ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Time Series Forecasting in the Presence of Explosive Bubbles

Julien Peignon[1], Fabrice Rossi[1] and Arthur Thomas[2]

1- CEREMADE, CNRS, UMR 7534, Université Paris-Dauphine
PSL University, Paris, France

2- LEDa, CNRS, UMR 8007, Université Paris-Dauphine
PSL University, Paris, France

Abstract. Neural forecasting methods typically assume Gaussian distributions, focusing on point prediction via MSE minimization. This overlooks heavy-tailed, locally explosive time series where predictive densities exhibit multimodality. We propose a Mixture Density Network with skewed Student-t components for density forecasting. To address extreme event rarity, we develop a dual reweighting strategy with post-hoc recalibration correcting distributional shift. Experiments on noncausal autoregressive processes demonstrate competitive point prediction with well-calibrated uncertainty quantification.

1 Introduction

Forecasting has been a central pursuit in time series analysis for decades. Following the deep learning revolution, neural network-based architectures have emerged as state-of-the-art approaches, ranging from recurrent neural networks [11] to transformer-based models [2] and foundation models [6]. However, most work focuses on point prediction through mean squared error minimization, a paradigm ill-suited for processes exhibiting strong nonlinearities and multimodal predictive densities.

A growing body of literature has emerged on uncertainty quantification, including quantile regression [12], conformal prediction [1], probabilistic forecasting [14], and density forecasting [13]. While density forecasting represents the most comprehensive approach – providing the complete predictive distribution from which all risk metrics can be derived – it has received limited attention in the deep learning community. This gap is particularly pronounced for highly non-Gaussian processes, where predictive densities may exhibit heavy tails, asymmetry, and multimodality.

In this paper, we address these challenges through four contributions: (i) a Mixture Density Network to capture multimodal predictive distributions, (ii) skewed Student-t components for flexible modeling of asymmetry and heavy tails, (iii) a dual reweighting strategy ensuring adequate learning from rare extreme events, and (iv) post-hoc recalibration correcting the distributional shift induced by reweighting.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

2 Methodology

2.1 Mixture Density Networks with Skewed Student-t Components

Mixture Density Networks (MDNs) [4] parameterize conditional densities through neural networks. Standard MDNs use Gaussian mixtures, but their exponentially decaying tails systematically underestimate extreme event probabilities. We instead employ the skewed Student-t distribution [3], which provides both heavy tails through the degrees of freedom parameter and asymmetry through the skewness parameter. The density is given by:

$$f_{\mathrm{ST}}(x;\mu,\sigma,\nu,\lambda) = \frac{2}{\sigma} f_t\left(\frac{x-\mu}{\sigma};\nu\right) F_t\left(\lambda\frac{x-\mu}{\sigma}\sqrt{\frac{\nu+1}{\nu+(x-\mu)^2/\sigma^2}};\nu+1\right), \quad (1)$$

where $f_t(\cdot;\nu)$ and $F_t(\cdot;\nu)$ denote the PDF and CDF of the standard Student-t distribution with ν degrees of freedom, $\mu \in \mathbb{R}$ controls location, $\sigma > 0$ controls scale, $\nu > 0$ controls tail heaviness, and $\lambda \in \mathbb{R}$ controls skewness.

Our conditional density is modeled as:

$$p(X_{t+h}|\mathbf{X}_t) = \sum_{j=1}^{K} \pi_j(\mathbf{X}_t) \cdot \mathrm{ST}(\mu_j(\mathbf{X}_t), \sigma_j(\mathbf{X}_t), \nu_j(\mathbf{X}_t), \lambda_j(\mathbf{X}_t)), \quad (2)$$

where $\mathbf{X}_t = (X_t, X_{t-1}, \ldots, X_{t-L+1})$ denotes a vector of L past observations, and the network predicts location μ_j, scale σ_j, and critically, shape parameters ν_j and λ_j for each of the K mixture components, enabling adaptive tail behavior and asymmetry.

Architecture. Our implementation employs a multilayer perceptron (MLP) with two hidden layers of dimension 128, weight normalization, and ReLU activations. While this simple architecture suffices for our univariate setting, the framework naturally accommodates more sophisticated encoders – such as recurrent neural networks or transformer-based architectures – for capturing complex temporal dependencies in multivariate contexts. The hidden representation feeds into five parallel output heads with constrained activations: softmax for mixture weights $\boldsymbol{\pi} \in \Delta^{K-1}$, softplus for scales $\boldsymbol{\sigma} \in \mathbb{R}_+^K$ and degrees of freedom $\boldsymbol{\nu} \in \mathbb{R}_+^K$, and linear activations for locations $\boldsymbol{\mu} \in \mathbb{R}^K$ and skewness parameters $\boldsymbol{\lambda} \in \mathbb{R}^K$. During training, we minimize the negative log-likelihood. This formulation offers significantly improved numerical stability compared to Tukey g-and-h based approaches [9], avoiding the numerical inverse transform required for density evaluation in the Tukey case. Moreover, the skewed Student-t distribution admits natural multivariate extensions, facilitating future generalization to vector-valued time series.

