Indexed Semi-Markov Model to forecast volatility in cryptocurrency market

Guglielmo D'Amico[1], Flavio Prattico and Filippo Petroni[2]
f.petroni@univpm.it

[1]Universitá "G. d'Annunzio" di Chieti-Pescara
[2]Universitá Politecnica delle Marche

June, 2022

Motivation

- We propose a Weighted Indexed Semi-Markov Chain (WISMC) model for price return.
- The aim of the models is to reproduce stylized facts for univariate high-frequency data in finance:
  asymmetry in the first passage time distribution
  persistence of volatility of real data
- Dependence structure for multivariate high-frequency data in finance.

The financial model

- Let us assume that the value of the asset under study is described by the time varying asset price $S(t)$.
- The return at time t calculated over a time interval of length 1, is defined as

$$R(t) = \log \frac{S(t+1)}{S(t)} \tag{1}$$

- The return are mapped into a discrete return $R_d(t)$ which assumes values into a finite set of states

$$E = \{-z_{min}\Delta, ..., -\Delta, 0, \Delta, ..., z_{max}\Delta\}, \tag{2}$$

where Δ is the grid amplitude of E.

The mapping is made by means of the following relation:

$$
\begin{aligned}
&R_d(t) = i\Delta \iff R(t) \in \left((i - \frac{1}{2})\Delta, (i + \frac{1}{2})\Delta \right], \ \forall i \notin \{-z_{min}\Delta, z_{max}\Delta\} \\
&R_d(t) = z_{max}\Delta \iff R(t) > (z_{max} - \frac{1}{2})\Delta, \\
&R_d(t) = -z_{min}\Delta \iff R(t) \leq (-z_{min} + \frac{1}{2})\Delta.
\end{aligned}
\tag{3}
$$

The map (3) produces a sequence of discrete returns $\{R_d(t)\}_{t \in \mathbb{N}}$ which are converted into a series of states and times.

To this purpose, we set $T_0 = 0$ and $J_0 = R_d(0)$ to denote that at current time (zero) the initial value of the return process is denoted by J_0. Then, by defining

$$T_n = \inf\{t > T_{n-1} : R_d(t) \neq R_d(T_{n-1})\}, \quad (4)$$

$$J_n = R_d(T_n), \quad (5)$$

we obtain a sequence $\{J_n\}_{n\in\mathbb{N}}$ with values in E describing the value of the return process at the n-th change and a sequence $\{T_n\}_{n\in\mathbb{N}}$ with values in $\mathbb{N}$ describing the time in which the n-th change of the return occurs.

ASSUMPTION: the sequences (7) and (8) form a MRP

To this purpose, we set $T_0 = 0$ and $J_0 = R_d(0)$ to denote that at current time (zero) the initial value of the return process is denoted by J_0. Then, by defining

$$T_n = \inf\{t > T_{n-1} : R_d(t) \neq R_d(T_{n-1})\}, \quad (4)$$

$$J_n = R_d(T_n), \quad (5)$$

we obtain a sequence $\{J_n\}_{n\in\mathbb{N}}$ with values in E describing the value of the return process at the n-th change and a sequence $\{T_n\}_{n\in\mathbb{N}}$ with values in $\mathbb{N}$ describing the time in which the n-th change of the return occurs.

ASSUMPTION: the sequences (7) and (8) form a MRP

Weghted-Indexed Semi-Markov Chain Model

- Following the approach by (AMM-2011) we consider the stochastic process

$$J_{-(m+1)}, J_{-m}, J_{-(m-1)}, \ldots, J_{-1}, J_0, J_1, \ldots$$

with a finite state space E.

- Let consider the stochastic process

$$T_{-(m+1)}, T_{-m}, T_{-(m-1)}, \ldots, T_{-1}, T_0, T_1, \ldots$$

with values in $\mathbb{N}$.

- Let consider also the stochastic process

$$V_{-(m+1)}, V_{-m}, V_{-(m-1)}, \ldots, V_{-1}, V_0, V_1, \ldots$$

with values in $\mathbb{R}$. The random variable V_n describes the value of the index process at the nth transition.

The index process is defined as follows:

$$V_n^\lambda = \sum_{k=0}^{n-1} \sum_{a=T_{n-1-k}}^{T_{n-k}-1} f(J_{n-1-k}, a, \lambda), \tag{6}$$

where $f : E \times \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ is a Borel measurable bounded function and V_0^λ is known and non-random.

The process V_n^λ can be interpreted as an accumulated reward process with the function f as a measure of the weighted rate of reward per unit time. The function f depends on the current time, on the state visited at current time and on the parameter λ that represents our weight.

