



Common dynamic factors for cryptocurrencies and multiple pair-trading statistical arbitrages

Gianna Figá-Talamanca¹ · Sergio Focardi² · Marco Patacca³ 

Received: 31 March 2020 / Accepted: 4 January 2021 / Published online: 5 February 2021
© The Author(s) 2021

Abstract

In this paper, we apply dynamic factor analysis to model the joint behaviour of Bitcoin, Ethereum, Litecoin and Monero, as a representative basket of the cryptocurrencies asset class. The empirical results suggest that the basket price is suitably described by a model with two dynamic factors. More precisely, we detect one integrated and one stationary factor until the end of August 2019 and two integrated factors afterwards. Based on this evidence, we define a multiple long-short trading strategy which proves profitable when the second factor is stationary.

Keywords Cryptocurrencies · Cointegration · Dynamic factor models · Forecasting analysis · Pair-trading

JEL Classification C32 · C38 · C53 · C58

1 Introduction

Since their initial creation, most cryptocurrencies, in particular Bitcoin, proved to be highly volatile investment assets. For example, at the beginning of 2 January 2013, Bitcoin had a value of about 13 USD, reached a value of more than 19,000 USD in December 2017 and fell down to 7000 USD by the end of 2019.

✉ Marco Patacca
marco.patacca@univr.it

Gianna Figá-Talamanca
gianna.figatalamanca@unipg.it

Sergio Focardi
sergio.focardi@devinci.fr

¹ Department of Economics, University of Perugia, via A. Pascoli 20, 06123 Perugia, Italy

² Léonard de Vinci Pôle Universitaire, Research Center, 92916 Paris, La Défense Cedex, France

³ Department of Economics, University of Verona, Via Cantarane 24, 37129 Verona, Italy

Statistical analysis of cryptocurrencies has revealed a number of stylized facts, i.e. statistical findings that appear regularly in the time series under analysis; a non-exhaustive list of empirical studies is represented by Zhang et al. (2018, 2019), Bariviera et al. (2017), Hu et al. (2019) and Giudici and Pagnottoni (2020) which mainly focus on individual behaviour of cryptocurrencies. Findings similar to those of traditional financial assets include the non-stationarity of prices, which display an integrated $I(1)$ behaviour and, as a consequence, the stationarity of price differences; besides, returns show weak autocorrelation, while their absolute values are strongly autocorrelated.

The abnormal and unexpected returns of Bitcoin after 2016 have produced a surge in scientific research on the economics of cryptocurrencies. Several studies investigate the relationship between cryptocurrencies and other financial assets, such as commodities, currencies and market indexes, e.g. Dyhrberg (2016), Ciaian et al. (2016), Kang et al. (2019); besides, hedging and safe-haven properties of cryptocurrencies against the risk of stock markets has been investigated by Tiwari et al. (2019), Shahzad et al. (2019) and Bouri et al. (2017).

Many empirical papers also address the dependence on other candidate factors; Kristoufek (2015) elaborates on the relationship of Bitcoin prices with the number of transactions as well as with several crypto-related factors, such as the mining difficulty and the hashrate. Ciaian et al. (2016), Figà-Talamanca and Patacca (2019, 2020), Cretarola et al. (2020), Ahn and Kim (2019) and Eom et al. (2019) investigate whether Bitcoin returns and volatility are associated with investor attention, sentiment or by specific measures of market attractiveness.

Structural breaks and/or bubbles in the dynamics have been evidenced, among others, in Garcia et al. (2014), Cheah and Fry (2015), Fry and Cheah (2016), Corbet et al. (2018), Bouri et al. (2019), Cretarola and Figà-Talamanca (2019, 2020), Chaim and Laurini (2019) and Agosto and Cafferata (2019).

Most of the above contributions focus on Bitcoin either because of data availability or as a benchmark for the whole sector. Indeed, the price dynamics of several cryptocurrencies display common movements to that of Bitcoin, e.g. Ciaian and Rajcaniova (2018), Blau et al. (2020). Notably, in Yaya et al. (2019), the dependence between the price dynamics of Bitcoin and other cryptocurrencies is identified within the theory of fractional cointegration. Furthermore, Figà-Talamanca et al. (2020) evidence common regimes in the dynamics of several cryptocurrencies, by applying the theory of Hidden Markov models.

By taking advantage of the mutual dependence across several cryptocurrencies, it is reasonable to explore the possibility of forming market neutral strategies, where gains and losses depend only on the relative behaviour of assets. Market neutral strategies are well known to investors in stocks and other conventional assets. They are based on the principle that although the behaviour of each individual asset might not be forecastable, the relative behaviour of assets can be forecasted. Several statistical techniques help in forming market neutral strategies. Most strategies are based on some variant of cointegration or factor models. See Pole (2011) and Avellaneda and Lee (2010) for an introduction to market neutral strategies.

An attempt in this direction is given in Leung and Nguyen (2019): the authors, after evidencing the presence of statistical cointegration among several cryptocurren-

cies, detect the linear relationship between the cryptocurrencies under analysis and investigate the output of a spread statistical arbitrage which takes advantage of this specific association. Cointegration-based and other arbitrage strategies are analysed in Lintilhac and Tourin (2017), Bistarelli et al. (2018, 2019) by exploiting Bitcoin price differences across online trading exchanges, rather than considering multiple cryptocurrencies.

In this paper, we also build on statistical cointegration in order to create a market neutral strategy by investing in a basket of cryptocurrencies. However, our trading investment is based on the assumption that the multivariate price dynamics of the assets is suitably modelled by a dynamic factor model and proves profitable when the dynamics is described by one integrated and one stationary factor.

Precisely, we estimate the model on moving windows which include three years of previous daily observations starting on January 2019, of Bitcoin, Ethereum, Litecoin and Monero prices. Building on the above result, we suggest a trading strategy on multiple pair spreads, and we provide an empirical investigation of the final value of the strategy as well as of its time changes over the period from January 2019 to November 2019. Our findings suggest that the strategy is particularly profitable when the second factor is stationary, i.e. until the end of August 2019. The paper is organized as follows: in Sect. 2 we perform the preliminary statistical analysis to confirm stylized facts, in particular cointegration, for the basket under investigation; in Sect. 3, we introduce dynamic factor models as a general framework while in Sect. 4 we suggest and estimate a specific price dynamics. In Sect. 5, we define the proposed marked neutral strategy, detail its theoretical properties and provide empirical results. Finally, in Sect. 6 we give some concluding remarks.

2 Preliminary analysis

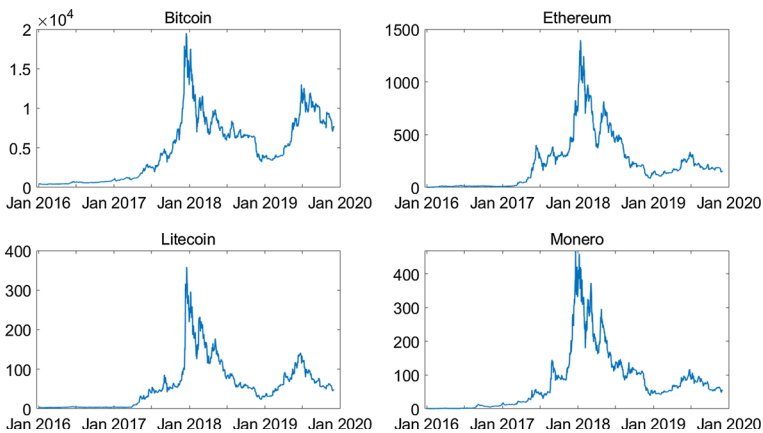
We consider, among the 20 cryptocurrencies with the highest market caps according to <https://coinmarketcap.com/> on December 2019, those which existed and traded by January 2016: we end up with a sample for the price of four cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC) and Monero (XMR), observed from January 2016 to the end of November 2019. The first three years of data are considered for estimation purposes only, while observations from January to November, 2019 serve as test-dates in order to evaluate the profitability of the market neutral strategy to be defined thereafter.

