A Markov-switching Generalized Additive Model for Compound Poisson Processes, with Applications to Operational Losses Models — Supplementary Material

J. Hambuckers, T. Kneib, R. Langrock and A. Silbersdorff

Simulation study — coverage of the 99.9% quantile

In this section, we look at the in-sample and out-of-sample coverages of the Monte Carlo 99.9% quantile estimates.

In-sample coverage

We first look at the in-sample coverage. To do so, for each sample of the Monte Carlo simulations presented in Section 3 of the paper, we use the estimated parameters and the decoded states (obtained with the Viterbi algorithm) to generate 100,000 realizations of the total loss (obtained by convolution of the frequency and severity realizations) at each point in time. The empirical 99.9% quantile of these realizations gives our final estimate. Then, we count the observed losses that are larger than their respective estimated quantile. Using estimated parameters $\hat{\theta}$ and a given sample $\{l_t\}_{t=1,...,T}$, we are interested in

$$p = T^{-1} \sum_{t=1}^{T} \mathbf{1}(\hat{Q}_{0.999}(L_t; \hat{\Theta}) < l_t), \tag{1}$$

where T is the size of the time series, $\mathbf{1}(\cdot)$ is an indicator function taking value one when the condition in parentheses is met, zero otherwise, and $\hat{Q}_{0.999}(L_t; \hat{\theta})$ is the Monte Carlo estimate of the 99.9% quantile of the total loss, obtained with estimated parameters $\hat{\Theta}$. We call this quantity the proportion of exceedances, which corresponds to one minus the coverage probability [see Christoffersen, 1998, for additional theoretical explanations]. A similar calculation is performed using the true model and the true states, to obtain Monte Carlo estimates of the proportion of exceedances (i.e. replacing $\hat{\Theta}$ by Θ). In that case, the proportion of exceedances is denoted p_0. Taking the average of the proportions computed over the B samples of the Monte Carlo simulations, we obtain our final estimates, $\bar{p}$ and $\bar{p}_0$. Table 1 gives the obtained results for the different values of the parameters used in

the simulation study. We expect values close to 0.1%. In addition, we also compute the proportion of samples where we observe at least once a total loss larger than its estimated quantile. These proportions are denoted ψ and ψ_0 for the estimated model and the true model, respectively. They can be used as critical values, for a given sample size, to assess the correct specification of a MS-GAMLSS model.

Regarding the DGP with a constant γ, our results show proportions of exceedances not far from the expected 0.1%, both with the true model and the estimated model. The short length of the time series, however, is the source of an important variability. Results seem a bit better with $T = 100$ and when $\beta_{\lambda,0}$ increases (except for $x_t^\lambda = x_t^\sigma$ and $T = 100$). Worst results are obtained when the number of events per time period is the lowest (i.e. when $\beta_{\lambda,0} = 2$). The proportion of samples with at least one total loss larger than the estimated quantile varies between 2% and 8.5% for $T = 50$, and between 6% and 11.5% for $T = 100$. Regarding the results with γ depending on covariates, we also observe acceptable rates. Worst results are obtained for $T = 50$ and $x_t^\lambda = x_t^\sigma = x_t^\gamma$. For that case, we overestimate the quantile (we observe no rejections). The high collinearity between explanatory variables can explain some of these observations. Besides, in the other cases, we are close to the expected proportion of 0.1%.

Lastly, we perform the same exercise using the model estimated on UniCredit data. Using the covariates' value of the original time series and the estimated parameters, we generate 1000 (parametric) bootstrap time series of the state, the frequency and the severity processes. Then, we re-estimate the parameters and use the Viterbi algorithm to decode the states. These quantities are used to compute Monte Carlo estimates of the 99.9% quantile and the related proportions of exceedances. We observe a proportion of exceedances close to 0.1% and a low proportion of samples with at least one breach (see Table 2). These results suggest that on such a short time series, it is quite unlikely to observe breaches and that these events should be considered as indicators of a potential misspecification.

$T = 50$	$x_t^\lambda = x_t^\sigma$				$x_t^\lambda \neq x_t^\sigma$			
$\beta_{\lambda,0}$	2	3	4	4*	2	3	4	4*
$\bar{p}$	0.04	0.05	0.13	0.00	0.08	0.05	0.08	0.02
$\bar{p}_0$	0.09	0.08	0.13	0.11	0.17	0.1	0.1	0.13
ψ	2.0	2.5	6.5	0.0	4.0	2.5	4.0	1.0
ψ_0	4.5	4.0	6.5	5.5	8.5	5.0	5.0	6.5
$T = 100$	$x_t^\lambda = x_t^\sigma$				$x_t^\lambda \neq x_t^\sigma$			
$\beta_{\lambda,0}$	2	3	4	4*	2	3	4	4*
$\bar{p}$	0.065	0.08	0.065	0.02	0.085	0.09	0.11	0.13
$\bar{p}_0$	0.08	0.08	0.065	0.09	0.12	0.09	0.12	0.07
ψ	6.5	8.0	6.0	2.0	8.5	9.0	10.5	11.0
ψ_0	7.0	7.5	6.5	8.5	11.5	8.5	11.0	7.0

Table 1: In-sample proportions of exceedances (in %) obtained with the estimated model ($\bar{p}$) and the true model ($\bar{p}_0$). Data are the ones of the simulation study. ψ and ψ_0 give the proportion of samples with at least one observed loss larger than its Monte Carlo estimate of the 99.9% quantile.

