Price discovery in cryptocurrency markets

Paolo Giudici, Iman Abu Hashish, Paolo Pagnottoni, Kamonchai Rujirarangsan[1]

[1]Data Science Laboratory, University of Pavia
@Facebook: Data Science Network

giudici@unipv.it

17 November 2017

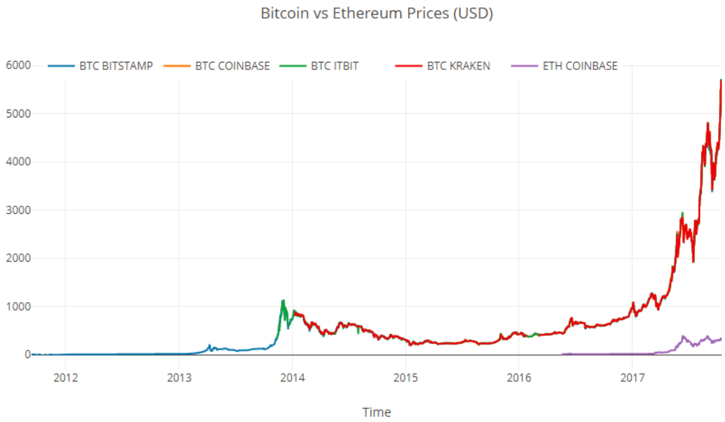
The project

Financial: To estimate the multivariate correlation structure among crypto currency prices, and among different markets.

Data: Daily, hourly and 5-minutes data of the most important crypto currencies, in different markets, scraped with Ptyhon, processed in R/Matlab

Science: Correlation Network models, VECM models, Market Information Share.

Data - crypto prices

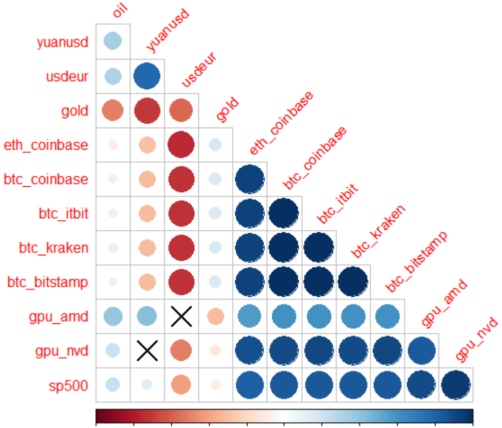


Data - summary statistics

Table: Daily prices

Price	Mean	St. Dev.	Min	Max
Coinbase Ethereum	40.17	34.12	6.75	397.37
Coinbase Bitcoin	1666.97	1401.25	438.38	6024.86
Kraken Bitcoin	1659.09	1394.47	433.50	6100.00
Bitstamp Bitcoin	1656.65	1391.15	439.62	6087.77
ItBit Bitcoin	1659.63	1391.63	438.61	6087.61
GPU AMD	10.16	3.30	3.77	15.2
GPU Nnvidia	107.15	44.54	43.36	201.86
Oil	48.67	3.16	39.51	54.45
SP500	2299.87	144.41	2000.54	2581.07
YuanUSD	6.75	0.12	6.48	6.96
USDEur	0.90	0.03	0.83	0.96
Gold	1263.06	52.81	1128.42	1366.38

Correlation structure



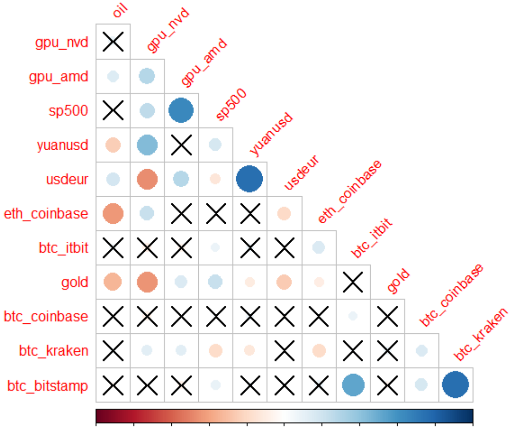
Partial Correlations

- From the inverse of the covariance matrix (A^{-1}, elements σ^{mn}), partial correlations can be derived:

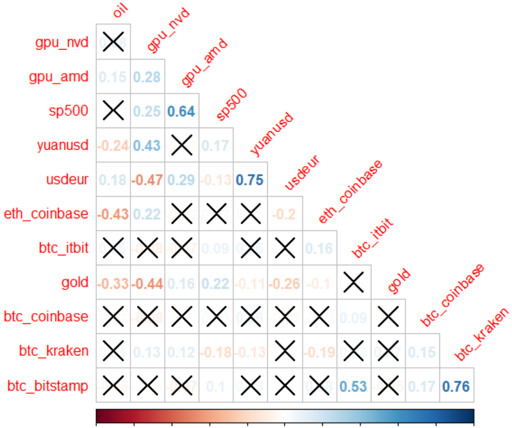
$$\rho_{mn|rest} = \frac{-\sigma^{mn}}{\sqrt{\sigma^{mm}\sigma^{nn}}}. \quad (1)$$

- They represent correlations between two prices, conditional on the remaining prices of the system (*rest*).

Partial correlation structure - I



Partial correlation structure - II



Correlation network models - I

Let i be a market price (crypto, forex, share). We assume prices y_t^i follow a structural VAR process:

$$y_t^i = \sum_{p=1}^{T} \beta_p^i y_{t-p}^i + \sum_{j \neq i} \gamma^j y_t^j + \varepsilon_t^i. \qquad (2)$$

or

$$Y_t = \sum_{p=1}^{T} A_p Y_{t-p} + B_0 Y_t + u_t, \qquad (3)$$

where B_0 has null diagonal elements.

Correlation network models - II

We can transform the VAR to its reduced form:

$$Y_t = \Gamma_1 Y_{t-1} + \ldots + \Gamma_p Y_{t-p} + u'_t, \tag{4}$$

where

$$\begin{cases} \Gamma_1 = (\mathbb{I} - B_0)^{-1} A_1, \\ \ldots \\ \Gamma_p = (\mathbb{I} - B_0)^{-1} A_p, \\ u'_t = (\mathbb{I} - B_0)^{-1} u_t. \end{cases} \tag{5}$$

Need to estimate B_0 to derive $A_1, \ldots, A_p$. Note that $u'_t = B_0 u'_t + u_t$. Thus:

$$\begin{cases} [u'_t]^i = \sum_{j \neq i} b_0^j [u'_t]^j + [u_t]^i, \\ \gamma^{ij} = \sqrt{b_0^j b_0^i} = corr(y_i, y_j | rest) \end{cases} \tag{6}$$

Results: estimation

Table: Contemporaneous and Autoregressive component

Prices	Contemp	Autoreg
Coinbase Ethereum	496.9	95.7
Coinbase Bitcoin	644.1	1510.9
Kraken Bitcoin	1594.0	718.9
Bitstamp Bitcoin	2396.8	151.99
ItBit Bitcoin	997.6	276.4
GPU AMD	441.7	9.6
GPU Nvidia	410.2	103.3
SP500	-255.6	2186.0
YuanUSD	-128.4	6.6
USDEur	-437.0	0.8
Gold	-537.7	1270

Results: predictive performance

Table: Predictive performance summary

Prices	RMSE full	RMSE autoreg
Coinbase Ethereum	16.00	15.48
Coinbase Bitcoin	213.78	193.76
Kraken Bitcoin	197.05	85.38
Bitstamp Bitcoin	195.82	17.96
ItBit Bitcoin	196.38	32.70
Gpu AMD	0.38	0.35
Gpu Nvidia	3.11	3.00
Oil	0.61	0.55
YuanUSD	0.02	0.02
USDEur	0.01	0.02
SP500	7.53	7.15
Gold	7.93	7.71

Intra-day Data

Intra-day data: tests

(p-value)	Jarque-Bera†	Box.test‡	ADF(level)	ADF(1^{st} diff)
BitStamp	3583.295	241.0817	-1.7321	-43.9976
GDAX	3583.717	240.9614	0.0489	-43.8102
Kraken	3583.476	236.1264	0.0841	-43.031
hitbtc	3582.678	232.7309	0.0011	-44.613
CEX.IO	3584.357	245.8599	0.1781	-44.1391

1%, 5% and 10% significant for ADF(drift) are -3.43, -2.86 and -2.57, respectively. Thus, variables are not be able to cointegrated at I(0) and I(1), then cointegration is implemented.

