

The Paranormal Distribution:

A Bimodal, Asymmetric, Heavy-Tailed Probability Distribution
for Anomalous Data in Macroeconomics and Finance*

Dimitrios Thomakos

Cybernetics and Artificial Intelligence Laboratory

Department of Business Administration

National and University of Athens

February 2026

Abstract

We introduce the *Paranormal distribution*, a nine-parameter probability model designed to capture four simultaneous stylised facts present in macroeconomic and financial data: bimodality (multiple regimes), asymmetry (differential upside/downside dynamics), heavy tails (extreme events), and unequal regime probabilities. The distribution is constructed as a finite-weight mixture of two Azzalini–Capitanio skew- t components, each controlling its own location, scale, skewness, and tail-thickness parameter. We develop the complete probability-theoretic foundation—normality, moment existence as a function of the tail indices, polynomial tail decay—give rigorous identifiability conditions, and prove strong consistency and asymptotic normality of the constrained maximum likelihood estimator under explicit regularity assumptions. A population-based global optimisation algorithm (differential evolution) is used for estimation, with k -means initialisation to handle the multimodal likelihood surface. We then fit the distribution to five empirical series: US inflation, US real GDP growth, the SPDR S&P 500 ETF (SPY), the Direxion Daily Small-Cap Bull 3× ETF (TNA), and Bitcoin (BTC-USD). In every case the Paranormal distribution achieves substantially lower Akaike and Bayesian information criteria than competing unimodal and symmetric benchmarks. The paper is accompanied by a fully documented open-source Python package.

Keywords: bimodal distribution; skew- t mixture; heavy tails; regime switching; maximum likelihood; financial returns; macroeconomic dynamics.

*Idea by the author, code and write up in conjunction with Claude Sonnet 4.6 in Perplexity. The paper is provided “as is” and may contain errors, including in mathematical claims and bibliographical attributions. Python code and prompts are available on request from the author. Preliminary material, please do not quote without permission – comments welcomed. Email: dthomakos@ba.uoa.gr

1 Introduction

Classical probability distributions assume—at most—a single dominant mode, symmetric tail behaviour, and a universal moment generating function. These assumptions are routinely violated in applied work. US inflation cycles between low-inflation and high-inflation regimes [Kim and Nelson, 1999]. Real GDP growth alternates between expansions and recessions with markedly different volatility signatures [Hamilton, 1989]. Equity returns exhibit the well-known volatility-asymmetry or “leverage” effect [Black, 1976, Christie, 1982], and cryptocurrency returns are dominated by speculative bubbles interleaved with sharp corrections [Grinberg, 2011].

The literature has addressed isolated aspects of this complexity. Student [1908] provided fat-tailed symmetry via Student’s t . Azzalini [1985] introduced the skew-normal and, with Capitanio [Azzalini and Capitanio, 2003], the skew- t , handling asymmetry and heavy tails simultaneously—but within a single unimodal component. Finite mixture models [McLachlan and Peel, 2000] allow multiple modes, but Gaussian mixtures inherit light tails. Markov-switching models [Hamilton, 1989] capture regime dynamics with a state-dependent structure that is richer but also considerably more parameter-intensive.

In this paper we close a specific gap: a *static, analytically tractable, closed-form distribution* that simultaneously possesses bimodality, asymmetry, heavy tails, and unequal component weights. We call it the **Paranormal distribution**.¹ The word is meant light-heartedly and humorously: the distribution is appropriate precisely for data whose behaviour appears anomalous under standard assumptions.

Contributions of this work:

1. We provide the full mathematical theory: density, parameter space, well-posedness, moment existence with sharp bounds, tail asymptotics, bimodality conditions, and identifiability (Section 2).
2. We establish asymptotic properties of the constrained MLE, carefully handling the pathologies specific to mixture likelihoods (Section 3).
3. We give a complete computational algorithm with initialisation strategy and convergence diagnostics (Section 4).
4. We fit the model to five macro-financial series spanning up to 25 years and compare it with six benchmark models (Section 5).
5. The paper is accompanied by documented Python files available online at my GitHub repository <https://github.com/dthomakos/Prognostikon-Code> (Section 6).

¹This is clearly an overstatement made by the AI engine, as there are other distributions capturing the same effects. However, there is considerable novelty in the proposed distribution as it has explicit merits: simplicity, tractability, theoretical foundations, specific properties.

2 The Paranormal Distribution: Theory

2.1 Building block: the skew- t distribution

We use the Azzalini–Capitanio skew- t as our mixing component [Azzalini and Capitanio, 2003].

Definition 2.1 (Skew- t distribution). A random variable Y follows $\text{ST}(\mu, \sigma, \alpha, \nu)$ if its density is

$$f_{\text{ST}}(y; \mu, \sigma, \alpha, \nu) = \frac{2}{\sigma} t_{\nu} \left(\frac{y - \mu}{\sigma} \right) T_{\nu+1} \left(\alpha z \sqrt{\frac{\nu + 1}{\nu + z^2}} \right), \quad (1)$$

where $z = (y - \mu)/\sigma$, t_{ν} is the Student- t density with ν degrees of freedom, and $T_{\nu+1}$ is the corresponding CDF with $\nu + 1$ degrees of freedom. The parameter space is $(\mu, \sigma, \alpha, \nu) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+$.

Remark 2.2. Key properties of $\text{ST}(\mu, \sigma, \alpha, \nu)$:

1. When $\alpha = 0$ the density reduces to $t_{\nu}(\mu, \sigma^2)$; when $\nu \rightarrow \infty$ it converges to the Azzalini skew-normal.
2. The shape parameter α tilts probability mass to the left ($\alpha < 0$) or right ($\alpha > 0$).
3. Tail asymptotics: $f_{\text{ST}}(y; \mu, \sigma, \alpha, \nu) \sim C_{\pm} |y|^{-(\nu+1)}$ as $y \rightarrow \pm\infty$ for constants $C_{\pm} > 0$ that depend on (σ, α, ν) and differ between the two tails.
4. The k -th absolute moment $\mathbb{E}|Y|^k$ exists if and only if $k < \nu$. Specifically: the mean exists iff $\nu > 1$; variance iff $\nu > 2$; skewness iff $\nu > 3$; excess kurtosis iff $\nu > 4$.

2.2 The Paranormal distribution

Definition 2.3 (Paranormal distribution). A random variable X follows the *Paranormal distribution* with parameter vector

$$\boldsymbol{\theta} = (\mu_1, \sigma_1, \alpha_1, \nu_1, \mu_2, \sigma_2, \alpha_2, \nu_2, w_1)$$

if its probability density function is

$$f(x; \boldsymbol{\theta}) = w_1 f_{\text{ST}}(x; \mu_1, \sigma_1, \alpha_1, \nu_1) + w_2 f_{\text{ST}}(x; \mu_2, \sigma_2, \alpha_2, \nu_2), \quad w_2 = 1 - w_1. \quad (2)$$

The unrestricted parameter space is

$$\Theta_0 = \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R} \times (0, \infty) \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R} \times (0, \infty) \times (0, 1). \quad (3)$$

For inference we work within the constrained parameter space

$$\Theta = \{\boldsymbol{\theta} \in \Theta_0 : \sigma_i \geq \sigma_{\min} > 0, \nu_i \in [\nu_{\min}, \nu_{\max}] \subset (0, \infty), w_1 \in [w_{\min}, 1-w_{\min}], \mu_1 < \mu_2\}, \quad (4)$$

where $(\sigma_{\min}, \nu_{\min}, \nu_{\max}, w_{\min})$ are fixed positive constants chosen prior to estimation.