BS	$\bar{p}$	$\tilde{p}$	$\bar{p}_0$	$\tilde{p}_0$	ψ	ψ_0	$\bar{R}$	$\tilde{R}$
$T = 38$	0.05	0	0.07	0	1.9	2.5	94.4	97.4

Table 2: Bootstrap in-sample proportions of exceedances (in %). Data have been generated using the model estimated on the UniCredit data. The number of resamples is $B = 1000$. $\bar{p}$ (resp. $\tilde{p}$) and $\bar{p}_0$ (resp. $\tilde{p}_0$) denote mean (resp. median) estimates over the resamples, obtained with the estimated model and the true model. $\bar{R}$ and $\tilde{R}$ denote mean and median proportions of correctly classified decoded states, respectively.

Out-of-sample coverage

Here, we look at out-of-sample proportions of exceedances. It allows to assess the correctness of the quantile estimates, when we also have to predict the state. We use one-step

ahead predictions of the state probabilities, obtained by

$$\hat{\alpha}_{T+1} = (\hat{\alpha}_T / \hat{\alpha}_T \mathbf{1})\hat{\Gamma}$$

where $\mathbf{1}$ is a vector of ones, $\hat{\alpha}_T$ is the estimated forward probability at time T and $\hat{\Gamma}$ is the estimated transition matrix [Zucchini et al., 2016]. With this formula, we obtain easily our predicted state probabilities at time $T+1$. Then, replacing T by $T+1$ (but keeping the same estimates of Γ), we are able to repeat this prediction for $T+2, T+3, \ldots, T+H$. Here, we choose $H = 1000$ and extend the first 100 simulated samples generated in our simulation study with 1000 new observations. Then, the same operations as in the in-sample case are executed: using the predicted states and the estimated parameters, we generate 100,000 values of the total loss and use the 99.9% empirical quantile as an estimate. This operation is repeated for the H additional points in the sample, as well as with the true parameters and the true states. We only consider the second data generating process of the simulation, where $x_t^{\lambda} \neq x_t^{\sigma}$.

Results are displayed in Table 3. For the true model, we are very close to 0.1%. Now, the length of the time series is sufficient to reduce the variability of the (true) proportion of exceedances. Regarding the proportion of exceedances obtained with the estimated model, we observe values way higher than 0.1%. For small samples (i.e. $T = 50$ and $\beta_{\lambda,0} < 4$), we observe mean values higher than 5%. However, when the sample size increases, the proportion of exceedances decreases and becomes closer to 0.1%. We suspect that bad out-of-sample proportion of exceedances could be mainly due to bad estimates of the tpm, resulting in bad state prediction. To investigate this conjecture, we first compute the average and median out-of-sample correct classification rates (column $\bar{R}$ and $\tilde{R}$ in Table 3). We see that low rates are associated with proportion of exceedances far from 0.1%. Especially, in the small sample case, the classification rates are very bad (around 65%), explaining the bad coverage. Computing the correlation coefficient between the absolute estimation error of the tpm' diagonal elements and this rate, we find values between -0.13 and -0.81. It could indicate that small errors on the tpm tend to have big consequences on the classification rate, thus on the proportion of exceedances. A way of improving the out-of-sample coverage would be therefore to estimate the parameters of the model on samples with a large number of events per time period. Results with γ depending on covariates are similar. For $\beta_{\lambda,0} = 4$, we observe bigger proportions of exceedances with $T = 50$ and 100, compared to results with a constant γ. These results emphasize that when we increase the complexity of the dimension structure, a larger number of observations per time period is needed to achieve a quality of the prediction that is similar to the one of less complex models. Regarding the classification rate, though, few differences are observed.

$T = 50$			$x_t^\lambda \neq x_t^\sigma$					$T = 100$			$x_t^\lambda \neq x_t^\sigma$				
$\beta_{\lambda,0}$	$\bar{p}$	$\tilde{p}$	$\bar{p}_0$	$\tilde{p_0}$	$\bar{R}$	$\tilde{R}$	ρ	$\beta_{\lambda,0}$	$\bar{p}$	$\tilde{p}$	$\bar{p}_0$	$\tilde{p_0}$	$\bar{R}$	$\tilde{R}$	ρ
2	23.7	20.7	0.093	0.1	56.1	56.4	-0.18	2	19.4	17.55	0.1	0.1	57.9	57.5	-0.24
3	8.6	6.7	0.1	0.1	65.2	65.9	-0.50	3	7.6	7.0	0.10	0.1	67.4	66.9	-0.13
4	1.3	0.9	0.12	0.1	91.6	92.5	-0.81	4	0.8	0.7	0.11	0.1	93.1	93.6	-0.47
4*	7.8	5.7	0.09	0.1	87.7	93.5	-0.73	4*	5.6	4.7	0.11	0.1	0.919	94.15	-0.86

Table 3: Out-of-sample mean and median proportions of exceedances (in %) obtained with the estimated model and the true model. $\bar{R}$ and $\tilde{R}$ give the mean and median proportions of correctly predicted states, respectively. ρ is the (Pearson) correlation coefficient between the correct classification rate and the total absolute error on the diagonal elements of the tpm.