Table 1 summarizes descriptive statistics, and Figure 1 represents the prices behaviour of the four cryptocurrencies for the whole dataset.

In line with the study of stylized empirical facts for general financial markets (Cont 2001; Cont and Tankov 2004), we perform the Augmented Dickey Fuller (ADF) and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests on prices and price differences: the former tests the null hypothesis of unit root while the latter tests the null of trend stationarity against the alternative of a unit root. We then apply the Ljung-Box autocorrelation test to the whole basket of returns. Results, summed up in Table 2, show that all cryptocurrencies are integrated of order one ($I(1)$) while their

Table 1 Summary Statistics of Cryptocurrencies prices from January 1, 2016 to November 30, 2019

	BTC	ETH	LTC	XMR
Min.	364.33	0.94	3.00	0.43
Q_1	818.41	12.79	4.05	11.27
Median	4035.57	175.05	48.82	54.56
Mean	4830.50	226.47	57.59	79.10
Q_3	7833.04	298.33	78.77	105.11
Max.	19497.40	1396.42	358.34	469.20
St. Dev.	3974.16	248.35	60.12	87.70

**Fig. 1** Price behaviour for Bitcoin (top-left), Ethereum (top-right), Litecoin (bottom-left) and Monero (bottom-right) from January 1, 2016 to November 30, 2019

differences are stationary ($I(0)$). Besides, the weak autocorrelation of returns and the strong autocorrelation of their absolute values are confirmed by the Ljung-Box test.¹

As a preliminary check on whether a dynamic factor model may be suitable to describe the price dynamics of the basket, we also perform a cointegration analysis on the four cryptocurrencies; Table 3 displays the outcomes of the Johansen cointegration test², see Johansen (1991).

It is clear that the four cryptocurrencies are cointegrated with three cointegrating vectors and consequently, they share one common integrated $I(1)$ factor.

3 Dynamic factor models

Classical factor models have a long history which goes back to the formalization of psychometric models. Spearman (1904) introduced a one-factor model of mental

¹ We don't report the results in the paper but they are available upon request.

² We applied the function `jcitest.m`, with trace test specification, provided in the *Econometrics* Toolbox of Matlab®.

Table 2 ADF and KPSS tests of cryptocurrencies daily prices (Panel a) and price differences (Panel b) from January 1, 2016 to November 30, 2019

	BTC	ETH	LTC	XMR
Panel a: prices				
ADF-Test p value	0.4312	0.2319	0.1546	0.1382
KPSS-Test p value	<0.010	<0.010	<0.010	<0.010
Panel b: price differences				
ADF-Test p value	<0.001	<0.001	<0.001	<0.001
KPSS-Test p value	>0.100	>0.100	>0.100	>0.100

Note that 0.001 and 0.100 are, respectively, the minimum and the maximum p values provided by Matlab[®] *adftest.m* and *kpsstest.m* functions

Table 3 Johansen cointegration test between Bitcoin, Ethereum, Litecoin and Monero prices: data are from January 1, 2016 to November 30, 2019

r	stat value	c value	p value
0	105.0086	40.1751	0.0010
1	57.8639	24.2747	0.0010
2	15.7722	12.3206	0.0130
3	0.5850	4.1302	0.6356

abilities, Thurstone (1938, 1947) introduced the first multi-factor model, and Hotelling (1933) described principal components analysis. Classical factor models as described, for example, in Anderson (2003), are strict factor models with a finite number of variables. In a strict factor model, residuals are mutually uncorrelated and uncorrelated with factors. This implies that all correlations are due to factors.

In order to identify a strict factor model, additional assumptions are needed. The setting of classical strict static factor models is one of independent samples extracted from a population with a multivariate Gaussian distribution and observations are i.i.d. vectors. Time-varying behaviour is not accounted for when samples are taken at different points in time. Dynamic factor models generalize the above setting by allowing to specify dynamics for the factors and for the processes themselves.

While modern static multi-factor models were initially proposed in the early 30s by Hotelling (1933), the first dynamic factor models were introduced in econometrics, by Geweke (1977) and by Sargent and Sims (1977), more than forty years later. The subsequent development of dynamic factor models followed two lines:

- dynamic factor models of stationary processes;
- dynamic factor models of integrated processes.

3.1 Dynamic factor models of stationary processes

Sargent and Sims (1977) and Geweke (1977) proposed dynamic factor models to describe the behaviour of I stationary variables, when observed up to time T , by means of K factors. The authors assume that I is finite, $K \ll I$ and T very large.

In addition, factors and residuals are uncorrelated and that residuals are mutually uncorrelated though possibly autocorrelated.

The above assumptions were relaxed in Engle and Watson (1981), Sargent (1989) and Stock and Watson (1989) by allowing for a small number of variables; parameter estimates are obtained by maximizing the likelihood and factors are recovered through the Kalman filter. An alternative estimation procedure, when a small number of variables is considered together with a large number of observations, is to consider dynamic factor models as instances of state-space models (see Lütkepohl and Poskitt 1991).

In general, we can specify a dynamic factor model as follows:

$$\begin{aligned} r_t &= \beta f_t + \epsilon_t \\ \Phi(L) f_t &= \Theta(L) \eta_t \\ \Phi(L) &= 1 - \Phi_1 L - \dots - \Phi_p L^p \\ \Theta(L) &= 1 - \Theta_1 L - \dots - \Theta_q L^q \end{aligned} \quad (1)$$

where $r = \{r_t\}_{t=1,2,\dots,T}$ is a vector processes with I components, the β_i are $I \times K$ matrices, $f = \{f_t\}_{t=1,2,\dots,T}$ is the vector of K stationary process (the factors) and L is the lag operator. The error process $\epsilon = \{\epsilon_t\}_{t=1,2,\dots,T}$ is a white noise with a full covariance matrix, $\eta = \{\eta_t\}_{t=1,2,\dots,T}$ has a full-rank covariance matrix and is serially uncorrelated and ϵ and η are mutually uncorrelated at all lags. That is, the common dynamic structure comes only from the factors while the idiosyncratic components can be correlated but no autocorrelation is allowed.

3.2 Dynamic factor models of integrated processes

The model specification for dynamic factor models of integrated factors does not differ from (1) but allows for the possibility of considering non-stationary, i.e. integrated, factors. The estimation of such models introduced in Peña and Poncela (2006) generalizes the estimation methodology for the case of stationary factors put forward in Peña and Box (1987). The paper proposes a test for the number of common factors based on the analysis of the eigenvalues of the generalized covariance matrices and factors are estimated with maximum likelihood. Further, Peña and Poncela (2004) analyse the forecasting performance of such models.

The notion of a factor model of integrated processes is rooted in the concept of cointegration. There is a vast literature on cointegration and on determining the number of cointegrating relationships. Following Engle and Granger (1987), who were jointly awarded the 2003 Nobel Memorial Prize in Economic Sciences for the discovery of cointegration and autoregressive conditional heteroskedasticity (ARCH) behaviour, two or more integrated time series are cointegrated if there is a linear combination $\sum_{i=1}^I \alpha_i x_{it}$ of the series that is stationary. The linear combinations $\sum_{i=1}^I \alpha_i x_{it}$ that are stationary are called cointegrating relationships. As observed in Galeano and Peña (2000), the idea that two or more time series can be individually integrated but that a linear combination of the series is stationary had already been put forward by Box and Tiao (1977) in introducing canonical correlation analysis. The state-of-the-art

cointegration test is the Johansen test (Johansen 2000). A concise yet exhaustive presentation of cointegration is given in Hendry and Juselius (2000) and Hendry and Juselius (2001).

