Journal of
Risk and Financial Management

MDPI

Article

Forecasting the Returns of Cryptocurrency: A Model Averaging Approach

Hui Xiao [1,*] and Yiguo Sun [2,*]

1 Department of Economics, Saint Mary's University, Halifax, NS B3H 3C3, Canada
2 Department of Economics and Finance, University of Guelph, Guelph, ON N1G 2W1, Canada
* Correspondence: Hui.Xiao@smu.ca (H.X.); yisun@uoguelph.ca (Y.S.)

Received: 26 October 2020; Accepted: 10 November 2020; Published: 13 November 2020

check for updates

Abstract: This paper aims to enrich the understanding and modelling strategies for cryptocurrency markets by investigating major cryptocurrencies' returns determinants and forecast their returns. To handle model uncertainty when modelling cryptocurrencies, we conduct model selection for an autoregressive distributed lag (ARDL) model using several popular penalized least squares estimators to explain the cryptocurrencies' returns. We further introduce a novel model averaging approach or the shrinkage Mallows model averaging (SMMA) estimator for forecasting. First, we find that the returns for most cryptocurrencies are sensitive to volatilities from major financial markets. The returns are also prone to the changes in gold prices and the Forex market's current and lagged information. Then, when forecasting cryptocurrencies' returns, we further find that an ARDL(p,q) model estimated by the SMMA estimator outperforms the competing estimators and models out-of-sample.

Keywords: cryptocurrencies; Mallows criterion; model averaging; model selection; shrinkage; tuning parameter choice

JEL Classification: G12; C13; C21

---

1. Introduction

Ever since the advent of the first digital currency, Bitcoin (BTC), which was created under the anonymous identity Nakamoto (2008), cryptocurrencies are becoming increasingly popular among investors as they provide an alternative investment strategy to conventional financial assets. Unlike fiat money, which is usually issued by central authorities, the cryptocurrencies typically originate from decentralized virtual networks that store and distribute assets digitally. The cryptocurrencies are highly sought after as they stem from a peer-to-peer structure that allows efficiency, security and profitability. As they are becoming more widely accepted by investors, it has cemented their legitimacy as a medium of exchange and a means to liquidity for businesses, financial institutions and the general public alike.

In this paper, we concentrate on investigating the returns of the top seven cryptocurrencies by market capitalization in USD as of August 8, 2020, according to coinmarketcap.com. The top seven cryptocurrencies are Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), Tether (USDT), Bitcoin Cash (BCH), Bitcoin SV (BSV), and Litecoin (LTC). Not only do we aim to uncover the relationship between the cryptocurrency and the macroeconomic conditions, but we also explore the dynamic interactions between the cryptocurrency market and the financial market, foreign exchange (Forex) market, commodity market and precious metals market. This helps potential investors to gain more insights into the leading cryptocurrencies.

The relationship between the macroeconomic conditions, the conventional assets market, and the cryptocurrencies' returns have been extensively investigated in the current literature.

For example, Dyhrberg (2016a), Dyhrberg (2016b) and Wu et al. (2019) showed that the gold prices help explain the variations in the cryptocurrencies. Moreover, Bouri et al. (2017), Demir et al. (2018), Panagiotidis et al. (2019) and Wu et al. (2019) provided evidence that the economic policy uncertainty plays a role in explaining the cryptocurrencies. However, there is not yet a unifying underlying theory that guides the modelling of the cryptocurrencies. Moreover, the decentralized and volatile nature of the cryptocurrency, which is supported by Aharon and Qadan (2019), further aggravates the problem of model uncertainty that is already ripe in the field of asset pricing.

To handle model uncertainty, model selection and model averaging have long been the competing approaches that originate from dichotomous modelling philosophies. Model selection searches for the most relevant variables, while model averaging aims to smooth over a set of candidate models rather than committing to a single model. For model selection, Hoerl and Kennard (1970) and Tibshirani (1996) introduced the ridge and the least absolute shrinkage and selection operator (LASSO) estimator for simultaneous selection and estimation. In contrast, the traditional best subsets approach often leads to local solutions.

However, the ridge and LASSO estimators' asymptotic bias prompted the development of a class of penalized least squares estimators that also possess the oracle properties. In particular, Fan and Li (2001) developed the smoothly clipped absolute deviation (SCAD) estimator, Zou (2006) introduced the adaptive LASSO (AdaLASSO) estimator, and Zhang (2010) invented the minimax concave penalty (MCP) estimator.

The penalized estimators' performance also relies on the tuning parameter choice for the penalty function to achieve the optimal estimation. The cross-validation (CV) and the information criterion (IC) are the two conventional approaches for selecting the tuning parameter. Shi and Tsai (2002) have shown that the Bayesian information criterion (BIC) consistently identifies the true model in finite samples. For the algorithm generating the candidate tuning parameters, Tibshirani et al. (2010) and Breheny and Huang (2011) developed the cyclical coordinate descent (CCD) algorithm to compute the solution path for the penalized least squares estimators. They provided evidence that the tuning parameter choice affects the model selection outcomes.

In the spirit of the OLS post-LASSO estimator introduced by Belloni and Chernozhukov (2013), Xiao and Sun (2019) further studied the finite sample performances of the class of OLS post-selection estimators with the tuning parameters determined by different tuning parameter selection approaches. The OLS post-selection estimators avoid the complicated penalty functions in building inference, and the results from Xiao and Sun (2019) further validate the importance of the tuning parameter selection.

On the other hand, for model averaging, the Shrinkage Mallows Model Averaging (SMMA) estimator introduced by Xiao and Sun (2019) compliments the penalized estimators by allowing for more than one model selection outcome and hedges against the penalized estimators' sensitivity towards the tuning parameter choice. The SMMA estimator extends the current Mallows Model Averaging (MMA) framework from Hansen (2007) by introducing a reasonable way for selecting candidate models. The SMMA estimator also builds on the shrinkage averaging estimator (SAE) by Schomaker (2012) by incorporating the tuning parameter optimization problem for each candidate model. The SMMA estimator essentially combines model selection and model averaging to address model uncertainty.