Simulation study — additional Monte Carlo bands

Scenario 1

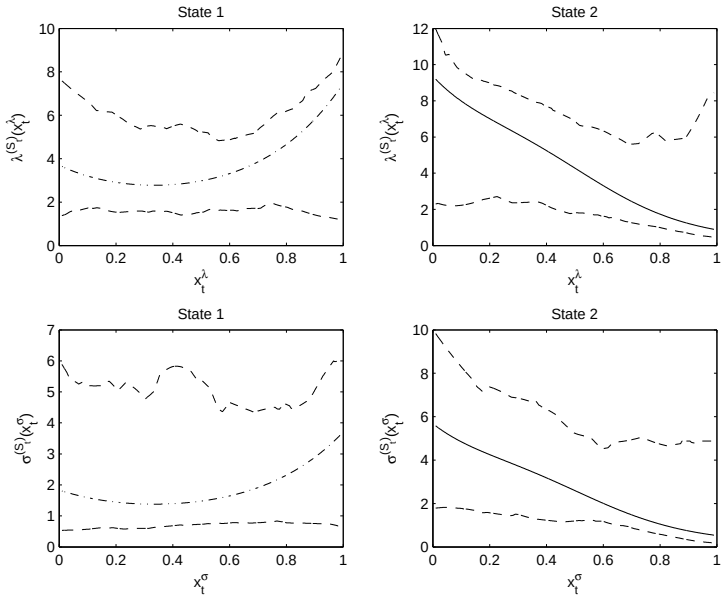

Figure 1: 95% coverage bands (dashed) for $\lambda^{(j)}(x_t^\lambda)$ (top) and $\sigma^{(j)}(x_t^\sigma)$ (bottom) when $T = 50$ and $\beta_{\lambda,0} = 2$. Dashed dotted (resp. solid): true function for $j = 1$ (resp. $j = 2$).

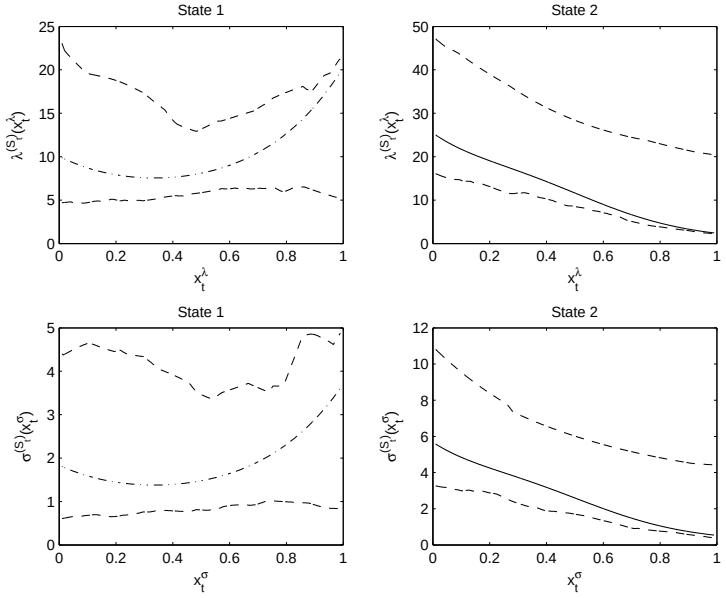

Figure 2: 95% coverage bands (dashed) for $\lambda^{(j)}(x_t^\lambda)$ (top) and $\sigma^{(j)}(x_t^\sigma)$ (bottom) when $T = 50$ and $\beta_{\lambda,0} = 3$. Dashed dotted (resp. solid): true function for $j = 1$ (resp. $j = 2$).

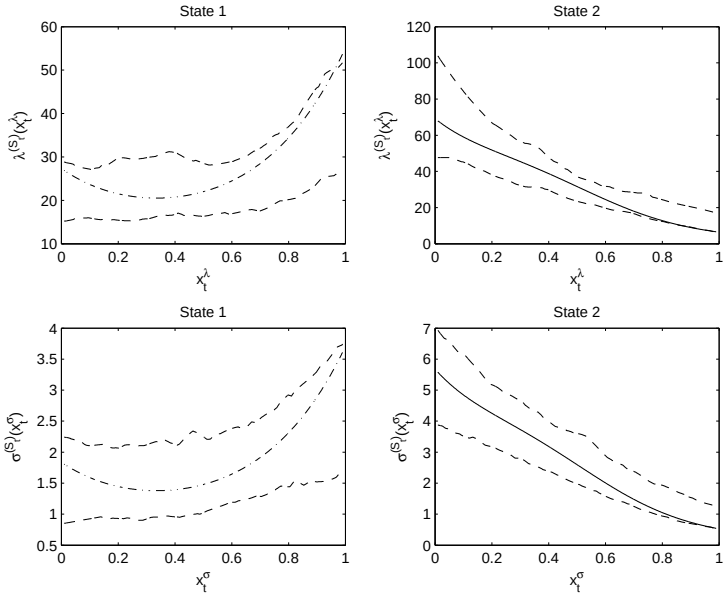

Figure 3: 95% coverage bands (dashed) for $\lambda^{(j)}(x_t^\lambda)$ (top) and $\sigma^{(j)}(x_t^\sigma)$ (bottom) when $T = 50$ and $\beta_{\lambda,0} = 4$. Dashed dotted (resp. solid): true function for $j = 1$ (resp. $j = 2$).

