A Score-Driven approach to Temperature Modelling in High Dimensions

Simone Serafini [1] Giacomo Bormetti [2] Francesco Calvori [3]

[1]Departments of Mathematics, University of Bologna

[2]Departments of Economics, University of Pavia

[3]Leithà, Unipol Group

Quantitative Methods in Finance - University of Technology Sydney (UTS)
December 17, 2024

Preview

- Climate (physical) Risk $\longrightarrow$ Increasing frequency and severity of adverse weather events
- Financial instruments such as weather derivatives and parametric insurance can be used to mitigate this risk. Is the modeling of these contracts adequate?
- Assessing climate risk requires high-quality climate data

Overview

- Study the time series of weather variables (maximum daily temperature) at N locations (**Univariate Modelling**)
- Generate distributional scenarios of climate variables taking into account the spatial dependence (**Multivariate Modelling**)
- Valuation of parametric insurance written on weather variables (**Valuation**)

Motivation and Contribution

- **Univariate Modelling:** Possibility of departing from the Normal assumption by considering distributions with a higher probability of extreme events. We rely on Score-Driven models that are robust to extreme events.

- **Multivariate Modelling:** When the cross-sectional dimension N is large, a parametric approach to estimate dependence leads to the *curse of dimensionality*. We tackle this issue by utilizing factor Copula model that reduce the dimension of parameter space and have closed-form likelihoods.

- **Multivariate Modelling:** Factor Copula model allows us to model the tail dependence of variables

- **Valuation:** Estimation of Risk Measure for a Portfolio of Parametric Insurance contracts

Temperature Modelling

The data generating process is assumed to be (2-step approach Šaltytė Benth and Benth, 2012)

$$T_t = S_t + Y_t$$

where

- $S_t = \beta_0 + \beta_1 t + \sum_{r=1}^{R} \beta_r^{\sin} \sin(2r\pi wt) + \sum_{r=1}^{R} \beta_r^{\cos} \cos(2r\pi wt)$
- $Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + \xi_t$
- S_t is a deterministic function that take into account the **trend** and the **seasonality**
- Y_t is an autoregressive equation that describe the **stochastic** part of the process $\rightarrow$ **Dynamic Score-Driven (DCS) Modelling**

Stochastic Modelling - Starting Point

The stochastic component of the process $Y_t = T_t - \hat{S}_t$ has been modelled in literature with an AR(p)-EGARCH(1,1) dynamic with external seasonal covariates (Erhardt and Engler, 2018)

$$Y_t = \sum_{p=1}^{P} \phi_p (Y_{t-p}) + \xi_t$$

The idiosyncratic term has the following EGARCH dynamic

$$\begin{aligned}
\xi_t &= \sigma_t \times \epsilon_t, \\
\epsilon_t &\sim \mathcal{N}(0,1) \\
\log\left(\sigma_t^2\right) &= \omega + \gamma \left(|\epsilon_t| - \sqrt{\frac{2}{\pi}} \right) + \alpha \epsilon_t + \beta \log\left(\sigma_{t-1}^2\right) \\
&\quad + \sum_{q=1}^{Q} \delta_q^{\sin} \times \sin\left(2q\pi w t\right) + \sum_{q=1}^{Q} \delta_q^{cos} \times \cos\left(2q\pi w t\right)
\end{aligned}$$

Score-Driven (SD) models

- SD models (Creal, Koopman, and Lucas, 2013, Harvey, 2013) are a class of *observation-driven* models
- They provide a general framework to adjust the parameters of the model based on past observations

We assume that the observations $y_t \in \mathbb{R}^p$ are generated by the conditional density $p(y_t|f_t, \Theta)$. The distribution parameters follow the **update rule**

$$f_{t+1} = \omega + Bf_t + AS_t\nabla_t$$

∇_t is the score of the *conditional log-likelihood* $\log p(y_t|f_t, \Theta)$:

$$\nabla_t = \frac{\partial \log p(y_t|f_t)}{\partial f_t}, \quad S_t = g(\mathcal{H}_{t|t-1}), \quad \mathcal{H}_{t|t-1} = \mathbb{E}[\nabla_t \nabla_t' | \mathcal{F}_{t-1}]$$