To construct an indexed model we have to specify a dependence structure between the variables. Toward this end we adopt the following assumption:

$$\begin{aligned} &P[J_{n+1} = j,\ T_{n+1} - T_n \leq t | \sigma(J_h, T_h, V_h^\lambda)_{h=0}^n, J_n = i, V_n^\lambda = v] \\ &= P[J_{n+1} = j,\ T_{n+1} - T_n \leq t | J_n = i, V_n^\lambda = v] := Q_{ij}^\lambda(v; t), \end{aligned} \tag{7}$$

where $\sigma(J_h, T_h, V_h^\lambda)$, $h \leq n$ is the natural filtration of the three-variate process.
The matrix of functions $\mathbf{Q}^\lambda(v; t) = (Q_{ij}^\lambda(v; t))_{i,j \in E}$ is called weighted-indexed semi-Markov kernel.

Specification of the function f:

$$f(J_{n-1-k}, a, \lambda) = \frac{\lambda^{T_n - a} J^2_{n-1-k}}{\sum_{k=0}^{n-1} \sum_{a=T_{n-1-k}}^{T_{n-k}-1} \lambda^{T_n - a}}$$

$$V_n^{\lambda} = \sum_{k=0}^{n-1} \sum_{a=T_{n-1-k}}^{T_{n-k}-1} \left(\frac{\lambda^{T_n - a} J^2_{n-1-k}}{\sum_{k=0}^{n-1} \sum_{a=T_{n-1-k}}^{T_{n-k}-1} \lambda^{T_n - a}} \right) \tag{8}$$

Motivated by the fact that volatility is long range positively autocorrelated and then clustered in time, this implies that there are periods of high and low volatility.

Multivariate WISMC model

Suppose that we have m stocks with return processes

$$Z_1^\lambda(t), Z_2^\lambda(t), \ldots, Z_m^\lambda(t)$$

We aim to specify a model such that the joint return distribution is formulated so that the marginals analysis yields model compatible with a WISMC of kernel $\mathbf{Q}_h^\lambda(v;t)$, $h = 1, 2, \ldots, m$.

Copula models are suitable for this purpose because they combine marginal distributions into multivariate distributions.

Multivariate WISMC model

Suppose that we have m stocks with return processes

$$Z_1^\lambda(t), Z_2^\lambda(t), \ldots, Z_m^\lambda(t)$$

We aim to specify a model such that the joint return distribution is formulated so that the marginals analysis yields model compatible with a WISMC of kernel $\mathbf{Q}_h^\lambda(v;t)$, $h = 1, 2, \ldots, m$.

Copula models are suitable for this purpose because they combine marginal distributions into multivariate distributions.

The construction of the multivariate model passes through the following steps:

- one step representation of the WISMC model and computation of the one step cumulative distribution function;

- extension of the cdf into a continuous setting (for example by linear interpolation);

- definition of a continuous return index;

- construction of the joint distribution of the returns by copula function.

Step 1. One step representation of WISMC

Alternative description of the SMP: introduce $B(t) := t - T_{N(t)}$ and describe the behavior of the Markov process $(Z(t), B(t))$ on the extended state space $E \times \mathbb{D}$ where $\mathbb{D} = \{0, 1, ..., D\}$ and D is the maximum length of stay of the states of the process. Let denote by

$$p_{((i,u)(j,d))}(v) := \mathbb{P}[Z(n+1) = j, B(n+1) = d \mid Z(n) = i, B(n) = u, V^{\lambda}(n) = v]. \quad (9)$$

((MPRF-2014))

Lemma

Let suppose that $V^{\lambda}(n) = v$, $T_{N(n)} = n - u$ *and* $T_{N(n)+1} > n$, *then*

$$V^{\lambda}_{N(n)} = v - \sum_{a=T_{N(n)}}^{n-1} f(J_{N(n)}, n-a, \lambda) + \sum_{k=0}^{N(n)-1} \sum_{a=T_{N(n)-1-k}}^{T_{N(n)-k}-1} \Delta f(J_{N(n)-k}, T_{N(n)}, n, a) \tag{10}$$

where $\Delta f(i, T_{N(n)}, n, a) := f(i, T_{N(n)} - a, \lambda) - f(i, n-a, \lambda)$.