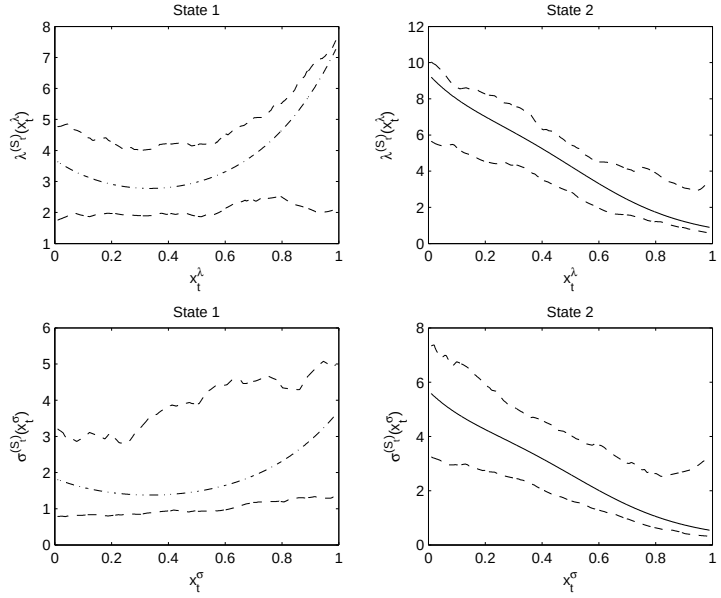

Figure 4: 95% coverage bands (dashed) for $\lambda^{(j)}(x_t^\lambda)$ (top) and $\sigma^{(j)}(x_t^\sigma)$ (bottom) when $T = 100$ and $\beta_{\lambda,0} = 2$. Dashed dotted (resp. solid): true function for $j = 1$ (resp. $j = 2$).

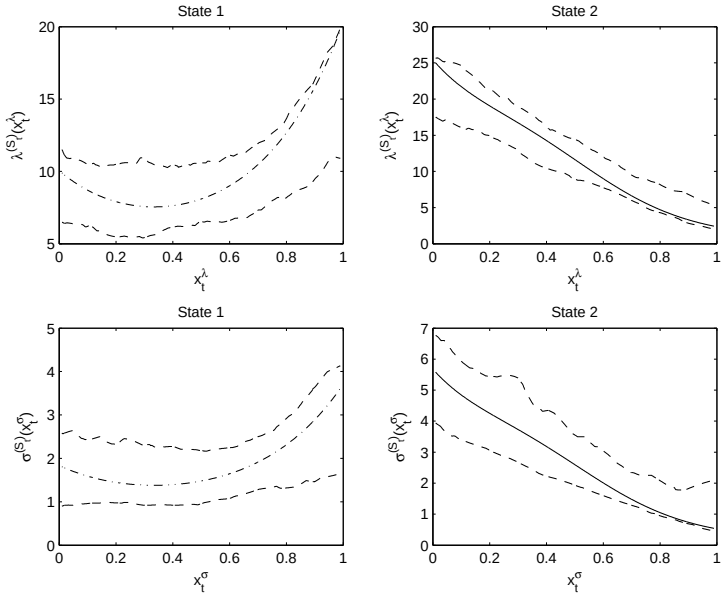

Figure 5: 95% coverage bands (dashed) for $\lambda^{(j)}(x_t^\lambda)$ (top) and $\sigma^{(j)}(x_t^\sigma)$ (bottom) when $T = 100$ and $\beta_{\lambda,0} = 3$. Dashed dotted (resp. solid): true function for $j = 1$ (resp. $j = 2$).

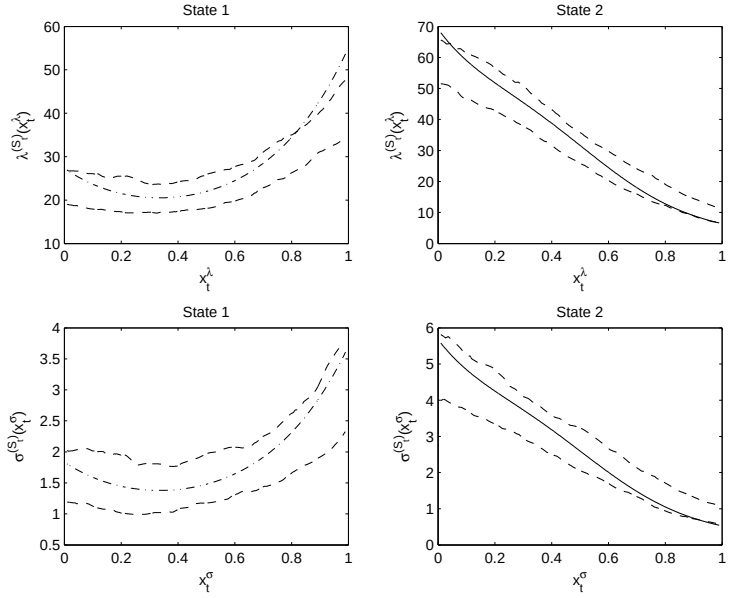

Figure 6: 95% coverage bands (dashed) for $\lambda^{(j)}(x_t^\lambda)$ (top) and $\sigma^{(j)}(x_t^\sigma)$ (bottom) when $T = 100$ and $\beta_{\lambda,0} = 4$. Dashed dotted (resp. solid): true function for $j = 1$ (resp. $j = 2$).