In this paper, we examine the determinants of the returns for major cryptocurrencies by applying the autoregressive distributed lag (ARDL) model with model selection via penalized least squares estimators. Then we average the candidate sub-models post model selection by the SMMA estimator for forecasting cryptocurrencies' returns. In particular, the issue of ARDL(p,q) models' lag selection, the discrete nature of traditional model selection algorithms, and the sensitivity of the tuning parameter choice affecting variable selection outcomes culminate in a call for a novel modelling approach. Therefore, we contribute to the existing literature on forecasting cryptocurrencies via a novel model averaging approach where both model selection and model averaging are combined by our SMMA estimator to address the cryptocurrencies' model uncertainty.

We find that most cryptocurrencies' returns are sensitive to fluctuations from major financial markets. They are also susceptible to the changes in gold prices and the Forex market's current and lagged information. We then forecast the returns of the cryptocurrencies via the SMMA estimator. We find that an ARDL(p,q) model estimated by the SMMA estimator outperforms competing estimators and competing models considered in this paper for forecasting cryptocurrencies' returns.

The rest of the paper is organized as follows. Section 2 outlines the methodology. Section 3 presents the model and Section 4 introduces the data respectively. Section 5 reports the empirical results and Section 6 conducts forecast evaluations. Section 7 concludes.

2. Methodology

To handle the model uncertainty that is common in asset pricing, especially for the cryptocurrencies, we consider both model selection and model averaging approaches. For model selection, we mainly consider the penalized least squares estimators studied in detail by Xiao and Sun (2019). According to their data examples, the six best-performing OLS post-selection estimators are listed in Table 1. They provide not only an estimation strategy, but more importantly, an appropriate data-driven approach for the ARDL(p,q) model's lag selection as there is not yet a no unified approach for its lag selection.

Table 1. Penalized least squares estimators. SCAD, smoothly clipped absolute deviation; MCP, minimax concave penalty.

Estimator	
OLS post-SCAD(BIC)	OLS post-SCAD(GCV)
OLS post-MCP(BIC)	OLS post-MCP(GCV)
OLS post-LASSO(BIC)	OLS post-LASSO(GCV)
OLS post-adaptive-LASSO(BIC)	OLS post-adaptive-LASSO(GCV)

For Table 1, the OLS post-SCAD(BIC) estimator is constructed with the tuning parameter in the SCAD penalty function selected by the BIC. Similarly, the OLS post-SCAD(GCV) is constructed with the tuning parameter in the SCAD penalty function selected by the generalized cross validation (GCV).

For example, an OLS post-SCAD(BIC) estimator where the tuning parameter in the SCAD penalty function is selected by the BIC for the ARDL(p,q) model in Equation (23) can be constructed with the tuning parameter in the penalty function selected by the BIC approach as follows.

Let $\Lambda = \{\lambda^1, \ldots, \lambda^Q\}$ be the set of candidate tuning parameters and $|\Lambda| = Q$ with $Q \in \mathbb{Z}^+$. Given any $\lambda \in \Lambda$, the SCAD estimator evaluated at λ gives:

$$\widehat{\gamma}^\lambda = \underset{\gamma}{\text{argmin}} \|y - Z\gamma\|^2 + \sum_{m=1}^{\mathcal{K}} F(|\gamma_m|; \lambda), \tag{1}$$

where F is the SCAD penalty function.

The BIC evaluated at this λ is defined as BIC_λ, which is given by:

$$BIC_\lambda = \frac{\|y - Z\widehat{\gamma}^\lambda\|^2}{n} + |S_\lambda| \frac{log(n)}{n} C_n, \tag{2}$$

where the values for λ originate from an exponentially decaying grid as in Tibshirani et al. (2010).

Let S_λ denote the set of nonzero parameters of the model when evaluated at λ so that $S_\lambda = \{m : \hat{\gamma}_m^\lambda \neq 0\}$. Then, $|S_\lambda|$ gives the number of nonzero parameters of the model when evaluated at λ and $C_n = 1$ following Shi and Tsai (2002).

The estimate of the optimal tuning parameter is denoted by $\widehat{\lambda}^{BIC}$ and solves the following problem:

$$\widehat{\lambda}^{BIC} = \underset{\lambda \in \{\lambda^1,\ldots,\lambda^Q\}}{\text{argmin}} \; BIC_\lambda. \tag{3}$$

Consequently, $\widehat{\gamma}^{\widehat{\lambda}^{BIC}}$ minimizes the SCAD penalized objective function given by Equation (1),

$$\widehat{\gamma}^{\widehat{\lambda}^{BIC}} = \underset{\gamma}{\text{argmin}} \|y - Z\gamma\|^2 + \sum_{m=1}^{\mathcal{K}} F(|\gamma_m|, \widehat{\lambda}^{BIC}). \tag{4}$$

Denoting $S_{\widehat{\lambda}^{BIC}} = \{m : \widehat{\gamma}_m^{\widehat{\lambda}^{BIC}} \neq 0\}$, we define the OLS post-SCAD(BIC) estimator as:

$$\widehat{\gamma}_{post} = \underset{\gamma}{\text{argmin}} \left\| y - \sum_{v \in S_{\widehat{\lambda}^{BIC}}} Z_v \gamma_v \right\|^2, \tag{5}$$

where Z_v is an $(n - max\{p,q\}) \times 1$ vector, which is the vth column of the predictor matrix Z, and γ_v is the vth parameter in $S_{\widehat{\lambda}^{BIC}}$.

Similarly, we construct other OLS post-selection estimators shown in Table 1 as the OLS post-selection(BIC or GCV) estimator. The OLS post-selection(BIC or GCV) estimator minimizes the BIC or the GCV criterion, respectively, to estimate the optimal tuning parameter.

For model averaging, we adopt a novel model averaging approach or the SMMA estimator introduced by Xiao and Sun (2019) as it combines model selection and model averaging to handle model uncertainty. The SMMA estimator performs significantly better when averaging high dimensional sparse models against model uncertainty.