((MPRF-2014))

Theorem

For all $i, j \in E$, $u, d \in \mathbb{N}$ and $v \in R$, the one step transition probabilities $p_{((i,u)(j,d))}(v) := \mathbb{P}[Z(n+1) = j, B(n+1) = d \mid Z(n) = i, B(n) = u, V^{\lambda}(n) = v]$ are given by

$$p_{((i,u)(j,d))}(v) = \begin{cases} \frac{\bar{H}_i^{\lambda}(v+\Delta V(N(n),n);1+u)}{\bar{H}_i^{\lambda}(v+\Delta V(N(n),n);1+u)} & \text{if } j = i,\ d = 1+u \\ \frac{q_{ij}^{\lambda}(v+\Delta V(N(n),n);1+u)}{\bar{H}_i^{\lambda}(v+\Delta V(N(n),n);1+u)} & \text{if } j \neq i,\ d = 0. \end{cases} \tag{11}$$

where $\bar{H}_i^{\lambda}(t) = 1 - H_i^{\lambda}(v;t) = \sum_{j \in E} Q_{ij}^{\lambda}(v;t)$ is the survival function of sojourn time in state i,
$q_{ij}^{\lambda}(x,t) = Q_{ij}^{\lambda}(x,t) - Q_{ij}^{\lambda}(x,t-1)$ and
$\Delta V(N(n),n) = V_{N(n)}^{\lambda} - V_n^{\lambda}$ is the opposit of the variation of the index process on the waiting time $n - N(n)$.

Step 2. Computation of the cumulative distribution function (cdf)

Define

$$P_{(i_h,v_h)(j_h,d_h)}(v_h) = \mathbb{P}[Z_h^\lambda(n+1) \leq j_h, B_h^\lambda(n+1) \leq d_h \mid Z_h^\lambda(n) = i_h, B_h^\lambda(n) = u_h, V_h^\lambda(n) = v_h] \tag{12}$$

$$= \begin{cases} \sum_{s<j_h} p_{(i_h,u_h)(s,0)}(v_h) & \text{if } i_h > j_h,\ i_h, j_h \in E \\ \sum_{s\leq j_h}^{s\neq i_h} p_{(i_h,u_h)(s,0)}(v_h) + p_{(i_h,u_h)(i_h,u_h+1)}(v_h) 1_{\{d_h \geq u_h+1\}} & \text{if } i_h \leq j_h,\ i_h, j_h \in E. \end{cases}$$

Step 3. Extension of the cdf into a continuous setting

For a non integer x, with $r < x < r+1$ for two successive return states $r, r+1 \in E$, we define:

$$P_{(i_h,u_h)(x,0)}(v_h)$$

$$= \begin{cases} P_{(i_h,u_h)(r,0)}(v_h) + (x-r)\Big(P_{(i_h,u_h)(r+1,0)}(v_h) - P_{(i_h,u_h)(r,0)}(v_h)\Big) & \text{if } i_h < r \\ P_{(r,u_h)(r,u_h+1)}(v_h) + (x-r)\Big(P_{(r,u_h)(r+1,0)}(v_h) - P_{(r,u_h)(r,u_h+1)}(v_h)\Big) & \text{if } i_h = r \\ P_{(r+1,u_h)(r,0)}(v_h) + (x-r)\Big(P_{(r+1,u_h)(r+1,u_h+1)}(v_h) - P_{(r+1,u_h)(r,0)}(v_h)\Big) & \text{if } i_h = r+1 \\ P_{(i_h,u_h)(r,0)}(v_h) + (x-r)\Big(P_{(i_h,u_h)(r+1,0)}(v_h) - P_{(i_h,u_h)(r,0)}(v_h)\Big) & \text{if } i_h > r+1. \end{cases}$$

Step 4. Inverting the continuous cdf

We denote the inverse function of $P_{(i_h,u_h)(\cdot,\cdot)}(v_h)$ by

$$P^{-1}_{(i_h,u_h)}(v_h) : [0,1] \to E \times \mathbb{D}$$

which exists whenever the elements

$$p_{(i_h,u_h)(j_h,0)}(v_h) \neq 0, \ \forall j_h \neq i_h, \forall v_h$$
$$p_{(i_h,u_h)(i_h,u_h+1)}(v_h) \neq 0, \ \forall v_h$$

Step 5. Probability distribution of next return

The function $P_{(i_h,u_h)(\cdot,\cdot)}(v_h)$ is the c.d.f. of a continuous return index $X_h^\lambda(n)$ of stock h at any point n $\forall x_h \in \mathbb{R}$ such that

$$P_{(i_h,u_h)(x_h,d_h)}(v_h) = \mathbb{P}[X_h^\lambda(n+1) \leq x_h, B_h^\lambda(n+1) \leq d_h \mid Z_h^\lambda(n) = i_h, B_h^\lambda(n) = u_h, V_h^\lambda(n) = v_h].$$

