Attention-based dynamics for Bitcoin price modeling and applications

Gianna Figà-Talamanca
joint work with Alessandra Cretarola and Marco Patacca

Department of Economics,
University of Perugia, Italy
gianna.figatalamanca@unipg.it

PoliMi Fintech Journey Workshop, May 10, 2018

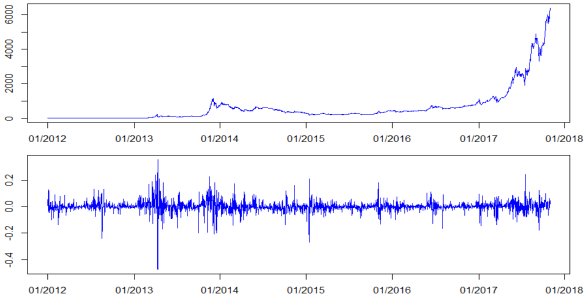


Figure: Daily average Bitcoin price and returns (Jan 2012 - Nov 2017).

- High volatility
- Speculative bubbles.

See. among others Yermack [10], Bukovina and Martiček [3]

Bitcoin price dynamics: 2012-2014

Time series analysis for serial correlation and time-varying variability of Bitcoin logarithmic returns.

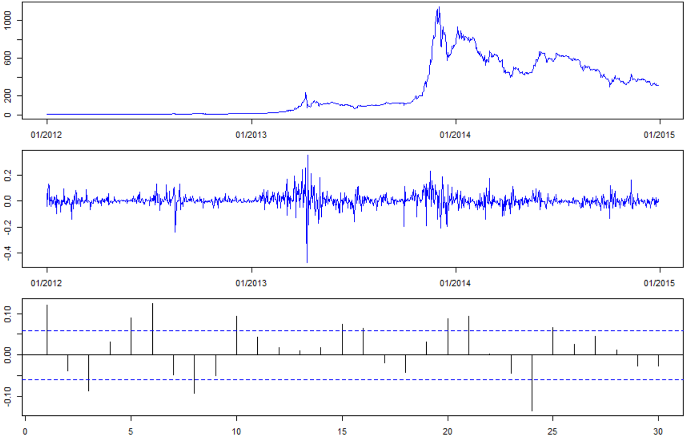


Figure: Bitcoin data from January 1, 2012 to December 31, 2014: price (top), logarithmic returns (center) and Autocorrelation function for Bitcoin logarithmic returns (bottom).

Bitcoin price dynamics: 2015-2017

Time series analysis for serial correlation and time-varying variability of Bitcoin logarithmic returns.

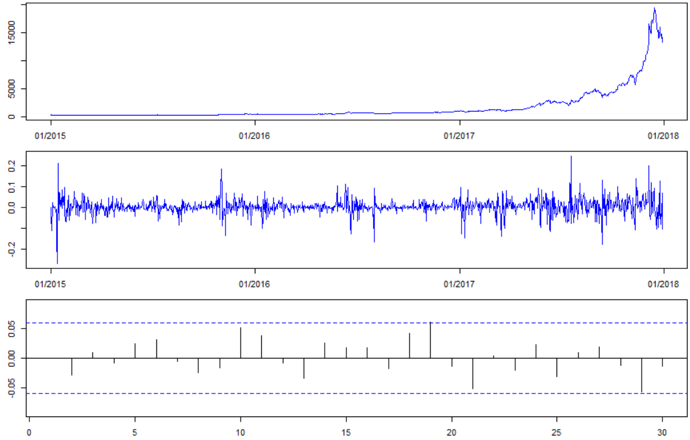


Figure: Bitcoin data from January 1, 2015 to December 31, 2017: price (top), logarithmic returns (center) and Autocorrelation function for Bitcoin logarithmic returns (bottom).

Basic statistical analysis of Bitcoin Returns

Table: Summary Statistics of daily returns

	Whole sample	1st subsample	2nd subsample
Min.	-0.4783	-0.4783	-0.2686
Median	0.0020	0.0011	0.0027
Mean	0.0036	0.0038	0.0035
Max.	0.3590	0.3590	0.2466
Standard Dev.	0.0457	0.0525	0.0377
Skewness	-0.7615	-0.9469	-0.1688
Kurtosis	20.2346	20.3505	10.9297
JB-Test p-value	0.0000	0.0000	0.0000

Annualized (historical) volatility around $86\%, 99.5\%, 71.5\%$ against typical values of $20\% - 30\%$ in the stock market and $5\% - 10\%$ in FX exchange rates.

