

Xie, Tian

**Article**

## Forecast Bitcoin volatility with least squares model averaging

Econometrics

**Provided in Cooperation with:**

MDPI – Multidisciplinary Digital Publishing Institute, Basel

*Suggested Citation:* Xie, Tian (2019) : Forecast Bitcoin volatility with least squares model averaging, *Econometrics*, ISSN 2225-1146, MDPI, Basel, Vol. 7, Iss. 3, pp. 1-20, <https://doi.org/10.3390/econometrics7030040>

This Version is available at:

<https://hdl.handle.net/10419/247540>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

Article

# Forecast Bitcoin Volatility with Least Squares Model Averaging

Tian Xie

College of Business, Shanghai University of Finance and Economics, Shanghai 200433, China; xietian001@hotmail.com

Received: 11 July 2019; Accepted: 11 September 2019; Published: 14 September 2019



**Abstract:** In this paper, we study forecasting problems of Bitcoin-realized volatility computed on data from the largest crypto exchange—Binance. Given the unique features of the crypto asset market, we find that conventional regression models exhibit strong model specification uncertainty. To circumvent this issue, we suggest using least squares model-averaging methods to model and forecast Bitcoin volatility. The empirical results demonstrate that least squares model-averaging methods in general outperform many other conventional regression models that ignore specification uncertainty.

**Keywords:** volatility forecasting; HAR; model uncertainty; model averaging; crypto currency

**JEL Classification:** C52; C53; G12; G17

## 1. Introduction

Bitcoin, the first and still one of the foremost applications of blockchain technology by far, was introduced early in 2008. Until the end of December 2018, the market capitalization of Bitcoin was roughly \$65 billion with \$3800 per token. As for the whole Bitcoin network, by the end of December 2018, there are more than 10,000 full nodes distributed across the world and roughly \$2.5 billion of value transacted on the main network. With the growth of the Bitcoin market, many investors are starting to view it as an emerging new asset class. In September 2015, the Commodity Futures Trading Commission (CFTC) in the United States officially designated Bitcoin as a commodity. Improved measures of Bitcoin volatility enable us to better gauge the current level of volatility and to understand its dynamics. Most importantly, Bitcoin volatility is now directly tradable,<sup>1</sup> which accredits the importance of Bitcoin volatility forecasting.

How to model and predict the volatility of financial assets is an interesting topic in risk management. Traditional approaches employ parametric models such as the generalized autoregressive conditional heteroskedasticity (GARCH) or stochastic volatility models. Recently a new approach to modeling volatility dynamics has relied on improved measures of ex post volatility composed from high-frequency intraday data. This new measure is called realized volatility (RV), which possesses a slowly decaying autocorrelation function, sometimes known as long-memory.<sup>2</sup> Various models have been proposed to capture stylized facts of realized volatility series, such as the fractionally integrated autoregressive moving average (ARFIMA) models<sup>3</sup> used in Andersen et al.

<sup>1</sup> The CME Group Inc. (Chicago Mercantile Exchange & Chicago Board of Trade) in December 2017 launched Bitcoin future (XBT), with Bitcoin as the underlying asset.

<sup>2</sup> This phenomenon has been documented by Dacorogna et al. (1993) and Andersen et al. (2001b) for the foreign exchange market and by Andersen et al. (2001a) for stock market returns.

<sup>3</sup> ARFIMA is designed to model time series with long memory at the beginning. It is now a popular tool for modeling volatility, since volatility exhibits long memory.

(2001b) and the heterogeneous autoregressive (HAR) model proposed by Corsi (2009). Compared with the ARFIMA model, the HAR model soon gained popularity because of its computational simplicity (e.g., ordinary least squares) and excellent out-of-sample performance.<sup>4</sup>

The HAR model can provide an intuitive economic interpretation that agents with three frequencies of trading (daily, weekly, and monthly) perceive and respond to, which changes the corresponding components of volatility.<sup>5</sup> Nevertheless, the suitability of such a specification is not subject to enough verification. Craioveanu and Hillebrand (2012) employed a parallel computing method to investigate all the possible combinations of lags in the additive model. Others tested the validity of the lag structure in the conventional HAR model from a model selection perspective; see, e.g., Audrino et al. (2015,2016); and Audrino et al. (2016), among others. While the lag terms in the HAR model survive the tests based on the least absolute shrinkage and selection operator (LASSO) and the adaptive LASSO (Audrino et al. 2015; Audrino and Knaus 2016) only in the case of simulated data by the HAR model, there is strong evidence in Audrino et al. (2016) that casts some doubts on the fixed choice of aggregation frequencies in the HAR model. In particular, Audrino et al. (2016) found that a conventional fixed lag structure was not statistically sustained by the group LASSO estimates for certain individual stocks in an unstable market environment such as the 2007–2009 crisis. They addressed the above issue with a proposed flexible HAR model, built dynamically from the group LASSO estimates.

The above conclusions may or may not hold in Bitcoin volatility forecasting considering the unique features of the crypto asset market. To tackle this question from a different angle, we consider the forecast implication of a flexible lag structure generated by the least squares model-averaging method. Unlike the model selection approach that picks only one winning model out of a pool of candidate models, model averaging calculates the weighted average of a group of candidate models. Barnard (1963) first discussed the concept of “model combination” in a paper studying airline passenger data. Buckland et al. (1997) suggested using the exponential Akaike information criterion (AIC) estimates as the model weights and proposed the model averaged AIC. There exists many other averaging-type approaches that provide a means to tackle model uncertainty, for instance, the Bayesian model-averaging method discussed in length in Hoeting et al. (1999), the weighted-average least squares method by Magnus et al. (2010), and the random forest method by Breiman (2001), among others.

The performance of the model-averaging method heavily relies on the weights chosen for the estimation process. In a pioneering study, Hansen (2007) proposed the Mallows model averaging (MMA) method that is asymptotically optimal in the sense of achieving the lowest possible mean squared errors. Wan et al. (2010) completed the theoretical foundation of the MMA. Extensions of the MMA that allow possible structural breaks, near unit root, and heteroskedasticity can be found in Hansen (2009,2010), and Hansen and Racine (2012), respectively. Xie (2015) proposed the prediction model averaging (PMA) method. Zhao et al. (2016) extended the PMA method to allow for heteroskedastic error terms (HPMA). Liu and Okui (2013) also proposed a heteroskedasticity-robust Mallows'  $C_p$  model-averaging method (HRCP).

There is a growing literature on solving the model uncertainty issue in volatility forecasting with least squares model averaging. Lehrer et al. (2018) proposed the model averaging HAR (MAHAR) method that optimally averages the forecasts of HAR models with different lag indexes. Qiu et al. (2019) showed that the above method can be extended to a more complicated HAR model with estimators of the variation of positive and negative returns (semi-variance components). Besides the above methods, we consider the approach designed by Qiu and Xie (2018), who proposed the heteroskedasticity-robust model averaging HAR method (H-MAHAR) that mainly applies the

---

<sup>4</sup> Corsi et al. (2012) provided a comprehensive review of the development of HAR-type models and their various extensions.

<sup>5</sup> Müller et al. (1993) referred to this interpretation as the Heterogeneous Market Hypothesis.

HPMA as the core model averaging estimator to exchange rate volatility. As a complement to the HPMA, we also include the jackknife model averaging (JMA) and the heteroskedasticity robust  $C_p$  (HRCP) model averaging estimators as companion methods in this paper.

In the empirical exercise, we consider a series of estimators including 9 conventional regression methods, 1 LASSO method, and 4 model-averaging methods to model and forecast the realized variance of Bitcoin prices. We show that the model-averaging methods that account for model uncertainty generally outperform the conventional regressions and the model-selection-based LASSO method. Moreover, the heteroskedasticity-robust methods tend to perform relatively better. Compared with non-model-averaging methods, the H-MAHAR method yields the highest forecasting accuracy in most of the exercises. The improvement that H-MAHAR provides is statistically significant at the 5% level, as confirmed by the Giacomini–White test (Giacomini and White 2006).

The remainder of the paper is arranged as follows. Section 2 provides a more detailed overview of existing HAR strategies. Section 3 discusses the way to model uncertainty under heteroskedasticity using least squares model averaging. Section 4 describes the data. Section 5 presents the empirical results, where we compared 14 methods in rolling window exercises. In all cases, model-averaging methods tended to have the dominating performance. To examine the robustness of the results, we tried different experimental settings in Section 6. Section 7 concludes this paper.

## 2. Prior HAR-Type Strategies to Forecast Volatility

Following Andersen and Bollerslev (1998), we estimate daily RV at day  $t$  ( $RV_t$ ) by summing the corresponding  $M$  equally spaced intra-daily squared returns  $r_{t,j}$ . Here, the subscript  $t$  indexes day  $t$  and  $j$  indicates the time within day  $t$ ,

$$RV_t \equiv \sum_{j=1}^M r_{t,j}^2 \quad (1)$$

where  $t = 1, 2, \dots, T$ ,  $j = 1, 2, \dots, M$ , and  $r_{t,j}$  define continuously compounded high-frequency returns by differing log-prices  $p_{t,j}$  ( $r_{t,j} = p_{t,j} - p_{t,j-1}$ ).

