On non-parametric calibration scheme for CGMY model on cryptocurrency markets by means of a Gaussian process regression

Vasily Rodochenko

ICSM-4, Divnomorskoye

June 2019

Contents

Academic interest

Over the past few years, "cryptocurrency" has become a well-used term in financial circles and news headlines:

Table: Number of papers with "cryptocurrency" keyword.

Engine	Publications (papers) found
Google Scholar	11 900
Researchgate	837
Springer link	722
Ssrn	454
ScienceDirect	268
BlockchainLibrary	190
ArXiv	189
Econbiz	167
Jstor	76

How to acquire historical data?

Sources (https://)

- Coinmarketcap (coinmarketcap.com)
- Poloniex API (docs.poloniex.com)
- Kaggle BTC (www.kaggle.com/mczielinski/bitcoin-historical-data)
- Custom data collectors

If you need to pick one

Kaggle

What top cryptocurrencies are (Coinmarketcap)

Position	Name	calitalization	avg price
1	Bitcoin (BTC)	$93,334,631,518	$5,285.70
2	Ethereum (ETH)	$17,848,982,526	$168.78
3	Ripple (XRP)	$13,436,161,868	$0.320132
4	Bitcoin Cash	$5,031,673,483	$283.62
5	EOS	$4,684,726,701	$5.17
6	Litecoin	$4,658,710,175	$75.83
7	Binance	$3,412,916,505	$24.17
8	Tether	$2,602,196,817	$1.01
9	Stellar	$2,157,410,101	$0.111416
10	Cardano	$1,875,473,273	$0.072336
11	TRON	$1,643,117,501	$0.024641
12	Monero	$1,157,920,100	$68.38
13	Dash	$1,055,288,634	$120.43
14	Bitcoin SV	$1,008,523,963	$56.85
15	Tezos	$873,571,780	$1.32

Top cryptocurrencies shares

Position	Name	Share	Cumulative share
1	Bitcoin Bitcoin	52.78%	52.78%
2	Ethereum Ethereum	10.09%	62.88%
3	XRP XRP	07.60%	70.48%
4	Bitcoin Cash	02.85%	73.32%
5	EOS EOS	02.65%	75.97%
6	Litecoin	02.63%	78.61%
7	Binance	01.93%	80.54%
8	Tether	01.47%	82.01%
9	Stellar	01.22%	83.23%
10	Cardano	01.06%	84.29%
11	TRON	00.93%	85.22%
12	Monero	00.65%	85.87%
13	Dash	00.60%	86.47%
14	Bitcoin SV	00.57%	87.04%
15	Tezos	00.49%	87.53%

In comparison to Russian economy

Cryptocurrency market vs MOEX

Total Russian stock market capitalization (MOEX): $645,233,992,728
Total cryptocurrency market cap: $176,822,984,800 (27.4%)

Cryptocurrency market summary

More than 50% of total the market capitalization is focused in Bitcoin

Contents

Problem description

General notes

- Cryptocurrency rates can be treated like “normal” stock prices.
- Cryptocurrency dynamics often has both high liquidity and volatility.
- Models with jumps in some cases reflects their dynamics better than purely Gaussian models.

Problem

- We try to calibrate GGMY model parameters using historical information on asset price and GGMY parameters.

Data preparation scheme

Data prepatation - logreturns

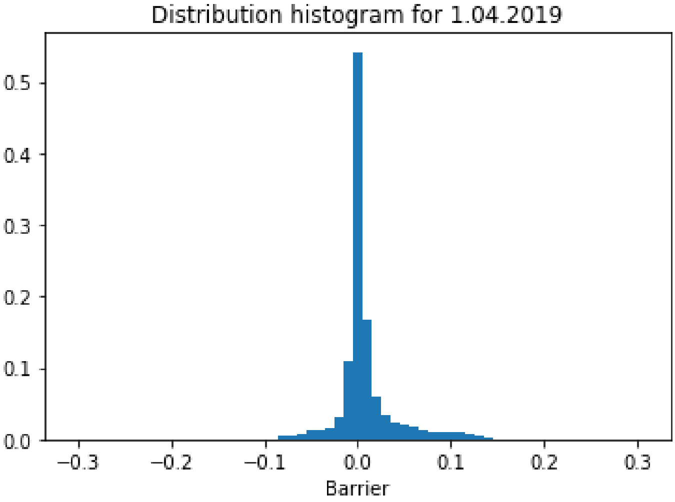
1. Consider price logreturns for 1 day (00:00 to 23:59 UTC).
2. To handle OHLCV: Consider the first available "open" price for an asset as a starting price and then use a volume-weighted average price for every next time step.
3. $x_i = \log(S_i/S_0), i = 1, 2, \ldots,$ where S_0 is the open price of a period and S_i is a volume-weighted average for each of the following 1-min periods.
4. At 00:00 renew S_0, reset the counter i and continue the process.

Bitcoin dynamics analysis

Data prepatation - barriers

1. Prepare a set of barriers $H_k = dh \cdot k$ for $k = -30, \ldots, -1$ and $k = 1, \ldots, 30$, with $dh = 0.01$. We obtain 30 positive barriers, which are multiples of 0.01, i.e. 1%: $0.01, 0.02, \ldots, 0.3$ and 30 negative ones: $-0.01, -0.02, \ldots, -0.3$. The number of barriers and dh were chosen experimentally.
2. Mark the barrier as H_k "crossed" at moment i if the respective logreturn x_i is greater or equal then H_k if $k > 0$ (less or equal, for $k < 0$).
3. Collect a statistics on "crossing" events for each barrier.

