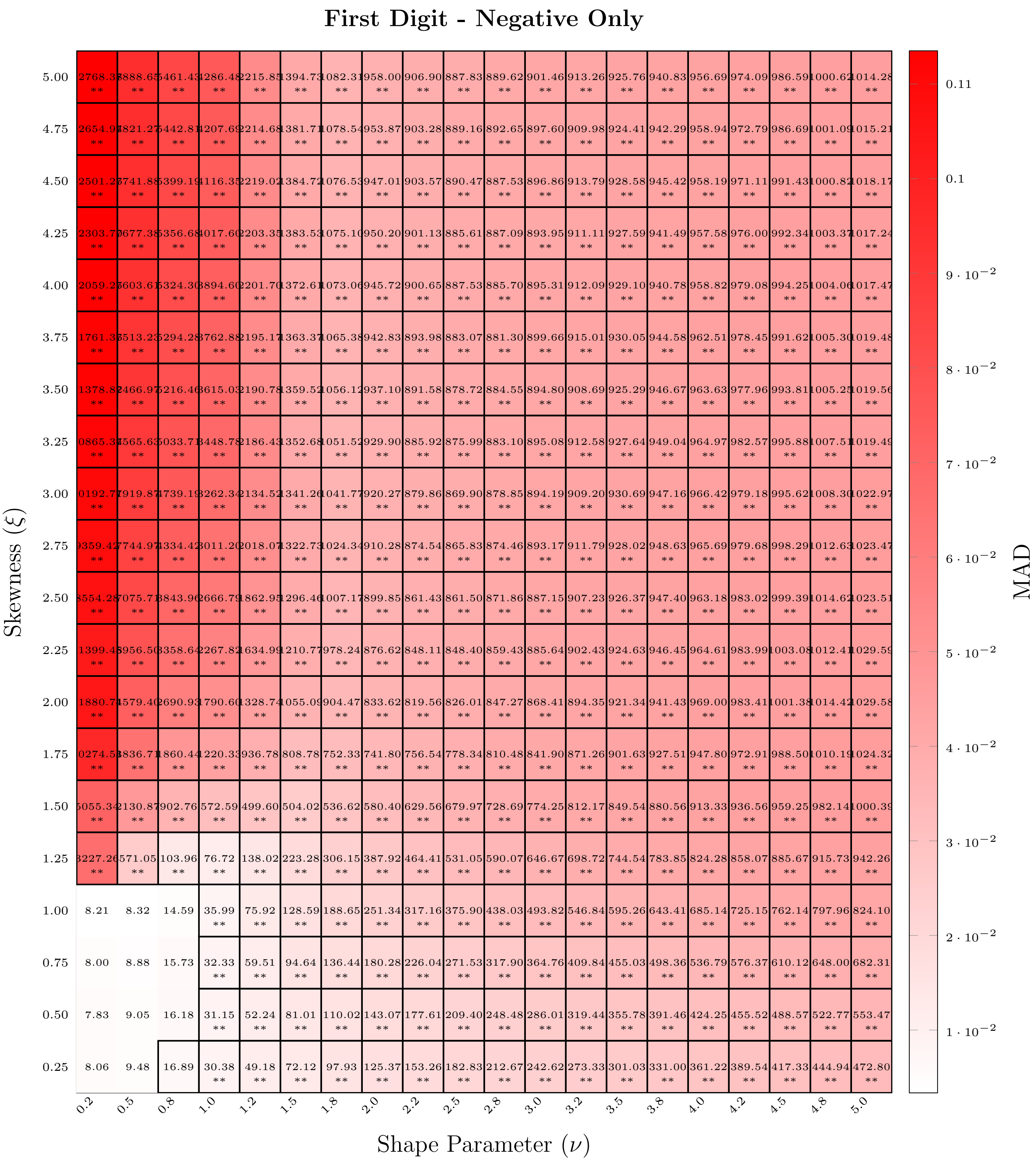

Figure 9: **First Digit Combined Heatmap - Negative Only.** Color scale represents Mean Absolute Deviation (MAD) across the (ξ, ν) grid. Each cell shows the χ^2 statistic with NPC significance asterisks below: (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$.