The first link between cointegration and dynamic factor models appeared in Stock and Watson (1988). This landmark paper demonstrates that if a set of I time series is cointegrated with K cointegrating relationships, then there are $Q = I - K$ integrated common trends and the I series can be described as regressions on the common trends. Later, Escribano and Peña (1994) established that common trends are equivalent to common dynamic factors in the sense that the existence of K cointegrating relationships is equivalent to the existence of $I - K$ dynamic integrated factors.

The outcomes of the cointegration analysis performed for the analysed cryptocurrencies together with the above remarks motivate the following section where a dynamic factor model with one integrated factor is suggested for modelling the basket price dynamics.

4 A dynamic factor model for cryptocurrencies

As already remarked, many cryptocurrencies are currently available for trading on online exchange platforms. It has been claimed in Ciaian and Rajcaniova (2018) and Blau et al. (2020) that the whole sector is possibly driven by the dynamics of Bitcoin (in particular in the short-run) or, alternatively, by a single common factor. Indeed, the price dynamics of several cryptocurrencies show a similar path, as evidenced in Fig. 1.

We build on these findings by assuming that the prices of a basket of cryptocurrencies may be described through a dynamic factor model. If this is the case, it would be reasonable to explore the possibility of forming market neutral strategies, where gains and losses depend only on the relative behaviour of the assets.

The dataset is split into two subsamples with the first three years of data (January 1, 2016–December 31, 2018) used only for model estimation and the remainder (January 1, 2019–November 30, 2019) as test dates for the performance of the market neutral strategy suggested in Sect. 5.

4.1 Model specification and fitting

Assume we are given with I different cryptocurrencies; based on our preliminary analysis, we model their mutual relationship through a dynamic factor model.

More precisely, we assume that, for $i = 1, 2, \dots, I$ and $t = 1, 2, \dots, T$:

$$\begin{aligned}
 p_{i,t} &= \alpha_i + \sum_{k=1}^K \beta_{ik} f_{k,t} + \epsilon_{i,t}, \\
 f_{k,t} &= \lambda_k f_{k,t-1} + \eta_{k,t}, \quad k = 1, 2, \dots, K \\
 \epsilon_{i,t} &= \phi_{i,1} \epsilon_{i,t-1} + \phi_{i,2} \epsilon_{i,t-2} + \dots + \phi_{i,p_i} \epsilon_{i,t-p_i} + u_{i,t}.
 \end{aligned} \tag{2}$$

where $p_{i,t}$ is the price at time t of the cryptocurrency i and $f_{k,t}$, for $k = 1, 2, \dots, K$ are the common factors. For each cryptocurrency i , $\epsilon_{i,t}$ is a centered process, representing the error term at time t , which may be autocorrelated and is suitably described by an autoregressive $AR(p_i)$ process; the error terms $u_{i,t}$ and $\eta_{k,t}$, $k = 1, 2, \dots, K$ are normally distributed i.i.d processes. We further assume, in line with Koop and Korobilis (2010), the independence between the factors and the error processes u_i, η_k, η_h , for $i = 1, 2, \dots, I$ and $k, h = 1, 2, \dots, K$, with $k \neq h$, and between the error process u_i and u_j , for $i, j = 1, 2, \dots, I$, with $i \neq j$.

Factors can be confidently estimated by means of principal components analysis in the limit case of an infinite market, in practice a very large market. When a limited number I of assets is given, dynamic factor models are usually estimated as state space models (Lütkepohl and Poskitt 1991). We used the Matlab[®] software for Bayesian models provided by Koop and Korobilis (2009), which adopts the latter methodology, since in the empirical application we have $I = 4$.

When focusing on the first three years of our specific sample ($T =$ December 31, 2018), the Johansen cointegration test confirms the existence of a common non-stationary factor, hence we expect one common factor, say f_1 , to be integrated $I(1)$, see Escribano and Peña (1994). In order to detect the correct number of common factors, we estimate, as a first step, the model in (2) with just one factor f_1 ; then we compute the covariance matrix of residuals $\epsilon_{i,t}, i = 1, 2, \dots, 4$ and corresponding eigenvalues, given by (1184763.89, 6946.14, 185.03, 14.32), which suggest the presence of a second relevant factor.

Therefore, we will assume $k = 2$ throughout our empirical investigation; the considered model specification is finally given by:

$$\begin{aligned}
 p_{i,t} &= \alpha_i + \beta_{i1}f_{1,t} + \beta_{i2}f_{2,t} + \epsilon_{i,t}, \\
 f_{1,t} &= \lambda_1 f_{1,t-1} + \eta_{1,t}, \\
 f_{2,t} &= \lambda_2 f_{2,t-1} + \eta_{2,t}, \\
 \epsilon_{i,t} &= \phi_{i,1}\epsilon_{i,t-1} + \phi_{i,2}\epsilon_{i,t-2} + \dots + \phi_{i,p_i}\epsilon_{i,t-p_i} + u_{i,t}.
 \end{aligned}
 \tag{3}$$

Stationarity or non-stationarity for the factors is established by the estimated values of the autoregressive parameters λ_1, λ_2 . As already noticed, the cointegration analysis suggests that f_1 is integrated $I(1)$, hence we expect $\lambda_1 = 1$ and $f_{1,t} = f_{1,t-1} + \eta_{1,t}$.

By iteration, we get $\mathbb{E}[f_{1,t}] = f_{1,0} := \mu$, which may be different from 0, and $Var(f_{1,t}) = t\sigma_{\eta_1}^2$ where $\sigma_{\eta_1}^2 = Var(\eta_{1,t})$. As a consequence $\mathbb{E}[p_{i,t}] = \alpha_i + \beta_{i1}\mu$.

Summing up, model (3) is fitted by applying the following estimation procedure:

- The model in (3) is estimated on demeaned prices to get parameters³ β_{ij} , and the time series for the two hidden factors.
- μ is estimated as the sample mean of the first factor and $\alpha_i, i = 1, 2, \dots, I$ are obtained by solving $\alpha_i = \bar{p}_{i,t} - \beta_{i1}\mu$ where \bar{p}_i is the sample mean of the process p_i .

³ We make use of Matlab[®] software for Bayesian models provided by Koop and Korobilis (2009); parameters are initialized with the values obtained by performing a standard Principle Component Analysis.

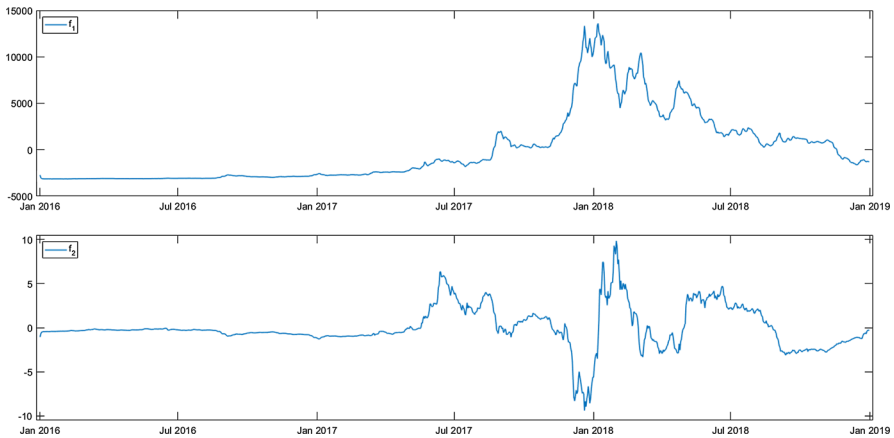


Fig. 2 Factors f_1 and f_2 from January, 2016 to December, 2018

Table 4 Parameter estimates from January, 2016 to December, 2018

	BTC ($i = 1$)	ETH ($i = 2$)	LTC ($i = 3$)	XMR ($i = 4$)
α_i	4045.7660	238.8579	53.1539	82.4727
β_{i1}	0.9911	0.0700	0.0170	0.0260
β_{i2}	0.1406	5.2255	0.0351	-0.0542

The estimated hidden factors f_1 , f_2 are plotted in Fig. 2: the first factor is an $I(1)$ process as suggested by our model specification, since the Augmented Dickey fuller test confirms that the estimated value of λ_1 is not significantly different from the unit ($\lambda_1=0.9962$, ADF p value= 0.2712^4) while the second factor is a stationary process ($\lambda_2 = 0.9823$, ADF p value= 0.0056^4), independent of f_1 . Corresponding parameter estimates α_i , β_{ij} , $i = 1, 2, 3, 4$ and $j = 1, 2$ are reported in Table 4; the estimated value for μ is negligible.

