Multi-step prediction of financial asset return probability density function using parsimonious autoregressive sequential model

Xiangru Fan, Xiaoqian Wei, Di Wang, Wen Zhang and Qi Wu

CityU-JD digits joint lab

Outlines

- Research background
- GARCH models
- Our algorithm
- Segregated BPTT approximate gradient training algorithm
- Result and conclusion

京东数科

GARCH family of models

$$r_t = \mu_t + \varepsilon_t, \qquad \varepsilon_t | \psi_{t-1} \sim \mathcal{N}(0, \sigma_t^2),$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2,$$

- r_t is asset return at time t
- μ_t is predicted mean
- ϵ_t is the residue
- $\sigma_t^2$is the predicted volatility

In stock/forex market, GARCH family of models can describe the conditional volatility given past daily price change of the asset. Investor may choose not to invest his/her money to the financial asset when predicted volatility is large.

京东数科

GARCH model multi-step prediction

GARCH model can be easily extended to conduct multi-step prediction

$$\sigma_t^2 = w + \alpha E_t\left[\epsilon_{t+h-1}^2\right] + \beta E\left[\sigma_{t+h-1}^2\right]$$

- r_t is asset return at time t
- μ_t is predicted mean
- ϵ_t is the residue
- σ_t^2 is the predicted volatility
- *h* is the prediction horizon

The analytical solution for multi-step GARCH prediction uses expected value of ϵ_{t+h-1}^2 in place of ϵ_{t+h-1}^2, Which is equal to σ_{t+h-1}^2

$$\sigma_t^2 = w + (\alpha + \beta)E\left[\sigma_{t+h-1}^2\right] \text{-} \alpha\mu_{t+h-1}$$

京东数科

Multi-step volatility prediction using deep learning

Unlike GARCH, deep learning can exploit more complex and richer features:

$$\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_L = \begin{bmatrix} r_1 \\ r_1^2 \\ r_1^3 \\ r_1^4 \end{bmatrix}, \begin{bmatrix} r_2 \\ r_2^2 \\ r_2^3 \\ r_2^4 \end{bmatrix}, ..., \begin{bmatrix} r_L \\ r_L^2 \\ r_L^3 \\ r_L^4 \end{bmatrix}$$

GRU is used for feature extraction and then a fully connected layer output the predicted mean and volatility :

$$\boldsymbol{h} = \mathbf{GRU}(\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_L; \Theta)$$

$$[\mu_{L+1}, \sigma_{L+1}]^T = \phi(\boldsymbol{W}_1 \boldsymbol{h} + \boldsymbol{b}_1)$$

京东数科

Multi-step volatility prediction using deep learning

Similar to GARCH, LSTM based multi-step prediction of volatility can also be rendered multi-step prediction

MS-DR

$$\boldsymbol{h} = \mathbf{GRU}(\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_L; \Theta) \tag{12}$$

$$[\mu_{L+1}, \sigma_{L+1}]^T = \phi(\boldsymbol{W}_1\boldsymbol{h} + \boldsymbol{b}_1) \tag{13}$$

$$[\mu_{L+2}, \sigma_{L+2}]^T = \phi(\boldsymbol{W}_2\boldsymbol{h} + \boldsymbol{b}_2) \tag{14}$$

$$[\mu_{L+3}, \sigma_{L+3}]^T = \phi(\boldsymbol{W}_3\boldsymbol{h} + \boldsymbol{b}_3) \tag{15}$$

Non-autoregressive version of the LSTM-based Volatility prediction model

PA-MS-DR

$$[\mu_{L+1}, \sigma_{L+1}]^T = \phi(\boldsymbol{W}\,\mathbf{GRU}(\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_L) + \boldsymbol{b}) \tag{27}$$

$$x_{L+1} = \begin{bmatrix} f_{E(r)}([\mu_{L+1}, \sigma_{L+1}]^T) \\ f_{E(r^2)}([\mu_{L+1}, \sigma_{L+1}]^T) \\ f_{E(r^3)}([\mu_{L+1}, \sigma_{L+1}]^T) \\ f_{E(r^4)}([\mu_{L+1}, \sigma_{L+1}]^T) \end{bmatrix} \tag{28}$$