Score-Driven Model: Beta-t-EGARCH

- Suppose that the conditional distribution of the adjusted Temperature process Y_t is t_ν
- Score-Driven model (Beta-t-EGARCH) with external seasonal covariates has the following dynamic

$$
\begin{aligned}
Y_t &= \sum_{i=1}^{p} \phi_i Y_{t-i} + \xi_t \\
\xi_t &= e^{f_t/2} \epsilon_t \\
f_t &= \omega + \beta f_{t-1} + \alpha \left(\frac{(\nu+1) Y_{t-1}^2}{\nu e^{f_{t-1}} + Y_{t-1}^2} - 1 \right) \\
&\quad + \sum_{k=1}^{Q} \delta_k^{\sin} \sin(2k\pi\omega t) + \sum_{k=1}^{Q} \delta_k^{\cos} \cos(2k\pi\omega t)
\end{aligned}
$$

Modelling Spatial Dependence: Factor Copulas

- After obtaining residuals from the two-step estimation at each location $(\epsilon_{1,t}, \ldots, \epsilon_{N,t})$, we incorporate spatial dependence using a factor copula model.

- We can generate correlated samples through the Copula function

$$F(\epsilon_{1,t}, \ldots, \epsilon_{N,t}) = C(u_{1,t}, \ldots, u_{N,t})$$

  where $u_{i,t} = F_i(\epsilon_{i,t})$ is the Probability Integral Transform (PIT)

- To avoid the *curse of dimensionality* we employ a Factor Copula model. This framework allows us to generate $u_{i,t}$ with a specific dependence structure. We can generate correlated multivariate samples in the following way

$$(\epsilon_{1,t}, \ldots, \epsilon_{N,t}) = \left(F_1^{-1}(u_{1,t}), \ldots, F_N^{-1}(u_{N,t})\right),$$

Modelling Spatial Dependence

- We apply the Factor Copula Model of Opschoor et al., 2021 that has the property to have closed MLE
- Copula function is obtained through the auxiliary variable $\boldsymbol{X}$

$$
\begin{aligned}
u_{i,t} &= T\left(x_{i,t};\nu\right), \quad i = 1, \cdots, N \\
x_{i,t} &= \sqrt{W_t}\left(\widetilde{\boldsymbol{\lambda}}_i' \mathbf{z}_t + \sigma_i \epsilon_{i,t}\right) \\
\mathbf{z}_t &\sim \text{ iid } N\left(\mathbf{0}, \mathbf{I}_k\right), \quad \epsilon_{i,t} \sim iid\mathcal{N}(0,1) \\
W_t &\sim \text{iid } IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \quad W_t \perp \mathbf{z}_t \perp \epsilon_{i,t}
\end{aligned}
$$

- The correlation matrix $\boldsymbol{R}$ satisfy the following condition

$$
\boldsymbol{R} = \tilde{\boldsymbol{L}}'\tilde{\boldsymbol{L}} + \boldsymbol{D}
$$

where $\tilde{\boldsymbol{L}} = (\widetilde{\boldsymbol{\lambda}}_1, \ldots, \widetilde{\boldsymbol{\lambda}}_N)$ e $\boldsymbol{D} = \text{diag}(\sigma_1^2, \ldots, \sigma_N^2)$

Factor Copula Model

- The assumption underlying the Factor Copula model is that our N variables can be clustered into G spatial groups.

- Specific factors describe the correlation between and within groups

- We use the MF and the MF-LT specifications for the factor loadings matrix

$$\tilde{\boldsymbol{L}}^\top = \begin{pmatrix} \tilde{\lambda}_{1,1} & 0 & 0 & 0 \\ \tilde{\lambda}_{1,2} & \tilde{\lambda}_{2,2} & 0 & 0 \\ \tilde{\lambda}_{1,3} & \tilde{\lambda}_{2,3} & \tilde{\lambda}_{3,3} & 0 \\ \tilde{\lambda}_{1,4} & \tilde{\lambda}_{2,4} & \tilde{\lambda}_{3,4} & \tilde{\lambda}_{4,4} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