Scenario 2

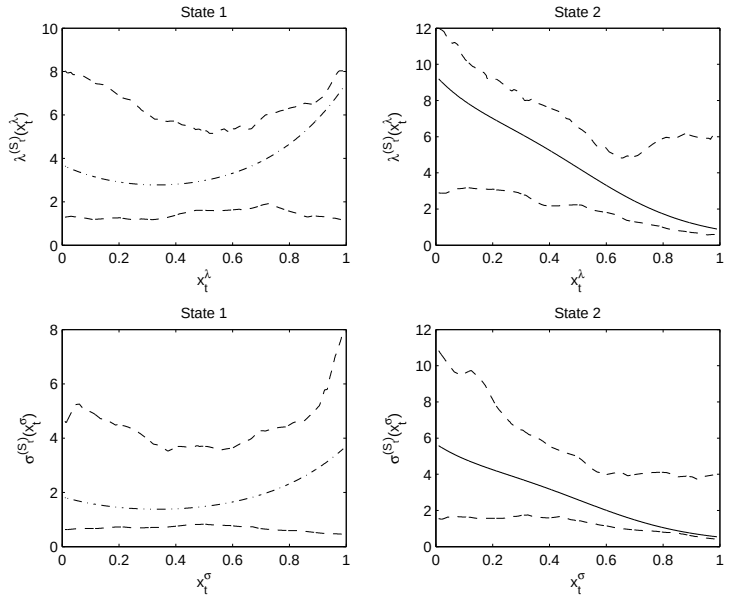

Figure 7: 95% coverage bands (dashed) for $\lambda^{(j)}(x_t^\lambda)$ (top) and $\sigma^{(j)}(x_t^\sigma)$ (bottom) when $T = 50$ and $\beta_{\lambda,0} = 2$. Dashed dotted (resp. solid): true function for $j = 1$ (resp. $j = 2$).

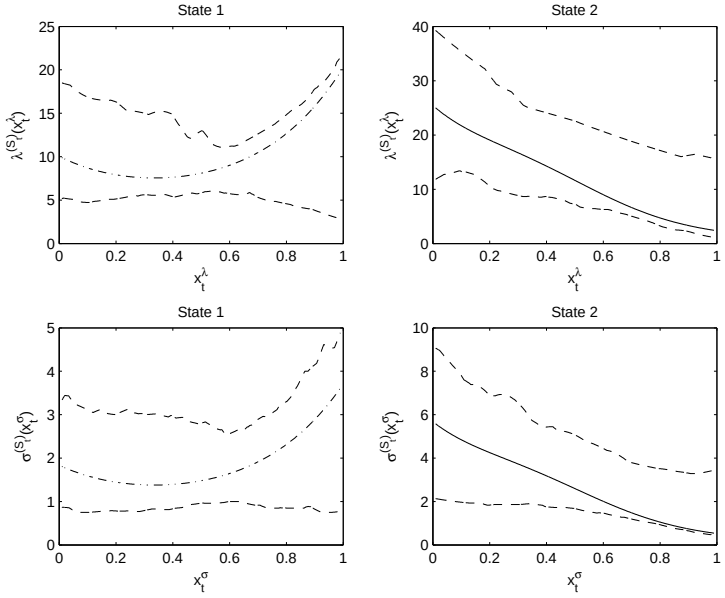

Figure 8: 95% coverage bands (dashed) for $\lambda^{(j)}(x_t^\lambda)$ (top) and $\sigma^{(j)}(x_t^\sigma)$ (bottom) when $T = 50$ and $\beta_{\lambda,0} = 3$. Dashed dotted (resp. solid): true function for $j = 1$ (resp. $j = 2$).

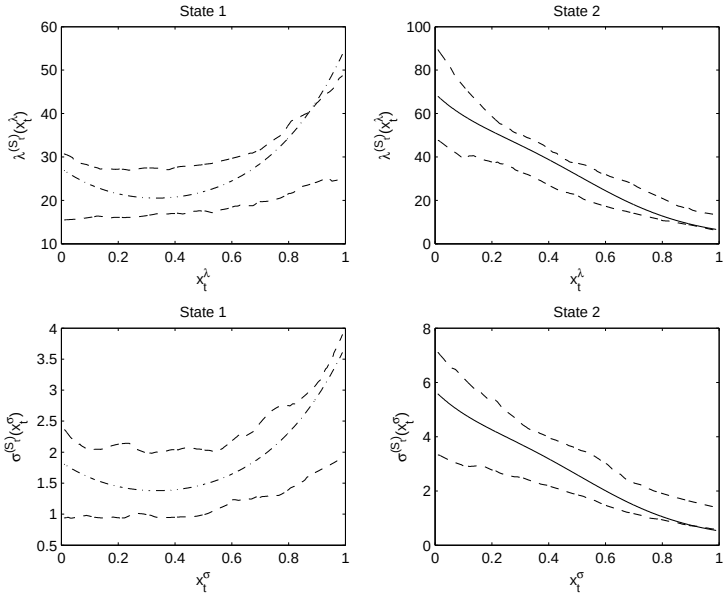

Figure 9: 95% coverage bands (dashed) for $\lambda^{(j)}(x_t^\lambda)$ (top) and $\sigma^{(j)}(x_t^\sigma)$ (bottom) when $T = 50$ and $\beta_{\lambda,0} = 4$. Dashed dotted (resp. solid): true function for $j = 1$ (resp. $j = 2$).

