Supplementary Appendix to:

Multivariate Dynamic Intensity Peaks-Over-Threshold Models

Nikolaus Hautsch[*] and Rodrigo Herrera[†]

1 Backtesting

The first three tests utilized in this paper are introduced by Christoffersen (1998). The first corresponds to a test of unconditional coverage (LR_{uc}), which evaluates the expected fraction of exceptions $I_t = \mathbb{I}\left(Z_t < -VaR_\alpha^t\right)$, with $\mathbb{I}$ denoting the indicator function and Z_t the log-returns. We test the null hypothesis that $I_t \mid \mathcal{H}_t \sim Bernoulli\left(\alpha\right)$ against the alternative that $I_t \mid \mathcal{H}_t \sim Bernoulli\left(\hat{\alpha}\right)$ (i.e., $H_0 : \alpha = \hat{\alpha}$). This can be tested by means of a likelihood-ratio test

$$LR_{uc} = 2\left[\mathcal{L}\left(\hat{\alpha}; I_1, \ldots, I_t\right) - \mathcal{L}\left(\alpha; I_1, \ldots, I_t\right)\right] \sim \chi_1^2,$$

where $\mathcal{L}$ is the binomial log likelihood. The maximum likelihood estimator $\hat{\alpha}$ is the ratio of the number of violations to the total number of observations. This test implicitly assumes that the exceptions are independent, an assumption which is tested in a second test. Here, we assume that the exceptions I_t follow a binary first order Markov chain with transition probability matrix

$$\Pi = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix}, \quad \pi_{ij} = P\left(I_t = j \mid I_{t-1} = i\right).$$

We test the null hypothesis that $H_0 : \pi_{01} = \pi_{11}$ (i.e., past VaR violations do not contain information about current and future violations). By denoting $\pi = \pi_{01} = \pi_{11}$, a likelihood ratio test is given by

$$LR_{ind} = 2 \ln \left(\frac{(1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1 - \pi_{11})^{n_{10}} \pi_{11}^{n_{11}}}{(1 - \pi)^{n_{00} + n_{10}} \pi^{n_{01} + n_{11}}} \right) \sim \chi_1^2,$$

with n_{ij} being the number of observations of an event i on day $t - 1$ following an event j on day t. The maximum-likelihood estimators under the alternative hypothesis are

$$\hat{\pi}_{01} = \frac{n_{01}}{n_{00} + n_{01}} \quad \text{and} \quad \hat{\pi}_{11} = \frac{n_{11}}{n_{10} + n_{11}}.$$

Christoffersen (1998) suggests to simultaneously test for the correct unconditional coverage and independence yielding a test for correct conditional coverage

$$LR_{cc} = LR_{uc} + LR_{ind} \sim \chi_2^2.$$

In a more general framework, Engle and Manganelli (2004) introduce a dynamic quantile test (DQ) to evaluate different types of dependence. Define $Hit_t(\alpha) = I_t - \alpha$ as the de-meaned exceptions. Jointly testing the hypothesis that $E[Hit_t(\alpha)] = 0$ and that $Hit_t(\alpha)$ is uncorrelated with variables included in some information set $\mathbf{X}$, can be done using the artificial regression,

$$Hit_t = \mathbf{X}\beta + \mathbf{u},$$

where $\mathbf{u}$ is an i.i.d mean zero random variable and $\mathbf{X}$ contains, for example, lags of Hit_t, $\mathbf{Hit} = \{Hit_1, \ldots, Hit_p\}$. The test statistic is given by

$$DQ = \frac{\mathbf{Hit}'\mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}\mathbf{Hit}'}{\alpha(1 - \alpha)},$$

where p is the rank of the matrix of explanatory variables $\mathbf{X}$. Under the null hypothesis, the regressors should have no explanatory power, i.e., the regressors are not correlated with the dependent variables. In the empirical application, we use the dynamic quantile hit (DQ_{hit}) test, where the regressors contain a constant and the lagged Hit variable, and the dynamic quantile VaR (DQ_{VaR})

test, utilizing in addition the contemporaneous VaR estimates.