$$[\mu_{L+2}, \sigma_{L+2}]^T = \phi(\boldsymbol{W}\,\mathbf{GRU}(\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_{L+1}) + \boldsymbol{b}) \tag{29}$$

$$x_{L+2} = \begin{bmatrix} f_{E(r)}([\mu_{L+2}, \sigma_{L+2}]^T) \\ f_{E(r^2)}([\mu_{L+2}, \sigma_{L+2}]^T) \\ f_{E(r^3)}([\mu_{L+2}, \sigma_{L+2}]^T) \\ f_{E(r^4)}([\mu_{L+2}, \sigma_{L+2}]^T) \end{bmatrix} \tag{30}$$

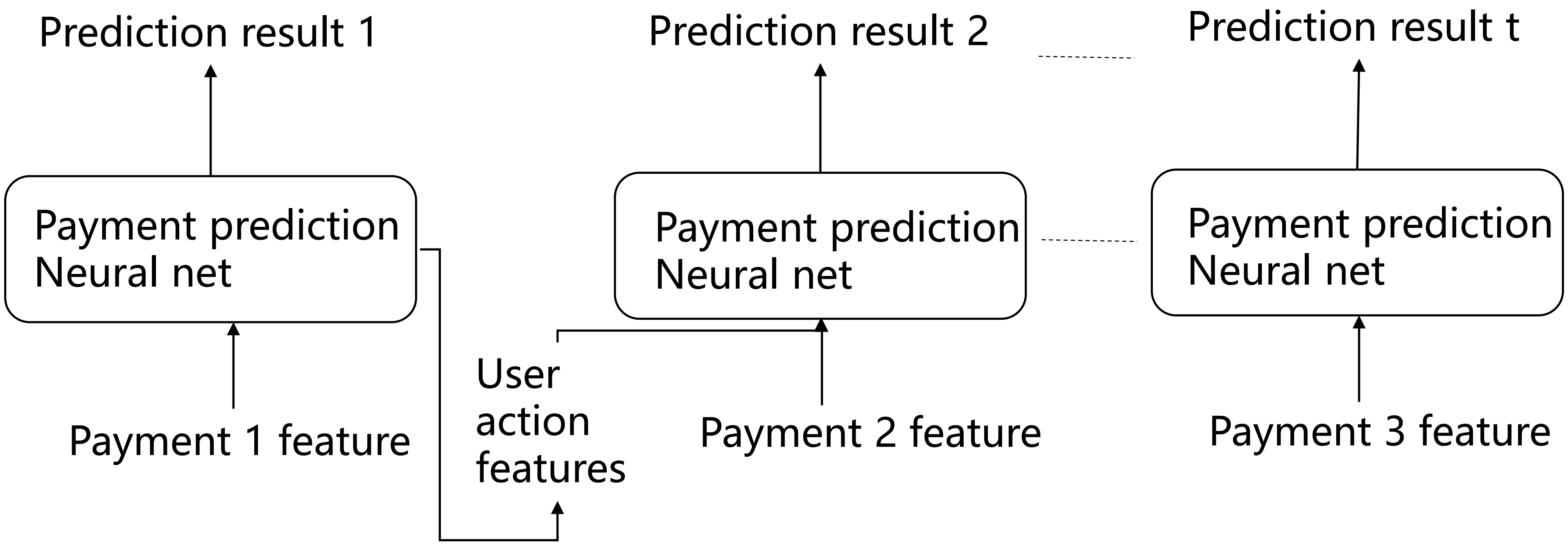
$$[\mu_{L+3}, \sigma_{L+3}]^T = \phi(\boldsymbol{W}\,\mathbf{GRU}(\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_{L+2}) + \boldsymbol{b}) \tag{31}$$

Autoregressive version of the LSTM-based Volatility prediction model

京东数科

t-distribution version of the model uses numerical integration to calculate $E(r^1)$, $E(r^2)$, $E(r^3)$ and $E(r^4)$, while analytical solution for them is aviable for normal distribution version of the model