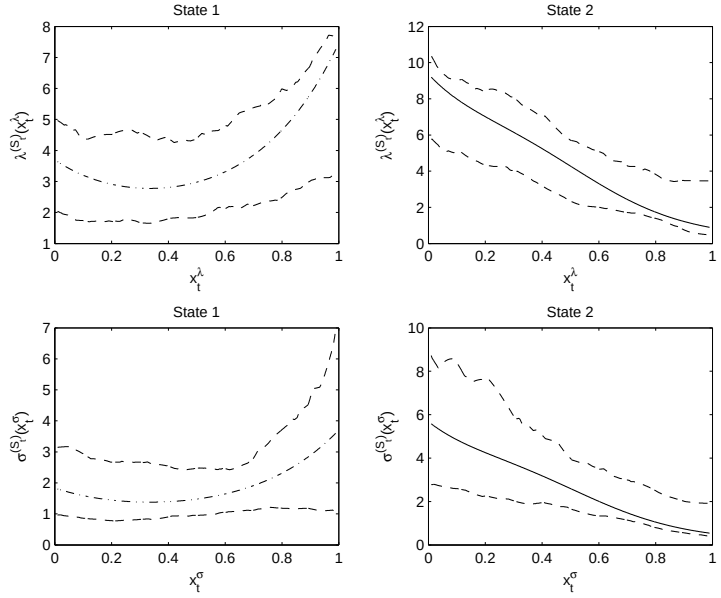

Figure 10: 95% coverage bands (dashed) for $\lambda^{(j)}(x_t^\lambda)$ (top) and $\sigma^{(j)}(x_t^\sigma)$ (bottom) when $T = 100$ and $\beta_{\lambda,0} = 2$. Dashed dotted (resp. solid): true function for $j = 1$ (resp. $j = 2$).

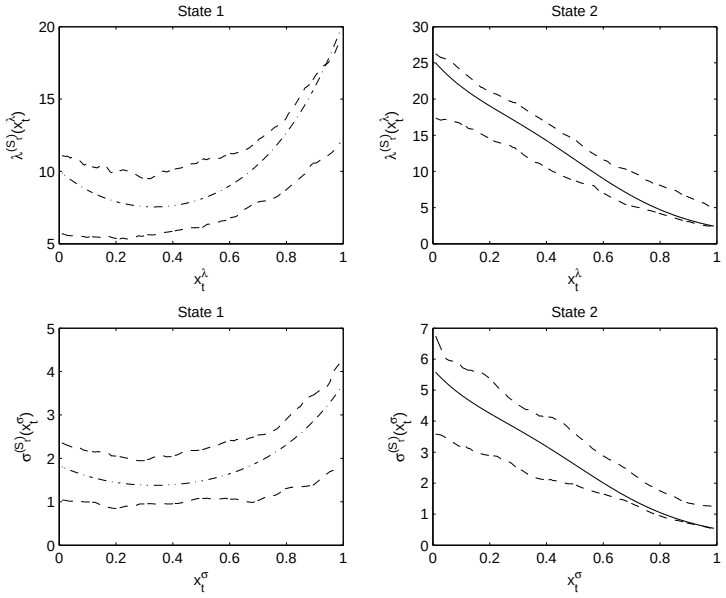

Figure 11: 95% coverage bands (dashed) for $\lambda^{(j)}(x_t^\lambda)$ (top) and $\sigma^{(j)}(x_t^\sigma)$ (bottom) when $T = 100$ and $\beta_{\lambda,0} = 3$. Dashed dotted (resp. solid): true function for $j = 1$ (resp. $j = 2$).

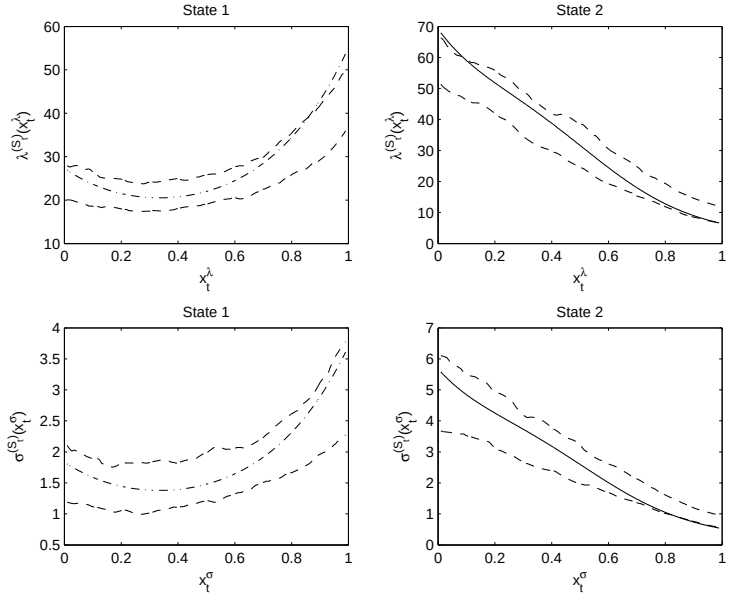

Figure 12: 95% coverage bands (dashed) for $\lambda^{(j)}(x_t^\lambda)$ (top) and $\sigma^{(j)}(x_t^\sigma)$ (bottom) when $T = 100$ and $\beta_{\lambda,0} = 4$. Dashed dotted (resp. solid): true function for $j = 1$ (resp. $j = 2$).

Remark: for the bootstrap confidence intervals (not displayed here), we use the initial time series of covariates for all bootstrap samples. However, the states as well as the bootstrap sample of the response variable have been drawn from the estimated model. This explains the large variability observed.