Among the RV models, the HAR model proposed by Corsi (2009) is quite prevalent. Not only is this because the HAR model accurately approximates the long-memory and multiscaling properties of RV but also this is very easy to implement in practice. The standard HAR model in Corsi (2009) postulates that the  $h$ -step-ahead daily  $RV_{t+h}$  can be described by

$$RV_{t+h} = \beta_0 + \beta_d RV_t^{(1)} + \beta_w RV_t^{(5)} + \beta_m RV_t^{(22)} + e_{t+h}, \quad (2)$$

where the explanatory variables can take the general form of  $RV_t^{(l)}$ .  $RV_t^{(l)}$  is defined by

$$RV_t^{(l)} \equiv l^{-1} \sum_{s=1}^l RV_{t-s} \quad (3)$$

where  $l$  is the period averages of daily RV,  $\beta$  is the coefficients, and  $\{e_t\}_t$  is a zero mean innovation process. The standard HAR model in Equation (2) is pinned down by some vector of lag index  $l = [1, 5, 22]$ .

Andersen et al. (2007) extended the standard HAR model two ways. First, they added the daily jump component  $J_t$  to Equation (2) to explicitly capture its impacts. The extended model is denoted the HAR-J model:

$$RV_{t+h} = \beta_0 + \beta_d RV_t^{(1)} + \beta_w RV_t^{(5)} + \beta_m RV_t^{(22)} + \beta^j J_t + e_{t+h}, \quad (4)$$

where the empirical measurement of the squared jumps is  $J_t = \max(RV_t - BPV_t, 0)$  and the standardized realized bipower variation (BPV) is defined as

$$BPV_t \equiv (2/\pi)^{-1} \sum_{j=2}^M |r_{t,j-1}| |r_{t,j}|.$$

Second, through a decomposition of RV into the continuous sample path and the jump component based on the  $Z_{1,t}$  statistic, Andersen et al. (2007) reconstructed the HAR-J model by explicitly incorporating the two types of volatility components mentioned above. The  $Z_{1,t}$  statistic identifies the “significant” jumps  $CJ_t$  and the continuous sample path components  $CSP_t$  respectively as

$$\begin{aligned} CSP_t &\equiv \mathbb{I}(Z_t \leq \Phi_\alpha) \cdot RV_t + \mathbb{I}(Z_t > \Phi_\alpha) \cdot BPV_t, \\ CJ_t &= \mathbb{I}(Z_t > \Phi_\alpha) \cdot \max(RV_t - BPV_t, 0), \end{aligned}$$

where  $Z_t$  is the ratio statistic in Huang and Tauchen (2005)<sup>6</sup> and  $\Phi_\alpha$  is the cumulative distribution function (CDF) of a standard Gaussian distribution with an  $\alpha$  level of significance. The daily, weekly, and monthly average components of  $CSP_t$  and  $CJ_t$  are then constructed in the same manner as  $RV_t^{(l)}$  in Equation (3). The model specification for the continuous HAR-J, in other words, the HAR-CJ, is given by

$$RV_{t+h} = \beta_0 + \beta_d^c CSP_t^{(1)} + \beta_w^c CSP_t^{(5)} + \beta_m^c CSP_t^{(22)} + \beta_d^j CJ_t^{(1)} + \beta_w^j CJ_t^{(5)} + \beta_m^j CJ_t^{(22)} + e_{t+h}. \quad (5)$$

Note the HAR-CJ model explicitly controls for the weekly and monthly effects of continuous jumps through the  $CJ_t^{(1)}$ ,  $CJ_t^{(5)}$ , and  $CJ_t^{(22)}$  terms, whereas the HAR-J model consists of only one aggregate jump term  $J_t$ . Thus, the HAR-J model can be regarded as a special and restrictive case of the HAR-CJ model for  $\beta_d = \beta_d^c + \beta_d^j$ ,  $\beta^j = \beta_d^j$ ,  $\beta_w = \beta_w^c + \beta_w^j$ , and  $\beta_m = \beta_m^c + \beta_m^j$ .

To capture the role of the “leverage effect” in predicting volatility dynamics, Patton and Sheppard (2015) developed a group of models using signed realized measures. The first model, denoted as HAR-RS-I, decomposes the daily RV in the standard HAR model (Equation (2)) into two asymmetric semi-variances:  $RS_t^+$  and  $RS_t^-$ .

$$RV_{t+h} = \beta_0 + \beta_d^+ RS_t^+ + \beta_d^- RS_t^- + \beta_w RV_t^{(5)} + \beta_m RV_t^{(22)} + e_{t+h}, \quad (6)$$

where  $RS_t^- = \sum_{j=1}^M r_{t,j}^2 \cdot \mathbb{I}(r_{t,j} < 0)$  and  $RS_t^+ = \sum_{j=1}^M r_{t,j}^2 \cdot \mathbb{I}(r_{t,j} > 0)$ . To verify whether the realized semi-variances add something beyond the classical leverage effect, Patton and Sheppard (2015) augmented the HAR-RS-I model with a term interacting the lagged RV with an indicator for negative lagged daily returns  $RV_t^{(1)} \cdot \mathbb{I}(r_t < 0)$ . The second model in Equation (7) is named HAR-RS-II.

$$RV_{t+h} = \beta_0 + \beta_1 RV_t^{(1)} \cdot \mathbb{I}(r_t < 0) + \beta_d^+ RS_t^+ + \beta_d^- RS_t^- + \beta_w RV_t^{(5)} + \beta_m RV_t^{(22)} + e_{t+h}, \quad (7)$$

<sup>6</sup> The ratio statistic is defined as

$$Z_t(\Delta) = \Delta^{-1/2} \times \frac{[RV_t(\Delta) - BPV_t(\Delta)]RV_t(\Delta)^{-1}}{[(\mu_1^{-4} + 2\mu_1^{-2} - 5) \max\{1, TQ_t(\Delta)BPV_t(\Delta)^{-2}\}]^{1/2}}$$

where  $\Delta$  is the notion of increasingly finer sampled returns,  $\mu_1 = \mathbb{E}(|\mathcal{Z}|)$  denotes the mean of the absolute value of standard normally distributed random variable, and  $TQ_t$  is the standardized realized tripower quarticity measure:  $TQ_t(\Delta) = \Delta^{-1} \mu_{4/3}^{-3} \sum_{j=3}^{1/\Delta} |r_{t+j\Delta, \Delta}|^{4/3} |r_{t+(j-1)\Delta, \Delta}|^{4/3} |r_{t+(j-2)\Delta, \Delta}|^{4/3}$  with  $\mu_{4/3} = \mathbb{E}(|\mathcal{Z}|^{4/3})$ .

where  $RV_t^{(1)} \cdot \mathbb{I}(r_t < 0)$  is designed to capture the effect of negative daily returns. As in the HAR-CJ model, the third and fourth models in Patton and Sheppard (2015), denoted as HAR-SJ-I and HAR-SJ-II respectively, disentangle the signed jump variations and the BPV from the volatility process.

$$RV_{t+h} = \beta_0 + \beta_d^j SJ_t + \beta_d^{bpv} BPV_t + \beta_w RV_t^{(5)} + \beta_m RV_t^{(22)} + e_{t+h}, \tag{8}$$

$$RV_{t+h} = \beta_0 + \beta_d^j SJ_t + \beta_d^{j+} SJ_t^+ + \beta_d^{j-} SJ_t^- + \beta_d^{bpv} BPV_t + \beta_w RV_t^{(5)} + \beta_m RV_t^{(22)} + e_{t+h}, \tag{9}$$

where  $SJ_t = RS_t^+ - RS_t^-$ ,  $SJ_t^+ = SJ_t \cdot \mathbb{I}(SJ_t > 0)$ , and  $SJ_t^- = SJ_t \cdot \mathbb{I}(SJ_t < 0)$ . The HAR-SJ-II model extends the HAR-SJ-I model by distinguishing the effect of a positive jump variation from that of a negative jump variation.

### 3. Model Uncertainty

It has been a tradition for the past literature to assume the lag structure of the HAR model to be  $l = [1, 5, 22]$ , which mimics the daily, weekly, and monthly traders in traditional financial markets that only open on workdays. On the other hand, given the 24/7 nonstop nature of bitcoin trading, it may not be appropriate to set the lag index at  $[1, 5, 22]$ . An initial guess for the lag index would be  $l = [1, 7, 30]$  that represents the tradition of daily, weekly, and monthly averages. However, the suitability of such a specification is subject to a statistical investigation, which is likely to cause evident model uncertainty.

Suppose the dependent variable is  $y = [RV_1, \dots, RV_T]^\top$  and the explanatory variable is  $X = [x_1, \dots, x_T]^\top$ ,<sup>7</sup> where the specification of  $x_t$  takes the general form of the HAR model

$$x_t = [1, RV_{t-h}^{(l_1)}, RV_{t-h}^{(l_2)}, \dots, RV_{t-h}^{(l_p)}]. \tag{10}$$

Here, we do not restrict the lag index  $l = [l_1, l_2, \dots, l_p]$  to be  $[1, 5, 22]$ . Instead, we acknowledge the specification uncertainty in  $l$  and consider a group of  $M$  candidate models to approximate the true data generating process. Following an usual approach in the model averaging literature, the set of  $M$  candidate models is constructed by taking a full permutation of all the lags from  $RV_{t-h}^{(l_1)}$  to  $RV_{t-h}^{(l_p)}$  ( $RV_{t-h}^{(l_1)}, \dots, RV_{t-h}^{(l_p)}$  and  $[l_1, \dots, l_p] = [1, \dots, 30]$ ). The maximum lag order  $l_p$  is chosen as 30. In this way, there are distinct model weights assigned to each HAR-type model with different lag combinations. Moreover, as the underlying data sets vary, this will alter the relevant model weights, which effectively makes the method dynamic and data-driven.