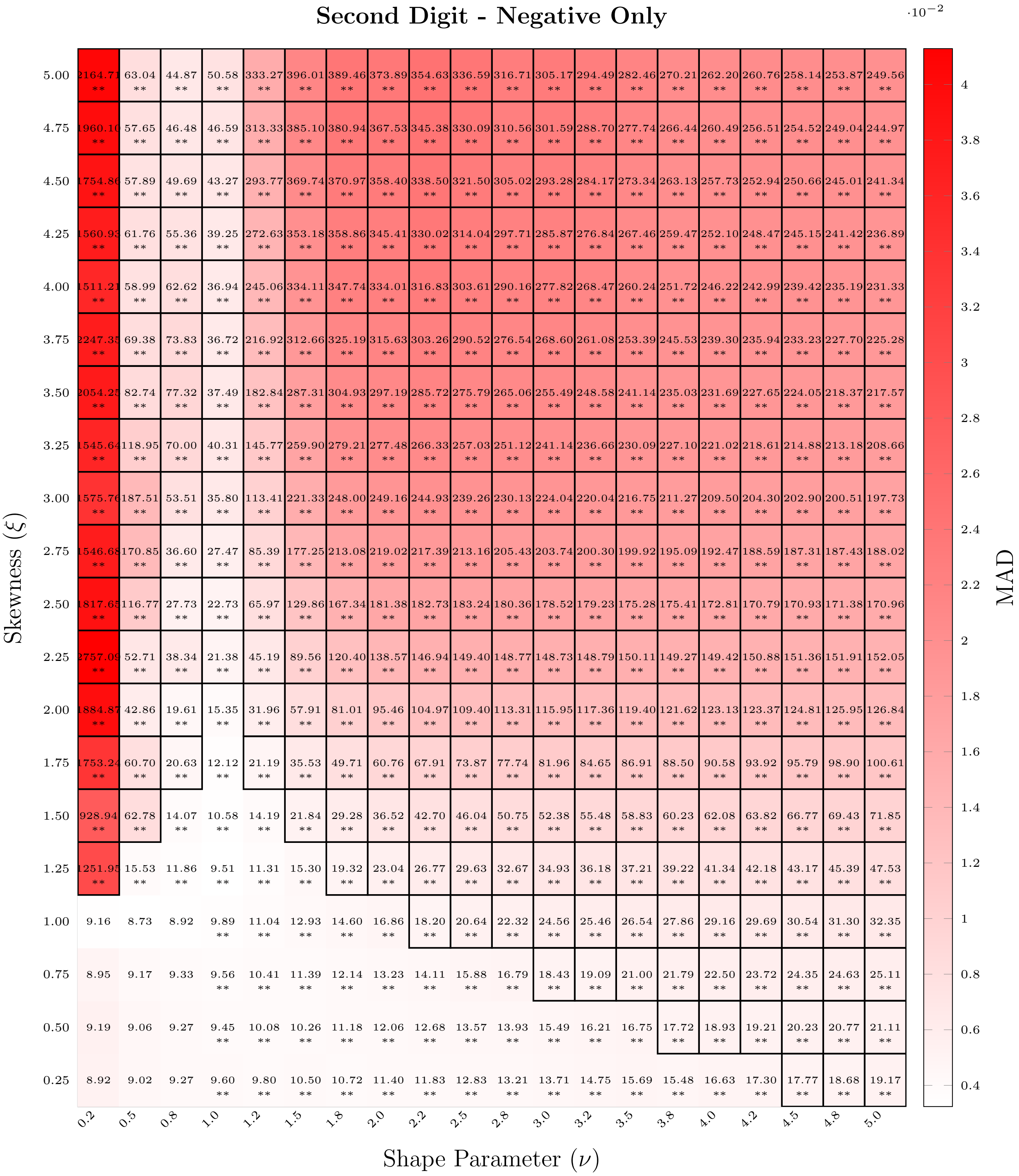

Figure 10: **Second Digit Combined Heatmap - Negative Only.** Color scale represents Mean Absolute Deviation (MAD) across the (ξ, ν) grid. Each cell shows the χ^2 statistic with NPC significance asterisks below: (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$.