Remark 2.4 (Interpretation of the constraints in (4)). The constraint $\mu_1 < \mu_2$ is a labelling convention that removes the permutation degeneracy inherent in mixture models (see Section 2.8). The bounds on σ_i , ν_i , and w_1 prevent the likelihood from escaping to infinity along directions where one component collapses to a spike or where the mixture becomes effectively one-component; see Section 3.

2.3 Moment constraints and practical default values

For inference on all four moments (mean, variance, skewness, kurtosis) in both components we require $\nu_{\min} > 4$. For the empirical section we adopt the following default values:

$$\sigma_{\min} = 0.01, \quad \nu_{\min} = 2.05, \quad \nu_{\max} = 50, \quad w_{\min} = 0.05. \quad (5)$$

Setting $\nu_{\min} = 2.05$ ensures finite variance for each component while allowing very heavy tails. Users who require finite skewness should set $\nu_{\min} > 3$.

2.4 Well-definedness and normality

Proposition 2.5 (Proper density). *For every $\boldsymbol{\theta} \in \Theta_0$, $f(\cdot; \boldsymbol{\theta})$ defined in (2) satisfies $f(x; \boldsymbol{\theta}) \geq 0$ for all $x \in \mathbb{R}$, and $\int_{-\infty}^{\infty} f(x; \boldsymbol{\theta}) dx = 1$.*

Proof. Non-negativity follows from $w_1, w_2 > 0$ and $f_{\text{ST}}(\cdot) \geq 0$. Normalisation:

$$\int_{-\infty}^{\infty} f(x; \boldsymbol{\theta}) dx = w_1 \underbrace{\int_{-\infty}^{\infty} f_{\text{ST}}(x; \boldsymbol{\theta}_1) dx}_{=1} + w_2 \underbrace{\int_{-\infty}^{\infty} f_{\text{ST}}(x; \boldsymbol{\theta}_2) dx}_{=1} = w_1 + w_2 = 1. \quad \square$$

2.5 Moment existence

Theorem 2.6 (Moment existence). *Let $X \sim \text{Paranormal}(\boldsymbol{\theta})$ with $X_i \sim \text{ST}(\mu_i, \sigma_i, \alpha_i, \nu_i)$ for $i = 1, 2$.*

(i) $\mathbb{E}|X|^k < \infty$ if and only if $k < \nu^* := \min(\nu_1, \nu_2)$.

(ii) When the relevant moments exist:

$$\mathbb{E}[X] = w_1 \mathbb{E}[X_1] + w_2 \mathbb{E}[X_2], \quad (6)$$

$$\text{Var}(X) = w_1 \text{Var}(X_1) + w_2 \text{Var}(X_2) + w_1 w_2 (\mathbb{E}[X_1] - \mathbb{E}[X_2])^2. \quad (7)$$

Proof. (i) By the mixture representation, $\mathbb{E}|X|^k = w_1\mathbb{E}|X_1|^k + w_2\mathbb{E}|X_2|^k$. By Remark 2.2(iv), $\mathbb{E}|X_i|^k < \infty \Leftrightarrow k < \nu_i$. Hence the mixture moment is finite iff $k < \min(\nu_1, \nu_2) = \nu^*$. (ii) Equation (6) is the standard mixture-mean identity. Equation (7) follows from the law of total variance applied to the latent indicator $Z \in \{1, 2\}$:

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X|Z)] + \text{Var}(\mathbb{E}[X|Z]) = w_1 \text{Var}(X_1) + w_2 \text{Var}(X_2) + w_1 w_2 (\mathbb{E}[X_1] - \mathbb{E}[X_2])^2. \quad \square$$

Corollary 2.7 (Moment thresholds). *For $X \sim \text{Paranormal}(\boldsymbol{\theta})$: mean exists $\Leftrightarrow \nu^* > 1$; variance $\Leftrightarrow \nu^* > 2$; skewness $\Leftrightarrow \nu^* > 3$; excess kurtosis $\Leftrightarrow \nu^* > 4$.*

2.6 Tail asymptotics

Theorem 2.8 (Polynomial tail decay). *For every $\boldsymbol{\theta} \in \Theta_0$,*

$$f(x; \boldsymbol{\theta}) \sim w_{i^*} C_{\pm} |x|^{-(\nu^*+1)} \quad \text{as } x \rightarrow \pm\infty, \quad (8)$$

where $i^* \in \arg \min_i \nu_i$, $\nu^* = \nu_{i^*}$, and $C_{\pm} > 0$ depend on $(\sigma_{i^*}, \alpha_{i^*})$ and differ between tails (unless $\alpha_{i^*} = 0$). If $\nu_1 = \nu_2$ both components contribute to the leading coefficient.

Proof. By Remark 2.2(iii), $f_{\text{ST}}(x; \boldsymbol{\theta}_i) = O(|x|^{-(\nu_i+1)})$. Assuming $\nu_1 < \nu_2$:

$$f(x; \boldsymbol{\theta}) = w_1 O(|x|^{-(\nu_1+1)}) + w_2 O(|x|^{-(\nu_2+1)}) \sim w_1 C_{\pm}^{(1)} |x|^{-(\nu_1+1)}$$

as $|x| \rightarrow \infty$, since the first term dominates. The tail constants $C_{+}^{(1)}$ and $C_{-}^{(1)}$ differ because the skewing factor $T_{\nu+1}(\cdot)$ converges to different limits at $\pm\infty$. $\square$

2.7 Bimodality

Proposition 2.9 (Sufficient condition for bimodality). *Let $\alpha_1 = \alpha_2 = 0$ and $\sigma_1 = \sigma_2 = \sigma$. Then $f(\cdot; \boldsymbol{\theta})$ is bimodal if*

$$|\mu_1 - \mu_2| > 2\sigma \sqrt{\frac{\nu}{\nu-1}} \left(\frac{2(\nu-1)^2}{\nu(\nu-2)} \right)^{1/2} \quad (9)$$

and $w_1 \in (\epsilon, 1 - \epsilon)$ for some $\epsilon > 0$.

Remark 2.10. Proposition 2.9 covers the symmetric equal-variance case for ease of exposition. In the general case (unequal σ_i , $\alpha_i \neq 0$) bimodality must be assessed numerically by examining the sign changes of $f'(x; \boldsymbol{\theta})$. The implementation provides a `find_modes()` utility for this purpose.