Finally, we employ the measure V^{ES} which evaluates the difference between the predicted ES $(\hat{ES}_\alpha^t)$ and the observed return Z_t, given that this return has exceeded the actual VaR, i.e.,

$$V^{ES} = \frac{\sum_{t=0}^{T} \left(Z_t - (-\hat{ES}_\alpha^t)\right) \mathbf{1}_{\left\{Z_t < -\hat{VaR}_\alpha^{t+1}\right\}}}{\sum_{t=0}^{T} \mathbf{1}_{\left\{Z_t < -\hat{VaR}_\alpha^{t+1}\right\}}}.$$

This statistic is close to zero if the model is appropriate Embrechts et al. (2005). However, its weakness is that it depends on the accuracy of the VaR estimates, since only returns below the VaR are taken into account.

2   Proof of Proposition 1

The intensity of the M-variate ground process according to a Hawkes model admits the matrix representation

$$\boldsymbol{\lambda}_{\mathbf{g}}\left(t \mid \mathcal{H}_t\right) = \boldsymbol{\mu} + \mathbf{B} \int_0^t \mathbf{H}\left(t - s\right) d\mathbf{N}(s),$$

where $\mathbf{B}$ and $\mathbf{H}$ denote the $M \times M$ branching and kernel matrix, respectively, and $\boldsymbol{\mu}$ is a $M \times 1$ vector of baseline intensities. Assuming that the spectral radius of $\mathbf{B}$ is strictly less than 1 and the expected value of the conditional intensity exists and is constant with $\mathbb{E}\left[\boldsymbol{\lambda}\left(t, y \mid \mathcal{H}_t\right)\right] = \boldsymbol{\nu}$, we obtain

$$
\begin{aligned}
\mathbb{E}\left[\boldsymbol{\lambda}_{\mathbf{g}}\left(t \mid \mathcal{H}_t\right)\right] &= \boldsymbol{\mu} + \mathbf{B} \int_{-\infty}^t \mathbf{H}\left(t - s\right) \mathbb{E}\left[d\mathbf{N}(s)\right] \\
&= \boldsymbol{\mu} + \mathbf{B} \int_{-\infty}^t \mathbf{H}\left(t - s\right) \boldsymbol{\lambda}\left(s, y \mid \mathcal{H}_s\right) ds \\
&= \boldsymbol{\mu} + \mathbf{B} \int_{-\infty}^t \mathbf{H}\left(t - s\right) \boldsymbol{\nu} ds \\
&= \boldsymbol{\mu} + \boldsymbol{\nu} \mathbf{B} \int_0^t \mathbf{H}\left(r\right) dr.
\end{aligned}
$$

Given that $\int_0^t \mathbf{H}\left(r\right) dr = 1$, we obtain

$$\mathbb{E}\left[\boldsymbol{\lambda}_{\mathbf{g}}\left(t \mid \mathcal{H}_t\right)\right] = \boldsymbol{\nu} = \left(\mathbf{1}_{\mathbf{M}} - \mathbf{B}\right)^{-1} \boldsymbol{\mu}.$$

3   Empirical Evidence for Hawkes-POT Processes

Table 1 reports the estimation results based on restricted and unrestricted trivariate Hawkes-POT models for extremes in DAX, S&P500, and FTSE100 returns. The restricted specification rules out feedback effects between individual exceedance intensities as well as between the magnitude of exceedances and their conditional ground intensities. In line with the corresponding results for ACI-POT models, as documented in the manuscript, we find that the unrestricted approaches exhibit a better fit in terms of the BIC and residual diagnostics. Hence, allowing for both clustering in extremes and for feedback between the size and the intensity of extremes is statistically strongly supported. However, the less flexible parametrization of the ground process in the Hawkes-POT specification compared to the ACI-POT specification results into lower explanatory power and is less appropriate to capture the dynamics in inter-exceedance times.

Table 2 gives the test outcomes for the in-sample and out-of-sample VaR and ES accuracy. Again, the highest VaR accuracy is provided by the *unrestricted* specifications that explicitly include mutual interactions between the point processes and the processes of exceedances. This confirms the results for ACI-POT models provided in the manuscript. Moreover, while the in-sample predictive performance is slightly higher for the ACI-POT model, both types of specifications are widely en par in an out-of-sample context.