Next discrete return $Z_h^\lambda(n+1)$ is determined by rounding up to the adjacent discrete return state, i.e.

$$(Z_h^\lambda(n+1), B_h(n+1)) = \begin{cases} (j_h, 0), & \text{if } \lceil X_h^\lambda(n+1) \rceil = j_h \neq i_h \\ (i_h, B_h^\lambda(n)+1), & \text{if } \lceil X_h^\lambda(n+1) \rceil = i_h \end{cases}$$

Step 6. The joint distribution of the returns

The joint distribution of the financial returns

$$(X_1^{\lambda}(n), X_2^{\lambda}(n), \dots X_m^{\lambda}(n))$$

is defined conditional on the returns at time $n-1$, on the values of the index process $v_h(n)$ and on the values of the backward recurrence time processes for each return $u_h(n)$.

$$R \begin{bmatrix} s_1(n) & u_1(n) & v_1(n) \\ s_2(n) & u_2(n) & v_2(n) \\ \vdots & \vdots & \vdots \\ s_m(n) & u_m(n) & v_m(n) \end{bmatrix} \begin{bmatrix} s_1(n+1) & u_1(n+1) \\ s_2(n+1) & u_2(n+1) \\ \vdots & \vdots \\ s_m(n+1) & u_m(n+1) \end{bmatrix}$$
$$= \mathbb{P}[\mathbf{X}(n+1) \leq \mathbf{x}, \mathbf{B}(n+1) \leq \mathbf{u} \mid \sigma(\mathbf{X}(s), \mathbf{X}(s)), s \leq n, \mathbf{X}(n) = \mathbf{s}, \mathbf{B}(n) = \mathbf{u}, \mathbf{V}(n) = \mathbf{v}]$$

The joint distribution of the financial returns is given by

$$R\begin{bmatrix} s_1 & u_1 & v_1 \\ s_2 & u_2 & v_2 \\ \vdots & \vdots & \vdots \\ s_m & u_m & v_m \end{bmatrix}\begin{bmatrix} x_1 & d_1 \\ x_2 & d_2 \\ \vdots & \vdots \\ x_m & d_m \end{bmatrix}$$
$$= C\left(P_{(s_1,u_1)(x_1,d_1)}(v_1), P_{(s_2,u_2)(x_2,d_2)}(v_2), \ldots, P_{(s_m,u_m)(x_m,d_m)}(v_m)\right)$$

and C is a copula function.

We define empirical estimators of the indexed probabilities $p_{ij}^{\lambda}(v)$, sojourn time distribution

$$g_{ij}^{\lambda}(v;t) := G_{ij}^{\lambda}(v;t) - G_{ij}^{\lambda}(v;t-1)$$

and indexed semi-Markov kernel

$$q_{ij}^{\lambda}(v;t) := Q_{ij}^{\lambda}(v;t) - Q_{ij}^{\lambda}(v;t-1) :$$

$$\hat{p}_{ij}^{\lambda}(v;L) := \frac{N_{ij}(v;L)}{N_i(v;L)}, \text{ if } N_i(v;L) \neq 0, \tag{13}$$

$$\hat{g}_{ij}^{\lambda}(v;t;L) := \frac{N_{ij}(v;t;L)}{N_{ij}(v;L)}, \tag{14}$$

$$\hat{q}_{ij}^{\lambda}(v;t;L) := \hat{p}_{ij}^{\lambda}(v;L)\hat{g}_{ij}^{\lambda}(v;t;L) = \frac{N_{ij}(v;t;L)}{N_i(v;L)}. \tag{15}$$

The multivariate model was applied to a portfolio of 4 cryptocurrency.

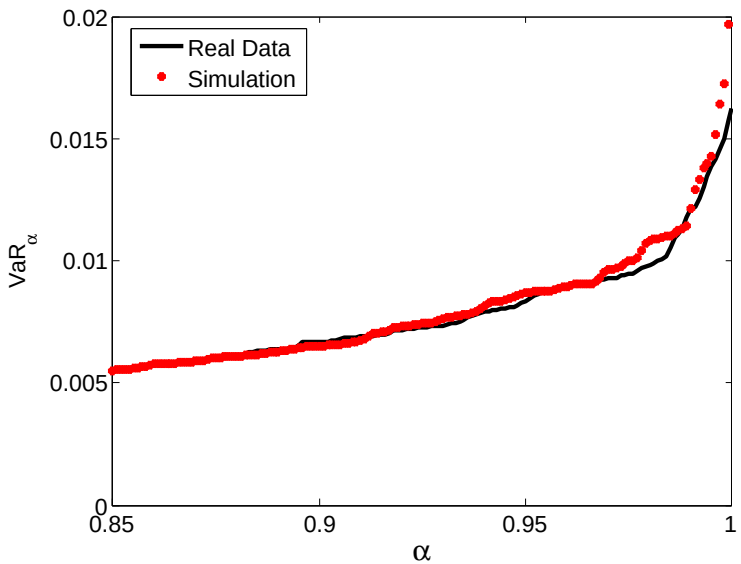
Stock	**WISMC**	**Persistence**
Bitcoin	26%	56%
Etherium	28%	63%
LiteCoin	23%	59%
Zcash	29%	48%