Note that the model averaging estimator with pre-screened candidate models is implemented in this paper, since keeping the total number of candidate models manageable or slowing its convergence to infinity is a necessary condition to maintain the asymptotic optimality of least square model averaging estimators. However, in the context of the HAR model with a maximum lag order of  $l_p$ , we could end up with  $2^{l_p}$  candidate models and the number of potential models grows exponentially with  $l_p$ . To solve this issue, we first apply the model screening method, for example, the adaptive regression by mixing with the model selection (ARMS) approach by Yuan and Yang (2005) or the hetero-robust model screening (HRMS) approach by Xie (2017). Both methods shrink the number of potential models by specifying model selection criteria before model averaging to an appropriate degree.

The true model is presumed to be

$$y = \mu + e, \tag{11}$$

where  $y = [y_1, \dots, y_T]^\top$ ,  $\mu = [\mu_1, \dots, \mu_T]^\top$ , and  $e = [e_1, \dots, e_T]^\top$ .  $\mu_t$  can be considered the conditional mean in the period  $t$ ,  $\mu_t = \mathbb{E}(y_t | y_{t-h}, y_{t-h-1}, \dots)$ , and the error term  $e_t$  has the zero conditional mean

<sup>7</sup> Although all the elements in  $x_t$  are  $h$ -period lags from the period  $t$ , we follow the conventional notation in time series and denote  $x_t$  as the explanatory variable corresponding to the period  $t$  dependent variable.

$\mathbb{E}(e_t|y_{t-h}, y_{t-h-1}, \dots) = 0$ . Note that the error term  $e$  is assumed to be heteroskedastic such that  $\mathbb{E}(e_t^2|x_t) = \sigma_t^2$ , which reflects a more realistic characterization of the realized volatility for a wide class of financial assets. In addition, we also hypothesize that  $e$  is not serially correlated and  $\mathbb{E}(e^\top e|X) = \Omega = \text{diag}\{\sigma_1^2, \dots, \sigma_T^2\}$ .<sup>8</sup> Let the  $m$ th candidate model be

$$y = X^m \beta^m + e^m,$$

where  $X^m$  are subsets of columns of  $X$ . With  $X^m$  at hand,  $\beta^m$  can be estimated by  $\hat{\beta}^m = (X^{m\top} X^m)^{-1} X^{m\top} y$ , and thus,  $\mu$  is estimated by

$$\hat{\mu}^m = X^m \hat{\beta}^m = X^m (X^{m\top} X^m)^{-1} X^{m\top} y = P^m y,$$

where  $P^m$  is a projection matrix for the model  $m$ . Extending from Hansen (2008), the optimal mean-square  $h$ -period ahead forecast is the conditional mean  $\mu_{T+h}$ . Therefore, the least-squares forecast of  $y_{T+h}$  from the  $m$ th approximation model is then  $\hat{y}_{T+h}^m = \hat{\mu}_{T+h}^m = x_{T+h}^{m\top} \hat{\beta}^m$ . Note that by the definition of Equation (10),  $x_{T+h}^m$  is observable in period  $t$ .

We obtain the forecasts of  $y_{T+h}$  from all approximation models and define the vector of forecasts  $\hat{y}_{T+h}$

$$\hat{y}_{T+h} \equiv [\hat{y}_{T+h}^1, \hat{y}_{T+h}^2, \dots, \hat{y}_{T+h}^M]^\top. \tag{12}$$

The model averaging forecast is simply the weighted average of  $\hat{y}_{T+h}$  such that

$$\hat{y}_{T+h}(w) \equiv w^\top \hat{y}_{T+h} = \sum_{m=1}^M w^m \hat{y}_{T+h}^m,$$

where  $w = [w^1, \dots, w^M]^\top$  is a weight vector in the unit simplex in  $\mathbb{R}^M$

$$\mathcal{H} \equiv \left\{ w \in [0, 1]^M : \sum_{m=1}^M w^m = 1 \right\}.$$

The performance of model averaging forecast crucially depends on the weight vector  $w$ . The model averaging estimator of the conditional mean is then given by

$$\hat{\mu}(w) \equiv P(w)y, \tag{13}$$

where  $P(w) \equiv \sum_{m=1}^M w^m P^m$  is the averaged projection matrix. The H-MAHAR method is the heteroskedasticity-robust version of the model averaging HAR (MAHAR) method proposed by Lehrer et al. (2018). The MAHAR criterion function is defined as follows:

$$\text{MAHAR}(w) = (y - \hat{\mu}(w))^\top (y - \hat{\mu}(w)) \left( \frac{T + k(w)}{T - k(w)} \right), \tag{14}$$

where  $k(w) \equiv \sum_{m=1}^M w^m k^m$  is the effective number of parameters and  $k^m$  is the number of regressors in the model  $m$ . We estimate the MAHAR weight estimator by minimizing the MAHAR criterion function under the restriction of  $w \in \mathcal{H}$ .

---

<sup>8</sup> Corsi et al. (2008) also demonstrated that the residuals of commonly used realized volatility models for the S&P 500 index exhibit non-Gaussianity and volatility clustering. They assessed its relevance for modeling and forecasting volatility in the proposed HAR-GARCH model.

Like most model selection and model averaging criteria, the H-MAHAR criterion balances between the fit and the complexity of a model:

$$\text{H-MAHAR}(w) = (y - \hat{\mu}(w))^\top (y - \hat{\mu}(w)) + 2\text{tr}(P(w)\hat{\Omega}(w)), \tag{15}$$

where  $\hat{\Omega}(w) \equiv \text{diag}\{\hat{e}_1^2(w), \dots, \hat{e}_T^2(w)\}$  is the averaged estimate of the  $\Omega$  matrix using model averaging residuals  $\hat{e}(w) = [\hat{e}_1^2(w), \dots, \hat{e}_T^2(w)]^\top = y - \hat{\mu}(w)$ .

The criterion in Equation (15) can be implemented to compute the empirical weight vector  $\hat{w}$  through

$$\hat{w} = \arg \min_{w \in \mathcal{H}} \text{H-MAHAR}(w).$$

Therefore, we obtain the model averaging forecast of  $y_{T+h}$  following  $\hat{y}_{T+h}(\hat{w}) = \hat{w}^\top \hat{y}_{T+h}$ . Note that the H-MAHAR estimator can be considered an extension to the model averaging with averaging covariance matrix (MAACM) estimator of Zhao et al. (2016) under the HAR framework, whereas the original MAACM estimator assumes no dynamic model structures.

Another heteroskedasticity-robust model-averaging method is the JMA estimator by Hansen and Racine (2012). The original JMA deals with cross-sectional data. Zhang et al. (2013) proved the asymptotic optimality of the JMA estimator under a dependent time-series. The JMA estimator is also known as leave-one-out cross-validation model averaging. As its name indicates, the JMA requires the use of jackknife residuals for the average estimator. The jackknife residual vector for model  $m$  can be conveniently expressed as  $\hat{e}_J^m = D^m \hat{e}^m$ , where  $\hat{e}^m$  is the least squares residual vector and  $D^m$  is the  $n \times n$  diagonal matrix with the  $i$ th diagonal element equal to  $(1 - h_i^m)^{-1}$ . The term  $h_i^m$  is the  $i$ th diagonal element of the projection matrix  $P^m$ . Define an  $n \times M$  matrix with all the jackknife residuals, in which  $\hat{E}_J = [\hat{e}_J^{(1)}, \dots, \hat{e}_J^{(M)}]$ . The least squares cross-validation criterion for the JMA is simply

$$\text{JMA}_n(w) = \frac{1}{n} w^\top \hat{E}_J^\top \hat{E}_J w$$

with model weights  $w$  estimated through  $\hat{w} = \text{argmin}_{w \in \mathcal{H}} \text{JMA}_n(w)$ .

Liu and Okui (2013) adopted the same model setup to propose the HRCP model averaging estimator for linear regression models with heteroskedastic errors. They demonstrated the asymptotic optimality of the HRCP estimator when the error term exhibits heteroskedasticity. They proposed estimating the model weights by the following feasible HRCP criterion:

$$\text{HRCP}(w) = \|y - P(w)y\|^2 + 2 \sum_{i=1}^n \hat{e}_i^2 p_{ii}(w) \tag{16}$$

with  $\hat{w} = \arg \min_{w \in \mathcal{H}} \text{HRCP}(w)$ . Obtaining  $w$  by minimizing Equation (16) under the condition  $w \in \mathcal{H}$  is a quadratic optimization process.

Equation (16) includes a preliminary estimate  $\hat{e}_i$  that must be obtained prior to estimation. Liu and Okui (2013) discussed several ways to obtain  $\hat{e}_i$  in practice. When the models are nested, Liu and Okui (2013) suggested using the residuals from the largest model. When the models are non-nested, they recommended building a model that contains all the regressors in the potential models and taking the corresponding predicted residuals. In addition, a degree-of-freedom correction on  $\hat{e}_i$  is recommended to improve finite-sample properties. For example, when the  $m$ th model is chosen to obtain  $\hat{e}_i$ , we may use

$$\hat{e} = \sqrt{n/(n - k^m)}(I - P^m)y$$

instead of  $(I - P^m)y$  to generate the preliminary estimate  $\hat{e}_i$ .