2.2 Dual Reweighting for Extreme Events

Extreme events are rare yet critical for forecasting performance. Standard training prioritizes abundant normal observations, yielding poor tail modeling. We address this through dual reweighting emphasizing extreme events.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Extreme event detection. We identify extremes using adaptive thresholds from the generalized boxplot [5], which accommodates asymmetry and tail heaviness by fitting a Tukey g-and-h distribution to rank-transformed data. Thresholds are defined as $\ell = Q^{\mathrm{TGH}}_{0.007/2}$ and $u = Q^{\mathrm{TGH}}_{1-0.007/2}$. Extreme events are $\mathcal{E} = \{t : X_t < \ell \text{ or } X_t > u\}$.

Reweighting strategy. We assign weights $\omega = \sqrt{|\mathcal{D}|/|\mathcal{E}|}$ to extremes and 1 otherwise. During training, we apply weights twice: (i) weighted sampling makes extremes ω times more likely in mini-batches, and (ii) weighted loss $\mathcal{L}(\theta) = -\frac{1}{\sum_{t\in\mathcal{B}}\omega_t}\sum_{t\in\mathcal{B}}\omega_t \log p_\theta(X_{t+h}|\mathbf{X}_t)$ emphasizes them during optimization. This yields effective weight $\omega^2 = |\mathcal{D}|/|\mathcal{E}|$, justifying the square-root scaling.

2.3 Post-hoc Calibration

Reweighting introduces distribution shift: the model trains on a distribution emphasizing extremes, causing miscalibration on the true distribution. We correct this using local recalibration [8]. On a held-out calibration set $\mathcal{D}_{\mathrm{cal}}$, we compute PIT values $\mathrm{PIT}_t = \hat{F}(X_{t+h}|\mathbf{X}_t)$, which should be $\mathcal{U}[0,1]$ under perfect calibration. We train XGBoost classifiers to estimate the conditional PIT distribution $\hat{\beta}(\alpha|\mathbf{X}_t) = \mathbb{P}(\mathrm{PIT} \le \alpha|\mathbf{X}_t)$ using I-spline bases for monotonicity. The calibration correction is $c(X_{t+h}|\mathbf{X}_t) = \frac{d\hat{\beta}}{d\alpha}\big|_{\alpha=\hat{F}(X_{t+h}|\mathbf{X}_t)}$, yielding recalibrated density:

$$\hat{p}_{\mathrm{cal}}(X_{t+h}|\mathbf{X}_t) = \frac{c(X_{t+h}|\mathbf{X}_t)\cdot\hat{p}(X_{t+h}|\mathbf{X}_t)}{\int c(z|\mathbf{X}_t)\cdot\hat{p}(z|\mathbf{X}_t)dz}. \tag{3}$$

This restores calibration while preserving improved extreme event modeling.

3 Experiments

To demonstrate the usefulness of density forecasting methods on processes exhibiting locally explosive behavior, we consider noncausal autoregressive models driven by α-stable innovations. Unlike purely causal processes, noncausal processes involve dependence on future innovations through the forward operator F, generating anticipatory dynamics that manifest as locally explosive episodes followed by abrupt mean reversions.

We focus on the simplest specification: the purely noncausal AR(1) model, defined by

$$(1-\psi F)X_t = \varepsilon_t, \quad |\psi| < 1, \quad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha,\beta,\sigma,\gamma), \tag{4}$$

where $\mathcal{S}(\alpha,\beta,\sigma,\gamma)$ denotes an α-stable distribution with tail index $\alpha \in (0,2]$, skewness parameter $\beta \in [-1,1]$, scale parameter $\sigma > 0$, and location parameter $\gamma \in \mathbb{R}$. We set $\psi = 0.9$, $\beta = 0$, $\sigma = 0.5$, $\gamma = 0$, and vary $\alpha \in \{1.0, 1.2, 1.4, 1.6, 1.8\}$. For each configuration, we generate $N_{\mathrm{train}} = N_{\mathrm{cal}} = N_{\mathrm{test}} = 10{,}000$ observations and use $K = 2$ mixture components. A typical realization is illustrated in Figure 1, exhibiting the characteristic locally explosive episodes followed by abrupt mean reversions.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