The SMMA estimator starts with a large general model and applies different penalty methods to select the candidate models for averaging. It is a two-stage estimator. For this paper, in the first stage, we apply different penalty estimators introduced in Table 1 with optimal tuning parameters selected via the GCV or BIC, thereby obtaining a sequence of candidate models. Then we apply the MMA criterion in the second stage to estimate the model parameters.

Let Λ^{Opt} represent the set of optimal tuning parameters selected either by the BIC or GCV for the model selection procedures introduced in Table 1 such that:

$$\Lambda^{Opt} = \{\widehat{\lambda}_1^{Opt}, \ldots, \widehat{\lambda}_s^{Opt}, \ldots, \widehat{\lambda}_{\mathcal{S}}^{Opt}\}, \tag{6}$$

where $\widehat{\lambda}_s^{BIC}$ represents the optimal tuning parameter for the sth candidate model given by the s^{th} penalized least squares estimator, and $|\Lambda^{Opt}| = \mathcal{S}$.

The model averaging estimator for the conditional mean of y is $\widehat{y}(\widehat{w}; \Lambda^{Opt})$, which is defined as:

$$\widehat{y}(\widehat{w}; \Lambda^{Opt}) = \sum_{s=1}^{\mathcal{S}^2} \widehat{w}_s P(\widehat{\lambda}_s^{Opt}) y = P(\widehat{w}; \Lambda^{Opt}) y, \tag{7}$$

where the projection matrix for the sth candidate model is defined as:

$$P(\widehat{\lambda}_s^{Opt}) = Z^{\widehat{\lambda}_s^{Opt}} \left(Z^{\widehat{\lambda}_s^{Opt}\,T} Z^{\widehat{\lambda}_s^{Opt}} \right)^{-1} Z^{\widehat{\lambda}_s^{Opt}\,T}, \tag{8}$$

and the estimator for the parameters in the sth candidate model is given by:

$$\widehat{\gamma}(\widehat{\lambda}_s^{Opt}) = \left(Z^{\widehat{\lambda}_s^{Opt}\,T} Z^{\widehat{\lambda}_s^{Opt}} \right)^{-1} \left(Z^{\widehat{\lambda}_s^{Opt}} \right)^T y. \tag{9}$$

Therefore, the SMMA estimator for the ARDL(p,q) model is solved as:

$$\widehat{\gamma}^{SMMA}(\widehat{w};\Lambda^{Opt}) = \sum_{s=1}^{\mathcal{S}} \widehat{w}_s \mathcal{F}_s \widehat{\gamma}(\widehat{\lambda}_s^{Opt}), \tag{10}$$

where $|\widehat{\gamma}(\widehat{\lambda}_s^{Opt})| = \mathcal{K}(\widehat{\lambda}_s^{Opt})$ is the number of nonzero values in $\widehat{\gamma}(\widehat{\lambda}_s^{Opt})$ and $\mathcal{F}_s$ is a $\mathcal{K}(\widehat{\lambda}_s^{BIC}) \times \mathcal{K}$ matrix specified as:

$$\mathcal{F}_s = \left(Z^T Z\right)^{-1} Z^T Z^{\widehat{\lambda}_s^{Opt}}, \tag{11}$$

whose rank is $\mathcal{K}(\widehat{\lambda}_s^{Opt})$, and each element in $\mathcal{F}_s$ is either one or zero to map $\widehat{\gamma}(\widehat{\lambda}_s^{Opt})$ to a $\mathcal{K} \times 1$ vector, while Z is the $(n - max\{p,q\}) \times \mathcal{K}$ regressor matrix.

The weight vector w is estimated by the MMA criterion,

$$\widehat{w} = \underset{w \in \mathcal{H}_{\mathcal{S}}}{\text{argmin}} \left(y - \widehat{y}(w;\Lambda^{Opt})\right)^T \left(y - \widehat{y}(w;\Lambda^{Opt})\right) + 2\sigma^2 \mathcal{K}(w;\Lambda^{Opt}), \tag{12}$$

where $w = \left[w_1, w_2 \ldots, w_{\mathcal{S}}\right]^T$ is a weight vector in the unit simplex in $\mathbb{R}^{\mathcal{S}}$ with $\mathcal{S} \in \mathbb{Z}^+$ such that:

$$\mathcal{H}_{\mathcal{S}} = \left\{w \in [0,1]^{\mathcal{S}} : \sum_{s=1}^{\mathcal{S}} w_s = 1\right\}, \tag{13}$$

and the effective number of parameters $\mathcal{K}(w;\Lambda^{Opt})$ is defined as:

$$\mathcal{K}(w;\Lambda^{Opt}) = \sum_{s=1}^{\mathcal{S}} w_s \mathcal{K}(\widehat{\lambda}_s^{Opt}). \tag{14}$$

Let the L index be the largest model in dimension from the set of the candidate models, i.e.,

$$L = \underset{s \in \mathcal{S}}{\text{argmax}} \{|\widehat{\gamma}(\widehat{\lambda}_s^{Opt})|\}, \tag{15}$$

and $\mathcal{K}_L$ be the number of nonzero parameters in the largest candidate model.

According to Hansen (2007), the σ^2 term will be estimated by $\widehat{\sigma}_L^2$, which is given below:

$$\widehat{\sigma}_L^2 = \frac{(y - Z_L \widehat{\gamma}_L)^T (y - Z_L \widehat{\gamma}_L)}{n - \mathcal{K}_L}. \tag{16}$$

Following Hansen (2008), an out-of-sample forecast combination by the SMMA estimator is generated as:

$$\widetilde{y}(\widehat{w};\Lambda^{Opt}) = \sum_{s=1}^{\mathcal{S}^2} \widehat{w}_s \widetilde{y}(\widehat{\lambda}_s^{Opt}), \tag{17}$$

where $\widetilde{y}(\widehat{\lambda}_s^{Opt})$ is the forecast from the sth candidate model given $\widehat{\lambda}_s^{Opt}$, and it is defined as:

$$\widetilde{y}(\widehat{\lambda}_s^{Opt}) = \widetilde{Z} \mathcal{F}_s \widehat{\gamma}(\widehat{\lambda}_s^{Opt}), \tag{18}$$

where $\widetilde{Z}$ is the out-of-sample regressor matrix that is $\left(n - max\{p,q\} + 1\right) \times \mathcal{K}$ in dimension.