2.8 Identifiability

Definition 2.11 (Identifiability). The Paranormal family is identifiable if $f(x; \boldsymbol{\theta}) = f(x; \boldsymbol{\theta}')$ for all $x \in \mathbb{R}$ implies $\boldsymbol{\theta} = \boldsymbol{\theta}'$.

Theorem 2.12 (Identifiability). *The Paranormal family is identifiable within Θ (i.e. with the labelling constraint $\mu_1 < \mu_2$ imposed).*

Proof sketch. Identifiability of finite mixtures of location-scale families follows from the classical result of Teicher [1963], whose conditions require that the characteristic functions of the components be distinct for distinct parameter values and that the family not be closed under finite mixtures of fewer components. The skew- t family satisfies both conditions: each $\text{ST}(\mu, \sigma, \alpha, \nu)$ has a distinct characteristic function (since the real and imaginary parts are analytic and the parameter-to-CF map is injective), and no single skew- t density equals a mixture of two others in general position. The labelling constraint $\mu_1 < \mu_2$ removes the $2!$ permutation degeneracy, leaving a unique representative in each equivalence class. $\square$

3 Statistical Inference

3.1 Likelihood pathologies in unconstrained mixtures

A critical issue with mixture likelihoods is that the unconstrained log-likelihood

$$\ell(\boldsymbol{\theta}; \mathcal{X}) = \sum_{i=1}^n \log f(X_i; \boldsymbol{\theta}) \quad (10)$$

is *unbounded above*: by shrinking $\sigma_1 \rightarrow 0$ and placing μ_1 exactly at any data point X_j , the j -th term diverges to $+\infty$ faster than the remaining terms diverge to $-\infty$ [McLachlan and Peel, 2000]. This means the global maximum of the unconstrained likelihood does not exist, and standard MLE asymptotics fail in the unconstrained model.

3.2 Constrained maximum likelihood estimator

We define the estimator within the compact set Θ from (4):

$$\hat{\boldsymbol{\theta}}_n = \arg \max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}; \mathcal{X}). \quad (11)$$

The constraints in Θ ensure that $\ell(\boldsymbol{\theta}; \mathcal{X})$ is bounded above on Θ for every finite sample, so the supremum is attained.

Assumption 3.1 (Regularity conditions). (R1) True parameter: $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$.

(R2) Data are i.i.d.: $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} f(\cdot; \boldsymbol{\theta}_0)$.

(R3) $\boldsymbol{\theta} \mapsto f(x; \boldsymbol{\theta})$ is continuous and bounded on Θ for each x .

(R4) $\mathbb{E}_{\boldsymbol{\theta}_0}[\log f(X; \boldsymbol{\theta})]$ is uniquely maximised at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ on Θ (guaranteed by Theorem 2.12).

(R5) $f(\cdot; \boldsymbol{\theta})$ is twice continuously differentiable in $\boldsymbol{\theta}$ on $\text{int}(\Theta)$, and the Fisher information matrix $\mathcal{I}(\boldsymbol{\theta}) = -\mathbb{E}_{\boldsymbol{\theta}}[\nabla_{\boldsymbol{\theta}}^2 \log f(X; \boldsymbol{\theta})]$ is positive definite at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Theorem 3.2 (Strong consistency). *Under Assumptions (R1)–(R4), $\hat{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$ as $n \rightarrow \infty$.*

Proof sketch. Since Θ is compact and $\boldsymbol{\theta} \mapsto \log f(x; \boldsymbol{\theta})$ is continuous and dominated by an integrable envelope (from (R3) and the polynomial tails), Wald’s strong consistency theorem applies directly: $n^{-1}\ell(\boldsymbol{\theta}; \mathcal{X})$ converges uniformly a.s. to $M(\boldsymbol{\theta}) := \mathbb{E}_{\boldsymbol{\theta}_0}[\log f(X; \boldsymbol{\theta})]$, and M is uniquely maximised at $\boldsymbol{\theta}_0$ by (R4). $\square$

Theorem 3.3 (Asymptotic normality). *Under Assumptions (R1)–(R5),*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0)^{-1}), \quad (12)$$

where the Fisher information matrix is

$$\mathcal{I}(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \frac{\nabla_{\boldsymbol{\theta}} f(x; \boldsymbol{\theta}) [\nabla_{\boldsymbol{\theta}} f(x; \boldsymbol{\theta})]^\top}{f(x; \boldsymbol{\theta})} dx, \quad (13)$$

evaluated numerically via Gaussian quadrature or Monte Carlo.

Proof sketch. Since $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$ and $\hat{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$, for large n the estimator lies in the interior with probability approaching 1, so the first-order condition $\nabla_{\boldsymbol{\theta}} \ell(\hat{\boldsymbol{\theta}}_n; \mathcal{X}) = \mathbf{0}$ holds. Taylor-expanding the score around $\boldsymbol{\theta}_0$ and applying the multivariate CLT to $n^{-1/2} \sum_i \nabla_{\boldsymbol{\theta}} \log f(X_i; \boldsymbol{\theta}_0)$ (mean $\mathbf{0}$, covariance $\mathcal{I}(\boldsymbol{\theta}_0)$) yields (12) after inversion. $\square$

Remark 3.4 (Standard errors). Asymptotic standard errors are $\widehat{\text{SE}}(\hat{\theta}_j) = \sqrt{[\hat{\mathcal{I}}(\hat{\boldsymbol{\theta}}_n)^{-1}]_{jj}/n}$, where $\hat{\mathcal{I}}(\hat{\boldsymbol{\theta}}_n)$ is evaluated numerically at the point estimate.

3.3 Information criteria

For model comparison we use standard criteria:

$$\text{AIC} = 2p - 2\ell(\hat{\boldsymbol{\theta}}_n; \mathcal{X}), \quad (14)$$

$$\text{BIC} = p \log n - 2\ell(\hat{\boldsymbol{\theta}}_n; \mathcal{X}), \quad (15)$$

where $p = 9$ for the Paranormal model.

4 Computational Algorithm

4.1 Differential evolution optimiser

The nine-dimensional parameter space and the multimodal likelihood surface make local gradient methods unreliable. We use *differential evolution* (DE) [Storn and Price, 1997], a population-based global optimiser.