It is clear from both the estimates and the plots that the first factor essentially emulates the dynamics of Bitcoin: indeed $\beta_{11} = 0.9911$ is the largest coefficient and it is very close to one. Figure 3, where the first factor and Bitcoin are overlaid, confirms this evidence. Besides, the second factor is strongly related to Ethereum ($\beta_{22} = 5.2255$) though its scale is larger.

In order to check whether the model in (3) is consistent with observed data we also perform usual diagnostics on factors and residuals u_i , $i = 1, 2, 3, 4$, detailed results are reported in Appendix 1 (Tables 7, 8). Specifically the correlations between two factors do not significantly differ from zero ($\rho_{f_1, f_2} = 0.005$) as well as the correlation among residuals. Besides, the Ljung-Box autocorrelation test, for different values of maximum lag H^5 , of the four residual processes is quite satisfactory, with the exception of Ethereum for which the process u_2 still shows some residual autocorrelation.

⁴ The p values are approximated since the common factors are estimated rather than observed variables.

⁵ We follow Tsay (2005) where it is suggested to select $H \leq \ln n$ where n =length of the time series.

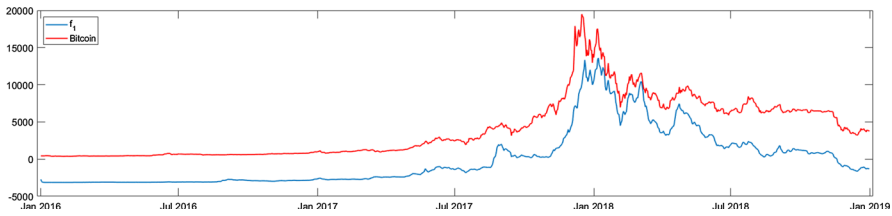


Fig. 3 First factor f_1 and Bitcoin price from January, 2016 to December, 2018

For completeness we also sum-up in Appendix 1 (Table 9) the estimates of the $AR(p_i)$ model for the error terms ϵ_i where the autoregression order p_i has been selected by minimizing the Bayesian Information Criteria, according to Rachev et al. (2007). It is worth noticing that above estimates are not necessary in order to build a market neutral strategy to exploit the relationship between the analysed cryptocurrencies, as it will be clear in the next section.

Since we are interested in building an investment strategy to take advantage of the dynamic model defined in (3), we can, without loss of generality, scale the price equations with the corresponding β_{i1} coefficients. Hence, we get:

$$p_{i,t}^* = \frac{p_{i,t}}{\beta_{i1}} = \frac{\alpha_i}{\beta_{i1}} + f_{1,t} + \frac{\beta_{i2}}{\beta_{i1}} f_{2,t} + \epsilon_{i,t} \tag{4}$$

Notably, the scaled prices are equal to the integrated factor $f_{1,t}$ plus a stationary process $f_{2,t}$ and a constant term, which are the same for all rescaled prices; in this way, the gain from a simple long-short strategy with any pair of assets is only driven by the stationary factor f_2 . Specifically, the difference (spread) between the prices of cryptocurrencies $i, k \in \{1, 2, \dots, I\}$ is given by the following equation:

$$d_{ik,t} = \frac{p_{i,t}}{\beta_{i1}} - \frac{p_{k,t}}{\beta_{k1}} = \left(\frac{\alpha_i}{\beta_{i1}} - \frac{\alpha_k}{\beta_{k1}} \right) + \left(\frac{\beta_{i2}}{\beta_{i1}} - \frac{\beta_{k2}}{\beta_{k1}} \right) f_{2,t} + \epsilon_{ik,t}. \tag{5}$$

The above difference process will be crucial in defining a market neutral strategy on the available basket of cryptocurrencies.

5 Market neutral strategy

Let us assume that we are at time τ and that the price of our basket of cryptocurrencies is described by the model in (3). Once model parameters have been estimated on a time series of prices observed up to a varying time $T = \tau$, we obtain one-day ahead forecasted prices as:

$$\hat{p}_{i,\tau+1} = \mathbb{E}_\tau (p_{i,\tau+1}) = \hat{\alpha}_i + \hat{\beta}_{i1} \mathbb{E}_\tau (f_{1,\tau+1}) + \hat{\beta}_{i2} \mathbb{E}_\tau (f_{2,\tau+1}) \tag{6}$$

where

$$\mathbb{E}_\tau (f_{1,\tau+1}) = \hat{\lambda}_1 f_{1,\tau} \quad , \quad \mathbb{E}_\tau (f_{2,\tau+1}) = \hat{\lambda}_2 f_{2,\tau}. \tag{7}$$

Hence, the one-day-ahead forecasted difference in $\tau + 1$ for any pair $i, k \in \{1, 2, \dots, I\}$ of cryptocurrencies is given by

$$\begin{aligned} \hat{d}_{ik,\tau+1} &= \left(\frac{\hat{\alpha}_i}{\hat{\beta}_{i1}} - \frac{\hat{\alpha}_k}{\hat{\beta}_{k1}} \right) + \left(\frac{\hat{\beta}_{i2}}{\hat{\beta}_{i1}} - \frac{\hat{\beta}_{k2}}{\hat{\beta}_{k1}} \right) \mathbb{E}_\tau (f_{2,\tau+1}) \\ &= \left(\frac{\hat{\alpha}_i}{\hat{\beta}_{i1}} - \frac{\hat{\alpha}_k}{\hat{\beta}_{k1}} \right) + \left(\frac{\hat{\beta}_{i2}}{\hat{\beta}_{i1}} - \frac{\hat{\beta}_{k2}}{\hat{\beta}_{k1}} \right) \hat{\lambda}_2 f_{2,\tau} \end{aligned} \tag{8}$$

If the above difference is strictly positive, a future revenue can be obtained by applying a long-short investment in the pair i, k . The forecasted difference only depends on the second factor at time τ . As the first factor may be interpreted as the market factor, the above pair trading strategy is market neutral. In order to maximize the revenue in $\tau + 1$, a multiple pair trading strategy is obtained by investing on several pairs $i, k \in \{1, 2, \dots, I\}$ which display a non-negative forecasted difference. More precisely, if we denote with $\hat{p}_{\tau+1}^{(i)}, i = 1, 2, \dots, I$ the ordinal statistics (in decreasing order) of the scaled forecasted prices for time $\tau + 1$, the multiple pair trading consists essentially of short positions on the first half of cryptocurrencies (with higher forecasted prices) and long positions in the second half of cryptocurrencies (with lower forecasted prices). Denote with v_τ the value at time τ of the above investment portfolio. The one-day-ahead expected value of the strategy, computed at time τ , is given by

$$g_{\tau+1} = \mathbb{E}_\tau [v_{\tau+1}] = \sum_{i=1}^{\lfloor I/2 \rfloor} \hat{p}_{\tau+1}^{(i)} - \sum_{i=\lceil I/2 \rceil+1}^I \hat{p}_{\tau+1}^{(i)} \tag{9}$$

where $\lfloor I/2 \rfloor, \lceil I/2 \rceil$ are, respectively, the floor and ceil rounding of $I/2$.