3. Model

We consider autoregressive distributed lag (ARDL)(p,q) model where p is the order of lags for the dependent variable y and q is the order of lags for the independent variable X. The rationale for applying the ARDL(p,q) model for modelling cryptocurrencies is that it nests more sub-models and, therefore, is more encompassing. As a result, we think that the ARDL(p,q) model is particularly helpful for modelling cryptocurrencies which feature model uncertainty.

The ARDL(p,q) model is defined as:

$$y_t = \alpha + \rho_h \sum_{h=1}^{p} y_{t-h} + \sum_{i=1}^{k} \sum_{l=0}^{q} x_{i,t-l}\beta_{i,l} + \varepsilon_t, \forall t > max\{p,q\}, \tag{19}$$

where y_t is the daily returns for any of the cryptocurrencies at period t and y_{t-h} represents the dependant variable lagged to period $t-h$. Further, $x_{i,t-l}$ is the observation for the ith explanatory variable at period $t-l$. Lastly, ρ_h is a scalar parameter for y_{t-h}, while $\beta_{i,l}$ is the parameter for $x_{i,t-l}$.

Let:

$$\gamma = \left[\alpha, (\rho_1, \rho_2, \ldots, \rho_p), (\beta_{1,0}, \ldots, \beta_{1,q}), \ldots, (\beta_{k,0}, \ldots, \beta_{k,q})\right]^T, \tag{20}$$

and Z be an $(n - max\{p,q\}) \times (1 + p + k \times q)$ regressor matrix defined as:

$$Z = \begin{bmatrix} Z_t \\ Z_{t+1} \\ \vdots \\ Z_n \end{bmatrix}, t = max\{p,q\} + 1, \tag{21}$$

where:

$$Z_t = [1, (y_{t-1}, \ldots, y_{t-p}), (x_{1,t}, \ldots, x_{1,t-q}), (x_{2,t}, \ldots, x_{2,t-q}) \ldots, (x_{k,t}, \ldots, x_{k,t-q})]. \tag{22}$$

Then, the model in Equation (19) can be rewritten as:

$$y = Z\gamma + \varepsilon, \tag{23}$$

where y is an $(n - max\{p,q\}) \times 1$ vector for the daily returns of any of the cryptocurrencies, Z is the $(n - max\{p,q\}) \times (1 + p + k \times q)$ regressor matrix and γ is a $(1 + p + k \times q) \times 1$ vector of parameters. To simplify the notations, let $\mathcal{K} = 1 + p + k \times q$.

4. Data

First, the dataset contains the daily close prices for the top seven cryptocurrencies by market capitalization in USD as of 8 August 2020 according to coinmarketcap.com, which is website that specializes in cryptocurrencies. The top seven cryptocurrencies are Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), Tether (USDT), Bitcoin Cash (BCH), Bitcoin SV (BSV) and Litecoin (LTC). The cryptocurrencies studied in the paper are summarized in Table 2.

Table 2. Cryptocurrency summary.

Name	Code	Release Date	Market Capitalization (USD)	Ranking	Sample Size
Bitcoin	BTC	3 January 2009	$214,216,361,823	1	1759
Ethereum	ETH	30 July 2015	$42,514,584,819	2	1545
Ripple	XRP	15 May 2018	$13,250,155,412	3	1281
Tether	USDT	25 February 2015	$10,029,337,359	4	1985
Bitcoin Cash	BCH	1 August 2017	$5,650,564,6948	5	1109
Bitcoin SV	BSV	19 November 2018	$4,255,662,214	6	628
Litecoin	LTC	12 November 2018	$3,745,359,046	7	635

Note: The market capitalizations of the top 7 cryptocurrencies in U.S. dollars as of 8 August 2020 are provided by coinmarketcap.com. The time series data for the daily close prices are from coinmarketcap.com and the respective samples dates from the initial public trading of the corresponding cryptocurrencies.

Table 2 presents the codes, release dates, market capitalizations (USD), rankings and sample sizes for the cryptocurrencies considered in this paper. It is worth noting that Ripple commands a relatively higher market capitalization than USDT and BCH despite a later release date. The USDT and BCH were released earlier in 2015 and 2017, respectively.

Following the literature, the returns for the cryptocurrency are defined as,

$$r_{t+1} = ln(\frac{P_{t+1}}{P_t}), \ t = 1, 2, \ldots, n, \tag{24}$$

where P_t stands for the daily close prices for respective cryptocurrencies at period t. To explain the cryptocurrency returns, we also incorporate several other independent variables into the dataset that comprise macroeconomic and financial market conditions.

To account for the macroeconomic conditions, we include the following independent variables. The independent variables are the daily Brent crude oil price following Aharon and Qadan (2019), the daily London Bullion Market Association (LBMA) gold price following Dyhrberg (2016a), Dyhrberg (2016b) and Wu et al. (2019). We also include U.S./Euro foreign exchange rate, the daily US federal funds rate, the daily economic policy uncertainty index for the US, and the monthly global economic policy uncertainty index following Bouri et al. (2017), Demir et al. (2018) and Panagiotidis et al. (2019). The daily time series data for the oil price, gold price, U.S./Euro foreign exchange rate and the daily US federal funds rate (FFR) are collected from quandl.com. The daily time series data for the US economic policy uncertainty index and the monthly time series data for the global economic policy uncertainty index are collected from policyuncertainty.com.

To control for the financial market conditions, we also introduce the Dow Jones NYSE index, the S&P 500 index, the Volatility Index (VIX), which is the daily Chicago Board Options Exchange (CBOE) volatility index in U.S. dollars, and Fama and French's Momentum Factor (Mom) following Aharon and Qadan (2019). The daily time series data for the Dow Jones NYSE index and S&P 500 index are from Yahoo finance. The daily time series data for VIX are collected from the CBOE's official website at cboe.com, and the daily time series data for Mom are collected from the Kenneth R. French Data Library.