#### 4. Data Description

Binance was founded in September 2017 and is now the largest crypto exchange around the world. Since the Bitcoin to U.S. dollar (BTC/USD) price data on Binance has only recently become available, we use the data from 1 January 2018 to 20 December 2018 for this exercise. The total number of daily observations is 352. We estimate the daily RV using Equation (1) at the 5-min interval.

The evolution of the RV data over this period is plotted by the solid line in the upper panel of Figure 1, whereas the horizontal axis represents the date and the vertical axis on the left-hand side stands for RV. Besides RV, the price of BTC/USD is also depicted by the dashed line with the vertical axis on the right-hand-side representing the price. We also list the corresponding daily trading volume in the lower panel of Figure 1. As seen in Figure 1, the dynamics of the RV follow the movements of price and volume: the RV increases as the price changes dramatically, which is usually accompanied by a noticeable peak in the trading volume.

Table 1 presents summary statistics for the data and  $p$ -values of both the Jarque–Bera (JB) test for normality and of the Augmented Dickey–Fuller (ADF) tests for unit root. Note that, for the JB and ADF test statistics that are outside tabulated critical values, we report the maximum (0.999) or minimum (0.001)  $p$ -values. In Table 1, we consider the first half, the second half, and full samples in columns 2–4, respectively. Each of the series exhibits tremendous variability and a large range across the respective sample period. Furthermore, none of the series are normally distributed or nonstationary at the 5% level.

**Table 1.** Descriptive statistics of the BTC/USD RV.

Statistics	First Half	Second Half	Full Sample
Mean	32.4200	12.3319	22.3760
Median	21.9565	6.4271	11.8865
Maximum	197.6081	115.6538	197.6081
Minimum	1.8285	0.5241	0.5241
Std. Dev.	33.7164	17.2047	28.5575
Skewness	2.4792	3.2186	2.9249
Kurtosis	10.6082	15.9842	14.3301
Jarque–Bera	0.0010	0.0010	0.0010
ADF Test	0.0010	0.0010	0.0010

Table 1 reports the mean, the sample mean, median, minimum, maximum, standard deviation, skewness, and kurtosis for the realized variance series of the BTC/USD returns. The  $p$ -values of the Jarque–Bera and the Augmented Dickey–Fuller (ADF) tests for RV are recorded in order to test their normality and stationarity, respectively. Note that, for JB and ADF test statistics that are outside tabulated critical values, we report the maximum (0.999) or minimum (0.001)  $p$ -values.

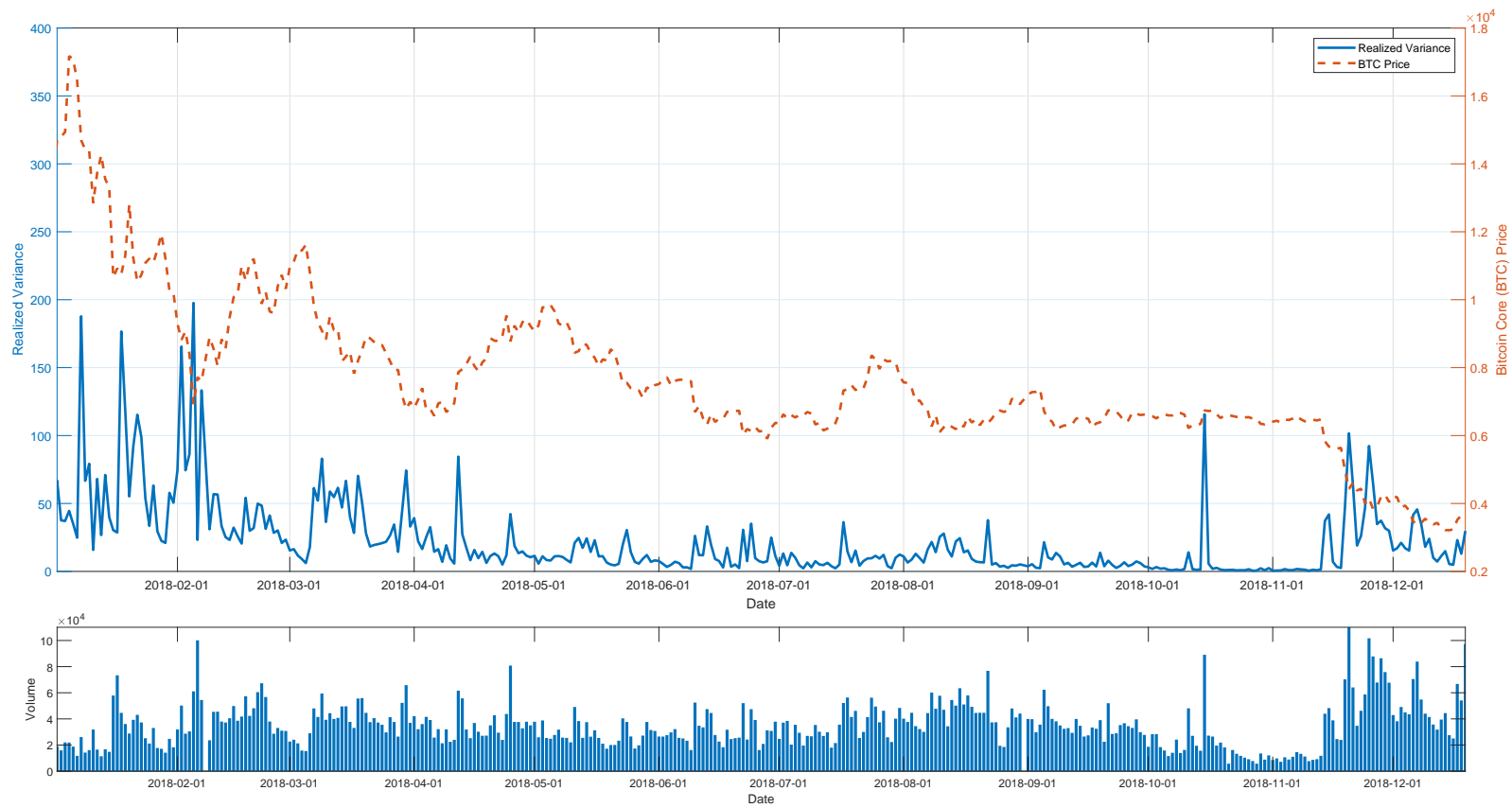


Figure 1. BTC/USD price, realized variance, and volume on Binance.

## 5. The Empirical Exercise

To investigate the relative prediction efficiency of the H-MAHAR estimator and its comparison methods, we conduct an  $h$ -step-ahead rolling window exercise of forecasting the BTC/USD RV for various forecasting horizons.<sup>9</sup> Table 2 lists each estimator considered in the exercise. For all the HAR-type estimators in Panel A, except the HAR-Full model with all the lagged covariates from 1 to 30, we set  $l = [1, 7, 30]$ . For the model-averaging methods in Panel B, our general unrestricted model that includes all covariates is the HAR-Full model which only replaces  $RV_t^{(1)}$ <sup>10</sup> with the semi-variance components from the HAR-RS-I. The candidate model set is first pre-screened by the ARMS method of Yuan and Yang (2005), and we only pick the top 10 models. The tuning parameter in LASSO is estimated through a 5-fold cross-validation.<sup>11</sup> Throughout the experiment, the window length is fixed at 100 observations. We also tried other window lengths and reached similar conclusions. See Section 6.2 for additional details.

**Table 2.** List of heterogeneous autoregressive (HAR)-type estimators.

<i>Panel A: Conventional Regressions</i>		
(1)	AR(1)	a simple autoregressive model
(2)	HAR-Full	the HAR model proposed in Corsi (2009) with $l = [1, 2, \dots, 30]$ , equivalent to a restricted AR(30)
(3)	HAR	the conventional HAR model proposed in Corsi (2009) with $l = [1, 7, 30]$
(4)	HAR-J	the HAR model with jump component proposed in Andersen et al. (2007)
(5)	HAR-CJ	the HAR model with continuous jump component proposed in Andersen et al. (2007)
(6)	HAR-RS-I	the HAR model with semi-variance components (Type I) proposed in Patton and Sheppard (2015)
(7)	HAR-RS-II	the HAR model with semi-variance components (Type II) proposed in Patton and Sheppard (2015)
(8)	HAR-SJ-I	the HAR model with semi-variance and jump components (Type I) proposed in Patton and Sheppard (2015)
(9)	HAR-SJ-II	the HAR model with semi-variance and jump components (Type II) proposed in Patton and Sheppard (2015)
<i>Panel B: Methods Acknowledging Model Uncertainty</i>		
(10)	LASSO	the LASSO HAR method proposed in Audrino and Knaus (2016)
(11)	MAHAR	the model averaging HAR method proposed in Lehrer et al. (2018)
(12)	HRCP	the hetero-robust model-averaging method proposed in Liu and Okui (2013)
(13)	JMA	the jackknife model-averaging method discussed in Zhang et al. (2013)
(14)	H-MAHAR	the hetero-robust model averaging HAR method proposed in Qiu and Xie (2018)

Table 2 lists all the HAR-type estimators included in the empirical exercise. For all the conventional HAR specifications without considering model uncertainty in Panel A, except the HAR-Full model (all the lagged covariates from 1 to 30), we set  $l = [1, 7, 30]$ . To build the candidate models for the model-averaging methods in Panel B, we take a general unrestricted model that includes all covariates in the HAR-Full model and only replace  $RV_t^{(1)}$  by the semi-variance components from HAR-RS-I.