4   Robustness Regarding Tail Threshold Selection

To determine the optimal number of observations k exceeding the tail threshold u and thus defining events as being "extreme", we follow the statistic proposed by Reiss and Thomas (2007). Here, we provide a robustness analysis in order to investigate the sensitivity of our results regarding the choice of u. In particular, we estimate the trivariate ACI-POT (1,1) specification for 100 different threshold values u, ranging from the 90% quantile to the 94.999% quantile. Table 3 reports how often the null hypothesis is rejected at the 5% significance level for each of the in-sample and out-of-sample VaR accuracy tests applied in the manuscript. We observe that the proportions of rejections widely resemble the p-values of Table 3 in the manuscript. We can thus conclude that the empirical findings reported in the manuscript are pretty stable with respect to the choice of the

thresholds u.

References

Christoffersen, P. (1998). Evaluating interval forecasts. *International economic review*, pages 841–862.

Embrechts, P., Kaufmann, R., and Patie, P. (2005). Strategic long-term financial risks: Single risk factors. *Computational Optimization and Applications*, 32(1-2):61–90.

Engle, R. and Manganelli, S. (2004). CAViaR. *Journal of Business and Economic Statistics*, 22(4):367–381.

Reiss, R.-D. and Thomas, M. (2007). *Statistical analysis of extreme values: with applications to insurance, finance, hydrology and other fields*. Birkhäuser Basel.

A Figures and Tables

Model	Hawkes-POT							Restricted Hawkes-POT					
Log-return m	FTSE		DAX		S&P500			FTSE		DAX		S&P500	
	par	(p-value)	par	(p-value)	par	(p-value)		par	(p-value)	par	(p-value)	par	(p-value)
					Ground process								
μ_m	0.011	(0.014)	0.018	(0.000)	0.012	(0.006)	μ_m	0.018	(0.000)	-0.026	(0.000)	0.016	(0.000)
b_{m1}	0.212	(0.004)	0.063	(0.272)	0.239	(0.002)	b_m	0.801	(0.000)	0.713	(0.000)	0.826	(0.000)
b_{m2}	0.007	(0.807)	0.105	(0.135)	0.001	(0.980)	a_m	0.036	(0.000)	0.044	(0.000)	0.025	(0.000)
b_{m3}	0.654	(0.009)	0.527	(0.024)	0.611	(0.006)							
a_m	0.086	(0.000)	0.078	(0.004)	0.016	(0.000)							
δ_m	0.287	(0.006)	0.563	(0.015)	-0.149	(0.474)							
LL_1			-4366.848							-4409.005			
Spr			0.856							0.826			
					Ground process residuals								
Mean (ε_m)	0.121		0.196		0.129			0.201		0.289		0.177	
$\tilde{\sigma}_\varepsilon$	0.838		0.819		0.834			0.961		0.925		0.976	
Excess.dis	-2.260	(0.024)	-2.498	(0.012)	-2.320	(0.020)		-0.581	(0.561)	-1.090	(0.276)	-0.355	(0.723)
LB_ε	0.660	(0.417)	0.017	(0.898)	5.111	(0.024)		0.000	(0.999)	0.010	(0.920)	3.612	(0.057)
					Mark process								
w^m	0.062	(0.001)	0.111	(0.000)	0.068	(0.001)	w^m	0.046	(0.002)	0.072	(0.000)	0.048	(0.000)
ρ_m	0.100	(0.000)	0.077	(0.001)	0.115	(0.000)	ρ_m	0.115	(0.000)	0.107	(0.000)	0.109	(0.000)
β_m	0.733	(0.000)	0.739	(0.000)	0.768	(0.000)	β_m	0.778	(0.000)	0.780	(0.000)	0.819	(0.000)
γ_m	-0.003	(0.010)	-0.006	(0.006)	-0.003	(0.115)	ξ_m	0.032	(0.485)	0.015	(0.738)	0.015	(0.741)
ξ_m	0.056	(0.221)	0.019	(0.684)	0.014	(0.753)							
	Diagnostics												
LL_2	-370.704		-487.889		-385.169			-375.237		-492.672		-386.742	
AIC	11287.220							11369.312					
BIC	11459.914							11479.208					

Table 1: Estimates of the trivariate Hawkes-POT models applied to the negative log-returns of the FTSE 100, DAX and S&P 500 indexes from January 3, 1994 to December 30, 2014. P-values are in parentheses. LL_1 corresponds to the log-likelihood of the Hawkes part, while LL_2 to the POT part. Spr : Spectral radius of the persistence matrix. Mean (ε_m): mean of residuals, $\tilde{\sigma}_\varepsilon$ standard deviation of the residuals, LB_ε: Ljung-Box statistic, *Excess.dis:* excess dispersion test.