Table: Percentage square root mean error on one step ahead volatility forecasting of each currency for the presented model WISMC and for a simple persistence model

Value at Risk

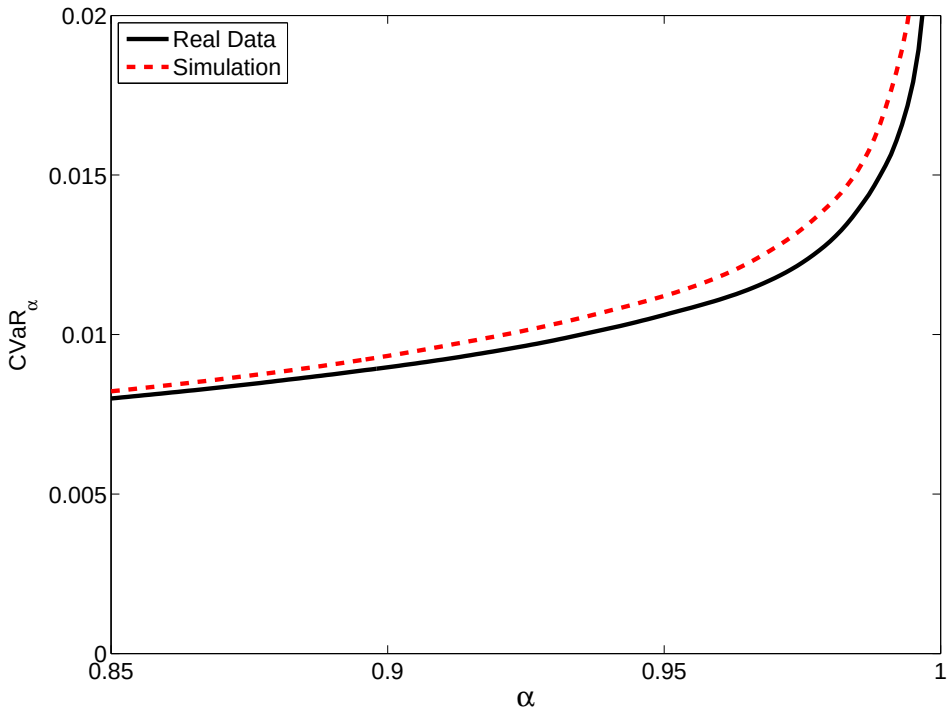
Given a confidence level $\alpha \in (0, 1)$, the VaR of the portfolio is the smallest number x such that the probability that the loss L exceeds x is at most $(1 - \alpha)$, i.e.

$$\text{VaR}_\alpha(L) = \inf\{x \in \mathbb{R} : P(L > x) \leq 1 - \alpha\}$$



At last we also compare the Conditional Value-at-Risk (CVaR), also know as expected shortfall, for a level α. The formal definition is the following:

$$CVar_\alpha = \frac{1}{1-\alpha}\int_\alpha^\infty VaR_\gamma \, d\gamma.$$



Conclusion

- We modeled high-frequency cryptocurrency data via semi-Markov chains;
- The model is compared with real data by means of Monte Carlo simulations;
- The model is able to reproduce the statistical properties of univariate and multivariate real data;
- The model can be used for portfolio selection problem and risk management.

Selected papers on Semi-Markov in Finance

- G. D'Amico(2011), Age-usage semi-Markov models, *Applied Mathematical Modelling*
- G. D'Amico and F. Petroni (2012), Weighted-indexed semi-Markov models for modeling financial returns, *Journal of Statistical Mechanics: Theory and Experiment*
- G. D'Amico, F. Petroni (2018), Copula Based Multivariate Semi-Markov Models with Applications in High-Frequency Finance. *European Journal of Operational Research*.
- G. D'Amico, A. Lika, F. Petroni (2019), Change point dynamics for financial data: an indexed Markov chain approach, *Annals of Finance*
- G. D'Amico, F. Petroni (2021), A micro-to-macro approach to returns, volumes and waiting times, *Applied Stochastic Models in Business and Industry*