Similarly, the daily returns for the independent variables in this dataset are defined following Equation (24). The descriptive statistics for the dataset are shown in Table 3.

Table 3. Descriptive statistics for cryptocurrencies, stocks, commodities, gold and Forex market.

Variable	Description	Mean	Std.Dev.	Min	Max	Skewness	Kurtosis
BTC	Bitcoin daily returns	0.001	0.02	−0.22	0.11	−1.22	17.94
ETH	Ethereum daily returns	0.001	0.03	−0.25	0.12	−0.81	10.75
XRP	Ripple daily returns	0.002	0.03	−0.17	0.26	1.84	12.24
USDT	Tether daily returns	0.000	0.01	−0.30	0.22	−10.61	621.76
BCH	Bitcoin Cash daily returns	0.001	0.04	−0.24	0.19	0.26	9.51
BSV	Bitcoin SV daily returns	−0.001	0.04	−0.27	0.39	2.61	33.25
LTC	Litecoin daily returns	0.001	0.03	−0.21	0.11	−0.93	14.02
Oil	daily % change in oil price	0.00	0.02	−0.28	0.18	−3.02	86.68
Gold	daily % change in gold price	0.0001	0.00	−0.02	0.02	0.17	4.30
Euro	daily % change in U.S./Euro exchange rate	0.00	0.00	−0.01	0.01	0.13	2.60
FFR	U.S. daily federal funds rate	1.09	0.81	0.04	2.45	0.26	−1.38
USEPU	daily % change in US policy uncertainty index	−0.0035	0.22	−1.51	1.13	−0.15	2.79
GlobalEPU	monthly % change in global policy uncertainty index	0.01	0.09	−0.21	0.27	0.51	1.12
NYSE	Dow Jones NYSE daily returns	0.0001	0.01	−0.06	0.05	−1.20	25.63
S&P 500	S&P 500 daily returns	0.0001	0.01	−0.06	0.04	−1.14	22.48
VIX	daily % change in CBOE's Volatility Index	0.00	0.04	−0.13	0.33	1.44	8.06
Mom	daily % change in Fama&French's Momentum Factor	0.02	0.99	−6.15	5.99	−0.59	6.55

Note: All of the variables, except for the federal funds rate (FFR) and GlobalEPU, are in daily returns. GlobalEPU is linearly interpolated from the original monthly global policy uncertainty index into the daily frequency.

We further conduct the normality test and the unit root test for the daily return series in the dataset, and the results are given in Table 4.

For the normality test, we presented p-values for the dataset variables in Table 4. We conducted the normality test to test if there is non-normality as it is common in financial datasets. To handle such non-normality in data, we have increased the sample size as large as possible to take advantage of the central limit theorem. As shown above, our cryptocurrencies sample starts from the day the said cryptocurrencies are openly traded till 8 August 2020, when this paper is written.

For the unit root test, we presented p-values for the variables in the dataset in Table 4. The unit root test is important as it tests whether a time series is stationary. We usually conduct unit root tests for the time series data first to avoid spurious regressions. The unit root test for FFR is omitted since the variable of interest is the FFR in the levels, rather than the first-order differenced FFR. We use the FFR in the levels in our subsequent regression models.

Table 4. Normality test and unit root test.

Variable	Jarque-Bera Test p-Value H_0: Normality	ADF Test p-Value H_0: Unit Root
BTC	0.00	0.01
ETH	0.00	0.01
XRP	0.00	0.01
USDT	0.00	0.01
BCH	0.00	0.01
BSV	0.00	0.01
LTC	0.00	0.01
USEPU	0.00	0.01
GlobalEPU	0.00	0.01
Oil	0.00	0.01
Gold	0.00	0.01
Euro	0.00	0.01
FFR	0.00	
NYSE	0.00	0.01
S&P 500	0.00	0.01
VIX	0.00	0.01
Mom	0.00	0.01

Note: ADF test refers to the augmented Dickey–Fuller test test. All variables except the FFR are in percentage changes and stationary while non-normally distributed.

5. Empirical Results

To determine the optimal order of lags for the ARDL(p,q) model for each cryptocurrency listed in Table 2, we shall apply the penalized least squares (PLS) estimators due to the lack of a unifying approach for the ARDL(p,q) model's lag selection and the discrete nature of the traditional model selection algorithms. In contrast, the PLS estimators provide a more holistic approach for simultaneous lag selection and parameter estimation.

Since the true data generating process (DGP) is not observed, we will also consider competing estimators and competing models. The competing estimators for estimating the ARDL(p,q) model are the Naive OLS, all of the penalized least squares estimators in Table 1, and the SMMA estimator.

To determine which model better fits the data, the competing models are the traditional asset pricing model (APM) $y = \alpha + X\beta + \varepsilon$, the constant model $y = \alpha + \varepsilon$, the artificial neural network (ANN) model with *tanh* activation function and the back-propagation neural network (BPNN) model.

To evaluate the performances, we compare the respective mean squared prediction error (MSPE), adjusted R^2, and model size (MS) that gives the number of non-zero estimated parameters. MSPE measures the fit and adjusted R^2 gauges the performance in identifying the most relevant regressors, and MS measures parsimony and efficiency. The evaluation is presented in Table 5.

Table 5. Cryptocurrency model and estimator performance evaluation. MSPE, mean squared prediction error; MS, model size; SMMA, shrinkage Mallows model averaging; APM, asset pricing model.