We first consider the case of one-day-ahead forecast ( $h = 1$ ). The results of the prediction experiment are reported in Table 3. The estimation strategies are listed in the first column, and the remaining columns present alternative criteria to evaluate the forecast performance. The criteria include (i) the mean squared forecast error (MSFE), (ii) the mean absolute forecast error (MAFE), (iii) the standard deviation of the forecast error (SDFE), and (iv) the Mincer–Zarnowitz pseudo  $R^2$ .

<sup>9</sup> Additional results using both the GARCH(1, 1) and the ARFIMA( $p, d, q$ ) models are available upon request. These estimators performed poorly relative to the HAR model and thus are not included for space limitation.

<sup>10</sup> The reason we have to exclude  $RV_t^{(1)}$  is because the summation of semi-variance terms equals  $RV_t^{(1)}$ .

<sup>11</sup> We also tried the 10-fold cross-validation and fixed tuning parameter  $\sqrt{\frac{\log n \log(k-1)}{n}}$ . The results remain qualitatively intact.

**Table 3.** Out-of-sample forecast comparison for the BTC/USD RV.

Method	MSFE	MAFE	SDFE	Pseudo R <sup>2</sup>
<i>Panel A: Conventional Regressions</i>				
AR(1)	239.1504	10.0717	15.4645	0.4106
HAR-Full	302.3662	10.8925	17.3887	0.2548
HAR	204.6532	8.3302	14.3057	0.4956
HAR-J	208.7348	8.5570	14.4477	0.4856
HAR-CJ	215.9540	8.3766	14.6954	0.4678
HAR-RS-I	193.2083	8.1705	13.8999	0.5238
HAR-RS-II	197.3354	8.2618	14.0476	0.5137
HAR-SJ-I	193.7362	8.2167	13.9189	0.5225
HAR-SJ-II	201.1249	8.3640	14.1819	0.5043
<i>Panel B: Method Acknowledging Model Uncertainty</i>				
LASSO	247.8799	8.2628	15.7442	0.3891
MAHAR	191.9673	7.1735	13.8552	0.5269
HRCP	196.8785	7.3539	14.0313	0.5148
JMA	191.9862	7.1772	13.8559	0.5269
H-MAHAR	<b>191.3624</b>	<b>7.1621</b>	<b>13.8334</b>	<b>0.5284</b>

Table 3 compares the out-of-sample performance of the H-MAHAR estimator relative to its comparison methods. The sample period for the Bitcoin RV spans from 1 January 2018 to 20 December 2018 (a total of 352 observations). We use a rolling window of 100 observations to estimate the coefficients of all the models and evaluate the out-of-sample forecast performance at  $h = 1$ . Bold numbers indicate the best performing model by each criterion.

To ease interpretation, the results that identify the estimator with the best performance in each column of Table 3 is marked in bold. The performance of autoregressive models, represented by the AR(1) and HAR-Full models, is weak. For each panel, the HAR-type methods demonstrate noticeably improved performances relative to the autoregressive models. In the case of Bitcoin volatility, there is not much gain from including the jump and/or semi-variance components in the standard HAR model. The above set of results suggests that the heterogeneity in modeling Bitcoin volatility cannot be fully accommodated by simply adding extra covariates to the linear model. The least squares model-averaging methods that acknowledge model uncertainty show superior forecasting accuracy under all the evaluation criteria. Among the averaging methods, H-MAHAR displays the best performance. On the other hand, the model-selection-based LASSO method has the worst performance in this situation.

To examine if the improvement from the least squares model-averaging methods is statistically significant, we perform the modified Giacomini–White (GW) test (Giacomini and White 2006)<sup>12</sup> of the null hypothesis that the *column method* performs equally as well as the *row method* in terms of MAFE. The corresponding  $p$ -values are presented in Table 4 for  $h = 1$ . We see that the gains in forecast accuracy from the model-averaging methods relative to other strategies are statistically significant at the 5% level.

By exploring weight estimates of the H-MAHAR estimator on the full dataset, we can shed light on both the relative importance of the candidate models and the inclusion of various HAR-type lagged components. The models that are assigned the five highest weights by the H-MAHAR estimator are described in Table 5 (presented in the 2nd row of Table 5 in a descending fashion). The “x” sign indicates that the corresponding covariate (listed in the first column) is contained in the model. Certain variables, like  $RV_t^-$  and  $RV_t^{(30)}$ , are included in every model, but variables like  $RV_t^+$  or  $RV_t^{(10)}$  are excluded from each of the top five dominant models.

<sup>12</sup> Giacomini and White (2006) proposed a framework for out-of-sample predictive ability testing and forecast selection designed for use in the realistic situation in which the forecasting model is possibly misspecified due to unmodeled dynamics, unmodeled heterogeneity, incorrect functional form, or any combination of these. The null hypothesis of the GW test is that the two models we want to compare are equally accurate on average based on certain criterion.

**Table 4.** Results of the Giacomini–White test for  $h = 1$ .

Method	AR(1)	Full	HAR	J	CJ	RS-I	RS-II	SJ-I	SJ-II	LASSO	MAHAR	HRCP	JMA
<i>Panel A: Conventional Regressions</i>													
AR(1)	-	-	-	-	-	-	-	-	-	-	-	-	-
HAR-Full	0.1617	-	-	-	-	-	-	-	-	-	-	-	-
HAR	0.0000	0.0000	-	-	-	-	-	-	-	-	-	-	-
HAR-J	0.0000	0.0000	0.0888	-	-	-	-	-	-	-	-	-	-
HAR-CJ	0.0000	0.0000	0.8449	0.4070	-	-	-	-	-	-	-	-	-
HAR-RS-I	0.0000	0.0000	0.4376	0.1074	0.4895	-	-	-	-	-	-	-	-
HAR-RS-II	0.0000	0.0000	0.7418	0.2239	0.7047	0.1137	-	-	-	-	-	-	-
HAR-SJ-I	0.0000	0.0000	0.5839	0.1283	0.5811	0.3202	0.5546	-	-	-	-	-	-
HAR-SJ-II	0.0000	0.0000	0.8739	0.4253	0.9667	0.0969	0.4063	0.1413	-	-	-	-	-
<i>Panel B: Methods Acknowledging Model Uncertainty</i>													
LASSO	0.0009	0.0000	0.8632	0.4805	0.7957	0.8245	0.9980	0.9113	0.8042	-	-	-	-
MAHAR	0.0000	0.0000	0.0001	0.0000	0.0007	0.0003	0.0001	0.0002	0.0001	0.0029	-	-	-
HRCP	0.0000	0.0000	0.0008	0.0001	0.0037	0.0035	0.0018	0.0023	0.0012	0.0150	0.0255	-	-
JMA	0.0000	0.0000	0.0001	0.0000	0.0007	0.0003	0.0001	0.0002	0.0001	0.0030	0.5774	0.0251	-
H-MAHAR	0.0000	0.0000	0.0000	0.0000	0.0006	0.0002	0.0001	0.0001	0.0001	0.0024	0.3481	0.0227	0.3119

The modified Giacomini–White test (Giacomini and White 2006) is implemented to test the null hypothesis that the *row method* (in vertical headings) performs equally as well as the *column method* (in horizontal headings) in terms of the absolute forecast error. Corresponding  $p$ -values for each method are reported in Panels A to B of Table 4.

**Table 5.** Top 5 models from the heteroskedasticity-robust model averaging HAR (H-MAHAR) estimator.

	Model 1	Model 2	Model 3	Model 4	Model 5
Weight	0.3441	0.3355	0.2546	0.0488	0.0170
<i>Panel A: HAR-RS Components</i>					
$RV_t^+$					
$RV_t^-$	x	x	x	x	x
<i>Panel B: Selected HAR Covariates</i>					
$RV_t^{(15)}$	x		x	x	
$RV_t^{(16)}$	x		x	x	
$RV_t^{(18)}$		x	x		
$RV_t^{(22)}$	x				x
$RV_t^{(23)}$	x				x
$RV_t^{(28)}$	x		x	x	x
$RV_t^{(29)}$	x	x	x	x	x
$RV_t^{(30)}$	x	x	x	x	x

Table 5 describes the models that are assigned the five highest weights by the H-MAHAR estimator. Note that x denotes that the explanatory variable is included in the specific model.

Throughout our analysis, we find that the incorporation of negative semi-variances improves the prediction accuracy and explains a large fraction of the variation in RV, which is consistent with the finding of the literature (Patton and Sheppard 2015). The H-MAHAR method places large weights on models with HAR components of lag indices greater than 15, which may be in part due to the strong short-term performance of the  $RV_t^-$  variable. We also observe that HAR components with high lag indices (for example,  $RV_t^{(29)}$  and  $RV_t^{(30)}$ ) mimicking the long-term dynamics of RV are intensively picked by the model averaging process. Most importantly, none of the top 5 models has the conventional lag index specification of [1, 7, 30]. The above exercise uncovers the sheer existence of model uncertainty for Bitcoin volatility and accredits the use of model-averaging methods.