		Hawkes-POT							Restricted Hawkes-POT						
		VaR in-sample													
Stock Index	α	excep.	LRuc	LRind	LRcc	DQ_{hit}	DQ_{VaR}	V^{ES}	excep.	LRuc	LRind	LRcc	DQ_{hit}	DQ_{VaR}	V^{ES}
FTSE	0.975	128	0.99	0.66	0.91	0.66	0.4	0.15	113	0.17	0.75	0.37	0.75	0.05	0.13
	0.98125	88	0.40	0.65	0.63	0.66	0.37	0.13	87	0.34	0.68	0.59	0.68	0.29	0.14
	0.9875	48	0.03	0.34	0.07	0.34	0.32	0.00	49	0.05	0.33	0.09	0.33	0.27	0.07
	0.99	37	0.04	0.46	0.08	0.47	0.73	-0.06	37	0.04	0.46	0.08	0.47	0.76	0.00
	0.99375	25	0.20	0.62	0.38	0.62	0.73	-0.11	25	0.20	0.62	0.38	0.62	0.67	0.00
DAX	0.975	118	0.36	0.46	0.50	0.46	0.19	0.22	120	0.47	0.91	0.76	0.91	0.05	0.24
	0.98125	76	0.03	0.14	0.03	0.14	0.11	0.17	77	0.04	0.47	0.10	0.47	0.11	0.17
	0.9875	43	0.00	0.05	0.00	0.06	0.16	0.00	44	0.01	0.06	0.00	0.06	0.11	0.07
	0.99	32	0.00	0.20	0.01	0.20	0.18	-0.19	32	0.00	0.02	0.00	0.02	0.05	-0.10
	0.99375	22	0.06	0.66	0.15	0.66	0.51	-0.21	21	0.04	0.08	0.02	0.08	0.08	-0.19
S&P 500	0.975	118	0.36	0.45	0.49	0.46	0.42	0.09	119	0.41	0.02	0.04	0.02	0.05	0.08
	0.98125	90	0.53	0.61	0.72	0.62	0.68	0.10	94	0.83	0.06	0.17	0.06	0.16	0.12
	0.9875	54	0.19	0.28	0.24	0.29	0.55	0.02	53	0.15	0.29	0.21	0.30	0.58	0.01
	0.99	40	0.10	0.43	0.19	0.43	0.24	-0.09	35	0.02	0.49	0.04	0.49	0.76	-0.14
	0.99375	21	0.04	0.68	0.10	0.68	0.69	-0.25	22	0.06	0.66	0.15	0.66	0.90	-0.24
		VaR out-sample													
Stock Index	α	excep.	LRuc	LRind	LRcc	DQ_{hit}	DQ_{VaR}	V^{ES}	excep.	LRuc	LRind	LRcc	DQ_{hit}	DQ_{VaR}	V^{ES}
FTSE	0.975	16	0.21	0.11	0.13	0.12	0.08	0.01	11	0.87	0.25	0.52	0.26	0.40	0.15
	0.98125	8	0.82	0.12	0.29	0.12	0.03	0.14	9	0.90	0.16	0.37	0.16	0.10	0.25
	0.9875	5	0.74	0.04	0.11	0.04	0.02	0.26	5	0.74	0.04	0.11	0.04	0.02	0.28
	0.99	4	0.77	0.79	0.93	0.79	0.34	0.15	3	0.42	0.01	0.03	0.01	0.01	0.14
	0.99375	1	0.20	0.95	0.44	0.95	0.82	0.24	1	0.20	0.95	0.44	0.95	1.00	-0.60
DAX	0.975	12	0.89	0.31	0.59	0.32	0.07	-0.17	9	0.43	0.55	0.62	0.55	0.39	-0.17
	0.98125	9	0.90	0.55	0.83	0.55	0.36	-0.09	7	0.56	0.64	0.76	0.65	0.43	-0.14
	0.9875	5	0.74	0.74	0.90	0.74	0.13	-0.03	6	0.92	0.69	0.92	0.69	0.48	0.17
	0.99	5	0.86	0.74	0.93	0.74	0.14	0.02	5	0.86	0.74	0.93	0.74	0.14	0.20
	0.99375	2	0.58	0.89	0.85	0.90	0.39	-0.04	2	0.58	0.89	0.85	0.90	0.95	-0.11
S&P 500	0.975	12	0.89	0.03	0.10	0.03	0.02	0.13	14	0.47	0.01	0.02	0.01	0.00	0.13
	0.98125	9	0.90	0.16	0.37	0.16	0.13	0.16	10	0.65	0.01	0.04	0.02	0.01	0.16
	0.9875	6	0.92	0.69	0.92	0.69	0.78	0.23	7	0.62	0.09	0.20	0.09	0.02	0.14
	0.99	6	0.53	0.69	0.76	0.69	0.81	0.26	6	0.53	0.69	0.76	0.69	0.16	0.23
	0.99375	3	0.94	0.84	0.98	0.84	0.02	0.46	4	0.53	0.79	0.79	0.79	0.84	0.37