Modeling Bitcoin price dynamics and its drivers: state of the art

The recent increasing trend in Bitcoin prices has pushed a new interest in the modeling of its returns.

- In Dyhrberg [5] a linear model is suggested to describe Bitcoin returns by taking into account other financial risk factors such as Stock market indexes, fiat currency exchange rates and Gold spot and future prices. No strong significance is found but for Gold.
- In Katsiampa [6] the author compares several GARCH model specifications to model Bitcoin returns and volatility without adding any explanatory variable.
- In Blau [2] the author identifies the dynamic relation between speculation activity (suitably measured) in a linear model with GARCH errors to account for heteroschedasticity.
- In Ciaian et al. [4] the authors analyse the dependence of Bitcoin price on several market forces such as supply and demand for Bitcoins, Stock market indices and oil price, and also include several attractiveness factors.

Modeling Bitcoin price dynamics and its drivers: state of the art II

Many papers have suggested that Bitcoin price and returns are affected by market attention

- In Ciaian et al. [4] attractiveness of Bitcoin is measured by means of the number of Wikipedia inquiries on the topic, the number of new users and the number of posts in a the online forum `https://bitcointalk.org/`. By estimating Vector AutoRegressive and Vector Error Correction models, they find that such variables are significant in explaining Bitcoin price behavior.
- In Bukovina and Martiček [3] it is claimed that Bitcoin price is driven by investors sentiment measured as positive, neutral or negative accordng to sentiment data obtained from `http://sentdex.com/`, a online platform specialized on natural language processing algorithms.
- Kristoufek [8, 7] the author applies techniques from econophysics, such as wavelets analysis, to prove that possible driving factors for Bitcoin price are Google searches, Wikipedia requests or more traditional indicators such as the number or volume of transactions

A discrete time proposal

Our first goal is to investigate whether and to which extent market attention influences the dynamics of Bitcoin returns.

- Serial dependence and Heteroschedasticity of returns suggest the choice of a linear model within the AutoRegressive Moving Average framework (ARMA) where innovations are described in the Generalized AutoRegressive Conditional Heteroschedastic (GARCH) family;
- We include a measure of attention is added as an explanatory variable in the model specification.
- We perform a econometric model selection within the suggested setting by considering different measures for the Attention Factor

Models withing the GARCH family have also been applied in this field in [5, 6].

A discrete time proposal: model specfcation

We model Bitcoin returns as an ARMA(1,1) process where an attention factor $A = \{A_t, t \geq 0\}$ is added as explanatory variable:

$$R_t = a_0 + a_1 R_{t-1} + b_1 \epsilon_{t-1} + cA_t + \epsilon_t, \quad t \geq 0,$$

where $\epsilon = \{\epsilon_t, t \geq 0\}$, the error process, is assumed to vary as a i.e. $\epsilon_t = \sqrt{h_t}\eta_t$, where $\eta = \{\eta_t, t \geq 0\}$ is a strong Gaussian noise and

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} + \gamma A_t$$

or

$$\log h_t = \alpha_0 + \alpha_1 \eta_{t-1} + \beta_1 \log h_{t-1} + \lambda \left(|\eta_{t-1}| - \mathbb{E}\left[|\eta_{t-1}|\right]\right) + \gamma A_t$$

known as GARCH(1,1) and EGARCH(1,1) model, respectively.
The attention process may also affect process h_t i.e. the variance/volatility of Bitcoin returns.

Measures of market attention

How to measure market attention?

- The total *trading volume* has been proven to measure for investors' attention (Barber and Odean (2008), Gervais, Kaniel, and Mingelgri (2001), and Hou, Peng, and Xiong (2008))

- The *Google SVI search volume index* is also proven as an alternative direct measure of investors' attention (Da et. al., 2011)

Quoting Da et al. (2011): "the search volume is likely to be representative of the internet search behavior of the general population and more critically, search is a revealed attention measure: if you search for a stock in Google, you are undoubtedly paying attention to it. Therefore, aggregate search frequency in Google is a direct and unambiguous measure of attention"

We consider daily data for the average price of Bitcoin across main exchanges, obtained by `https://blockchain.info/`, from January 1, 2012 to December 31, 2017.