In general, we will adopt the above pair trading strategy (go long with the multiple pair trading) whether $g_{\tau+1} > v_\tau$ or the opposite strategy (going short in the pair trading) in case $g_{\tau+1} < v_\tau$. In addition, in order to avoid huge transaction costs, the above multiple pairs strategy can be optimized by trading only when the difference between the forecasted and current value of the investment is above a fixed threshold.

Specifically, we suggest the following strategy:

- if $g_{\tau+1} > v_\tau + c\sigma_\tau^v$, go long with the multiple pair trading,
- if $g_{\tau+1} < v_\tau - c\sigma_\tau^v$, go short with the multiple pair trading,
- if $v_\tau - c\sigma_\tau^v \leq g_{\tau+1} \leq v_\tau + c\sigma_\tau^v$, hold the current positions (no trade),

where c is an arbitrary chosen constant and σ_τ^v is the standard deviation of the trading position value corresponding to the basket price time series observed up to time τ . If $c = 0$ the trading strategy reduces to the multiple pair trading defined above.

Table 5 Observed and forecasted data available at time τ

	BTC ($i = 1$)	ETH ($i = 2$)	LTC ($i = 3$)	XMR ($i = 4$)
β_{i1}	0.9911	0.0700	0.0170	0.0260
$p_{i,\tau}$	9000.00	300.00	40.00	75.00
$\bar{p}_{i,\tau}$	5000.00	150.00	20.00	40.00
$\hat{p}_{i,\tau+1}$	8950.00	295.00	45.00	78.00
$\hat{p}_{i,\tau}^*$	3985.47	2071.43	1470.59	1461.54

Example 1 Assume we are in time τ , we have estimated the model in (3) up to time τ and we have the data reported in Table 5 where all prices are given in USD.

Recall that, for $i = 1, 2, 3, 4$, β_{i1} is the coefficient of the first factor in (3); $\bar{p}_{i,\tau}$ is the mean price up to time τ , $p_{i,\tau}$ is the observed price at time τ while $\hat{p}_{i,\tau+1}$ is the one-day ahead price forecast (computed at time τ). Finally, $\hat{p}_{i,\tau}^*$ is the scaled price defined in (4).

After computation of the ordinal statistics $\hat{p}_{\tau+1}^{(i)}$, $i = 1, 2, \dots, I$, the suggested multiple pair trading strategy is based on selling the highest two ranked cryptocurrencies (BTC, ETH) and buying the lowest two (LTC, XMR). The forecasted one-day ahead gain $g_{\tau+1}$ of above positions is given by:

$$g_{\tau+1} = \mathbb{E}_{\tau} [v_{\tau+1}] = (8950.00 + 295.00 - 45.00 - 78.00)\$ = 9122.00\$ \quad (10)$$

The decision on whether to trade the above strategy and its direction (long or short) depends on the comparison between $g_{\tau+1}$ and $v_{\tau} \pm c\sigma_{\tau}^v$ where $v_{\tau} = (9000.00 + 300.00 - 40.00 - 75.00) = 9185.00$. If we have $c = 0.1$, $\sigma_{\tau}^v = 500.00$, we get $g_{\tau+1} = 9122.00 < v_{\tau} - c\sigma_{\tau}^v = 9135.00$; hence, we short the suggested multiple pair trading and collect 9185.00. If we further assume that future prices at time $\tau + 1$ are given by $p_{1,\tau+1} = 8900$, $p_{2,\tau+1} = 290$, $p_{3,\tau+1} = 43.00$ and $p_{4,\tau+1} = 80.00$, we get $v_{\tau+1} = 9067.00$ and we are allowed to close the strategy with a net gain $G = (9185 - 9067) = 118.00$.

The suggested investment selection can be generalized to a dynamical setting where the model in (3) is estimated on moving windows and every long-short portfolio is liquidated on the following date. The above strategy provides a non-negative payoff, for each date $s \in \{\tau + 1, \tau + 2, \dots, \tau + M\}$, on the condition that the basket prices are properly described by model (3) in the corresponding moving window. Hence, usual diagnosis tests should be repeated for each moving window.

If the above trading strategy is repeated for m consecutive days $\{\tau + 1, \tau + 2, \dots, \tau + m\}$ then the expected cumulative gain in $\tau + m$ is given by

$$G_{\tau+m} = \sum_{l=\tau}^{\tau+m-1} [g_{l+1} - v_l] \mathbb{1}_{trade(l)}, \quad (11)$$

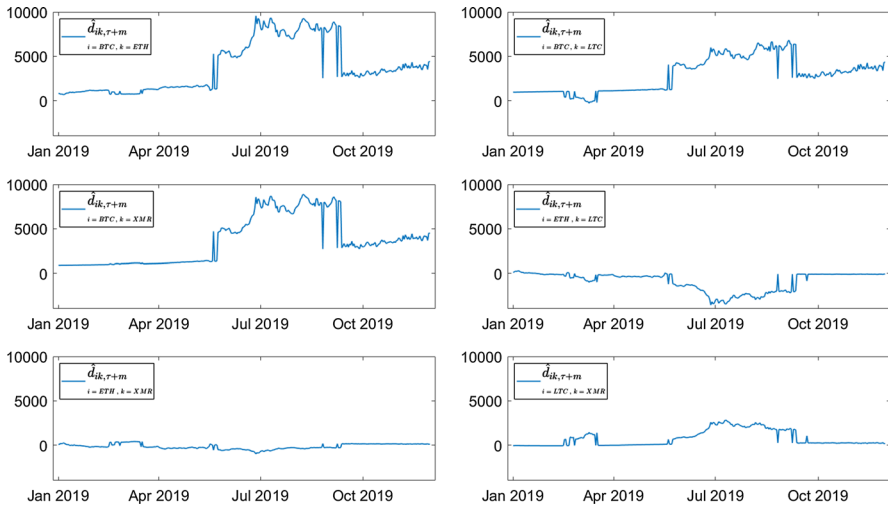


Fig. 4 Forecasted value of differences $\hat{d}_{ik, \tau+m}$ from $\tau + 1 = \text{January 1, 2019}$ to $\tau + M = \text{November 30, 2019}$

where the indicator function $\mathbb{1}_{trade(l)}$ is defined by

$$\mathbb{1}_{trade(l)} = \begin{cases} 1 & \text{if there is trade at time } l \\ 0 & \text{if there is no trade at time } l. \end{cases} \tag{12}$$

5.1 Empirical results

In this subsection, we provide the empirical results obtained by applying the market neutral strategy proposed above to the daily prices of Bitcoin, Ethereum, Litecoin and Monero, from January 1, 2019–November 30, 2019, i.e. a total of $M = 334$ days. Each day the dynamic factor model is estimated on a moving window of daily observations available for the previous three years ($T = 1096$ observations). Precisely, the first window starts in January 2016 and ends in December 2018, then the window moves one-day-ahead so that the last one runs from November 30, 2016 to November 29, 2019.

Once parameters are estimated on each moving window, the forecasted prices for the four cryptocurrencies in the basket are computed using the rule in (6). Since our strategy depends on the forecasted differences in (5), Fig. 4 displays these values for all pairs in the basket.