Empirical analysis: additional results

In this section, we provide the results obtained from fitting a MS-GAMLSS model where the VIX has been replaced by the VSTOXX on the same period. Indeed, as pointed out by a reviewer, UniCredit has its core business in Europe (mostly in western countries) and could be differently impacted by uncertainty on European stock market, compared to US stock market. Estimated parameters are given in Table 4 whereas the estimated function forms in the different states are given by Figure 13. The vector of smoothing parameters (obtained by the same CV procedure) is {2, 2, 15, 15, 2, 2}. For comparison purposes, we reproduced the estimated parameter of the MS-GAMLSS model with the VIX. We see that all parameters are very close one to another. In term of decoded state, we observe the same switches as well (Figure 14). In term of AIC, the model based on VIX is slightly better. From a statistical perspective, these results can be explained by the fact that VIX and VSTOXX are highly correlated. Using daily prices over the period 01/2005-06/2014,

we found a correlation coefficient of 0.958 (Figure 15), whereas using quarterly data, it is 0.952.

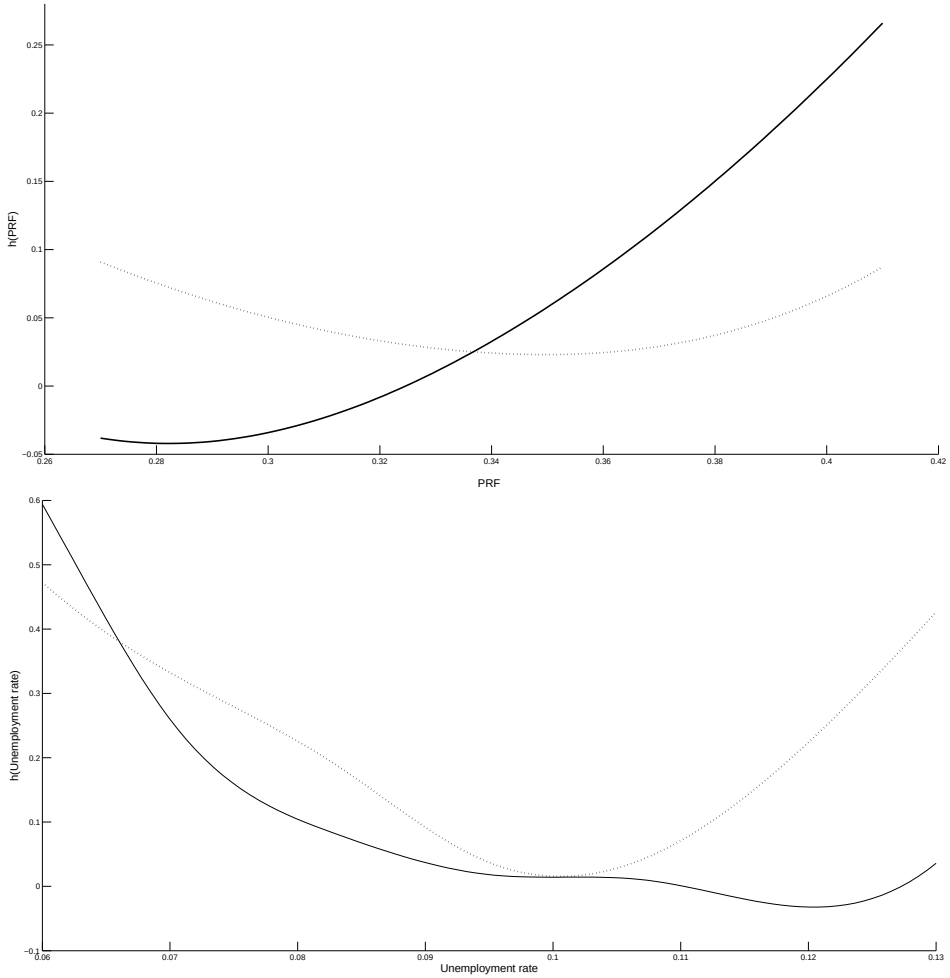

Figure 13: Estimated nonparametric functional forms between the PRF and λ and between the unemployment rate and σ (respectively, from top to bottom). Solid (resp. dotted) line: state 1 (resp. 2).

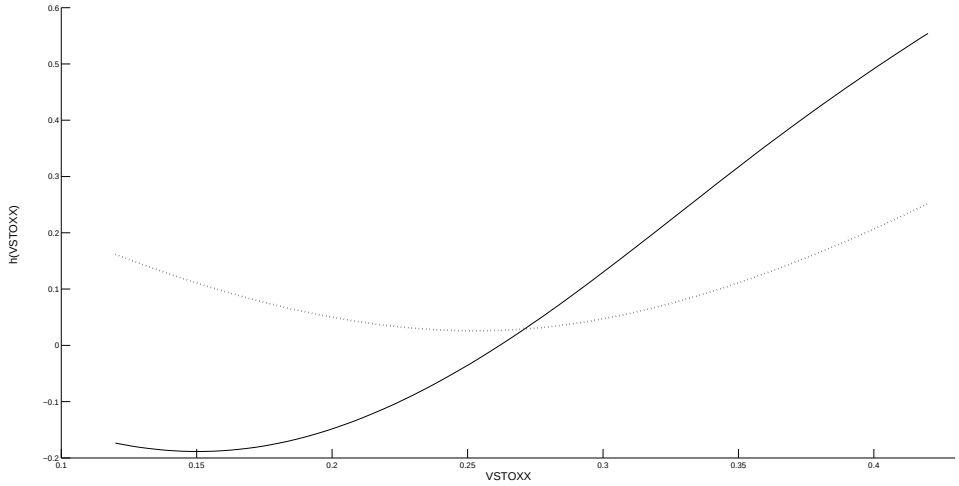

Figure 13: Estimated nonparametric functional forms between the VSTOXX index and γ (respectively, from top to bottom). Solid (resp. dotted) line: state 1 (resp. 2).