Model	Estimator	MSPE	Adjusted R^2	MS
Cryptocurrency		**BTC**		
ARDL(3,3)	Naive OLS	0.02	0.05	48
	OLS post-adaptive-LASSO(GCV)	0.02	0.05	3
	SMMA	0.02	0.05	7
APM	Naive OLS	0.02	0.04	13
Constant		0.02		
ANN		0.12		
BPNN		0.02		
		ETH		
ARDL(3,3)	Naive OLS	0.03	0.07	48
	OLS post-LASSO(GCV)	0.03	0.08	3
	SMMA	0.03	0.07	5
APM	Naive OLS	0.03	0.03	13
Constant		0.03		
ANN		0.06		
BPNN		0.03		
		XRP		
ARDL(3,3)	Naive OLS	0.03	0.06	48
	OLS post-LASSO(GCV)	0.03	0.06	5
	SMMA	0.03	0.06	7
APM	Naive OLS	0.03	0.02	13
Constant		0.03		
ANN		0.21		
BPNN		0.03		
		USDT		
ARDL(3,3)	Naive OLS	0.01	0.55	48
	OLS post-SCAD(BIC)	0.01	0.55	6
	SMMA	0.01	0.54	8
APM	Naive OLS	0.01	0.02	13
Constant		0.01		
ANN		0.05		
BPNN		0.01		
		BCH		
ARDL(3,3)	Naive OLS	0.03	0.05	48
	OLS post-MCP(GCV)	0.04	0.06	2
	SMMA	0.04	0.05	4
APM	Naive OLS	0.04	0.02	13
Constant		0.04		
ANN		0.14		
BPNN		0.04		
		BSV		
ARDL(3,3)	Naive OLS	0.04	0.08	48
	OLS post-LASSO(GCV)	0.04	0.09	4
	SMMA	0.04	0.08	5
APM	Naive OLS	0.04	0.02	13
Constant		0.04		
ANN		0.13		
BPNN		0.04		

Table 5. *Cont.*

Model	Estimator	MSPE	Adjusted R^2	MS
		LTC		
ARDL(3,3)	Naive OLS	0.02	0.1	48
	OLS post-adaptive-LASSO(GCV)	0.02	0.11	7
	SMMA	0.02	0.08	8
APM	Naive OLS	0.02	0.07	13
Constant		0.03		
ANN		0.08		
BPNN		0.03		

As shown in Table 5, the best performing penalized least squares estimators for each cryptocurrency are selected based on MSPE and adjusted R^2. The respective penalized least squares estimators and the SMMA estimator outperforms the naive OLS terms of parsimony while yielding similar MSPE and adjusted R^2.

For the residual diagnostics shown in Table 6, we conducted the Breusch–Godfrey test up to the order of three for detecting any remaining serial autocorrelation in the model residuals. We found no serial autocorrelation in the model residuals for all the models considered in Table 6. This further supports the application of the ARDL(p,q) model, and we conclude that based on the empirical evidence, there is no need for us to further consider the class of ARMA models.

Table 6. Residual diagnostics.

Dependent Variable	Model	Penalized Estimator	Breusch–Godfrey Test p-Value
BTC	ARDL(3,3)	OLS post-adaptive-LASSO(GCV)	0.99
ETH	ARDL(3,3)	OLS post-LASSO(GCV)	0.99
XRP	ARDL(3,3)	OLS post-LASSO(GCV)	0.98
USDT	ARDL(3,3)	OLS post-SCAD(BIC)	0.99
BCH	ARDL(3,3)	OLS post-MCP(GCV)	0.95
BSV	ARDL(3,3)	OLS post-LASSO(GCV)	0.99
LTC	ARDL(3,3)	OLS post-adaptive-LASSO(GCV)	0.99

The ARDL(p,q) model estimated by penalized least squares estimators and the SMMA estimator outperforms the competing models in MSPE. This supports applying the ARDL(p,q) model and our estimation strategies to better explain the cryptocurrencies' returns. But the ARDL(p,q) model estimated by penalized least squares estimators outperforms the SMMA estimator in MS. Therefore, to explain the returns, we shall focus on the estimation results of the ARDL(p,q) model by the corresponding penalized least squares estimators for parsimony. Table 7 presents the determinants of the returns for each cryptocurrency.

In summary, given the sparsity of the parsimonious ARDL(p,q) model after model selection, we do not consider including more lags in the ARDL(p,q) model, and the true ARDL(p,q) model is likely sparse as well. We then examine the returns determinants for each cryptocurrency, respectively.

For the BTC's daily returns, daily returns from the S&P 500 and the gold market have positive effects. This finding is consistent with Aharon and Qadan (2019), who suggested that the BTC's daily returns are positively correlated with the S&P 500 and Gold. This indicates that the BTC market responds promptly to the performances of the S&P 500 index and the gold market.

In comparison, daily returns from the NYSE and the gold market have positive effects on the ETH's daily returns. This finding is supported by Deniz and Stengos (2020), who also discovered a positive correlation between the ETH market and the NYSE and the gold market.

Lagged XRP returns, NYSE and the Mom introduced by Fama and French (1993), have positive effects on the current daily returns of the XRP. This finding coincides with Deniz and Stengos (2020), where a positive correlation between the XRP market and the NYSE index is also discovered.

However, the FFR has a negative effect on the XRP's daily returns, which suggests that the XRP is sensitive to the price of liquidity. Therefore, as the liquidity becomes cheaper, investors divest from the XRP for other assets. Deniz and Stengos (2020) also uncovered the negative relationship between the U.S. federal funds market and the XRP market.

Table 7. Cryptocurrency returns determinants.

y_t = Cryptocurrency's Returns	BTC	ETH	XRP	USDT	BCH	BSV	LTC
Variable	Estimated Coefficient						
y_{t-1}						−0.12	
y_{t-2}			0.12	−0.44		0.15	
NYSE		0.36	0.60	−0.11			
S&P500	0.53						
S&P500$_{t-2}$							−0.01
Gold	0.54	0.47				0.71	0.77
Mom			0.45	−0.07	0.93		0.52
Mom$_{t-1}$							−0.75
FFR			−0.01				
Euro				−0.12			0.92
Euro$_{t-1}$							−0.93
Euro$_{t-2}$				−0.11			
Intercept	0.001	0.001	0.006	0.001	0.001	0.001	0.002

Note: After model selection, all of the coefficients in the table above are statistically significant at the 1% significance level.

For the USDT, its lagged returns, NYSE and the Mom negatively affect the current daily returns. The Euro, which represents the daily percentage change from the foreign exchange rate, and the lagged Euro negatively affect the USDT daily returns. It suggests that the USDT is more exposed to the volatility in the Forex market. Based on the estimation results, the USDT could provide an investment strategy that hedges the risk exposures from the XRP market as they react contrarily to the fluctuations in financial markets.