## 6. Robustness Check

In this section, we perform three robustness checks on our results in Section 5. We first extend the exercises to relatively longer forecast horizons. Specifically, we consider  $h = 2, 3$ , and 4. In the second robustness check, we consider alternative window lengths. In the last robustness check, the H-MAHAR method is compared with Model 1 from Table 5, the one with the highest model weight among all candidate models.

### 6.1. Various Forecast Horizons

Table 6 represents the forecast performance of the considered estimators for  $h = 2, 3$ , and 4 periods ahead.<sup>13</sup> Table 7 examines the statistical significance of the forecasting accuracy improvement. For all  $h$  periods, the forecasts by least squares model averaging estimators dominate those by other methods in general. Among all the model averaging estimators, the HRCP method is seen to perform the best in most times according to the criteria we used, although such improvement is not statistically significant according to the results in Table 7.

<sup>13</sup> Note that the forecasting horizons we considered in this paper are all short. The HAR-type models which our model-averaging methods build upon do not perform well in the long forecasting horizons. One possible explanation is that the Bitcoin market is relatively small compared to conventional stock markets; therefore, it is more sensitive to various policy shocks, information impact, and even social media sentiment changes. Most of these shocks are short-lived, and it seems that the momentum effect does not last long in Bitcoin realized volatility. How to model Bitcoin volatility in a long forecasting horizon is beyond the scope of this paper and guarantees future research.

Table 6. Forecast performance comparison for various horizons.

Method	MSFE	MAFE	SDFE	Pseudo $R^2$	MSFE	MAFE	SDFE	Pseudo $R^2$	MSFE	MAFE	SDFE	Pseudo $R^2$
	$h = 2$				$h = 3$				$h = 4$			
<i>Panel A: Conventional Regressions</i>												
AR(1)	262.9289	10.3286	16.2151	0.3062	277.6988	10.6643	16.6643	0.2643	283.2303	10.9911	16.8294	0.2502
HAR-Full	334.5891	11.6691	18.2918	0.1171	346.2303	12.1207	18.6073	0.0828	346.7523	12.2453	18.6213	0.0821
HAR	224.2440	8.8998	14.9748	0.4083	234.6886	9.3141	15.3196	0.3783	241.8037	9.2817	15.5500	0.3599
HAR-J	226.2970	8.9210	15.0432	0.4028	235.0144	9.2912	15.3302	0.3774	245.8197	9.5108	15.6786	0.3493
HAR-CJ	221.2180	8.8997	14.8734	0.4162	223.1287	9.1104	14.9375	0.4089	244.7930	9.6451	15.6459	0.3520
HAR-RS-I	226.5678	8.9060	15.0522	0.4021	258.0224	9.5877	16.0631	0.3164	230.6331	9.0845	15.1866	0.3895
HAR-RS-II	231.1408	9.0014	15.2033	0.3901	262.3614	9.7318	16.1976	0.3050	242.0860	9.3867	15.5591	0.3591
HAR-SJ-I	228.7837	8.9241	15.1256	0.3963	260.7971	9.6530	16.1492	0.3091	232.6273	9.1173	15.2521	0.3842
HAR-SJ-II	233.0146	9.1070	15.2648	0.3851	290.9509	9.6781	17.0573	0.2292	239.4566	9.3162	15.4744	0.3661
<i>Panel B: Methods Acknowledging Model Uncertainty</i>												
LASSO	265.2809	8.7407	16.2874	0.3000	270.9054	9.0213	16.4592	0.2823	270.9619	9.2020	16.4609	0.2827
MAHAR	<b>216.7814</b>	8.4566	<b>14.7235</b>	<b>0.4280</b>	228.6793	8.4375	15.1221	0.3942	225.0127	8.1343	15.0004	0.4043
HRCP	217.2918	<b>8.4100</b>	14.7408	0.4266	<b>228.5923</b>	<b>8.3638</b>	<b>15.1193</b>	<b>0.3944</b>	<b>220.1249</b>	<b>7.9448</b>	<b>14.8366</b>	<b>0.4173</b>
JMA	216.9337	8.4577	14.7287	0.4276	228.7295	8.4372	15.1238	0.3940	227.8901	8.1982	15.0960	0.3967
H-MAHAR	216.9262	8.4746	14.7284	0.4276	228.7536	8.4606	15.1246	0.3940	223.9671	8.1245	14.9655	0.4071

Table 6 compares the out-of-sample performance of the H-MAHAR estimator relative to its comparison methods. The sample period for the Bitcoin RV spans from 1 January 2018 to 20 December 2018 (a total of 352 observations). We use a rolling window of 100 observations to estimate the coefficients of all the models and evaluate the out-of-sample forecast performance at  $h = 2, 3$ , and 4. The results for each  $h$  are reported in the left, middle, and right blocks, respectively. Bold numbers indicate the best performing model by each criterion.

**Table 7.** Results of the Giacomini–White test for various forecast horizons.

Method	AR(1)	Full	HAR	J	CJ	RS-I	RS-II	SJ-I	SJ-II	LASSO	MAHAR	HRCP	JMA
<i>Panel A: h = 2</i>													
AR(1)	-	-	-	-	-	-	-	-	-	-	-	-	-
HAR-Full	0.0688	-	-	-	-	-	-	-	-	-	-	-	-
HAR	0.0008	0.0000	-	-	-	-	-	-	-	-	-	-	-
HAR-J	0.0013	0.0000	0.7804	-	-	-	-	-	-	-	-	-	-
HAR-CJ	0.0065	0.0000	0.9999	0.9478	-	-	-	-	-	-	-	-	-
HAR-RS-I	0.0011	0.0000	0.9548	0.9016	0.9848	-	-	-	-	-	-	-	-
HAR-RS-II	0.0033	0.0000	0.4109	0.5236	0.7577	0.1555	-	-	-	-	-	-	-
HAR-SJ-I	0.0011	0.0000	0.8446	0.9817	0.9418	0.6648	0.3559	-	-	-	-	-	-
HAR-SJ-II	0.0048	0.0000	0.2016	0.2599	0.5321	0.0556	0.4204	0.0425	-	-	-	-	-
LASSO	0.0149	0.0004	0.7522	0.7190	0.7794	0.7414	0.6079	0.7187	0.4705	-	-	-	-
MAHAR	0.0005	0.0000	0.2397	0.2196	0.2723	0.2467	0.1613	0.2397	0.1070	0.5903	-	-	-
HRCP	0.0003	0.0000	0.1952	0.1811	0.2413	0.2040	0.1304	0.1987	0.0868	0.5285	0.6363	-	-
JMA	0.0005	0.0000	0.2430	0.2231	0.2768	0.2500	0.1640	0.2428	0.1090	0.5930	0.9213	0.6168	-
H-MAHAR	0.0006	0.0000	0.2616	0.2404	0.2964	0.2685	0.1780	0.2606	0.1188	0.6154	0.3316	0.4998	0.2370
<i>Panel B: h = 3</i>													
AR(1)	-	-	-	-	-	-	-	-	-	-	-	-	-
HAR-Full	0.0974	-	-	-	-	-	-	-	-	-	-	-	-
HAR	0.0121	0.0000	-	-	-	-	-	-	-	-	-	-	-
HAR-J	0.0149	0.0000	0.8366	-	-	-	-	-	-	-	-	-	-
HAR-CJ	0.0281	0.0001	0.6213	0.6402	-	-	-	-	-	-	-	-	-
HAR-RS-I	0.0477	0.0001	0.2135	0.3144	0.3559	-	-	-	-	-	-	-	-
HAR-RS-II	0.0979	0.0002	0.0511	0.1174	0.2237	0.1753	-	-	-	-	-	-	-
HAR-SJ-I	0.0580	0.0002	0.1491	0.2454	0.3068	0.0937	0.4849	-	-	-	-	-	-
HAR-SJ-II	0.0933	0.0005	0.2751	0.3349	0.3545	0.5870	0.7948	0.8733	-	-	-	-	-
LASSO	0.0410	0.0014	0.6278	0.6430	0.9034	0.3877	0.2914	0.3399	0.3513	-	-	-	-
MAHAR	0.0011	0.0000	0.0393	0.0431	0.1216	0.0218	0.0102	0.0174	0.0327	0.3106	-	-	-
HRCP	0.0007	0.0000	0.0236	0.0257	0.0872	0.0136	0.0059	0.0108	0.0230	0.2573	0.4987	-	-
JMA	0.0011	0.0000	0.0402	0.0443	0.1260	0.0223	0.0105	0.0178	0.0333	0.3115	0.9892	0.4993	-
H-MAHAR	0.0012	0.0000	0.0471	0.0515	0.1411	0.0258	0.0124	0.0206	0.0372	0.3333	0.4041	0.3999	0.2248
<i>Panel C: h = 4</i>													
AR(1)	-	-	-	-	-	-	-	-	-	-	-	-	-
HAR-Full	0.1863	-	-	-	-	-	-	-	-	-	-	-	-
HAR	0.0063	0.0000	-	-	-	-	-	-	-	-	-	-	-
HAR-J	0.0150	0.0000	0.1822	-	-	-	-	-	-	-	-	-	-
HAR-CJ	0.0272	0.0006	0.3321	0.6941	-	-	-	-	-	-	-	-	-
HAR-RS-I	0.0022	0.0000	0.2372	0.0900	0.1312	-	-	-	-	-	-	-	-
HAR-RS-II	0.0109	0.0001	0.6134	0.6596	0.5091	0.0238	-	-	-	-	-	-	-
HAR-SJ-I	0.0021	0.0000	0.2823	0.0800	0.1454	0.4460	0.0413	-	-	-	-	-	-
HAR-SJ-II	0.0038	0.0000	0.8608	0.4391	0.3665	0.0606	0.5947	0.0570	-	-	-	-	-
LASSO	0.0723	0.0047	0.9119	0.6823	0.5731	0.8570	0.7716	0.8976	0.8607	-	-	-	-
MAHAR	0.0003	0.0000	0.0242	0.0067	0.0027	0.0370	0.0057	0.0317	0.0122	0.0649	-	-	-
HRCP	0.0004	0.0000	0.0118	0.0027	0.0009	0.0172	0.0046	0.0149	0.0061	0.0484	0.5237	-	-
JMA	0.0004	0.0000	0.0385	0.0124	0.0051	0.0593	0.0100	0.0516	0.0207	0.0866	0.1211	0.4208	-
H-MAHAR	0.0002	0.0000	0.0200	0.0049	0.0025	0.0331	0.0055	0.0281	0.0112	0.0673	0.8426	0.5390	0.3280