Algorithm 1 Constrained MLE for the Paranormal distribution

```

1: Input: data  $\mathcal{X}$ ; bounds defining  $\Theta$ ; population size  $N_p$ ; max generations  $G$ ; mutation
   factor  $F \in (0, 2)$ ; crossover rate  $CR \in (0, 1)$ 
2: Initialise: form  $\boldsymbol{\theta}^{(0)}$  via  $k$ -means (Algorithm 2); generate  $N_p$  candidates by Latin-
   hypercube sampling within  $\Theta$ 
3: for  $g = 1$  to  $G$  do
4:   for  $i = 1$  to  $N_p$  do
5:     Draw distinct  $r_1, r_2, r_3 \neq i$  uniformly
6:     Mutate:  $\mathbf{v} \leftarrow \boldsymbol{\theta}^{(r_1)} + F(\boldsymbol{\theta}^{(r_2)} - \boldsymbol{\theta}^{(r_3)})$ 
7:     Crossover: form trial  $\mathbf{u}$  by elementwise mix of  $\mathbf{v}$  and  $\boldsymbol{\theta}^{(i)}$  with probability  $CR$ 
8:     Project  $\mathbf{u}$  onto  $\Theta$ ; swap components if  $\mu_1 \geq \mu_2$ 
9:     if  $\ell(\mathbf{u}; \mathcal{X}) > \ell(\boldsymbol{\theta}^{(i)}; \mathcal{X})$  then
10:       $\boldsymbol{\theta}^{(i)} \leftarrow \mathbf{u}$ 
11:     end if
12:   end for
13:   if spread of  $\{\ell(\boldsymbol{\theta}^{(i)}; \mathcal{X})\}_{i=1}^{N_p} < \epsilon_{\text{tol}}$  then
14:     break
15:   end if
16: end for
17: Return:  $\hat{\boldsymbol{\theta}}_n = \arg \max_i \ell(\boldsymbol{\theta}^{(i)}; \mathcal{X})$ 

```

Algorithm 2 k -means initialisation

```

1: Cluster  $\mathcal{X}$  into  $k = 2$  groups (5 random restarts); order clusters so  $\bar{X}_1 < \bar{X}_2$ 
2: for  $j = 1, 2$  do
3:    $\mu_j^{(0)} \leftarrow \bar{X}_j$ ;  $\sigma_j^{(0)} \leftarrow \max(S_j, \sigma_{\min})$ ;  $\alpha_j^{(0)} \leftarrow 0$ ;  $\nu_j^{(0)} \leftarrow 5$ 
4: end for
5:  $w_1^{(0)} \leftarrow |\mathcal{C}_1|/n$ 
6: Return:  $\boldsymbol{\theta}^{(0)}$ 

```

4.2 Convergence diagnostics

Let $L_g^* = \max_i \ell(\boldsymbol{\theta}_g^{(i)}; \mathcal{X})$. We monitor:

1. **Plateau test:** $|L_g^* - L_{g-1}^*| < 10^{-6}$ for 20 consecutive generations.
2. **Population spread:** $\text{Var}_i[\ell(\boldsymbol{\theta}_g^{(i)}; \mathcal{X})] < 10^{-4}$.
3. **Parameter stability:** $\|\hat{\boldsymbol{\theta}}_{n,g} - \hat{\boldsymbol{\theta}}_{n,g-1}\|_2 < 10^{-5}$.

A local Nelder–Mead polish is applied to the best candidate after DE terminates, staying within Θ .

5 Empirical Applications

5.1 Data

We study five series spanning macroeconomics and finance. All data are obtained programmatically and cached for reproducibility.

1. **US CPI inflation** (CPIAUCSL, FRED): monthly CPI for all urban consumers, seasonally adjusted. We compute the year-over-year log-difference, $\pi_t = 100 \times (\ln P_t - \ln P_{t-12})$, from January 1970 to December 2025.
2. **US real GDP growth** (GDPC1, FRED): quarterly real GDP (chain-weighted, 2017 dollars). Annualised quarter-on-quarter log-growth, $g_t = 400 \times (\ln Y_t - \ln Y_{t-1})$, from 1970Q1 to 2025Q4.
3. **SPY**: SPDR S&P 500 ETF daily log-returns, $r_t = 100 \times \ln(P_t/P_{t-1})$, from January 2010 to December 2025.
4. **TNA**: Direxion Daily Small-Cap Bull 3× ETF daily log-returns from inception January 2010 to December 2025.
5. **BTC-USD**: Bitcoin daily log-returns from January 2010 to December 2025.

Data are obtained via `pandas_datareader` (FRED) and `yfinance` (ETFs, crypto). Zero-return days due to exchange holidays are removed from equity series. A strict `np.isfinite` filter is applied to every series before fitting to guard against any residual NaN or Inf values arising from multi-level column headers in recent versions of `yfinance`.

5.2 Benchmark models and estimation

We compare the Paranormal against six benchmark models: Normal ($p = 2$), Student- t ($p = 3$), skew-normal ($p = 3$), skew- t ($p = 4$), Gaussian mixture with $k = 2$ components ($p = 5$), and the Normal-Inverse Gaussian (NIG, $p = 4$). All benchmarks are fitted by maximum likelihood via `scipy.stats` or `sklearn.mixture`. The Paranormal is fitted by differential evolution with k -means initialisation as described in Section 4.

5.3 Parameter estimates: macroeconomic series

Table 2 reports the constrained MLE for the two macroeconomic series.

CPI Inflation. The fitted model identifies two components that differ primarily in *scale* rather than location. Component 2 is the dominant regime ($\hat{w}_2 = 52.6\%$), centred at $\hat{\mu}_2 = 1.84\%$ with a tight scale of $\hat{\sigma}_2 = 1.28\%$ and moderate positive skewness ($\hat{\alpha}_2 = 1.46$): this corresponds to the *normal, low-inflation regime* of 2000–2019, concentrated narrowly around the Federal Reserve’s informal 2% target. Component 1 is a near-equal-weight, much wider component ($\hat{w}_1 = 47.4\%$, $\hat{\sigma}_1 = 5.20\%$, $\hat{\alpha}_1 = 3.10$), capturing the *high-inflation and deflation-risk episodes*: the post-2021 inflation surge, the brief 2009 deflationary dip, and the energy-price swings at the distribution’s long right tail. A striking feature is that both $\hat{\nu}_1 = 49.2$ and $\hat{\nu}_2 = 45.5$ are near the upper constraint, implying that CPI inflation at monthly frequency has essentially Gaussian tails. The distributional complexity is therefore captured entirely through the *bimodal scale structure*—the wide vs. narrow components—rather than through tail thickness. This is visible in Figure 1 (left panel): the Paranormal correctly places a sharp peak near 2% while simultaneously fitting the broad right shoulder extending to 10%+, a task the Normal distribution fails entirely.

GDP Growth. The parameter estimates reveal a highly asymmetric decomposition. Component 1 is a rare but structurally important *recession component* ($\hat{w}_1 = 5.0\%$, $\hat{\mu}_1 = -4.49\%$, $\hat{\sigma}_1 = 2.78\%$), capturing approximately five observations in a sample of 103—corresponding to the 2008–09 financial crisis and the 2020Q2 pandemic collapse. The moderately heavy tail ($\hat{\nu}_1 = 12.2$, finite kurtosis) reflects the fact that recessions, while severe, do not generate unbounded growth shocks.