- Spatial Groups: Geographic Approach vs **K-means** Clustering

Dataset and Empirical Application

- 20 Weather station from US Airport (from 1 January 1962 to 31 December 2012). $T = 18250$
- ERA5 (Reanalysis data) of 107 Italian provinces (from 1 January 1976 to 31 December 2021). $T = 16790$
- We determine the best model based on the AIC using this metric:

$$\Delta_{\text{AIC}} = \frac{\text{AIC}_{\text{Beta-t-EGARCH}} - \text{AIC}_{\text{EGARCH}}}{\text{AIC}_{\text{EGARCH}}} \times 100$$

AIC Result I

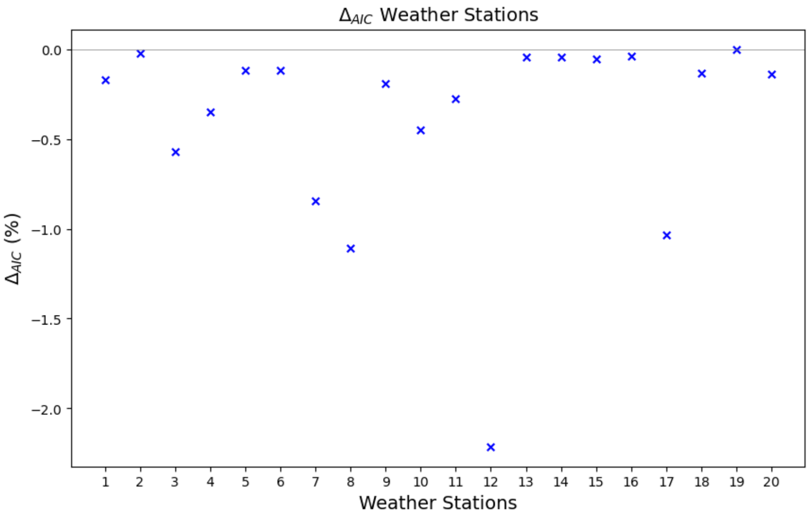

Figure: Δ_{AIC} for weather stations data

AIC Result II

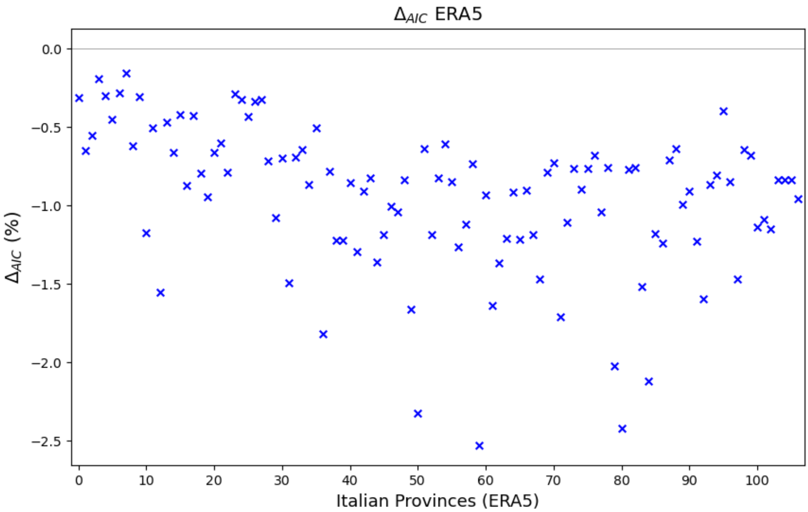

Figure: Δ_{AIC} for Italian provinces

K-means Spatial Group

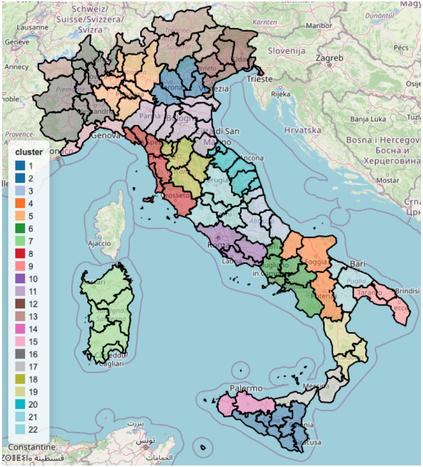

Figure: Clusters of standardized residuals

Factor copula model estimates

Table: Factor copulas on G=22, Parametric Residuals

Model	ν_C	LogL	AIC	# par
Gaussian factor copulas				
MF		1,221,016	-2,441,944	44
MF-LT		1,357,531	-2,714,556	253
t-factor copulas				
MF	9.08	1,302,560	-2,605,030	45
MF-LT	9.55	**1,430,323**	**-2,860,139**	254

Factor Copula Model estimaets

Table: Geographic Clustering, Parametric Residuals

Model	ν_C	LogL	AIC	# par
Gaussian factor copulas				