In order to account for time variability on outcomes we also split the available data in two subsamples:

- from January 1, 2012 to December 31, 2014;
- from January 1, 2015 to December 31, 2017.

Analysis of Bitcoin time series

- Serial dependence and heteroschedasticity of returns suggests the ARMA GARCH models;
- we add an exogenous process in the model specification representing market attention, ARMA-X GARCH-X;
- we estimate an ARMA(1,1)-X to start with and move to higher number of lagged variables when necessary;
- we also estimate nested models as well as a standard regression model (LR) on the attention process as a benchmark.

Models withing the GARCH family have also been applied in this field in [5, 6].

Model specification

The attention measure A_t is based on two sources of data:

- volume of transactions (`https://blockchain.info/`) ;
- SVI index (`https://googletrends/`).

Considering alternatively with P the volume of transactions or the SVI index, we define as possible measures of attention the variables:

- $X_1 := \log(P)$;
- $X_2 := \Delta \log(P)$;
- $X_3 := |X_2|$;

and we estimate the model replacing A_t respectively with X_1, X_2 and X_3. Best models are selected according to AIC and BIC criteria.

Empirical results using traded volume

Table: AIC and BIC model selection analysis (whole series)

Model		AIC	BIC	X_m	X_v
LR	EGARCH(1,1)	-8507.43	-8479.17	-	-
ARMA(1,1)	EGARCH(1,1)	-8510.50	-8470.63	-	-
LR-X_1	**EGARCH(1,1)-X_2**	**-8647.00**	**-8607.12**	****	****
ARMA(1,1)-X_1	**EGARCH(1,1)-X_2**	**-8648.10**	**-8596.83**	****	****
AR(6)^2-X_1	**EGARCH(1,1)-X_2**	**-8648.32**	**-8597.05**	****	****

Columns from left to right represent the mean equation model, the variance equation model, the AIC, the BIC, the significance of the explanatory variable in the mean X_m, the significance of the explanatory variable in the variance X_v.

* $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$.

† AR(6)-X with parameters $a_2 = a_3 = a_4 = a_5 = 0$.

Empirical results using traded volume

Table: Diagnostics for competing non-linear models (whole series)

Test		p-value			
		lag=5	lag=10	lag=15	lag=20
LR-X_1 EGARCH(1,1)-X_2	Ljung-Box Q	0.0006	0.0002	0.0001	0.0001
	Ljung-Box Q^2	0.0052	0.1176	0.3588	0.6294
	Engle's Arch	0.0378	0.3030	0.5920	0.8102
ARMA(1,1)-X_1 EGARCH(1,1)-X_2	Ljung-Box Q	0.0170	0.0039	0.0029	0.0024
	Ljung-Box Q^2	0.0071	0.1422	0.4030	0.6832
	Engle's Arch	0.0475	0.3421	0.6368	0.8469
AR$(6)^2$-X_1 EGARCH(1,1)-X_2	Ljung-Box Q	0.0217	0.0343	0.0256	0.0200
	Ljung-Box Q^2	0.0057	0.1218	0.3615	0.6452
	Engle's Arch	0.0507	0.3408	0.6347	0.8536

† AR(6)-X with parameters $a_2 = a_3 = a_4 = a_5 = 0$.

Empirical results using traded volume

In all of the three competing models the residuals are rejected to be heteroschedastic. Since the gain in explaining serial correlation is not much across the three models and might worsen the BIC values, we believe that the overall best is the simple LR-X_1 EGARCH(1,1)-X_2 model.

Table: Parameter estimates for the LR-X_1 EGARCH(1,1)-X_2 model (whole series)

	Estimate	Std. Error	t value	Pr($>|t|$)
a_0	-0.061991	0.000854	-72.5654	0.00000
c	0.005296	0.000071	74.1544	0.00000
α_0	-0.173573	0.031312	-5.5434	0.00000
α_1	0.018756	0.010924	1.7169	0.08599
β_1	0.970030	0.004689	206.8636	0.00000
λ	0.263746	0.020464	12.8885	0.00000
γ	2.380453	0.195636	12.1678	0.00000

Empirical results using traded volume

Performing the same analysis separately for each subsample we select:

- AR(1)-X_1 EGARCH(1,1)-X_2 (1st subsample);
- LR EGARCH(1,1)-X_2 (2nd subsample).