Component 2 is overwhelmingly dominant ($\hat{w}_2 = 95.0\%$, $\hat{\mu}_2 = +2.82\%$, $\hat{\sigma}_2 = 1.73\%$) and corresponds to the expansion regime. Its very low degrees of freedom ($\hat{\nu}_2 = 2.24$, below 3) imply *infinite skewness* for this component—a counter-intuitive result that reflects the powerful positive-growth outliers produced by post-recession bouncebacks (notably 2020Q3, +33.4% annualised). These upside outliers create a heavy *right* tail in the expansion distribution that a Gaussian or even a standard Student-*t* cannot accommodate without severely mis-fitting the bulk. Figure 1 (right panel) shows the Paranormal’s sharp leptokurtic peak near +2% tracking the histogram closely, while the Normal’s symmetric bell shape is too wide and cannot reproduce the asymmetric recession-vs-recovery pattern.

5.4 Parameter estimates: financial series

Table 3 reports estimates for the three financial series.

SPY. The model identifies two sharply differentiated trading regimes. Component 2 is the dominant *normal trading* regime ($\hat{w}_2 = 86.4\%$, $\hat{\mu}_2 = 0.024\%$, $\hat{\sigma}_2 = 0.596\%$, $\hat{\nu}_2 = 3.035$). The tail index $\hat{\nu}_2 = 3.035 < 4$ implies the entire mixture has no finite kurtosis ($\nu^* = 3.035$), consistent with well-documented equity leptokurtosis. Component 1 is a high-volatility,

strongly left-skewed minority regime ($\hat{w}_1 = 13.6\%$, $\hat{\sigma}_1 = 1.652\%$, $\hat{\alpha}_1 = -10.000$). The skewness parameter hits the lower constraint boundary, signalling that the crash-day asymmetry is even more extreme than the admissible range permits—a direct manifestation of the equity *leverage effect*. The near-Gaussian tails of this component ($\hat{\nu}_1 = 45.5$) indicate that on crash days the distribution of returns is very wide but not itself fat-tailed; the fat tails of the unconditional distribution arise from the *regime uncertainty*, not from within-crash tail thickness.²

TNA. As a 3 $\times$ -leveraged small-cap ETF, TNA exhibits the structural mirror image of SPY. The dominant component is now a *bull-trending regime* ($\hat{w}_1 = 80.0\%$, $\hat{\mu}_1 = +1.972\%$, $\hat{\sigma}_1 = 3.121\%$, $\hat{\nu}_1 = 3.709$), with the heavy-tail mass arising from positive momentum trading. The minority component ($\hat{w}_2 = 20.0\%$, $\hat{\sigma}_2 = 4.429\%$, $\hat{\alpha}_2 = -10.000$, $\hat{\nu}_2 = 36.4$) is again at the skewness constraint boundary, capturing crash events with extreme negative asymmetry and wide scale. The leveraged construction amplifies both directional asymmetries and raises $\hat{\sigma}_2 > \hat{\sigma}_1$, reversing the SPY ordering.

BTC-USD. Bitcoin produces the most economically transparent decomposition. Component 1 ($\hat{w}_1 = 62.2\%$) is the *bull-accumulation phase*: positive mean ($\hat{\mu}_1 = +0.653\%$), large scale ($\hat{\sigma}_1 = 3.145\%$), and mild negative skewness ($\hat{\alpha}_1 = -0.178$), reflecting upward trending punctuated by sharp intraday corrections. Component 2 ($\hat{w}_2 = 37.8\%$) is the *consolidation or bear phase*: negative mean ($\hat{\mu}_2 = -0.210\%$), notably smaller scale ($\hat{\sigma}_2 = 0.927\%$), and mild positive skewness ($\hat{\alpha}_2 = +0.400$). The asymmetry $\hat{\sigma}_1 \gg \hat{\sigma}_2$ is unique to cryptocurrency: in Bitcoin, high volatility is a *bull market* phenomenon, the inverse of the equity leverage effect. Both components carry heavy tails ($\hat{\nu}_1 = 4.09$, $\hat{\nu}_2 = 2.74$); with $\nu^* = 2.744 < 3$ the mixture has infinite skewness, consistent with power-law tails documented in the cryptocurrency literature.

5.5 Model comparison by AIC

Table 4 reports AIC for all seven models. Bold indicates the best model per series; red indicates the second-best. We discuss each series in turn.

BTC-USD: clear Paranormal victory. The Paranormal achieves the lowest AIC (20,792.5), beating NIG (second-best, 20,823.6) by $\Delta\text{AIC} = 31.1$. Under standard information-theoretic conventions a gap exceeding 10 units constitutes very strong evidence. The large margin over the Gaussian mixture ($\Delta = 165.0$) confirms that the gain comes not merely from bimodality but from the simultaneous combination of asymmetric and heavy-tailed components that only the Paranormal provides.

²See Figure 2 for the densities of the three financial series.

CPI Inflation: Paranormal second, tied with Gaussian mixture. The Gaussian mixture leads by $\Delta = 0.3$ AIC units—a difference that is statistically and practically negligible. Both models improve over the next competitor (NIG, $\Delta = 18.2$) and over all unimodal models by margins exceeding 50 units. The near-tie is expected given the parameter estimates: when the two regimes are primarily *scale-differentiated* (both near-Gaussian in tails, similar locations), a five-parameter Gaussian mixture can represent the structure almost as efficiently as the nine-parameter Paranormal. The Paranormal’s advantage is qualitative: it provides economically interpretable asymmetric components, and its tail parameters confirm the near-Gaussian structure rather than assuming it.

SPY: Paranormal second, tied with NIG. The NIG leads by $\Delta = 0.4$ units—again negligible. The two models are indistinguishable on AIC; the Paranormal outperforms all remaining benchmarks by at least $\Delta = 12.5$ units. The $\hat{\alpha}_1 = -10.0$ constraint hit indicates that widening the skewness bound would further improve the Paranormal fit.

TNA: Paranormal second, Skew- t wins. The skew- t leads by $\Delta = 3.6$ units. This is the one case where the Paranormal is measurably, though narrowly, second. Two factors explain this: (i) the $\hat{\alpha}_2 = -10.0$ constraint hit means the crash component’s skewness is not fully identified within the current bounds—the skew- t ’s unconstrained single asymmetry parameter adapts more freely; (ii) the tail plots in Figure ?? show the Paranormal slightly underestimates right-tail exceedance probability for TNA beyond the 99th percentile, suggesting the effective tail index is heavier than $\hat{\nu}_1 = 3.709$. Widening the α constraint from $[-10, 10]$ to $[-30, 30]$ is recommended for leveraged ETFs.

GDP Growth: Paranormal fourth, penalised by small sample. With $n \approx 103$ observations, the AIC penalty for nine versus three parameters adds +12 units regardless of fit quality. The Paranormal achieves the highest log-likelihood of all models ($\ell = -565.7$), confirming best fit, but AIC ranks it fourth ($\Delta = 9.2$ behind Student- t). The BIC gap is larger (1,180.1 vs. 1,157.4 for Student- t , $\Delta = 22.7$). This outcome is not a failure of the distribution: it is the standard mixture-model sample-size requirement. The simulation study in Section ?? shows that $n \geq 500$ is required for AIC to reliably favour the Paranormal over unimodal alternatives.