The suggested trading strategy is repeated each day from January 1, 2019 to November 30, 2019, after fitting the model in (3) on a three years long moving window; the suggested trading strategy has been shown to be (theoretically) profitable under specific assumptions: the existence of a non-stationary and a stationary factor and their mutual independence (or, at least, weak correlation). In order to consider a dynamic generalization of the strategy, the above hypothesis should be verified on each moving window and the actual trading should be conditioned on their validity. Hence, we

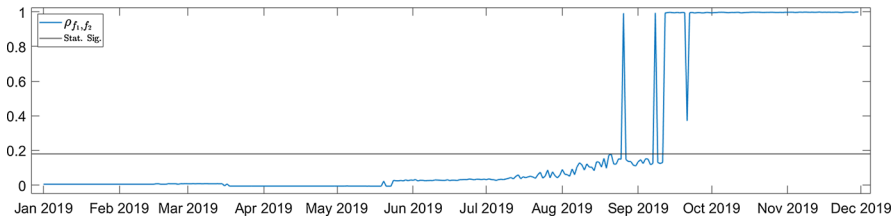


Fig. 5 Correlation between factors f_1, f_2 for all moving windows ($m = 1, 2, \dots, M$). The threshold level 0.18 for no trade is also plotted as an horizontal line

Table 6 Summary statistics of cumulative gain $G_{\tau+M}$ and net cumulative gain $G_{\tau+M}^*$

	$c = 0.00$	$c = 0.10$	$c = 0.20$	$c = 0.30$	$c = 0.50$	$c = 1.00$
Panel a: $G_{\tau+M}$						
Trade n.	252.00	232.00	222.00	200.00	172.00	72.00
Mean	7625.65	7620.59	7623.34	7615.08	7554.18	7467.75
St. Dev.	5474.23	5472.13	5474.02	5467.97	5477.52	5580.96
Panel b: $G_{\tau+M}^*$, taking into account transaction fees						
Trade n.	252.00	232.00	222.00	200.00	172.00	72.00
Mean	3031.17	3028.80	3032.97	3027.38	2973.89	2883.38
St. Dev.	2596.05	2594.68	2598.39	2593.16	2608.96	2720.23

avoid trading when both factors f_1 and f_2 are integrated and/or highly correlated. In the empirical application, the suggested market neutral strategy is applied only when the mutual correlation between factors is below 0.18⁶. Figure 5 shows the value of correlation between factors f_1, f_2 for all moving windows ($m = 1, 2, \dots, M$) and the threshold level.

In order to appreciate the influence of the arbitrary constant c on the cumulative gain of our strategy, Table 6, Panel a, shows the summary statistics of the cumulative gain at time $\tau + M, G_{\tau+M}$, corresponding to several choices of c . In addition, to evaluate the performance of the suggested trading strategy in a real market, where transaction cost is associated to each trade, the net cumulative gain $G_{\tau+M}^*$ is also reported in Table 6, Panel b. The net gain is computed under the assumption that transaction fees are given by the 0.10% of the investment, which corresponds to the *maker fee* of Coinbase⁷ for the pricing tier from \$100k to \$1m of USD trading volume over the trailing 30-day period.

As expected, the number of trades decreases when c increases and the cumulative gain is maximized when $c = 0$ (Panel a); a wise selection appears to be $c = 0.20$, which maximizes the net cumulative gain, once transactions fees are accounted for (Panel b).

⁶ The correlation parameter is significant at the $\alpha = 0.05$ level when above the threshold 0.06. We choose 0.18 that is equal to three times the threshold to have more tolerance.

⁷ <https://pro.coinbase.com/fees>.

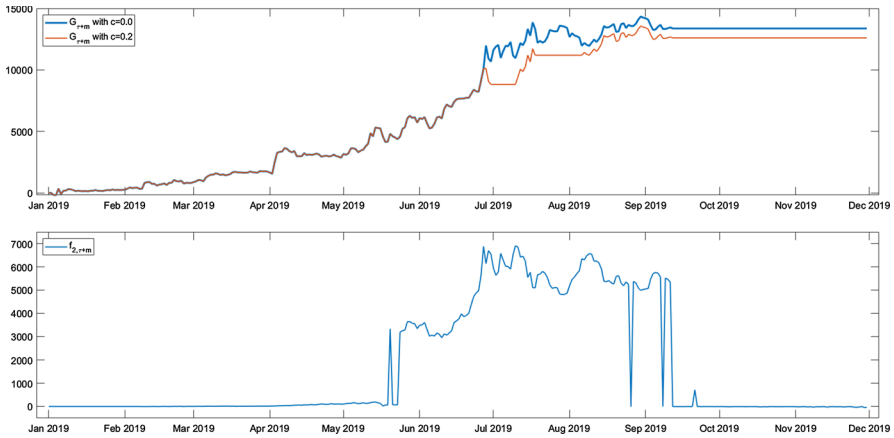


Fig. 6 Cumulative gain $G_{\tau+m}$ with parameter $c = 0.00$ and $c = 0.20$ (top) and forecasted value of factor $f_{2,\tau+m}$ (bottom) from $\tau + 1=$ January 1, 2019 to $\tau + M=$ November 30, 2019

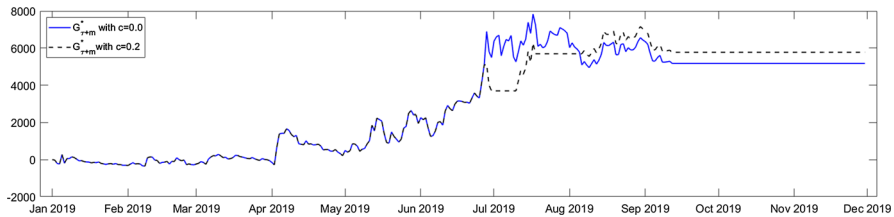


Fig. 7 Net cumulative gain $G_{\tau+m}^*$ with parameter $c = 0.00$ and $c = 0.20$ from $\tau + 1=$ January 1, 2019 to $\tau + M=$ November 30, 2019.

The time-varying dynamics of the cumulative gain $G_{\tau+m}$, $m = 1, 2, \dots, M$ when $c = 0.00$ and $c = 0.20$ is illustrated in the top-picture of Fig. 6. Notably, no transactions appear after September 2019; indeed, the outcomes from the Johansen cointegration test, applied within the time frame of each moving window, suggest the existence of two common integrated factors (rather than one), starting from August 20, 2019; moreover, the correlation between the two estimated factors is above the fixed threshold (see Fig. 5). These two evidences are in contrast with the basic assumptions underlying the proposed strategy and suggest no trading. Finally, it is worth noticing that the strategy is as much profitable as higher is the forecasted value of f_2 , plotted in Fig. 6 (bottom picture).

Similarly, in Fig. 7 we plot the time-varying net cumulative gain $G_{\tau+m}^*$, $m = 1, 2, \dots, M$ when $c = 0.00$ and $c = 0.20$. It is evident that, in both cases, the proposed strategy provides a net positive gain even when taking into account transaction costs.

6 Concluding remarks

In this paper, we suggest a dynamic factor model to describe the price dynamics of a basket of cryptocurrencies including Bitcoin, Ethereum, Litecoin and Monero,

observed daily from January 1, 2016–November 30, 2019. These were selected among the 20 cryptocurrencies with highest market cap, according to <https://coinmarketcap.com/> on December 2019, which existed and traded for at least three years. The outcomes of our analysis confirm the presence of cointegration as already evidenced in Ciaian and Rajcaniova (2018) and Blau et al. (2020) and show the appropriateness of dynamic factor models to describe the price process of the whole basket. Indeed, it is evidenced that the basket is driven by two common, dynamic factors, the first of which is a non-stationary $I(1)$ process. In order to check the consistency over time of the suggested model specification, the estimation is repeated on three-year long moving windows. By applying usual diagnostic checks, it is proven that dynamic factor models provide a satisfactory fit throughout the analysed period. It is worth noticing that the second factor displays a stationary behaviour on moving windows series observed until the end of August 2019 while it is integrated $I(1)$ afterwards. Besides, the correlation between the two factors increases within the same period. We stress that our analysis is limited to a basket of four cryptocurrencies since dynamic factor models rely on a large set of parameters which need to be estimated on a sufficiently long time series. Nevertheless, we conjecture that analogous results would be obtained on a larger basket if longer time series were available.