Test		p-value			
		lag=5	lag=10	lag=15	lag=20
1st subsample	Ljung-Box Q	0.5603	0.5020	0.4419	0.4438
	Ljung-Box Q^2	0.4131	0.7642	0.9492	0.9913
	Engle's Arch	0.7411	0.9045	0.9846	0.9975
2nd subsample	Ljung-Box Q	0.2067	0.0427	0.0640	0.1114
	Ljung-Box Q^2	0.0030	0.0471	0.1385	0.1507
	Engle's Arch	0.0227	0.1238	0.4585	0.4748

Empirical results using SVI index

Table: AIC and BIC model selection analysis (whole series)

	Model	AIC	BIC	X_m	X_v
LR	EGARCH(1,1)	-8507.43	-8479.17	-	-
ARMA(1,1)	EGARCH (1,1)	-8510.50	-8470.63	-	-
LR-$\mathbf{X_2}$	**EGARCH(1,1)-$\mathbf{X_2}$**	**-8933.80**	**-8893.93**	**	****
AR(6)^2-$\mathbf{X_2}$	**EGARCH(1,1)-$\mathbf{X_2}$**	**-8937.09**	**-8880.12**	****	****

Table: Columns from left to right represent the mean equation model, the variance equation model, the AIC, the BIC, the significance of the explanatory variable in the mean X_m, the significance of the explanatory variable in the variance X_v.
* $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$.
† AR(6)-X with parameters $a_2 = a_3 = a_4 = 0$.

Empirical results using SVI index

Table: Diagnostics for competing non-linear models (whole series)

Test		p-value			
		lag=5	lag=10	lag=15	lag=20
LR-X_2 EGARCH(1,1)-X_2	Ljung-Box Q	0.0014	0.0004	0.0006	0.0005
	Ljung-Box Q^2	0.3397	0.3446	0.6982	0.9090
	Engle's Arch	0.6530	0.5630	0.8250	0.9481
AR(6)^2-X_2 EGARCH(1,1)-X_2	Ljung-Box Q	0.0095	0.0484	0.0723	0.0564
	Ljung-Box Q^2	0.3284	0.4332	0.7822	0.9449
	Engle's Arch	0.6328	0.6467	0.8748	0.9648

† AR(6)-X with parameters $a_2 = a_3 = a_4 = 0$.

The overall best in this case is the AR(6)^2-X_2 EGARCH(1,1)-X_2.

Empirical results using SVI index

Table: Parameter estimates for the AR(6)^2-X_2 EGARCH(1,1)-X_2 model (whole series)

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
a_0	0.001585	0.000453	3.5016	0.000462
a_1	0.029688	0.020891	1.4211	0.155287
a_5	0.023004	0.020752	1.1085	0.267636
a_6	0.048577	0.017025	2.8533	0.004327
c	0.010218	0.002471	4.1355	0.000035
α_0	-0.157260	0.024078	-6.5313	0.000000
α_1	-0.050700	0.011899	-4.2610	0.000020
β_1	0.974485	0.003479	280.0698	0.000000
λ	0.259720	0.020271	12.8121	0.000000
γ	2.934440	0.150161	19.5420	0.000000

† AR(6)-X with parameters $a_2 = a_3 = a_4 = 0$.

Empirical results using SVI index

Performing the same analysis separately for each subsample we select:

- LR EGARCH(1,1)-X_2 (1st subsample);
- LR-X_1 EGARCH(1,1)-X_2 (2nd subsample).