The BCH's daily returns rely heavily on the Mom factor, and the Mom has a positive effect on the current daily returns. As a derivative cryptocurrency from the original cryptocurrency Bitcoin, the BCH market is much more speculative based on this evidence. BCH investors tend to focus on the financial markets ' momentum as the central reference point for developing investment strategies.

However, for the BSV, which is also a derivative cryptocurrency from the original cryptocurrency Bitcoin, its current daily returns are strongly correlated with its lagged returns. Besides, BSV heavily relies on the daily returns from the gold as the primary returns determinant. It suggests that the BSV market is more closely connected to the gold market, and it provides similar risk-hedging investment strategies to the gold.

For the LTC market, contemporaneous returns from Gold, Mom and Euro have positive effects on the LTC's contemporaneous daily returns. In contrast, the lagged returns from S&P 500, Mom and Euro have negative effects. This finding is also consistent with Deniz and Stengos (2020). It suggests that the LTC market is contemporaneously positively correlated with the gold market, the Forex market, and the financial market's momentum while being negatively affected by the lagged information from the markets above. The LTC market closely reacts to the fluctuations in the gold market and the Forex market, and the momentum of the financial market.

Overall, most of the cryptocurrencies (BTC, ETH, BSV and LTC) in Table 7 are sensitive to changes in the gold market as they provide a similar investment strategy to hedge against the risks from the real economy. In this aspect, the cryptocurrencies above provide similar functionalities to the precious metals market.

Besides Gold, NYSE, Mom, and Euro are also major deciding factors for the cryptocurrencies' daily returns. In particular, ETH, XRP, USDT, BCH and LTC are sensitive to the changes and the momentum in financial markets. It suggests these cryptocurrencies show more extensive exposures to the fluctuations in conventional financial assets. Furthermore, the dependence on the lagged returns for cryptocurrencies such as the XRP, USDT and BSV suggests that their returns follow an autoregressive process so that market memories from the lagged market information are a crucial factor to consider when modelling for their daily returns.

6. Forecast Evaluation

We generate one-step-ahead forecasts for the daily returns of each cryptocurrency by the recursive forecasting approach. Following Hansen (2008) and Xiao and Sun (2019), forecasts for the cryptocurrencies' daily returns from the ARDL(p,q) model are averaged by the SMMA estimator. Specifically, we forecast the next day's daily returns based on the market information accumulated up to the previous day.

To facilitate the forecast evaluation, we compare the ARDL(p,q) model's forecasting performance when estimated by the SMMA estimator against the aforementioned competing estimators and competing models. The competing models are the traditional asset pricing model (APM) $y = \alpha + X\beta + \varepsilon$, the constant model $y = \alpha + \varepsilon$, the artificial neural network (ANN) model with *tanh* activation function and the back-propagation neural network (BPNN) model.

For performance evaluation, we compare the forecast accuracy of respective forecasting approaches by the root mean squared forecast error (RMSFE), defined as:

$$RMSFE = \sqrt{\frac{1}{f}\sum_{t=1}^{f}(y_t - \widetilde{y}_t)^2}, \tag{25}$$

where $\widetilde{y}_t$ is the out-of-sample forecast for y_t generated by respective approaches and f represents the forecast sample size.

Similarly, the root mean absolute forecast error (RMAFE) is defined as:

$$RMAFE = \sqrt{\frac{1}{f}\sum_{t=1}^{f}|y_t - \widetilde{y}_t|}. \tag{26}$$

We used the "Rstudio" software to conduct the forecast evaluation, and the computation took 168 h on a computer with an Intel(R) Core(TM) i7-7700HQ CPU. The forecasting performance evaluation of the competing estimators and competing models is presented in Table 8 below.

Table 8. Cryptocurrency returns one-step ahead forecast evaluation. RMSFE, root mean squared forecast error; RMAFE, root mean absolute forecast error.

Cryptocurrency	BTC	ETH	XRP	USDT	BCH	BSV	LTC
Model				**RMSFE**			
ARDL(3,3) (PLS)	0.11	0.13	0.15	0.05	0.15	0.15	0.14
ARDL(3,3) (Naive OLS)	0.34	0.35	0.35	0.34	0.34	0.34	0.35
ARDL(3,3) (SMMA)	0.08	0.1	0.11	0.01	0.12	0.11	0.1
APM	0.11	0.14	0.15	0.06	0.16	0.16	0.14
Constant	0.11	0.13	0.14	0.05	0.15	0.14	0.13
ANN	0.33	0.34	0.35	0.29	0.36	0.33	0.32
BPNN	0.11	0.14	0.15	0.06	0.16	0.15	0.14
				RMAFE			
ARDL(3,3) (PLS)	0.02	0.03	0.03	0.02	0.04	0.04	0.03
ARDL(3,3) (Naive OLS)	0.24	0.15	0.15	0.14	0.14	0.15	0.15
ARDL(3,3) (SMMA)	0.01	0.02	0.02	0.01	0.03	0.03	0.02
APM	0.02	0.03	0.03	0.02	0.04	0.04	0.03
Constant	0.02	0.03	0.03	0.02	0.04	0.04	0.03
ANN	0.15	0.15	0.17	0.14	0.17	0.15	0.14
BPNN	0.02	0.03	0.03	0.01	0.04	0.04	0.03
Forecast Horizon	813	563	461	1172	929	453	715

Note: PLS stands for the corresponding best performing penalized least squares estimators for each cryptocurrency shown in Table 5. For forecast horizons, we do not attempt to forecast the weekend daily returns for all of the cryptocurrencies studied in this paper as major financial markets such as the NYSE and S&P 500 do not trade on weekends.

Since the actual returns for the cryptocurrencies y and the forecasts for the cryptocurrency returns $\tilde{y}$ are already in log terms, the RMSFE and the RMAFE give the approximate average squared percentage deviations[1] and average absolute percentage deviations from the actual returns by the respective forecasting method.