MF		1,015,086	-2,030,096	38
MF-LT		1,123,547	-2,246,714	190
t-factor copulas				
MF	8.92	1,152,581	-2,305,085	39
MF-LT	9.85	**1,255,820**	**-2,511,258**	191

Parametric Insurance Pricing

- We want to price Heat Wave Parametric Insurance (Larsson, 2023)
- We define the event

$$HW(n, a, \tau, T_m) = \{T_{mt} \geq a \quad \text{for } n \text{ consecutive days } t,\ t \in \tau\},$$

- The contract pays K if the event HW happen. The evaluation of the contract is

$$V_t = e^{-r(t-t_0)} K \zeta^{\mathbb{P}}(t, n, a, \tau, T_m)$$

where

$$\zeta^{\mathbb{P}}(t, n, a, \tau, T_m) = \mathbb{P}(HW(n, a, \tau, T_m) \geq 1),$$

Parametric Insurance Pricing - Example

- We consider a Portfolio with $N_c = 20$ locations of Italian Provinces. The HW definition takes as parameters $n = 5$ consecutive days and $a = 98th\text{-}historic\ Maximum\ Temperature\ percentile\ of\ the\ location$

- Actuarial perspective, we examine the aggregate "loss" of the portfolio

$$S = \sum_{i=1}^{Nc} \Phi_i$$

  where each Φ_i represents an insurance contract with a fixed claim size $K = 100$. The Loss is influenced only by the likelihood of the event HW associated with each location. We generate $M = 100,000$ Monte Carlo path to build the distribution of S