Test		p-value			
		lag=5	lag=10	lag=15	lag=20
1st subsample	Ljung-Box Q	0.0017	0.0035	0.0107	0.0116
	Ljung-Box Q^2	0.3158	0.1040	0.2488	0.3533
	Engle's Arch	0.0682	0.1569	0.3825	0.3674
2nd subsample	Ljung-Box Q	0.2006	0.4115	0.7125	0.7783
	Ljung-Box Q^2	0.0316	0.1175	0.1528	0.2347
	Engle's Arch	0.0227	0.1238	0.4585	0.4748

Bitcoin derivatives

- A market for derivatives on Bitcoin is available online in platforms such as `https://coinut.com` and `https://deribit.com` trading European Calls and Puts
- The Chicago Board Options Exchange (CBOE) has launched standardized Future contracts on the cryptocurrency in December 2017; this might give rise to a new era for Bitcoin trades and open the way to other standardized derivatives.

Motivated by the quoted papers and by potential interest in Bitcoin options, we build a model in continuous time, depending on an attention factor, to describe the dynamics of Bitcoin prices and provide a pricing formula for European style derivatives.

Building a model for price and market attention: our proposal

Motivated by the quoted papers and by new interest in Bitcoin contingent claims, we propose a continuous time model to describe Bitcoin price dynamics S, **depending on an attention factor** P, and also provide a pricing formula derivative assets.

A **possible delay** τ is considered between the attention factor and its effect on Bitcoin price changes.

$$\begin{cases} \mathrm{d}S_t = \mu_S P_{t-\tau} S_t \mathrm{d}t + \sigma_S \sqrt{P_{t-\tau}} S_t \mathrm{d}W_t, \quad S_0 = s_0 \in \mathbb{R}_+, \\ \mathrm{d}P_t = \mu_P P_t \mathrm{d}t + \sigma_P P_t \mathrm{d}Z_t, \quad P_t = \phi(t), \; t \in [-L, 0], \end{cases}$$

We prove **existence of a strong solution** to the above equations and we give (mild) conditions under which the market is **arbitrage-free**.

What we do: estimation I

Attention is measured by *trading volume* and *Google SVI index* and fit the model historical time series of Bitcoin prices from January 1, 2015 to March 31, 2017.

We derive **approximated closed formula for the likelihood** of a discrete sample obtained from the model and apply the **profile likelihood method** in order to estimate the *delay* parameter.

The *delay* parameter is **indeed relevant** in the model specification as we obtained $\tau = 5$ days for the *trading volume* and $\tau = 1$ week for the *Google SVI index*.

The volatility parameters σ_P and σ_S as well as the drift parameter μ_S are also **significant** for both *trading volume* and *SVI index*.

What we do: Option pricing I

A **quasi-closed formula** is derived for any European style derivative on Bitcoin and it does a quite good job!

K	Bid	Ask	CFP-Volume	CFP-Google	BS
2200	0.1662	0.2318	0.1981	0.2052	0.1967
2300	0.1670	0.2072	0.1674	0.1765	0.1655
2400	0.1390	0.1845	0.1429	0.1501	0.1369
2500	0.1142	0.1638	0.1182	0.1264	0.1112
2600	0.0922	0.1376	0.0965	0.1053	0.0887
2700	0.0749	0.1202	0.0776	0.0868	0.0695
2800	0.0572	0.1047	0.0616	0.0709	0.0535
2900	0.0442	0.0983	0.0483	0.0573	0.0405

Table: Option Prices with $S_t = 2710\$$, t=30 July and T=25 August: source `http://www.deribit.com` prices in BTC, strikes in USD

What we do: Option pricing II

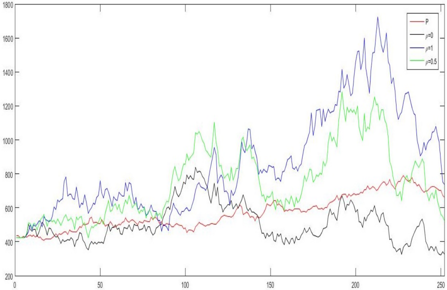
A deeper look...