The modified Giacomini–White test (Giacomini and White 2006) is implemented to test the null hypothesis that the *row method* (in vertical headings) performs equally as well as the *column method* (in horizontal headings) in terms of the absolute forecast error. Corresponding *p*-values for each *h* are reported in Panels A to C of Table 7.



## 6.2. Alternative Window Lengths

In the main exercises, we set the window length at  $L = 100$ . In this section, we also tried other window lengths such as  $L = 50$  and  $200$ . We present the estimation results for  $h = 1$ . Although not reported here, we also tried other forecast horizons and the robustness remains intact.

Table 8 shows the forecast performance of all the methods for various window lengths. In all the cases, the H-MAHAR estimator yields the smallest MSFE, MAFE, and SDFE and the largest Pseudo  $R^2$ . We examine the statistical significance of the forecast accuracy improvement in Table 9. The small  $p$ -values on the H-MAHAR method against other methods, especially that with no model averaging estimators, indicate that the improvement is significant at the 5% level in most cases.

**Table 8.** Forecast performance comparison under different window lengths.

Method	MSFE	MAFE	SDFE	Pseudo $R^2$	MSFE	MAFE	SDFE	Pseudo $R^2$
	$L = 50$				$L = 150$			
<i>Panel A: Conventional Regression</i>								
AR(1)	249.6477	10.0160	15.8002	0.5658	238.3546	9.9032	15.4387	0.4277
HAR-Full	1637.2399	22.4248	40.4628	-1.8473	243.7945	9.6486	15.6139	0.4147
HAR	276.8757	10.5585	16.6396	0.5185	212.9744	8.2480	14.5936	0.4887
HAR-J	293.4600	10.7562	17.1307	0.4896	210.7548	8.2194	14.5174	0.4940
HAR-CJ	374.7990	11.8093	19.3597	0.3482	208.9187	8.2310	14.4540	0.4984
HAR-RS-I	283.6314	10.6281	16.8414	0.5067	201.2067	8.1321	14.1847	0.5169
HAR-RS-II	295.0338	10.7334	17.1765	0.4869	204.8401	8.1870	14.3122	0.5082
HAR-SJ-I	284.4146	10.6127	16.8646	0.5054	200.6347	8.1156	14.1646	0.5183
HAR-SJ-II	299.5199	10.9971	17.3066	0.4791	206.0011	8.2656	14.3527	0.5054
<i>Panel B: Method Acknowledges Model Uncertainty</i>								
LASSO	806.0115	15.7427	28.3903	-0.4017	252.3836	7.7437	15.8866	0.3940
MAHAR	255.5945	9.3474	15.9873	0.5555	200.1770	7.3587	14.1484	0.5194
HRCP	298.5371	10.1002	17.2782	0.4808	203.2551	7.4930	14.2568	0.5120
JMA	257.5357	9.4021	16.0479	0.5521	200.4459	7.3838	14.1579	0.5187
H-MAHAR	<b>252.6259</b>	<b>9.2245</b>	<b>15.8942</b>	<b>0.5607</b>	<b>199.8410</b>	<b>7.3540</b>	<b>14.1365</b>	<b>0.5202</b>

Table 8 compares the out-of-sample performance of the H-MAHAR estimator relative to its comparison methods. The sample period for the Bitcoin  $RV$  spans from 1 January 2018 to 20 December 2018 (a total of 352 observations). We consider alternative rolling window lengths of 50 and 150 observations to estimate the coefficients of all the models and evaluate the out-of-sample forecast performance at  $h = 1$ . The results for each  $L$  are reported in the left and right blocks, respectively. Bold numbers indicate the best performing model by each criterion.

**Table 9.** Results of the Giacomini–White test for different window lengths.

Method	AR(1)	Full	HAR	J	CJ	RS-I	RS-II	SJ-I	SJ-II	LASSO	MAHAR	HRCP	JMA
<i>Panel A: L = 50</i>													
AR(1)	-	-	-	-	-	-	-	-	-	-	-	-	-
HAR-Full	0.0000	-	-	-	-	-	-	-	-	-	-	-	-
HAR	0.1474	0.0000	-	-	-	-	-	-	-	-	-	-	-
HAR-J	0.0681	0.0000	0.4518	-	-	-	-	-	-	-	-	-	-
HAR-CJ	0.0034	0.0000	0.0321	0.0603	-	-	-	-	-	-	-	-	-
HAR-RS-I	0.1389	0.0000	0.8267	0.7392	0.0471	-	-	-	-	-	-	-	-
HAR-RS-II	0.0890	0.0000	0.6123	0.9521	0.0642	0.5173	-	-	-	-	-	-	-
HAR-SJ-I	0.1486	0.0000	0.8691	0.7025	0.0448	0.7674	0.4758	-	-	-	-	-	-
HAR-SJ-II	0.0244	0.0000	0.1869	0.4996	0.1604	0.1067	0.3281	0.1017	-	-	-	-	-
LASSO	0.0000	0.0015	0.0000	0.0000	0.0009	0.0000	0.0000	0.0000	0.0000	-	-	-	-
MAHAR	0.1109	0.0000	0.0062	0.0016	0.0001	0.0043	0.0022	0.0051	0.0003	0.0000	-	-	-
HRCP	0.8760	0.0000	0.3767	0.2145	0.0160	0.3150	0.2315	0.3332	0.0954	0.0000	0.0045	-	-
JMA	0.1450	0.0000	0.0089	0.0024	0.0002	0.0062	0.0032	0.0073	0.0005	0.0000	0.0783	0.0056	-
H-MAHAR	0.0552	0.0000	0.0024	0.0005	0.0001	0.0015	0.0007	0.0019	0.0001	0.0000	0.1164	0.0059	0.0532
<i>Panel B: L = 150</i>													
AR(1)	-	-	-	-	-	-	-	-	-	-	-	-	-
HAR-Full	0.6126	-	-	-	-	-	-	-	-	-	-	-	-
HAR	0.0000	0.0003	-	-	-	-	-	-	-	-	-	-	-
HAR-J	0.0000	0.0003	0.7912	-	-	-	-	-	-	-	-	-	-
HAR-CJ	0.0002	0.0013	0.9390	0.9513	-	-	-	-	-	-	-	-	-
HAR-RS-I	0.0001	0.0004	0.5688	0.6630	0.7045	-	-	-	-	-	-	-	-
HAR-RS-II	0.0001	0.0008	0.7840	0.8864	0.8770	0.4173	-	-	-	-	-	-	-
HAR-SJ-I	0.0001	0.0004	0.5361	0.5855	0.6460	0.7331	0.4473	-	-	-	-	-	-
HAR-SJ-II	0.0001	0.0012	0.9318	0.8167	0.8949	0.1200	0.4995	0.0689	-	-	-	-	-
LASSO	0.0000	0.0005	0.1911	0.2364	0.3101	0.3522	0.2891	0.3814	0.2050	-	-	-	-
MAHAR	0.0000	0.0000	0.0014	0.0015	0.0028	0.0033	0.0024	0.0043	0.0007	0.3147	-	-	-
HRCP	0.0000	0.0000	0.0068	0.0079	0.0114	0.0158	0.0118	0.0196	0.0038	0.5165	0.0173	-	-
JMA	0.0000	0.0000	0.0018	0.0021	0.0037	0.0045	0.0033	0.0057	0.0009	0.3481	0.0001	0.0463	-
H-MAHAR	0.0000	0.0000	0.0013	0.0015	0.0027	0.0031	0.0023	0.0040	0.0006	0.3081	0.6172	0.0204	0.0166

The modified Giacomini–White test (Giacomini and White 2006) is implemented to test the null hypothesis that the *row method* (in vertical headings) performs equally as well as the *column method* (in horizontal headings) in terms of the absolute forecast error. Corresponding *p*-values for each *L* are reported in Panels A to B of Table 9, respectively.