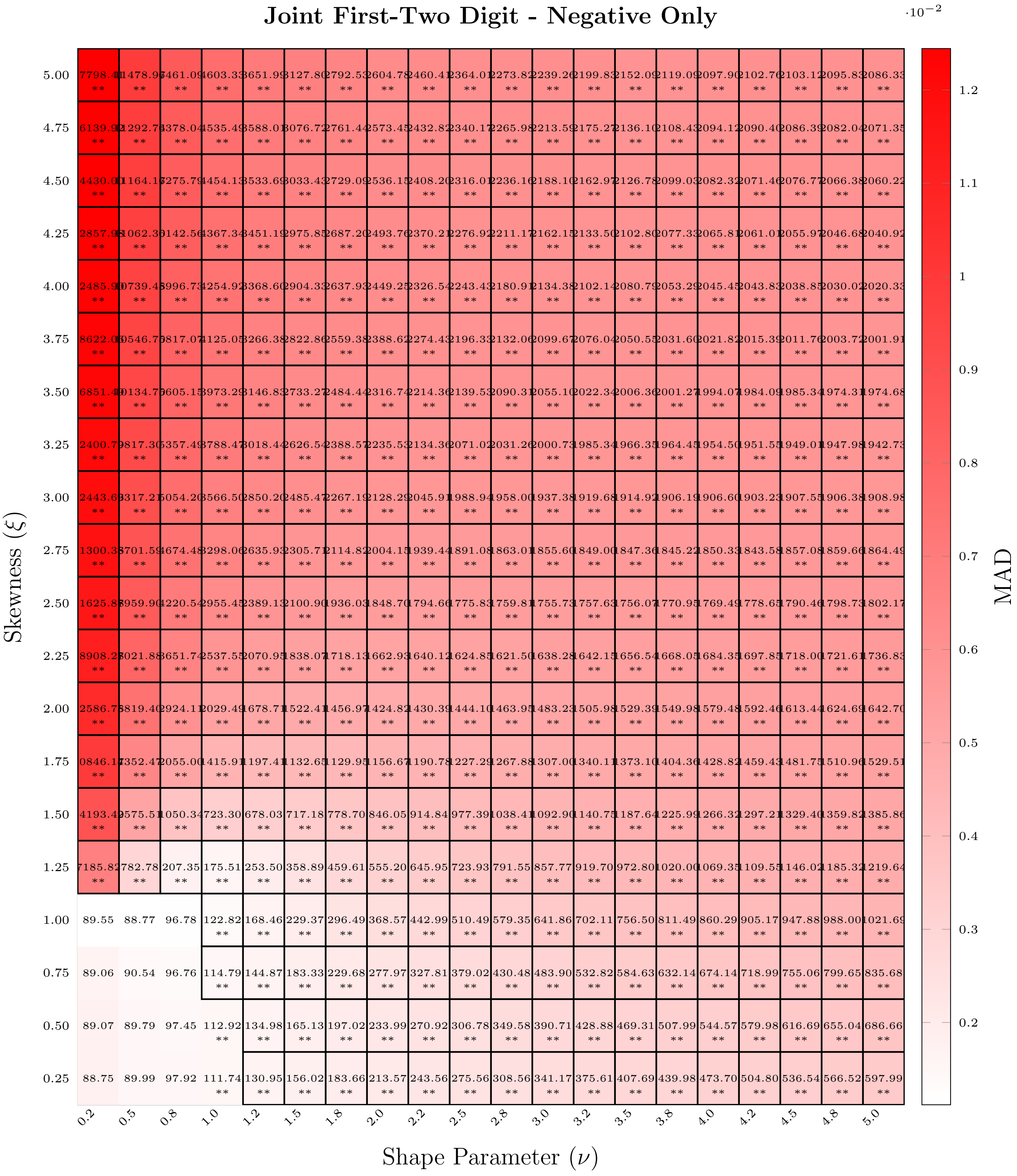

Figure 11: **Joint First-Two Digit Combined Heatmap - Negative Only.** Color scale represents Mean Absolute Deviation (MAD) across the (ξ, ν) grid. Each cell shows the χ^2 statistic with NPC significance asterisks below: (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$.