Johansen test	test	10pct	5pct	1pct
r <= 4	4.93	10.49	12.25	16.26
r <= 3	47.36 ***	16.85	18.96	23.65
r <= 2	159.76 ***	23.11	25.54	30.34
r <= 1	200.74 ***	29.12	31.46	36.65
r = 0	312.02 ***	34.75	37.52	42.36

At r<=3 test, the value is significant, so there is a cointegration at least 3 time series.

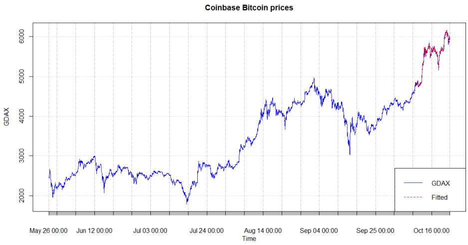
The VECM model

VECM model

$$D_{yt} = \mu + \Pi_{yt-1} + \sum_{i=1}^{p} \Gamma_i^* D_{yt-i} + \varepsilon_t$$

where y_t is a $m \times 1$ vector, D_{yt} is a $m \times 1$ vector of the first differences of y_t, μ is a $m \times 1$ vector of intercept coefficients, Π and Γ_i^* are m $\times$ m coefficient matrices and ε_t is a $m \times 1$ vector of errors.

VECM predictions



	ME	RMSE	MAE	MPE	MAPE
BitStamp	3.4507	55.6476	38.1270	0.0592	0.6862
GDAX	2.6386	51.5484	35.8713	0.0454	0.6452
Kraken	2.8961	48.0128	34.3433	0.0504	0.6176
hitbtc	2.0439	53.0548	37.3819	0.0344	0.6722
CEX.IO	3.0169	47.5627	32.4208	0.0524	0.5800
Average	2.8092	51.2565	35.6288	0.0484	0.6403

High frequency data

- 6 price series:
  - Bitstamp (USD)
  - Kraken (EUR)
  - BTC-e (USD)
  - Bitfinex (USD)
  - Okcoin (CNY)
  - BTC China (CNY)
- Time period analyzed: 2 January 2014 - 6 March 2017
- Sampling interval: 5 minutes

Information share

- Consider the VMA representation of the VECM:

$$y_t = \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \ldots = \Psi(L)\varepsilon_t \tag{7}$$

- We consider bivariate and trivariate models, accounting for the potential impact of exchange rate
- Cholesky decomposition: $\Sigma_\varepsilon = FF'$, with F lower triangular
- Hasbrouck (1995) information share:

$$IS_{i,j} = \frac{([\psi' F]_{i,j})^2}{[\psi \Sigma_\varepsilon \psi']_{i,i}} \tag{8}$$

represents the fraction of residual variance in the price of the market i which is due to shocks in the price of the market j

Results

Variable	Innovation					
	Okcoin	Btcn	Bitfinex	Bitstamp	Kraken	BTC-e
Okcoin	-	43.44	48.87	47.11	44.48	44.90
Btcn	56.66	-	48.98	46.85	45.16	38.87
Bitfinex	50.25	50.63	-	23.30	21.58	13.88
Bitstamp	52.28	52.87	76.70	-	33.19	28.42
Kraken	51.32	49.16	73.81	61.24	-	38.02
BTC-e	54.29	60.57	86.12	71.58	55.55	-

- Clear ranking: CNY markets seem to mostly drive prices
- The exchange rate does not exert a great influence on the Bitcoin price discovery mechanism