The common factor dynamics depicted in the first part of the paper is exploited in order to define a market neutral strategy consisting of suitably scaled long short investments on suitably selected cryptocurrency pairs. Theoretical properties of the suggested strategies are analysed: in particular the time-varying value for the cumulative gain is computed daily from January to November, 2019. Our findings suggest that the proposed trading strategy is particularly profitable until the second factor remains stationary. By taking into account transaction fees, according to the trading rules imposed by the Coinbase exchange, we also prove that the net cumulative gain remains positive throughout the considered period. A yet more profitable trading strategy might be defined by investing in multiple (rather than two) pairs, once a larger basket is available.

Funding Open Access funding provided by Università degli Studi di Verona. Fondazione Cassa di Risparmio di Perugia Grant no. 2018.0427.021.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

See Tables 7, 8 and 9.

Table 7 Correlation matrix between the two factors and residuals u from January, 2016 to December, 2018

	f_1	f_2	u_1	u_2	u_3	u_4
f_1	1.0000	0.0050	-0.0568	-0.0204	-0.0075	0.0472
f_2	0.0050	1.0000	0.0600	-0.0413	0.0256	-0.0279
u_1	-0.0568	0.0600	1.0000	0.1876	0.2297	-0.0673
u_2	-0.0204	-0.0413	0.1876	1.0000	0.1753	0.0402
u_3	-0.0075	0.0256	0.2297	0.1753	1.0000	-0.0744
u_4	0.0472	-0.0279	-0.0673	0.0402	-0.0744	1.0000

Table 8 Ljung-Box p values for u residuals from January, 2016 to December, 2018

	lag=1	lag=3	lag=5	lag=7
u_1	0.7417	0.9810	0.9098	0.9573
u_2	0.9369	0.3721	0.0008	0.0007
u_3	0.9552	0.9433	0.9825	0.0543
u_4	0.8121	0.9933	0.9926	0.9988

Table 9 Parameter estimates of ϕ coefficients from January, 2016 to December, 2018

	Estimate	SE	t -statistic	p value
Panel a: Bitcoin				
$\phi_{i,1}$	1.0649	0.0087	122.0500	0.0000
$\phi_{i,2}$	-0.2402	0.0153	-15.6990	0.0000
$\phi_{i,3}$	0.2078	0.0157	13.2280	0.0000
$\phi_{i,4}$	-0.1638	0.0167	-9.8030	0.0000
$\phi_{i,5}$	0.2676	0.0184	14.5420	0.0000
$\phi_{i,6}$	-0.3373	0.0188	-17.9330	0.0000
$\phi_{i,7}$	0.1784	0.0130	13.7630	0.0000
Panel b: Ethereum				
$\phi_{i,1}$	-0.5759	0.0127	-45.2080	0.0000
$\phi_{i,2}$	-0.3691	0.0146	-25.2620	0.0000
$\phi_{i,3}$	-0.0436	0.0172	-2.5329	0.0113
$\phi_{i,4}$	0.1789	0.0172	10.3780	0.0000
$\phi_{i,5}$	0.4026	0.0150	26.7940	0.0000
$\phi_{i,6}$	0.4812	0.0139	34.7000	0.0000
$\phi_{i,7}$	0.3299	0.0163	20.2530	0.0000
$\phi_{i,8}$	0.1542	0.0155	9.9343	0.0000
$\phi_{i,9}$	0.0812	0.0173	4.7023	0.0000
Panel c: Litecoin				
$\phi_{i,1}$	0.9706	0.0100	97.2300	0.0000
$\phi_{i,2}$	-0.1012	0.0106	-9.5439	0.0000
$\phi_{i,3}$	0.1866	0.0139	13.4390	0.0000
$\phi_{i,4}$	-0.1508	0.0084	-17.9020	0.0000

Table 9 continued

	Estimate	SE	<i>t</i> -statistic	<i>p</i> value
Panel d: Monero				
$\phi_{i,1}$	0.0041	0.0111	0.3669	0.7137
$\phi_{i,2}$	0.1080	0.0121	8.9081	0.0000
$\phi_{i,3}$	0.0818	0.0142	5.7506	0.0000
$\phi_{i,4}$	-0.0051	0.0122	-0.4188	0.6754
$\phi_{i,5}$	0.0444	0.0124	3.5718	0.0004
$\phi_{i,6}$	0.2814	0.0137	20.5970	0.0000
$\phi_{i,7}$	0.0940	0.0124	7.5686	0.0000
$\phi_{i,8}$	-0.0154	0.0151	-1.0236	0.3061
$\phi_{i,9}$	0.0046	0.0144	0.3202	0.7488
$\phi_{i,10}$	-0.1393	0.0132	-10.5760	0.0000

References

- Agosto, A., Cafferata, A.: Financial bubbles: a study of co-explosivity in the cryptocurrency market. *Risks* **8**(2), 34 (2020)
- Ahn, Y., Kim, D.: Sentiment disagreement and bitcoin price fluctuations: a psycholinguistic approach. *Appl. Econ. Lett.* 1–5 (2019)
- Anderson, T.W.: *An Introduction to Multivariate Statistical Analysis*. Wiley, New York (2003)
- Avellaneda, M., Lee, J.-H.: Statistical arbitrage in the us equities market. *Quant. Finance* **10**(7), 761–782 (2010)
- Bariviera, A.F., Basgall, M.J., Hasperué, W., Naiouf, M.: Some stylized facts of the bitcoin market. *Physica A Stat. Mech. Appl.* **484**, 82–90 (2017)
- Bistarelli, S., Cretarola, A., Figà-Talamanca, G., Mercanti, I., Patacca, M.: Is arbitrage possible in the bitcoin market?(work-in-progress paper). In: *International Conference on the Economics of Grids, Clouds, Systems, and Services*, pp. 243–251. Springer (2018)
- Bistarelli, S., Cretarola, A., Figà-Talamanca, G., Patacca, M.: Model-based arbitrage in multi-exchange models for bitcoin price dynamics. *Digital Finance* **1**(1–4), 23–46 (2019)
- Blau, B., Griffith, T., Whitby, R.: Comovement in the cryptocurrency market. *Econ. Bull.* **40**(1), 448–455 (2020)
- Bouri, E., Molnár, P., Azzi, G., Roubaud, D., Hagfors, L.I.: On the hedge and safe haven properties of bitcoin: Is it really more than a diversifier? *Finance Res. Lett.* **20**, 192–198 (2017)
- Bouri, E., Gil-Alana, L.A., Gupta, R., Roubaud, D.: Modelling long memory volatility in the bitcoin market: Evidence of persistence and structural breaks. *Int. J. Finance Econ.* (2019)
- Box, G.E.P., Tiao, G.C.: A canonical analysis of multiple time series. *Biometrika* **64**(2), 355–365 (1977)
- Chaim, P., Laurini, M.P.: Is bitcoin a bubble? *Physica A Stat. Mech. Appl.* **517**, 222–232 (2019)
- Cheah, E.-T., Fry, J.: Speculative bubbles in bitcoin markets? An empirical investigation into the fundamental value of bitcoin. *Econ. Lett.* **130**, 32–36 (2015)
- Ciaian, P., Rajcaniova, M.: Virtual relationships: short-and long-run evidence from bitcoin and altcoin markets. *J. Int. Financ. Mark. Inst. Money* **52**, 173–195 (2018)
- Ciaian, P., Rajcaniova, M., Kancs, A.: The economics of bitcoin price formation. *Appl. Econ.* **48**(19), 1799–1815 (2016)
- Cont, R.: Empirical properties of asset returns: stylized facts and statistical issues. *Quant. Finance* **1**(2), 223–236 (2001)
- Cont, R., Tankov, P.: *Financial Modelling With Jump Processes*. Chapman & Hall/CRC, Boca Raton (2004)
- Corbet, S., Lucey, B., Yarovaya, L.: Datestamping the bitcoin and ethereum bubbles. *Finance Res. Lett.* **26**, 81–88 (2018)
- Cretarola, A., Figà-Talamanca, G.: Detecting bubbles in bitcoin price dynamics via market exuberance. *Ann. Oper. Res.* 1–21 (2019)