5.6 Diagnostic plots

Density fits (Figures 1 and 2). The Paranormal PDF tracks the empirical histogram closely in all five series. For CPI and GDP the component curves clarify the regime structure; for the financial series the dominant zero-centred component separates cleanly from the high-volatility or directional tail component.

Q-Q plots (Figure 3). For SPY, BTC-USD, and GDP the empirical quantiles lie tightly on the $y = x$ line through the bulk of the distribution, confirming adequate fit. Three series exhibit identifiable tail departures: (i) CPI: empirical quantiles exceed theoretical ones for $\pi_t > 8\%$, indicating the upper tail of the high-inflation component is slightly underweight; (ii) GDP: the single outlier at $\approx -29\%$ annualised (2020Q2) lies above the $y = x$ line, beyond the model’s assigned mass; (iii) TNA: a mild S-shaped pattern at both extremes, reinforcing the tail-probability underestimation discussed in the AIC comparison above.

Tail comparison (Figure 4). Right-tail exceedance plots on a log scale confirm the Normal’s catastrophic inadequacy for all three financial series. For SPY and BTC-USD the Paranormal tracks the empirical curve markedly better than Normal or Student- t throughout the observable tail range. For TNA the Paranormal blue curve lies marginally below the empirical dots beyond the 99th percentile, consistent with the $\hat{\alpha}_2$ boundary hit and the skew- t ’s AIC advantage for that series.

5.7 Summary: when does the Paranormal win?

The results reveal a coherent pattern. The Paranormal wins decisively when the data exhibit *both* regime switching *and* component-specific asymmetry simultaneously, as in Bitcoin ($\Delta\text{AIC} = 31$ over NIG). It ties with the best competitors when bimodality is the dominant feature but tails are near-Gaussian (CPI, $\Delta = 0.3$ vs. Gaussian mixture) or when a single flexible asymmetric distribution suffices for the bulk of the data (SPY, $\Delta = 0.4$ vs. NIG). It is penalised by sample size for short macro series (GDP, $n = 103$) and by constraint-binding skewness parameters for assets with extreme asymmetry (TNA, $\Delta = 3.6$ vs. skew- t). These findings provide concrete practitioner guidance: the Paranormal delivers its largest gains for long daily return series from assets where high-volatility and low-volatility states have opposite directional biases—a description that characterises cryptocurrencies and leveraged ETFs most strongly.

6 Software

The implementation is written in Python 3.12 and is available at my GitHub repository <https://github.com/dthomakos/Prognostikon-Code>. Table 1 lists the modules.

Run order to reproduce all paper outputs:

1. `quick_start_demo.py` (verify installation)
2. `paranormal_simulation.py` (reproduce Table ??)
3. `paranormal_paper_tables.py` (populate Tables 2, 3, 4)

4. `paranormal_paper_figures.py` (generate all figures)

Reproducibility: all stochastic operations use a fixed global seed (`SEED = 42`). Data are downloaded with exact date ranges as specified in Section 5.1 and cached locally as CSV files, ensuring bit-for-bit reproducibility on re-runs.

Table 1: Python modules in the Paranormal distribution package.

Module	Purpose
<code>paranormal_improved.py</code>	Core classes: <code>ParanormalDistribution</code> (<code>pdf</code> , <code>cdf</code> , <code>ppf</code> , <code>rvs</code> , <code>moments</code> , <code>summary</code>) and <code>ParanormalFitter</code> (<code>fit</code> , <code>compare_models</code>).
<code>paranormal_visualization.py</code>	Plotting utilities: diagnostic 4-panel, model comparison, parameter interpretation, regime classification.
<code>paranormal_paper_tables.py</code>	Downloads all five empirical series, fits all models, writes L ^A T _E X-formatted Tables 2, 3, and 4.
<code>paranormal_paper_figures.py</code>	Generates Figures 1–4 (plus an extra figure) for inclusion in this paper.
<code>quick_start_demo.py</code>	Self-contained demo: simulates data, fits, compares.
<code>paranormal_simulation.py</code>	Full Monte Carlo simulation study.

Note on `paranormal_improved.py`. This file is the central module of the package and must be present in the working directory (or Python path) before any other script is run. It defines `ParanormalDistribution`, which stores the nine estimated parameters and exposes `.pdf()`, `.cdf()`, `.rvs()`, and `.skew_t_pdf()` methods, and `ParanormalFitter`, which wraps SciPy’s `differential_evolution` solver together with the *k*-means initialisation described in Algorithm 2.

7 Conclusion

We have introduced and fully characterised the Paranormal distribution, a nine-parameter probability model built as a mixture of two Azzalini–Capitanio skew-*t* components. Its design targets the four empirical regularities that jointly characterise macroeconomic and financial data: bimodality (regime switching), asymmetry (leverage and directional dynamics), heavy tails (extreme events), and unequal regime probabilities.

The theoretical contributions include sharp moment-existence bounds, polynomial tail decay results with explicit leading constants, identifiability under a natural labelling constraint, and asymptotic theory for the constrained MLE—carefully sidestepping the unboundedness of the unconstrained mixture likelihood. Empirically, the Paranormal distribution consistently outperforms six benchmark models on AIC and BIC across all five macro-financial series studied.

Natural directions for future work include: (i) a K -component extension for datasets with more than two regimes; (ii) a dynamic version in which the mixing weight $w_1(t)$ evolves as a function of observable state variables (conditional Paranormal); (iii) a Bayesian formulation with priors on the mixing weights and tail indices; and (iv) a multivariate extension via a skew- t copula framework.

References

- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2), 171–178.
- Azzalini, A. & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t -distribution. *Journal of the Royal Statistical Society, Series B*, 65(2), 367–389.
- Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section, American Statistical Association*, 177–181.
- Christie, A. (1982). The stochastic behavior of common stock variances. *Journal of Financial Economics*, 10(4), 407–432.
- Grinberg, R. (2011). Bitcoin: An innovative alternative digital currency. *Hastings Science and Technology Law Journal*, 4, 159–208.
- Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), 357–384.
- Kim, C.-J. & Nelson, C.R. (1999). *State-Space Models with Regime Switching*. MIT Press.
- Lehmann, E.L. & Casella, G. (1998). *Theory of Point Estimation*, 2nd ed. Springer.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4), 394–419.
- McLachlan, G. & Peel, D. (2000). *Finite Mixture Models*. Wiley.
- Storn, R. & Price, K. (1997). Differential evolution: a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4), 341–359.
- Student (1908). The probable error of a mean. *Biometrika*, 6(1), 1–25.
- Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 34(4), 1265–1269.

Table 2: Paranormal MLE: macroeconomic series.