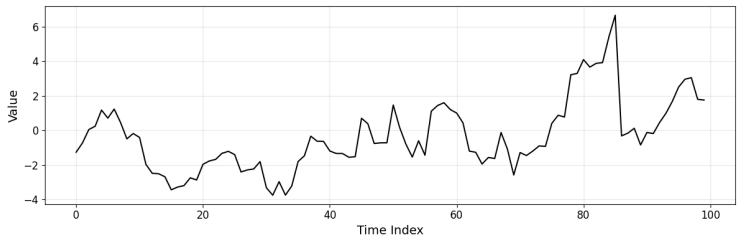

Figure 1: Simulation of $X_t = \psi X_{t+1} + \varepsilon_t, \quad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1.4, 0, 0.5, 0)$

A key feature of this model is the bimodality of its predictive densities in the tails. Conditioning on extreme realizations of X_t, the conditional distribution of X_{t+h} exhibits two distinct modes corresponding to continuation of the explosive trajectory versus reversion toward the mean [7]. We define a *crash* as an observation in the distributional tail that is followed by a sudden drop toward the unconditional mean (zero in our setting).

Point predictions become inadequate when the true predictive density is bimodal, as any single-valued forecast misrepresents the distributional structure (see Figure 2). However, density forecasting typically incurs a cost in predictive accuracy. Our experimental design evaluates whether this trade-off can be mitigated. We assess: (i) whether during unimodal regimes our MDN achieves comparable RMSE to XGBoost and a MLP (with the same architecture and training procedure as our MDN, differing only in the output layer) (ii) whether it provides reliable confidence interval coverage when predictive uncertainty increases in the tails (*i.e.,* when the predictive density becomes bimodal), and (iii) whether crash probabilities are well-calibrated.

We detect bimodality using Hartigan & Hartigan's dip test for unimodality [10], rejecting at the 5% significance level. For unimodal densities, we compute standard $(1-\alpha)$ confidence intervals using the $\alpha/2$ and $(1-\alpha/2)$ quantiles; the MDN point forecast is taken as the mode, enabling direct RMSE comparison. For bimodal densities, we locate the antimode, *i.e.,* the minimum density between the two peaks, and compute separate confidence intervals for each mode. Coverage is assessed by checking whether the realized value falls within at least one mode-specific interval. Crash probabilities are computed as the probability mass assigned to the mode nearest zero (see Figure 2), and calibration is verified by checking that actual crash rates match predicted probabilities within predefined bins (*e.g.,* $[0.1, 0.2]$).

4 Results

Table 1 demonstrates two key findings. First, the MDN achieves competitive point prediction on unimodal densities, consistently outperforming XGBoost and remaining close to the MLP baseline. Second, while point prediction methods exhibit RMSE increases of $3-6\times$ on bimodal densities, the MDN maintains 95%

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Table 1: Simulation Results

Horizon	α	MDN		XGBoost RMSE		MLP RMSE		n_{bimodal}
		Coverage 95% CI (bimodal)	RMSE (unimodal)	unimodal	bimodal	unimodal	bimodal	
$h=1$	1.0	–	13.8000	**12.4906**	–	12.5276	–	0
	1.2	–	4.5687	4.4974	–	**4.1724**	–	0
	1.4	0.9298	1.8617	1.7796	7.0466	**1.7536**	6.7039	242
	1.6	0.9211	1.1030	1.0736	6.5850	**1.0532**	5.9232	114
	1.8	0.8571	0.9008	0.8720	3.0776	**0.8484**	3.0834	28
$h=2$	1.0	0.9211	11.4333	**9.8089**	44.9730	10.8781	45.1919	1014
	1.2	0.9071	2.7313	2.6823	19.1768	**2.4984**	17.5981	861
	1.4	0.9104	1.8007	1.8277	9.4992	**1.7290**	8.9494	614
	1.6	0.9122	1.4470	1.4412	6.0608	**1.3695**	5.6203	296
	1.8	0.8416	1.1576	1.1837	4.0994	**1.1097**	3.4856	101
$h=3$	1.0	0.9080	15.4491	12.7823	57.6449	**12.6783**	57.1396	837
	1.2	0.9022	3.0787	3.1773	18.6831	**3.0073**	17.7424	1196
	1.4	0.9140	2.1048	2.1546	9.3995	**2.0139**	8.6925	907
	1.6	0.9424	1.6533	1.7060	6.1715	**1.5830**	5.5164	451
	1.8	0.9038	1.3378	1.3831	4.0401	**1.2974**	3.3611	156

Note : RMSE separated by density type. MDN coverage for bimodal cases. Bold shows best unimodal RMSE.

confidence interval coverage between 0.90 and 0.95. This degradation is expected: when the true predictive density is bimodal, any point forecast misrepresents the distributional structure. This confirms that density forecasting provides competitive RMSE during unimodal regimes while delivering well-calibrated uncertainty quantification precisely when point forecasts become unreliable.