Parametric Insurance Pricing - Results

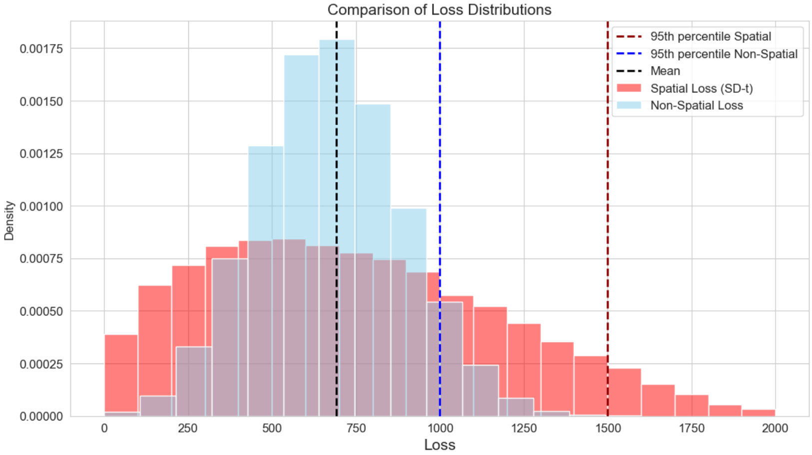

Figure: Spatial vs Non-Spatial Loss

Parametric Insurance Pricing - Results

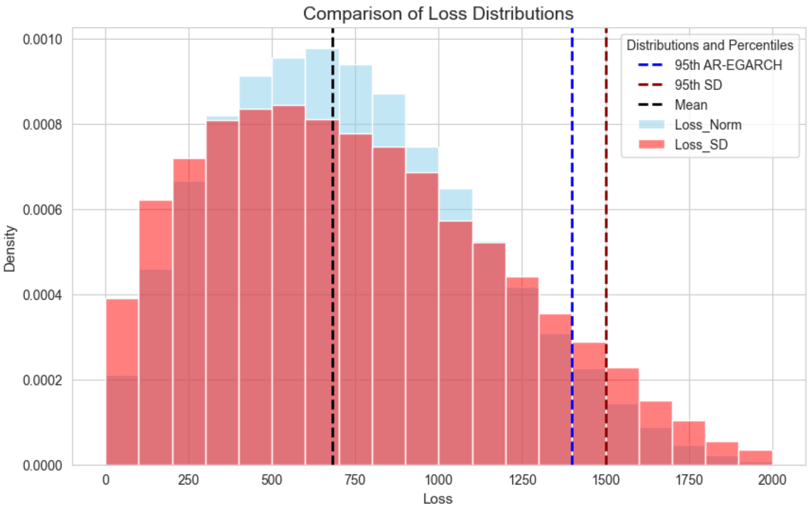

Figure: Spatial vs Non-Spatial Loss

Study the severity of HW - EDD index (Preliminary)

- New extreme climate indexes through the use of Reanalysis data are emerging (ACI in US, E3CI in Europe)
- One challenge of parametric insurance is creating indices that strongly correlate with losses (basis risk)
- We define the Extreme Degree Days (EDD) for the summer season as follows:

$$\text{EDD} = \max(T_m - \hat{T}_m^{95\text{th}}, 0)$$

where $\hat{T}_m^{95\text{th}}$ represents the 95th percentile of maximum temperature during the summer period based on the reference level 1981 - 2010

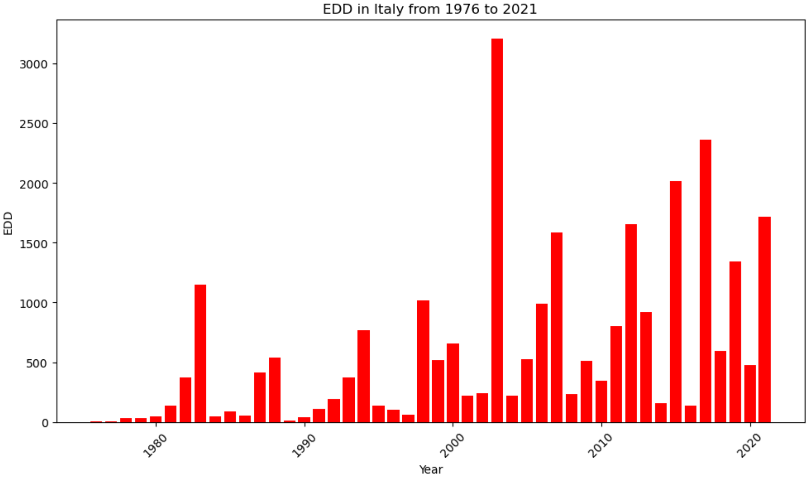

Figure: cEDD Italy

Conclusion

- In light of the increasing importance of climate risk, we propose a methodology for evaluating the impact of physical risk on financial and insurance products, considering the spatial dependence of weather variables
- An important aspect is the flexibility of the two-stage procedure, where we first model time series at a specific location and then establish non-trivial spatial dependencies through factor copula model
- The methodology can be expanded by exploring different models for various weather variables and incorporating the SD framework for the correlation

Thank you for your attention!

Appendix - Score Driven Correlation

- The factor loadings have the following DCS dynamic

$$\lambda_{i,t+1} = \omega_i + \alpha \frac{\partial \log \mathbf{c}_t \left(\mathbf{x}_t ; \mathbf{R}_t, \Theta \right)}{\partial \lambda_{i,t}} + \beta \lambda_{i,t}$$

- The copula can be Gaussian, t, skew-t (Oh and Patton, 2023)

Table: Factor copulas on G=22, Parametric Residuals

Model	ν_C	LogL	AIC	# par
Gaussian factor copulas				
MF		1,221,016	-2,441,944	44
MF-LT		1,357,531	-2,714,556	253
MF-LT (SD)		1,357,606	-2,714,702	255
t-factor copulas				
MF	9.08	1,302,560	-2,605,030	45
MF-LT	9.55	1,430,323	-2,860,139	254
MF-LT (SD)	8.41	**1,440,365**	**-2,880,218**	256