Histogram sample



CGMY model

CGMY *(The fine structure of asset returns: An empirical investigation. 2002)* model has 4 parameters, namely:

1. $\nu > 0$, "overall measure of activity"
2. $\lambda_- < 0$, "rate of left tail's exponential decay"
3. $\lambda_+ > 0$, "rate of right tail's exponential decay"
4. $0 < c < 2, c \neq 1$ "controls monotonicity, activity and variation of Lèvy density"

Lévy density of the associated Lévy process:

$$\pi(x) = cx_+^{-\nu-1}e^{\lambda_- x} + cx_-^{-\nu-1}e^{\lambda_+ x}$$

Characteristic exponent

$$\phi(\xi) = -i\mu\xi + c\Gamma(-\nu)\left[(-\lambda)_-^{\nu} - (-\lambda_- - i\xi)^{\nu} + \lambda_+^{\nu} - (\lambda_+ + i\xi)^{\nu}\right]$$

Model calibration (CGMY)

- We interpret historical probabilities calculated above as first touch digital options prices with a payoff equals to 1 for each of H_k under consideration.
- Make an initial guess on parameter values, use a fast Wiener-Hopf-factorization-based pricing method to calculate option prices.
- We use Nelder-Mead (NM) algorithm, which starts from equally spaces set of points and place additional linear penalties, to restrict the set of parameter values it should stop on. The penalty adds the value $|b_n - p_n|$, $n = 1, .., 4$ where b_n is a boundary condition and p_n is a respective parameter value guess to error function.

Gaussian process regression (GPR)

Inputs and outputs

Consider $(X, y) = \{(x_i, y_i), i = 1, .., n\}$,

- x_i are GCMY parameters, daily logreturns and barrier crossing probabilities,
- y_i is one of CGMY model parameters.

Then assume

$$y_i = f(x_i) + \varepsilon_i$$

,
where $f(x)$ is Gaussian process, $f \sim N(0, K(X, X))$, and i.i.d. $\varepsilon_i \sim N(0, \sigma^2)$, $\sigma^2 > 0$ is the noise in the data. In our case $\sigma = 0$ produced degenerate approximation.

Learning (GPR)

Covariance matrix inversion

Let's denote the testing set as X_*. To evaluate unknown f_* GP posterior we use observations of y as follows:

$$f_*|X_*, X, y \sim \mathcal{N}\Big(K(X_*, X)\big[K(X,X) + \sigma^2 I\big]^{-1} y,$$
$$K(X_*, X_*) - K(X_*, X)\big[K(X,X) + \sigma^2 I\big]^{-1} K(X, X_*)\Big)$$

Then we generate pointwise predictions as its mean value.

Square exponential (SE) kernel

$$k(x, x') = \sigma_f^2 \exp\left(\frac{-|x - x'|^2}{2l^2}\right) = \sigma_f^2 \exp\left(\frac{-\sum_{k=1}^{d}(x_k - x_k')^2}{2l^2}\right) \quad (1)$$

l (smoothness of the fit) and σ_f^2 (highest possible covariance) are hyperparameters.

Estimating hyperparameters

A common approach is to maximize the marginal log-likehood:

$$-\frac{1}{2}\log(|K(X, X)|) - \frac{1}{2}y^T K(X, X)^{-1} y + const$$

Summary

Result obtained

GPR with a square exponential kernel function can be used to calibrate CGMY model.

Important notes

- GPR can speed up CGMY model calibration
- GPR is more accurate than linear and quadratic interpolation on a reasonably dense grid
- GPR works best for the parameters away from the border values of parameters of training set.
- Unlike neural networks, GPR does not always try to use the same values for CGMY parameters.

THANK YOU FOR ATTENTION!

Contents

References

1. *Carr P., Geman H., Madan D., Yor M.* (2002) The fine structure of asset returns: An empirical investigation. The Journal of Business, 75(2):305–333.
2. *Kim T., Kim H. Y.* (2019) Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data. PLoS ONE. Vol. 14, No. 2.
3. *Kudryavtsev O., Grechko S.* (2018) Statistical methods for calibrating models of cryptocurrencies prices. Accounting and Statistics. 2018. Vol. 4, No. 52, pp. 67–76.
4. *Nelder, J. A., Mead R.* (1993) A simplex method for function minimization. 1965. Computer Journal. 7: 308–313. doi:10.1093/comjnl/7.4.308.
5. *Nelson D. M.Q., Pereira A. C.M.* (2017) Stock market's price movement prediction with LSTM neural networks. International Joint Conference on Neural Networks (IJCNN). 2017. DOI: 10.1109/IJCNN.2017.7966019
6. *Serrano, W.* (2018) Fintech Model: The Random Neural Network with Genetic Algorithm. Procedia Computer Science, 126, 537–546.
7. *De Spiegeleer J., Madan D.B., Reyners S., Schoutens W.* (2018) Machine Learning for Quantitative Finance: Fast Derivative Pricing, Hedging and Fitting. av. at SSRN: ssrn.com/abstract=3191050, DOI http://dx.doi.org/10.2139/ssrn.3191050.

The research was supported by the RFBR (project 18-01-00910)