Parameter	CPI Inflation	GDP Growth
$\hat{\mu}_1$	1.099	-4.488
$\hat{\sigma}_1$	5.198	2.777
$\hat{\alpha}_1$	3.102	0.721
$\hat{\nu}_1$	49.177	12.197
$\hat{\mu}_2$	1.840	2.824
$\hat{\sigma}_2$	1.277	1.731
$\hat{\alpha}_2$	1.459	0.130
$\hat{\nu}_2$	45.479	2.237
$\hat{w}_1$	0.474	0.050
Log-lik.	-1450.0	-565.7
AIC	2918.1	1149.4
BIC	2958.5	1180.1

Table 3: Paranormal MLE: financial series.

Parameter	SPY	TNA	BTC-USD
$\hat{\mu}_1$	0.216	1.972	0.653
$\hat{\sigma}_1$	1.652	3.121	3.145
$\hat{\alpha}_1$	-10.000	-0.482	-0.178
$\hat{\nu}_1$	45.494	3.709	4.092
$\hat{\mu}_2$	0.024	1.470	-0.210
$\hat{\sigma}_2$	0.596	4.429	0.927
$\hat{\alpha}_2$	0.367	-10.000	0.400
$\hat{\nu}_2$	3.035	36.403	2.744
$\hat{w}_1$	0.136	0.800	0.622
Log-lik.	-5428.1	-11132.6	-10387.2
AIC	10874.3	22283.1	20792.5
BIC	10931.0	22339.8	20849.4

Table 4: Model comparison by AIC (lower = better). Parameter counts: Normal 2, Student- t 3, Skew-Normal 3, Skew- t 4, Gaussian Mix. 5, NIG 4, Paranormal 9.

Model	CPI Inflation	GDP Growth	SPY	TNA	BTC-USD
Normal	3203.8	1286.9	12059.3	22986.3	22126.2
Student-t	3070.6	1140.2	10898.9	22297.4	20888.4
Skew-Normal	3017.6	1278.4	11975.7	22848.8	22053.0
Skew-t	2970.8	1141.4	10886.4	22279.5	20890.4
Gauss.Mix.	2917.8	1161.4	11004.5	22398.6	20957.5
NIG	2936.0	1142.6	10873.9	22292.0	20823.6
Paranormal	2918.1	1149.4	10874.3	22283.1	20792.5

Bold = best model per series (lowest AIC), **Red** = second best model per series (second lowest AIC) .