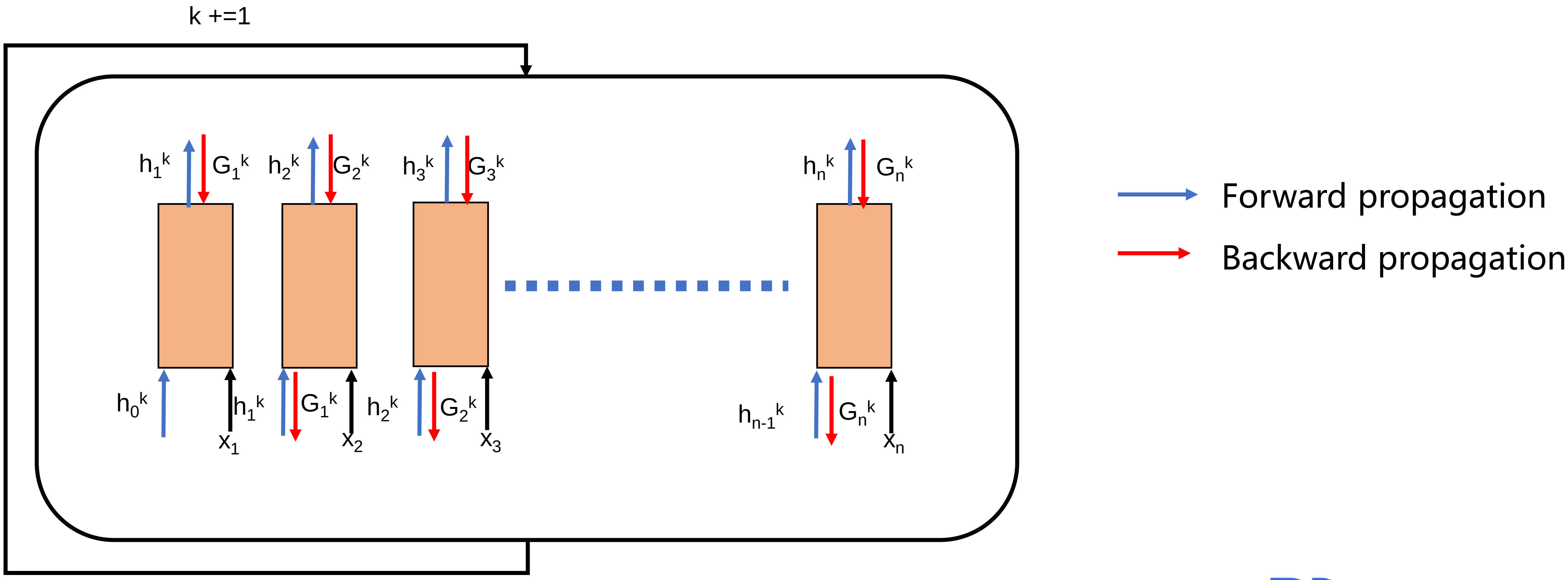
Inspirations of this training algorithm



In JDD, we frequently encounter long-term loans with 12 or 24 monthly installments,
So, we invented this recursive prediction model to mode
l the dynamics of loan performance during its lifetime.