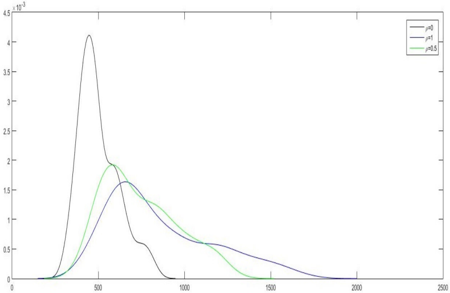
Options	N	CFP Volume	CFP Google	BS
All	144	**0.0257**	0.0290	0.0407
Very Shorts	16	0.0225	**0.0130**	0.0204
1 Months	32	0.0200	**0.0110**	0.0256
2 Months	32	**0.0231**	0.0426	0.0310
ITM	54	**0.0209**	0.0268	0.0356
ATM	36	0.0281	**0.0277**	0.0437
OTM	54	**0.0283**	0.0318	0.0432

Table: Root mean squared error (RMSE) between model and market prices

where

$$RMSE^2 = \frac{1}{N}\sum_{i=1}^{N}\left(C_i^{mod} - C_i^{mk}\right)^2$$



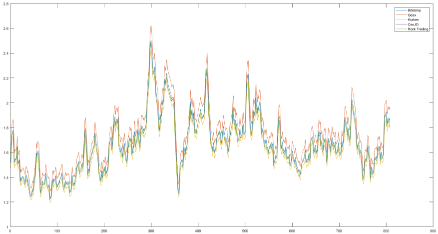


Referring to the mathematical theory of Bubbles (Protter [9], Bernard et al. [1]) we are able to prove, under suitable conditions, that the **above model is a bubble if and only if** the correlation parameter is above some threshold.

This is consistent with our conjecture: if the correlation parameter is too high there is a **stronger dependence among market attention and the Bitcoin price which boosts in a bubble**!

In progress: multi-exchange framework

Consider $\mathcal{I}$ exchanges trading Bitcoin in the same fiat currency and assume the price dynamics of $S^{(i)}$ are perfectly correlated and described by the above suggested model, with μ_i, σ_i, τ_i, $\forall\ i = 1, 2, \ldots, \mathcal{I}$.



No arbitrage in such a market is achieved iff the **market price of risk is the same** for all exchanges.

Not the case, as shown in the picture....

Figure: Dynamics for the market price of risk across exchanges

... so, at lest theoretically, **arbitrage opportunities are possible**, and this is indeed the case in real exchanges.

[1] Carole Bernard, Zhenyu Cui, and Don McLeish. On the martingale property in stochastic volatility models based on time-homogeneous diffusions. *Mathematical Finance*, 27(1):194–223, 2017.

[2] Benjamin M. Blau. Price dynamics and speculative trading in bitcoin. *Research in International Business and Finance*, 41(Supplement C): 493 – 499, 2017. ISSN 0275-5319. doi: https://doi.org/10.1016/j.ribaf.2017.05.010. URL http://www.sciencedirect.com/science/article/pii/S0275531917303057.

[3] Jaroslav Bukovina and Matúš Martiček. Sentiment and bitcoin volatility. Technical report, Mendel University in Brno, Faculty of Business and Economics, 2016.

[4] Pavel Ciaian, Miroslava Rajcaniova, and d'Artis Kancs. The economics of bitcoin price formation. *Applied Economics*, 48(19): 1799–1815, 2016.

References II

[5] Anne Haubo Dyhrberg. Bitcoin, gold and the dollar - a garch volatility analysis. *Finance Research Letters*, 16(Supplement C):85 – 92, 2016. ISSN 1544-6123. doi: https://doi.org/10.1016/j.frl.2015.10.008. URL http://www.sciencedirect.com/science/article/pii/S1544612315001038.

[6] Paraskevi Katsiampa. Volatility estimation for bitcoin: A comparison of garch models. *Economics Letters*, 158(Supplement C):3 – 6, 2017. ISSN 0165-1765. doi: https://doi.org/10.1016/j.econlet.2017.06.023. URL http://www.sciencedirect.com/science/article/pii/S0165176517302501.

[7] Ladislav Kristoufek. Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific reports*, 3:3415, 2013.

References III

[8] Ladislav Kristoufek. What are the main drivers of the bitcoin price? Evidence from wavelet coherence analysis. *PLoS ONE*, 10(4): e0123923, 2015.

[9] Philip E. Protter. Stochastic differential equations. In *Stochastic Integration and Differential Equations*, pages 249–361. Springer, 2005.

[10] David Yermack. Is bitcoin a real currency? An economic appraisal. Technical report, National Bureau of Economic Research, 2013.