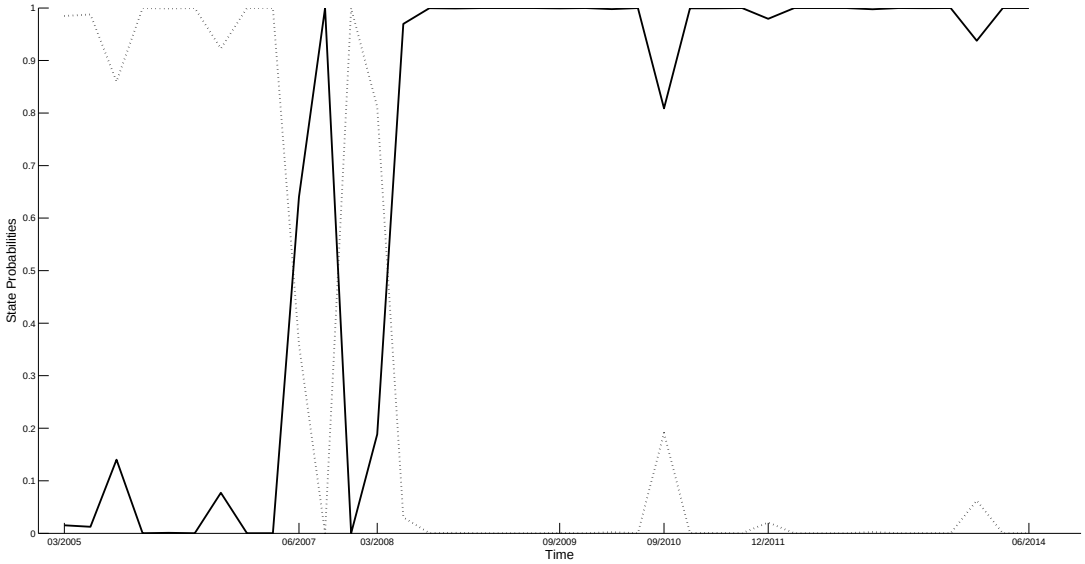

Figure 14: State probability over time, for MS-GAMLSS model with the VSTOXX index as explanatory variable for γ. Thin solid: first state; dotted: second state.

Final results	$\hat{\pi}_{11}$	$\hat{\pi}_{22}$	$\hat{\beta}_0^{(1)}(\sigma)$	$\hat{\beta}_0^{(2)}(\sigma)$	$\hat{\beta}_0^{(1)}(\gamma)$	$\hat{\beta}_0^{(2)}(\gamma)$	$\hat{\beta}_0^{(1)}(\lambda)$	$\hat{\beta}_0^{(2)}(\lambda)$	-LL	AIC
MS-VSTOXX	0.95	0.94	9.98	9.73	-0.94	-0.68	2.78	3.49	9624.8	19278
MS-VIX	0.95	0.93	10	9.81	-0.95	-0.70	2.8	3.47	9619.2	19272

Table 4: Estimations of the constant parameters for the MS-GAMLSS models with either VSTOXX or VIX as explanatory variable for γ.

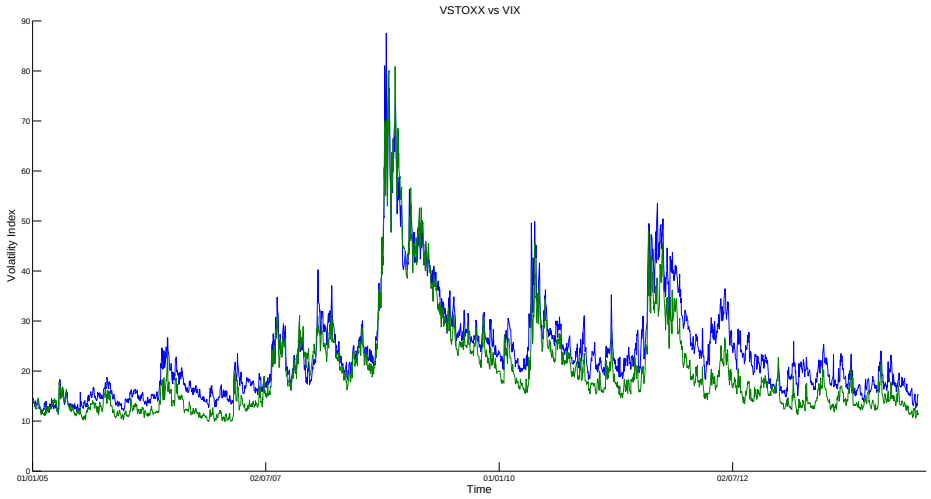

Figure 15: Daily VSTOXX (blue) versus VIX (green) volatility index over the period 03/01/2005 - 30/06/2014. Correlation coefficient: 0.958.

References

P. Christoffersen. Evaluating interval forecasts. *International Economic Review*, 39(4): 841–862, 1998.

W. Zucchini, I. MacDonald, and R. Langrock. *Hidden Markov Models for Time Series: An Introduction Using R, Second Edition*. Chapman & Hall, Boca Raton, 2016.