Table 2: Crash Probability Calibration.

h	α	Predicted Probability Interval			
		[0.1,0.2)	[0.2,0.3)	[0.3,0.4)	[0.4,0.6)
1	1.4	0.168	–	–	–
	1.6	0.207	–	–	–
2	1.0	0.179	0.240	0.476	–
	1.2	0.234	0.237	0.200	0.222
	1.4	0.221	0.310	0.613	0.549
	1.6	–	0.342	–	–
	1.8	–	0.323	–	0.381
3	1.0	0.241	0.301	0.486	–
	1.2	0.270	0.273	0.385	0.411
	1.4	0.446	0.365	0.325	0.406
	1.6	–	0.505	0.453	0.473
	1.8	–	–	0.450	0.582

Note : Cells report actual crash rates for bimodal densities grouped by predicted probability. Perfect calibration corresponds to interval midpoints. "–" indicates insufficient observations.

Figure 2: Examples of bimodal predictive densities ($\alpha = 1.4$, $h = 3$). Top: continuation scenario. Bottom: crash scenario.

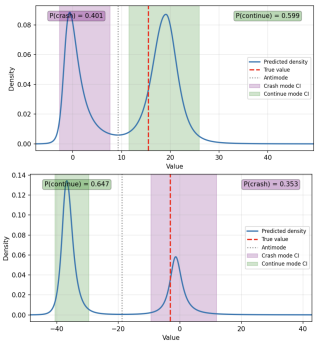

Table 2 demonstrates that the model successfully discriminates between crash risk levels. Higher predicted probabilities systematically correspond to higher realized crash rates, confirming that the MDN captures meaningful structure in the predictive distribution. The model produces well-distributed predictions across probability bins, indicating effective uncertainty quantification. While

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

some bins exhibit deviations from perfect calibration, the consistent ordering validates the model's ability to rank extreme event likelihood, a critical capability for risk management applications.

5 Conclusion

Our experiments confirm that the proposed MDN successfully addresses all three objectives: (i) it achieves competitive RMSE with XGBoost and MLP during unimodal regimes, demonstrating that density forecasting need not sacrifice point prediction accuracy; (ii) it maintains reliable 95% confidence interval coverage (0.90–0.95) precisely when predictive densities become bimodal and point forecasts fail; and (iii) the model effectively discriminates between crash risk levels, with higher predicted probabilities systematically corresponding to higher realized crash rates.

References

[1] A. N. Angelopoulos and S. Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification, 2022. arXiv:2107.07511.

[2] A. F. Ansari, L. Stella, C. Turkmen, X. Zhang, P. Mercado, H. Shen, O. Shchur, S. S. Rangapuram, S. P. Arango, S. Kapoor, J. Zschiegner, D. C. Maddix, H. Wang, M. W. Mahoney, K. Torkkola, A. G. Wilson, M. Bohlke-Schneider, and Y. Wang. Chronos: Learning the language of time series. *Transactions on Machine Learning Research*, 2024.

[3] A. Azzalini and A. Capitanio. 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, 2003.

[4] C. M. Bishop. Mixture density networks. *Aston University Technical Report*, 1994.

[5] C. Bruffaerts, V. Verardi, and C. Vermandele. A generalized boxplot for skewed and heavy-tailed distributions. *Statistics & Probability Letters*, 2014.

[6] A. Das, W. Kong, R. Sen, and Y. Zhou. A decoder-only foundation model for time-series forecasting. In *ICML*, 2024.

[7] G. de Truchis, S. Fries, and A. Thomas. Forecasting extreme trajectories using seminorm representations. Working paper, 2025.

[8] B. Dey, J. A. Newman, B. H. Andrews, R. Izbicki, A. B. Lee, D. Zhao, M. M. Rau, and A. I. Malz. Re-calibrating photometric redshift probability distributions using feature-space regression. In *NeurIPS Workshop on Machine Learning and the Physical Sciences*, 2022.

[9] A. P. Guillaumin and N. Efremova. Tukey g-and-h neural network regression for non-Gaussian data, 2024. arXiv:2411.07957.

[10] J. A. Hartigan and P. M. Hartigan. The dip test of unimodality. *The Annals of Statistics*, 13(1):70–84, 1985.

[11] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[12] R. Koenker and G. Bassett. Regression quantiles. *Econometrica*, 46:33–50, 1978.

[13] J. Rothfuss, F. Ferreira, S. Walther, and M. Ulrich. Conditional density estimation with neural networks: Best practices and benchmarks, 2019. arXiv:1903.00954.

[14] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020.