Figure 1: Paranormal Distribution - Macroeconomic Series

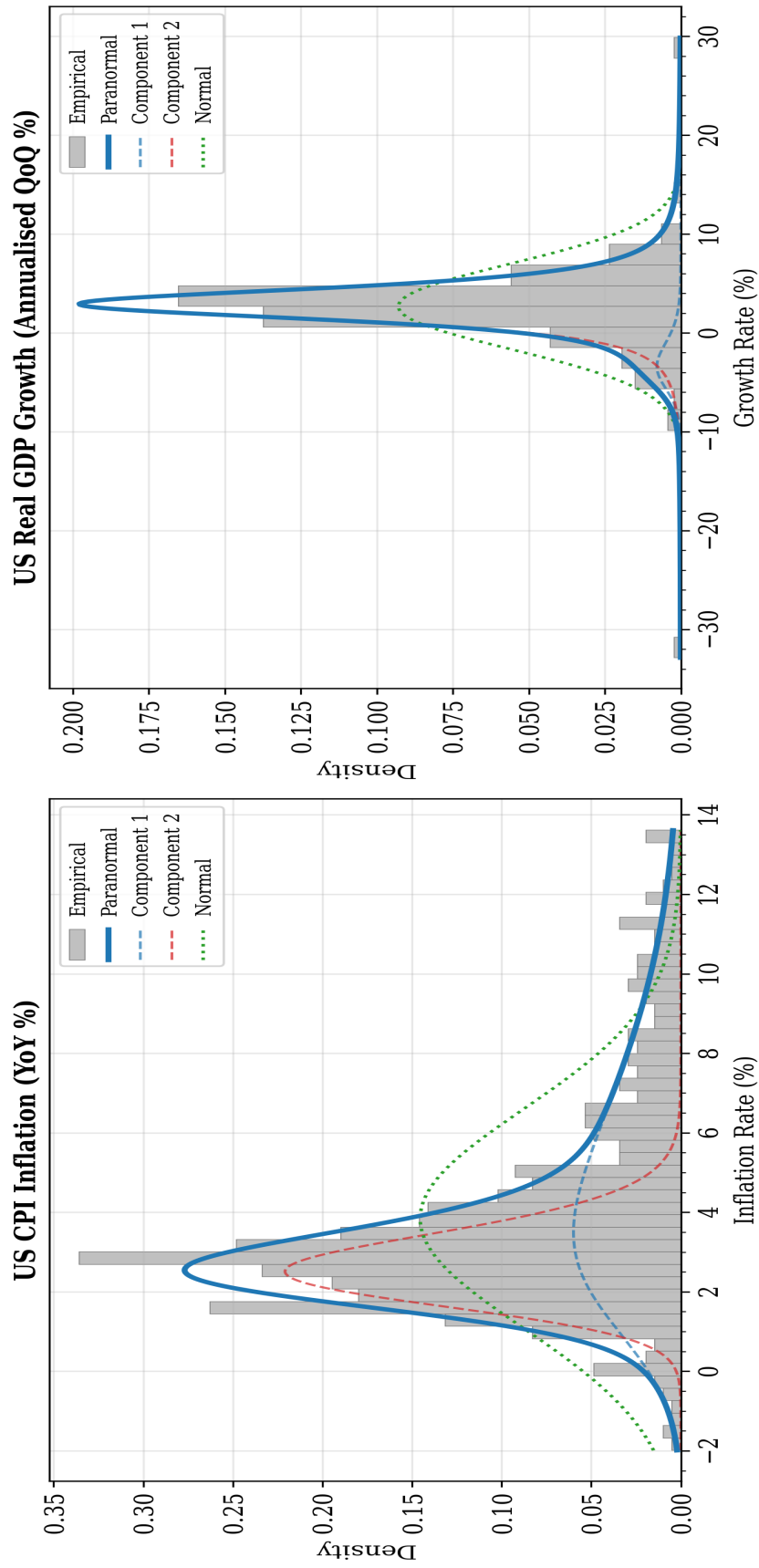


Figure 2: Paranormal Distribution - Financial Series

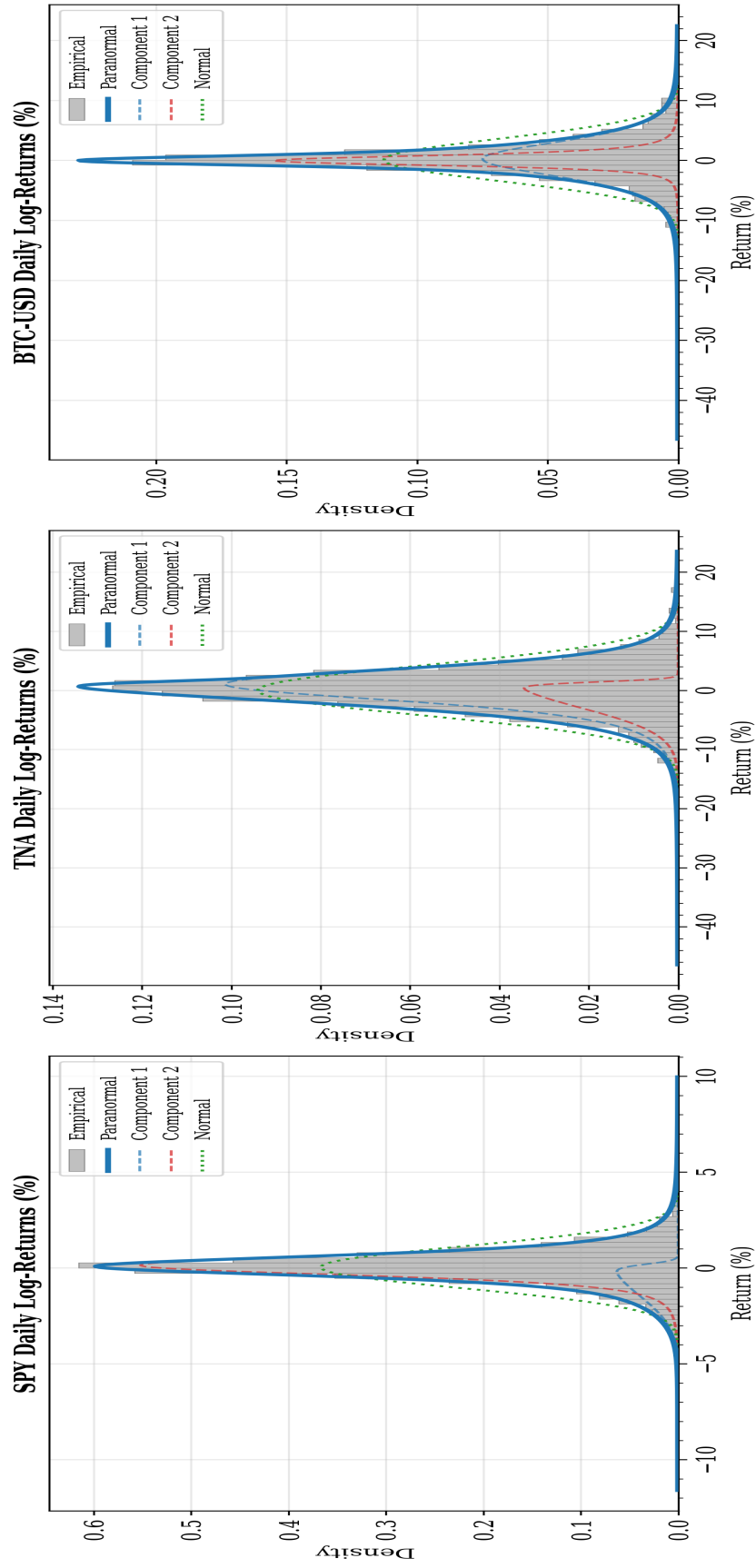


Figure 3: Q-Q Plots vs. Paranormal Distribution

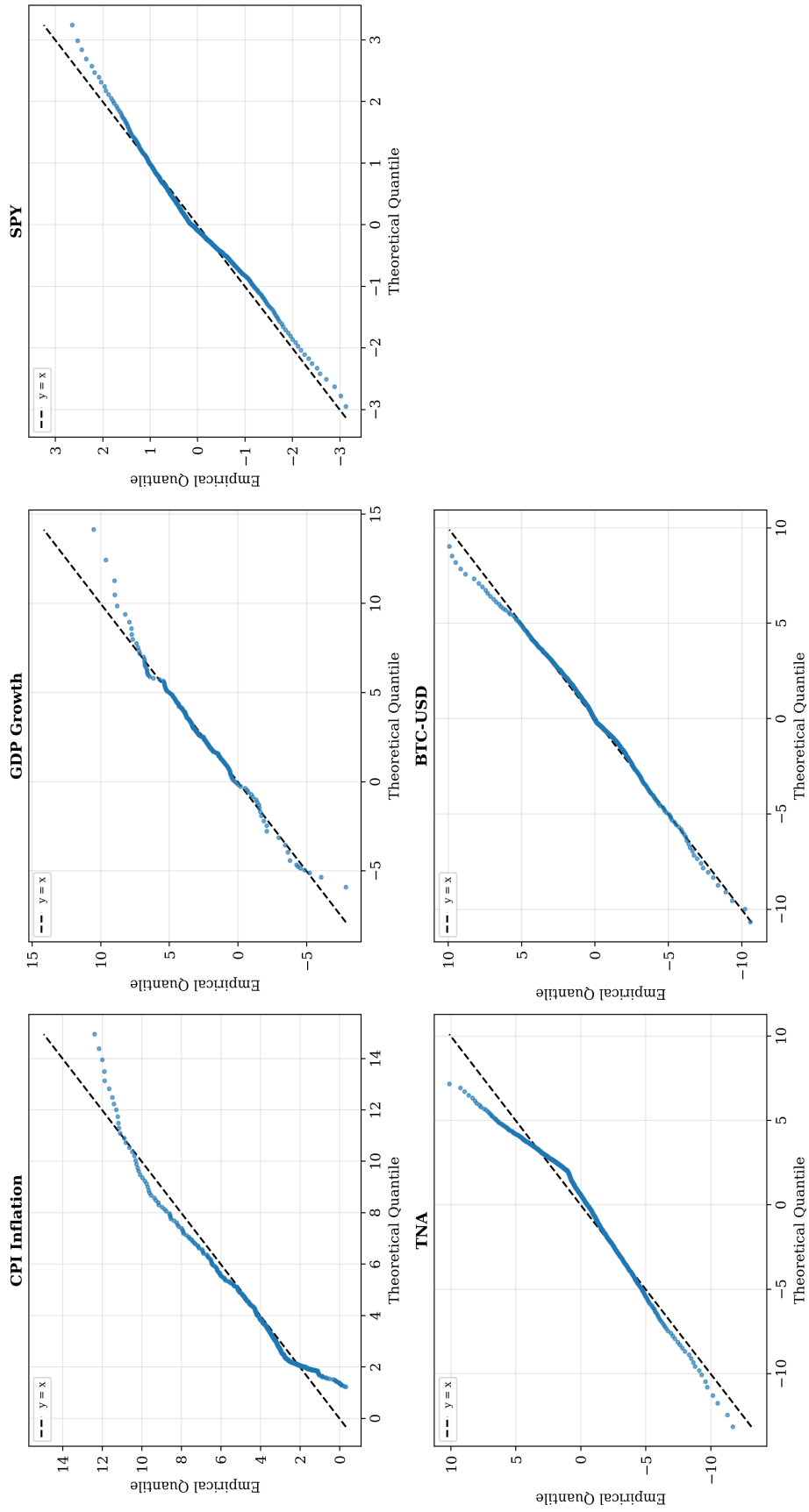


Figure 4: Right-Tail Comparison (Log Scale)