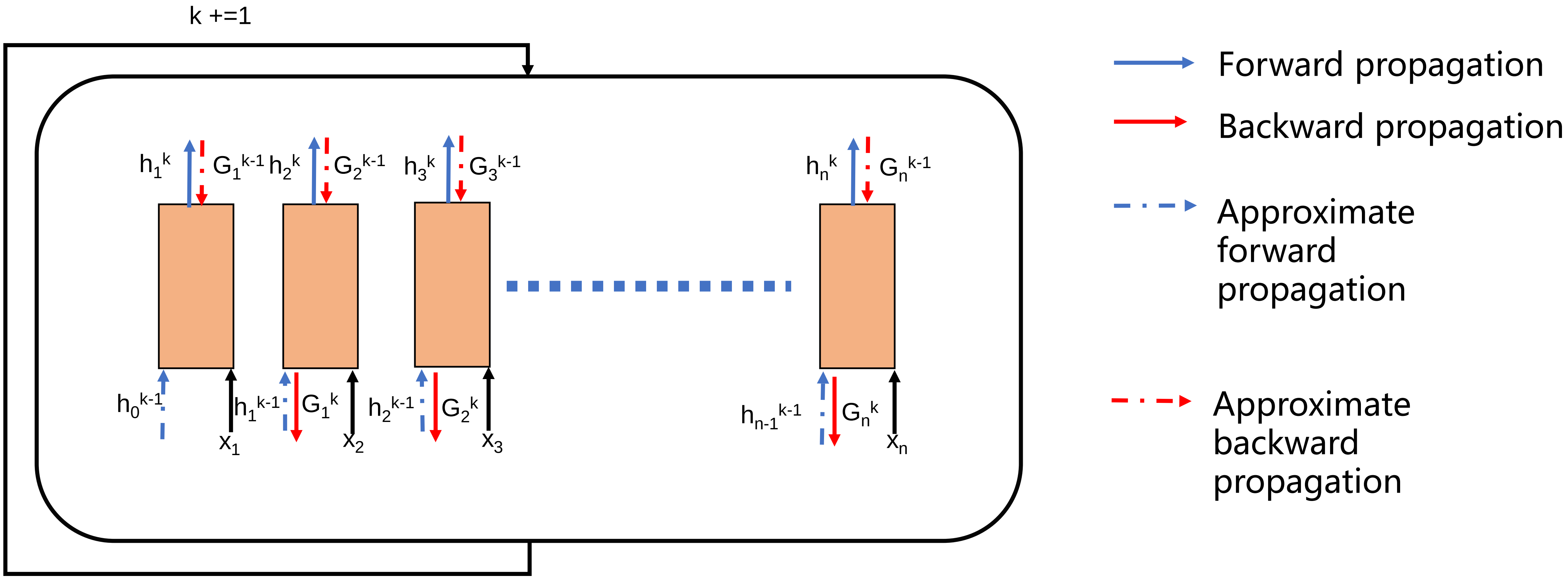
The specific algorithm for training this algorithm is then migrated to the stock-volatility prediction problem, as described in next slides.

Segregated back-propagation training algorithm



Normal back-propagation algorithm

Segregated back-propagation training algorithm



The new training method omitted back propagate through time process,
Hence greatly improved training speed

Proof can be found at https://github.com/imo1991/appendix4papers

Result of the model

	S&P 500	NASDAQ 100	Nikkei 225	EUR-USD	JPY-USD
AR-GJR-GARCH-t	1.3819†	1.6078 †	1.5634	0.6020	0.5212†
GJR-GARCH-t	1.3711	1.6033†	1.5599	0.6038	0.5176†
GARCH-t	1.3668	1.6052†	1.5609	0.6046	0.5183†
PA-MS-DR-*t*	**1.2165**	**1.4924**	**1.5057**†	**0.6008**⋆	**0.4950**†
MS-DR-*t*	1.2209	1.4971	1.5223†	0.6094	0.4983†
AR-GJR-GARCH	1.5318⋆	1.6838†	1.6152	0.5897	0.4818⋆†
GJR-GARCH	1.5058⋆	1.6805†	1.5979	0.5925⋆	0.4839
GARCH	1.5086⋆	1.6682†	1.6050	**0.5924**⋆	0.4840
PA-MS-DR	**1.2504**⋆	**1.5360**	**1.5189**†	0.6023	**0.4604**†
MS-DR	1.2564 †	1.5611	1.5403	0.6099	0.4740†

Table 2. The negative log-likelihood of the test sets of stock indexes S& P 500, NASDAQ 100 and Nikkei 225 as well as exchange rate EUR-USD and JPY-USD. The result is the mean of the negative log-likelihood for 5 future days. suffix-t indicate using t-distribution. The bolded letters indicate the lowest value. i.e. the best result. The threshold for rejecting the null hypothesis with 90% confidence level is 2.7060 for Christophersen's independence test. ⋆ represents the threshold is exceeded. The threshold for rejecting the null hypothesis with 95% confidence level is 3.8415 for Kupiec's proportion of failures coverage test. † sign indicates the threshold is exceeded.