Table 2: VaR accuracy test for Hawkes-POT approaches, for the in-sample period (from January 3, 1994 to December 30, 2014) and the backtesting period (January 2, 2015 to December 30, 2016). Entries in the rows are the number of observations exceeding the VaR level ("exceptions") and p-values of the corresponding accuracy tests.

Stock Index	α	VaR in sample						VaR out of sample					
		$\tilde{\alpha}$	LRuc	LRind	LRcc	DQ_{hit}	DQ_{VaR}	$\tilde{\alpha}$	LRuc	LRind	LRcc	DQ_{hit}	DQ_{VaR}
FTSE	0.975	0.97130	59	19	27	19	4	0.96808	73	51	59	51	25
	0.98125	0.97999	85	25	29	25	2	0.98096	100	88	88	88	33
	0.9875	0.98841	92	70	75	72	15	0.98967	100	55	93	55	59
	0.99	0.99085	97	85	89	86	41	0.99342	100	66	97	66	70
	0.99375	0.99436	100	100	100	100	85	0.99793	74	90	90	90	90
DAX	0.975	0.97527	100	100	100	100	30	0.97697	100	100	100	100	37
	0.98125	0.98312	76	100	92	100	20	0.98452	100	100	100	100	48
	0.9875	0.99027	52	100	61	100	22	0.99014	100	100	100	100	65
	0.99	0.99250	51	86	54	88	48	0.99333	100	100	100	100	82
	0.99375	0.99578	48	93	64	93	57	0.99603	100	100	100	100	100
S&P 500	0.975	0.97653	100	89	88	89	83	0.98123	100	69	74	69	18
	0.98125	0.98267	100	100	100	100	68	0.98681	100	85	87	85	29
	0.9875	0.98970	76	100	87	100	27	0.98955	100	30	100	30	26
	0.99	0.99237	58	97	78	97	51	0.99045	100	48	100	48	45
	0.99375	0.99568	55	100	69	100	90	0.99364	100	99	100	99	60

Table 3: Robustness of VaR accuracy for the ACI-POT approach for the in-sample period (from January 3, 1994 to December 30, 2014) and the backtesting period (January 2, 2015 to December 30, 2016). The numbers rest on 100 estimates of the ACI-POT(1,1) model based on different thresholds u varying on a regular grid from the 90% to the 94.999% quantile. The third column reports the averages of the relative proportions of VaR exceptions (i.e., exceedances of the respective VaR predictions) during the in-sample and out-of-sample periods, respectively, through the 100 estimates. We denote the average by $\tilde{\alpha}$ as it reflects the (average) empirical significance level. The remaining columns report the number of times for which the null hypothesis of the respective accuracy test is rejected at the 0.05 significance level.