## 7. Conclusions

In this paper, we study the forecast performance of least squares model-averaging methods when predicting Bitcoin volatility. Our method allows for a more general lag structure under the HAR framework, instead of restricting it to daily, weekly, and monthly frequencies. Specially, we estimate the semi-variance HAR models in Patton and Sheppard (2015) with the least squares model-averaging method and consider constructing the potential model set with a full permutation of all of the possible lags and the maximum lag order of 30. The H-MAHAR-embedded model is data-driven, as the empirical weights on potential models with different lag combinations vary with underlying volatility series and forecast horizons.

In the out-of-sample application to high-frequency data of the realized variance of BTC/USD, we provide suggestive evidence that there exists excessive model uncertainty when modeling the Bitcoin volatility by conventional regression methods. We further demonstrate that the model-averaging methods can generally outperform conventional regression methods under various forecast criteria as well as across all forecast horizons ( $h = 1, 2, 3, 4$ ). Specifically, we apply the GW test to examine the statistical significance of the improvement made by the model-averaging method. We reveal that the model-averaging method, especially the one robust to heteroskedasticity (the H-MAHAR), performs significantly better than conventional regressions at a 5% confidence level. Therefore, the least squares model-averaging methods adapt themselves remarkably well to a relatively short sample with evident model uncertainty.

This research also shed some light on future works related to the emerging asset class such as the cryptocurrency. When a new asset class is introduced, proper asset valuation theory is always invented with lags and institutional investors will hesitate to enter the market for risk control purposes. Regulations and technology developments are also likely to keep the market structure susceptible to shocks and to cause great price variations. Moreover, the lack of trading data of long durations is particularly a concern compared with other well-established asset classes. In this situation, model averaging contributes to alleviating model specification uncertainty and even to controlling for heteroskedasticity. There are still some interesting questions left to further research, for instance, the deep relationship between the crypto trading environment (i.e., the impact of sentiment) and volatility data structure.

**Funding:** This research was funded by the National Natural Science Foundation of China grant number 71701175 and the Humanities and Social Science Fund of Ministry of Education of China grant number 17YJC790174.

**Acknowledgments:** I wish to thank Yue Qiu, Guanxi Yi, and Jun Yu, seminar participants at the SoFiE 2019 Conference in Shanghai from Xiamen University, Shanghai University of Finance and Economics, and Singapore Management University, respectively, for their helpful comments and suggestions. The usual caveat applies.

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

## References

- Andersen, Torben G., and Tim Bollerslev. 1998. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review* 39: 885–905. [\[CrossRef\]](#)
- Andersen, Torben G., Tim Bollerslev, and Francis X. Diebold. 2007. Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. *The Review of Economics and Statistics* 89: 701–20. [\[CrossRef\]](#)
- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens. 2001a. The distribution of realized stock return volatility. *Journal of Financial Economics* 61: 43–76. [\[CrossRef\]](#)
- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys. 2001b. The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association* 96: 42–55. [\[CrossRef\]](#)
- Audrino, Francesco, and Simon D. Knaus. 2016. Lassoing the HAR Model: A Model Selection Perspective on Realized Volatility Dynamics. *Econometric Reviews* 35: 1485–521. [\[CrossRef\]](#)
- Audrino, Francesco, Huang Chen, and Okhrin Ostap. 2019. Flexible HAR Model for Realized Volatility. *Studies in Nonlinear Dynamics & Econometrics* 23: 1–22.

- Audrino, Francesco, Lorenzo Camponovo, and Constantin Roth. 2015. *Testing the lag Structure of Assets' Realized Volatility Dynamics*. Economics Working Paper Series 1501. St. Gallen: University of St. Gallen, School of Economics and Political Science.
- Barnard, George A. 1963. New Methods of Quality Control. *Journal of the Royal Statistical Society. Series A (General)* 126: 255–58. [\[CrossRef\]](#)
- Breiman, Leo. 2001. Random Forests. *Machine Learning* 45: 5–32. [\[CrossRef\]](#)
- Buckland, Steven T., Kenneth P. Burnham, and Nicole H. Augustin. 1997. Model Selection: An Integral Part of Inference. *Biometrics* 53: 603–18. [\[CrossRef\]](#)
- Corsi, Fulvio, Francesco Audrino, and Roberto Renò. 2012. HAR Modeling for Realized Volatility Forecasting. In *Handbook of Volatility Models and Their Applications*. Hoboken: John Wiley & Sons, Inc., pp. 363–82.
- Corsi, Fulvio, Stefan Mittnik, Christian Pigorsch, and Uta Pigorsch. 2008. The Volatility of Realized Volatility. *Econometric Reviews* 27: 46–78. [\[CrossRef\]](#)
- Corsi, Fulvio. 2009. A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7: 174–96. [\[CrossRef\]](#)
- Craioveanu, Mihaela, and Eric Hillebrand. 2012. *Why It Is OK to Use the HAR-RV (1,5,21) Model*. Technical Report. Missouri: University of Central Missouri.
- Dacorogna, Michael M., Ulrich A. Müller, Robert J. Nagler, Richard B. Olsen, and Olivier V. Pictet. 1993. A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance* 12: 413–38. [\[CrossRef\]](#)
- Giacomini, Raffaella, and Halbert White. 2006. Tests of Conditional Predictive Ability. *Econometrica* 74: 1545–78. [\[CrossRef\]](#)
- Hansen, Bruce E. 2007. Least Squares Model Averaging. *Econometrica* 75: 1175–89. [\[CrossRef\]](#)
- Hansen, Bruce E. 2008. Least-squares forecast averaging. *Journal of Econometrics* 146: 342–50. [\[CrossRef\]](#)
- Hansen, Bruce E. 2009. Averaging Estimators for Regressions with A Possible Structural Break. *Econometric Theory* 25: 1498–514. [\[CrossRef\]](#)
- Hansen, Bruce E. 2010. Averaging Estimators for Autoregressions with A Near Unit Root. *Journal of Econometrics* 158: 142–55. [\[CrossRef\]](#)
- Hansen, Bruce E., and Jeffrey S. Racine. 2012. Jackknife model averaging. *Journal of Econometrics* 167: 38–46. [\[CrossRef\]](#)
- Hoeting, Jennifer A., David Madigan, Adrian E. Raftery, and Chris T. Volinsky. 1999. Bayesian Model Averaging: A Tutorial. *Statistical Science* 14: 382–401.
- Huang, Xin, and George Tauchen. 2005. The Relative Contribution of Jumps to Total Price Variance. *Journal of Financial Econometrics* 3: 456–99. [\[CrossRef\]](#)
- Lehrer, Steven F., Tian Xie, and Xinyu Zhang. 2018. Wits versus Tweets: Does Adding Social Media Wisdom Trump Admitting Ignorance when Forecasting the CBOE VIX? Working Paper A0167, The City University of Hong Kong, Hong Kong, China.
- Liu, Qingfeng, and Ryo Okui. 2013. Heteroskedasticity-robust  $C_p$  Model Averaging. *The Econometrics Journal* 16: 463–72. [\[CrossRef\]](#)
- Magnus, Jan R., Owen Powell, and Patricia Prüfer. 2010. A comparison of two model averaging techniques with an application to growth empirics. *Journal of Econometrics* 154: 139–53. [\[CrossRef\]](#)
- Müller, Ulrich A., Michel M. Dacorogna, Rakhal D. Davé, Olivier V. Pictet, Richard B. Olsen, and J. Robert Ward. 1993. *Fractals and Intrinsic Time—A Challenge to Econometricians*. Technical Report. Zürich: Olsen & Associates.
- Patton, Andrew J., and Kevin Sheppard. 2015. Good Volatility, Bad Volatility: Signed Jumps and The Persistence of Volatility. *The Review of Economics and Statistics* 97: 683–97. [\[CrossRef\]](#)
- Qiu, Yue, and Tian Xie. 2018. Forecasting Foreign Exchange Realized Volatility: A Least Squares Model Averaging Approach. *Journal of Systems Science and Mathematical Sciences* 38: 725–44.
- Qiu, Yue, Xinyu Zhang, Tian Xie, and Shangwei Zhao. 2019. Versatile HAR model for realized volatility: A least square model averaging perspective. *Journal of Management Science and Engineering* 4: 55–73. [\[CrossRef\]](#)
- Wan, Alan TK, Xinyu Zhang, and Guohua Zou. 2010. Least Squares Model Averaging by Mallows Criterion. *Journal of Econometrics* 156: 277–83. [\[CrossRef\]](#)
- Xie, Tian. 2015. Prediction Model Averaging Estimator. *Economics Letters* 131: 5–8. [\[CrossRef\]](#)
- Xie, Tian. 2017. Heteroscedasticity-robust Model Screening: A Useful Toolkit for Model Averaging in Big Data Analytics. *Economics Letters* 151: 119–22. [\[CrossRef\]](#)

- Yuan, Zheng, and Yuhong Yang. 2005. Combining Linear Regression Models: When and How? *Journal of the American Statistical Association* 100: 1202–14. [[CrossRef](#)]
- Zhang, Xinyu, Alan TK Wan, and Guohua Zou. 2013. Model averaging by jackknife criterion in models with dependent data. *Journal of Econometrics* 174: 82–94. [[CrossRef](#)]
- Zhao, Shangwei, Xinyu Zhang, and Yichen Gao. 2016. Model Averaging with Averaging Covariance Matrix. *Economics Letters* 145: 214–17. [[CrossRef](#)]



© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).