It is clear from the forecasting performances that the ARDL(p,q) model estimated by the SMMA estimator consistently outperforms the competing estimators and competing models in both RMSFE and RMAFE. This result supports the application of the ARDL(p,q) model and the SMMA estimator in forecasting the cryptocurrency markets. It further highlights the importance of attending to model uncertainty in forecasting cryptocurrencies.

7. Conclusions

This paper has investigated the major cryptocurrencies' returns determinants and introduced a novel model averaging approach, namely the SMMA estimator, for forecasting cryptocurrencies' returns.

Model uncertainty features most asset pricing studies, and such model uncertainty proves especially challenging for cryptocurrencies in the absence of a unifying theory that should guide modelling. In particular, the issue of model's lag selection, the discrete nature of the traditional model selection algorithms, and the sensitivity of the tuning parameter choice affecting variable selection outcomes for penalized estimators all call for a novel modelling approach.

Therefore, to handle model uncertainty, we introduced the SMMA estimator to forecast the major cryptocurrencies' returns. The SMMA estimator combines model selection and model averaging to handle model uncertainty and outperforms when averaging high dimensional sparse models against model uncertainty. Our model averaging approach outperformed the conventional benchmark forecasting models for the cryptocurrencies in the literature.

We first investigated the cryptocurrencies' returns determinants, we find that most cryptocurrencies are sensitive to changes in the financial markets such as the NYSE and the S&P 500, and dynamics in the financial market momentum. They are also sensitive to the changes in gold prices and current and lagged information from the Forex market.

---

1 $log(y) - log(\tilde{y}) \approx \frac{y-\tilde{y}}{\tilde{y}}$.

From the forecast evaluation, it is clear that the ARDL(p,q) model estimated by the SMMA estimator consistently outperforms the competing estimators and models in RMSFE and RMAFE. The forecast evaluation result supports the application of the ARDL(p,q) model and the SMMA estimator in forecasting the cryptocurrency markets. It further highlights the importance of attending to model uncertainty in forecasting cryptocurrencies.

This paper remains limited in the following aspects. Due to the model uncertainty in cryptocurrencies, this paper could be further improved by considering more explanatory variables, especially the text-based variables that could help gauge the market sentiment for cryptocurrencies. Furthermore, more variants of machine learning approaches could be further introduced to compete against the SMMA estimator in forecasting the cryptocurrencies' returns.

Author Contributions: Both authors contributed to the project formulation and paper preparation. All authors read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We thank the anonymous referees for their constructive comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

Aharon, David Yechiam, and Mahmoud Qadan. 2019. Bitcoin and the day-of-the-week effect. *Finance Research Letters* 31. [CrossRef]

Belloni, Alexandre, and Victor Chernozhukov. 2013. Least squares after model selection in high-dimensional sparse models. *Bernoulli* 5: 521–47. [CrossRef]

Bouri, Elie, Rangan Gupta, Aviral Kumar Tiwari, and David Roubaud. 2017. Does bitcoin hedge global uncertainty? Evidence from wavelet-based quantile-in-quantile regressions. *Finance Research Letters* 23: 87–95. [CrossRef]

Breheny, Patrick, and Jian Huang. 2011. Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *The Annals of Applied Statistics* 3: 232–53. [CrossRef] [PubMed]

Demir, Ender, Giray Gozgor, Chi Keung Marco Lau, and Samuel A. Vigne. 2018. Does economic policy uncertainty predict the bitcoin returns? An empirical investigation. *Finance Research Letters* 26: 145–49. [CrossRef]

Deniz, Pinar, and Thanasis Stengos. 2020. Cryptocurrency Returns before and after the Introduction of Bitcoin Futures. *Journal of Risk and Financial Management* 13: 116. [CrossRef]

Dyhrberg, Anne Haubo. 2016a. Bitcoin, gold and the dollar—A garch volatility analysis. *Finance Research Letters* 16: 85–92. [CrossRef]

Dyhrberg, Anne Haubo. 2016b. Hedging capabilities of bitcoin. Is it the virtual gold? *Finance Research Letters* 16: 139–44. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3–56. [CrossRef]

Fan, Jianqing, and Runze Li. 2001. Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association* 456: 1348–60. [CrossRef]

Hansen, Bruce E. 2007. Least squares model averaging. *Econometrica* 2: 1175–89. [CrossRef]

Hansen, Bruce E. 2008. Least-squares forecast averaging. *Journal of Econometrics* 146: 342–50. [CrossRef]

Hoerl, Arthur E., and Robert W. Kennard. 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* 2: 55–67. [CrossRef]

Nakamoto, Satoshi. 2008. Bitcoin: A Peer-to-Peer Electronic Cash System. Available online: https://bitcoin.org/bitcoin.pdf (accessed on 8 August 2020).

Panagiotidis, Theodore, Thanasis Stengos, and Orestis Vravosinos. 2019. The effects of markets, uncertainty and search intensity on bitcoin returns. *International Review of Financial Analysis* 63: 220–42. [CrossRef]

Schomaker, Michael. 2012. Shrinkage averaging estimation. *Statistical Papers* 11: 1015–34. [CrossRef]

Shi, Peide, and Chih-Ling Tsai. 2002. Regression model selection-A residual likelihood approach. *Journal of the Royal Statistical Society Series B* 5: 237–52. [CrossRef]

Tibshirani, Robert. 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58: 267–88. [CrossRef]

Tibshirani, Rob, Trevor Hastie, and Jerome Friedman. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of statistical software* 2: 1–22.

Wu, Shan, Mu Tong, Zhongyi Yang, and Abdelkader Derbali. 2019. Does gold or bitcoin hedge economic policy uncertainty? *Finance Research Letters* 31: 171–78. [CrossRef]

Xiao, Hui, and Yiguo Sun. 2019. On Tuning Parameter Selection in Model Selection and Model Averaging: A Monte Carlo Study. *Journal of Risk and Financial Management* 12: 109. [CrossRef]

Zhang, Cun-Hui. 2010. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics* 4: 894–942. [CrossRef]

Zou, Hui. 2006. The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association* 2: 1418–29. [CrossRef]