EDD risk measure with $N_c = 20$ Italian Provinces

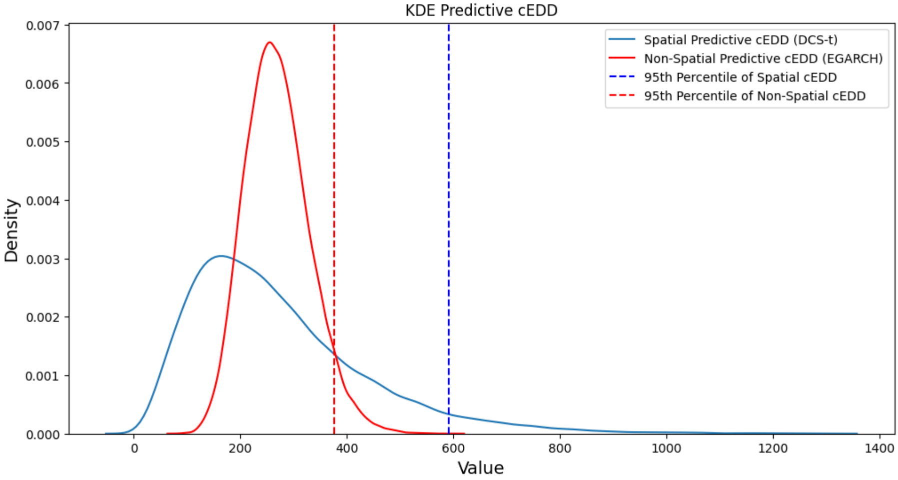

Figure: EDD Spatial vs Non-Spatial

Disclaimer

The views expressed in this note are the only responsibility of the authors and do not represent in any way those of the authors' current employer. All errors are the only responsibility of the authors

References I

Creal, Drew, Siem Jan Koopman, and André Lucas (2013). "GENERALIZED AUTOREGRESSIVE SCORE MODELS WITH APPLICATIONS". In: *Journal of Applied Econometrics* 28.5, pp. 777–795. DOI: https://doi.org/10.1002/jae.1279.

Erhardt, Robert and David Engler (2018). "An Extension of Spatial Dependence Models for Estimating Short-Term Temperature Portfolio Risk". In: *North American Actuarial Journal* 22.3, pp. 473–490. DOI: 10.1080/10920277.2018.1444496.

Harvey, Andrew C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Society Monographs. Cambridge University Press. DOI: 10.1017/CBO9781139540933.

References II

Larsson, Karl (2023). "Parametric heat wave insurance". In: *Journal of Commodity Markets* 31, p. 100345. ISSN: 2405-8513. DOI: https://doi.org/10.1016/j.jcomm.2023.100345. URL: https://www.sciencedirect.com/science/article/pii/S2405851323000351.

Oh, Dong Hwan and Andrew J. Patton (2023). "Dynamic factor copula models with estimated cluster assignments". In: *Journal of Econometrics* 237.2, Part C, p. 105374. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2022.07.012. URL: https://www.sciencedirect.com/science/article/pii/S0304407622002135.

References III

Opschoor, Anne et al. (2021). "Closed-Form Multi-Factor Copula Models With Observation-Driven Dynamic Factor Loadings". In: *Journal of Business & Economic Statistics* 39.4, pp. 1066–1079. DOI: 10.1080/07350015.2020.1763806. eprint: https://doi.org/10.1080/07350015.2020.1763806. URL: https://doi.org/10.1080/07350015.2020.1763806.

Šaltytė Benth, Jūratė and Fred Espen Benth (2012). "A critical view on temperature modelling for application in weather derivatives markets". In: *Energy Economics* 34.2, pp. 592–602. ISSN: 0140-9883. DOI: https://doi.org/10.1016/j.eneco.2011.09.012. URL: https://www.sciencedirect.com/science/article/pii/S0140988311002283.