- Cretarola, A., Figà-Talamanca, G.: Bubble regime identification in an attention-based model for bitcoin and ethereum price dynamics. *Econ. Lett.* **191**, 108831 (2020)
- Cretarola, A., Figà-Talamanca, G., Patacca, M.: Market attention and bitcoin price modeling: theory, estimation and option pricing. *Decis. Econ. Finance* **43**(1), 187–228 (2020)
- Dyhrberg, A.H.: Bitcoin, gold and the dollar—a garch volatility analysis. *Finance Res. Lett.* **16**(Supplement C), 85–92 (2016). <https://doi.org/10.1016/j.frl.2015.10.008>
- Engle, R., Watson, M.: A one-factor multivariate time series model of metropolitan wage rates. *J. Am. Stat. Assoc.* **76**(376), 774–781 (1981)
- Engle, R.F., Granger, C.W.J.: Co-integration and error correction: representation, estimation, and testing. *Econometrica J. Econ. Soc.* 251–276 (1987)
- Eom, C., Kaizoji, T., Kang, S.H., Pichl, L.: Bitcoin and investor sentiment: statistical characteristics and predictability. *Physica A Stat. Mech. Appl.* **514**, 511–521 (2019)
- Escribano, A., Peña, D.: Cointegration and common factors. *J. Time Ser. Anal.* **15**(6), 577–586 (1994)
- Figà-Talamanca, G., Patacca, M.: Does market attention affect bitcoin returns and volatility? *Decis. Econ. Finance* **42**(1), 135–155 (2019)
- Figà-Talamanca, G., Patacca, M.: Disentangling the relationship between Bitcoin and market attention measures. *Economia e Politica Industriale* **47**(1), 71–91 (2020)
- Figà-Talamanca, G., Focardi, S., Patacca, M.: Regime switches and commonalities of the cryptocurrencies asset-class. Working paper (2020). <https://ssrn.com/abstract=3388642>
- Fry, J., Cheah, E.-T.: Negative bubbles and shocks in cryptocurrency markets. *Int. Rev. Financ. Anal.* **47**, 343–352 (2016)
- Galeano, P., Peña, D.: Multivariate analysis in vector time series. *Resenhas do Instituto de Matemática e Estatística da Universidade de São Paulo* **4**(4), 383–403 (2000)
- Garcia, D., Tessone, C.J., Mavrodiev, P., Perony, N.: The digital traces of bubbles: feedback cycles between socio-economic signals in the bitcoin economy. *J. R. Soc. Interface* **11**(99), 20140623 (2014)
- Geweke, J.: The dynamic factor analysis of economic time series. *Latent variables in socio-economic models* (1977)
- Giudici, P., Pagnottoni, P.: Vector error correction models to measure connectedness of bitcoin exchange markets. *Appl. Stoch. Models Bus. Ind.* **36**(1), 95–109 (2020)
- Hendry, D.F., Juselius, K.: Explaining cointegration analysis: part I. *Energy J.* **21**(1), 1–42 (2000)
- Hendry, D.F., Juselius, K.: Explaining cointegration analysis: part II. *Energy J.* **22**(1), 75–120 (2001)
- Hotelling, H.: Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24**(6), 417 (1933)
- Hu, A.S., Parlour, C.A., Rajan, U.: Cryptocurrencies: stylized facts on a new investible instrument. *Financ. Manag.* **48**(4), 1049–1068 (2019)
- Johansen, S.: Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica J. Econ. Soc.* 1551–1580 (1991)
- Johansen, S.: Modelling of cointegration in the vector autoregressive model. *Econ. Model.* **17**(3), 359–373 (2000)
- Kang, S.H., McIver, R.P., Hernandez, J.A.: Co-movements between bitcoin and gold: a wavelet coherence analysis. *Physica A Stat. Mech. Appl.* **536**, 120888 (2019)
- Koop, G., Korobilis, D.: Manual to accompany matlab package for bayesian var models. Retrieved **10**, 2012 (2009)
- Koop, G., Korobilis, D.: Bayesian multivariate time series methods for empirical macroeconomics. *Found. Trends Econ.* **3**(4), 267–358 (2010)
- Kristoufek, L.: What are the main drivers of the bitcoin price? Evidence from wavelet coherence analysis. *PLoS ONE* **10**(4), e0123923 (2015)
- Leung, T., Nguyen, H.: Constructing cointegrated cryptocurrency portfolios for statistical arbitrage. *Stud. Econ. Finance* (2019)
- Lintilhac, P.S., Tourin, A.: Model-based pairs trading in the bitcoin markets. *Quant. Finance* **17**(5), 703–716 (2017)
- Lütkepohl, H., Poskitt, D.S.: Estimating orthogonal impulse responses via vector autoregressive models. *Econ. Theory* **7**(4), 487–496 (1991)
- Peña, D., Box, G.E.P.: Identifying a simplifying structure in time series. *J. Am. Stat. Assoc.* **82**(399), 836–843 (1987)
- Peña, D., Poncela, P.: Forecasting with nonstationary dynamic factor models. *J. Econ.* **119**(2), 291–321 (2004)

- Peña, D., Poncela, P.: Nonstationary dynamic factor analysis. *J. Stat. Plan. Inference* **136**(4), 1237–1257 (2006)
- Pole, A.: *Statistical Arbitrage: Algorithmic Trading Insights and Techniques*, vol. 411. Wiley, New York (2011)
- Rachev, S.T., Mittnik, S., Fabozzi, F.J., Focardi, S.M.: *Financial Econometrics: From Basics to Advanced Modeling Techniques*, vol. 150. Wiley, New York (2007)
- Sargent, T.J.: Two models of measurements and the investment accelerator. *J. Polit. Econ.* **97**(2), 251–287 (1989)
- Sargent, T.J., Sims, C.A.: Business cycle modeling without pretending to have too much a priori economic theory. *New Methods Bus. Cycle Res.* **1**, 145–168 (1977)
- Shahzad, S.J.H., Bouri, E., Roubaud, D., Kristoufek, L., Lucey, B.: Is bitcoin a better safe-haven investment than gold and commodities? *Int. Rev. Financ. Anal.* **63**, 322–330 (2019)
- Spearman, C.: “general intelligence,” objectively determined and measured. *Am. J. Psychol.* **15**(2), 201–292 (1904)
- Stock, J.H., Watson, M.W.: Testing for common trends. *J. Am. Stat. Assoc.* **83**(404), 1097–1107 (1988)
- Stock, J.H., Watson, M.W.: New indexes of coincident and leading economic indicators. *NBER Macroecon. Ann.* **4**, 351–394 (1989)
- Thurstone, L.L.: *Primary mental abilities*. Psychometric monographs (1938)
- Thurstone, L.L.: *Multiple-factor Analysis a Development and Expansion of The Vectors of Mind*. University of Chicago Press, Chicago (1947)
- Tiwari, A.K., Raheem, I.D., Kang, S.H.: Time-varying dynamic conditional correlation between stock and cryptocurrency markets using the Copula-adcc-egarch model. *Physica A Stat. Mech. Appl.* **535**, 122295 (2019)
- Tsay, R.S.: *Analysis of Financial Time Series*, vol. 543. Wiley, New York (2005)
- Yaya, O.S., Ogbonna, A.E., Olubusoye, O.E.: How persistent and dynamic inter-dependent are pricing of bitcoin to other cryptocurrencies before and after 2017/18 crash? *Physica A Stat. Mech. Appl.* 121732 (2019)
- Zhang, W., Wang, P., Li, X., Shen, D.: Some stylized facts of the cryptocurrency market. *Appl. Econ.* **50**(55), 5950–5965 (2018)
- Zhang, Y., Chan, S., Chu, J., Nadarajah, S.: Stylised facts for high frequency cryptocurrency data. *Physica A Stat. Mech. Appl.* **513**, 598–612 (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.