Duration Dependent Volatility Models with Value-weighted Approach

Fernando H P. S. Mendes[a,*], Douglas Eduardo Turatti[b], João H. G. Mazzeu[c,**], João F. Caldeira[d]

[a]*Federal University of Rio Grande do Sul, Department of Economics - Brazil*
[b]*Aalborg University, Aalborg University Business School - Denmark*
[c]*University of Campinas, Department of Statistics - Brazil*
[d]*Federal University of Santa Catarina, Department of Economics - Brazil*

Abstract

This paper attempts to revisit the duration-dependent Markov-switching model of Maheu & McCurdy (2000b) through a model combination approach. The existing literature did not fully explore the issues of using a single duration value to characterize the conditional variance returns. Hence, the capabilities of the model to predict conditional volatility have not been well investigated from an out-of-sample forecasting context. We highlight the foundations of adopting different duration values to conduct a one-step-ahead weighing volatility forecast. To evaluate our empirical study's contribution, we conducted a statistical and risk prediction evaluation for the daily bitcoin returns. In general, our results outperform GARCH-type models using different volatility proxies and robust loss functions, both for point volatility forecasting and Value-at-Risk analysis.

Keywords: Duration Dependence, Markov-switching model, Model Combination, Volatility Forecasting, Value-at-Risk.
JEL: C5, C22, G1.

---

*Corresponding author: fernandohpsm@hotmail.com
**The third author acknowledges financial support from Coordination for the Improvement of Higher Level Personnel (CAPES), Brazil grant 88882.305837/2018-01.

1. Introduction

Since the seminal paper of Hamilton (1989), many authors have proposed variations of the Markov-switching model. In this regard, Durland and McCurdy (1994) introduced the duration-dependent Markov-switching (DDMS) model, based on a higher-order Markov chain that allowed state transition probabilities to be duration dependent. Initially applied to investigate business cycle asymmetries, this modeling approach includes bull and bear market stock identification (Maheu and McCurdy, 2000a) and FX volatility estimation (Maheu and McCurdy, 2000b), since duration can be included as a conditioning variable in conditional mean and variance. The empirical literature has also explored multivariate generalization (Pelagatti, 2007), leading indicator inclusion (Layton and Smith, 2007), and, more recently, multi-state expansions (Bejaoui and Karaa, 2016) for the baseline DDMS specification. However, these applications were primarily focusing on the in-sample analysis, and none of the previous studies have conducted detailed point forecasting exercises, especially for the conditional volatility process. This paper thus aims to investigate the out-of-sample properties of the DDMS model to forecast the conditional volatility while allowing for uncertainty in its duration process.

More precisely, duration selection is fundamental for the success of a DDMS conditional volatility model. In contrast to the other parameters of the model, duration selection is rather arbitrary and has been overlooked in empirical work, especially in a forecasting context. This is particularly relevant for conditional volatility models, where it is difficult to set the parameter on theoretical grounds. The researcher may think that the impact of the duration-dependence vanishes after selecting a reasonably large value. However, as we show in the empirical application, this idea hardly holds in a conditional volatility context given the high persistence in volatility. Since the DDMS model is able to generate different conditional volatility forecasts for every duration parameter selected, the empirical question then becomes how to set the duration parameter to generate more precise forecasts.

In this paper, we tackle the problem of duration selection for volatility forecasting in a real-time setting. We recognize that the duration parameter may be subject to a variance and estimation error. Duration variability arises as there is uncertainty in the autocorrelation pattern of the volatility process, leading to parameter breaks in its process. This explains the success of Markov-switching models for volatility estimation (see among others Cai, 1994; Gray, 1996). Moreover, estimated GARCH models tend to be highly persistent or very close to being non-stationary (for an explanation see Mikosch and Stărică, 2004). These stylized facts mean that the temporal dependence in the volatility process is likely not constant, but rather changes over time. The direct implication for DDMS models is that the duration parameter may also change over time, as the duration implies dependence between two separate observations in time.

The usual approach for selecting the duration is by setting an arbitrarily large threshold, or by applying a search grid for the highest likelihood value at the full sample. However, by overlooking changes in the duration, these methods are not able to mimic online (i.e. real-time) predictions. This is especially relevant in light of inevitable specification issues when applying real data, given the latency of conditional volatility. As a consequence, the forecasting power of the DDMS has still not been evaluated in an out-of-sample context.

Furthermore, the usual approach of fixing duration may generate less precise forecasts. Past research has considered all estimated parameters and filtered values as a function of a deterministic duration parameter (see Maheu and McCurdy, 2000b). The problem escalates as all other parameters are estimated conditional on a deterministic duration. Moreover, the standard error of the forecasts do not take into account possible errors in the duration. This means that if an inappropriate duration parameter is selected, these errors will be dragged into the forecasting functions rendering them less precise. Furthermore, it is also relevant for interval forecasting, which implies that tests for forecasting superiority may fail to reject the null.

From this perspective, we aim at incorporating duration uncertainty in forecasting conditional volatility. However, it is not straightforward to incorporate duration uncertainty as the estimation of the duration parameter lacks a formal classical maximum likelihood or Bayesian treatment. We tackle this problem by generating forecast functions from DDMS models with different duration parameters and dynamically forming forecasts. This allow us to circumvent the problem of estimating duration in the forecasting context, which simplifies the procedure. It is worth mentioning that our statistical approach is sufficient for our empirical objective, which is to evaluate the volatility forecasting power of DDMS models. We consider simple averages and data-driven methods, allowing volatility forecasts to incorporate duration uncertainty in different ways.

Finally, our empirical application extends the research concerned with evaluating alternative volatility modeling and forecasting methods for bitcoin log-returns by broadening the GARCH-type models to include the DDMS specification. More precisely, we conducted a pair-wise statistical test for the one-day ahead predicted bitcoin volatility from April 2018 to January 2020, compriseing 641 out-of-sample observations. We apply different volatility proxies and loss functions commonly found in the literature for robustness checking. In addition, we perform a value-at-risk (VaR) comparison between the DDMS and several GARCH and MS-GARCH models. Overall results indicate that our modeling approach outperforms benchmark specifications. In addition, taking into account the uncertainty about the duration dependence parameter is important to improve point forecasts and VaR estimation.

The remainder of this paper is organized as follows: Section 2 presents the model methodology; Section 3 describes the data and our empirical findings; Section 4 concludes.

2. Methodology

2.1. Duration Dependence

Based on Maheu and McCurdy (2000b), we begin by assuming that the time series process is governed by a discrete mixture of distributions, where the state mixing variable $S_t \in \{0,1\}$ depends only on S_{t-1} and its duration, D_{t-1}. The duration variable depicts the length of a run of realizations of a particular state and is given by:

$$D_t = \min\left(D_{t-1} I\left(S_t, S_{t-1}\right) + 1, \tau\right), \tag{1}$$

where $I(S_t, S_{t-1}) = 1$ if $S_t = S_{t-1}$ and 0 otherwise. To make the estimation possible, the duration memory is defined by τ.[1] Moreover, it is necessary to parametrize the transition matrix using a function with restricted range between zero and one so that all the probabilities are well defined. Several parametrizations are available in the literature (see Maheu and McCurdy, 2000a). In this paper, we employ the sine-squared function, which was found to be computationally more efficient to find the maximum likelihood estimate while proving similar results when comparing to alternative parametrizations. Using i and d to index the state and duration, where $\gamma_1(i)$ e $\gamma_2(i)$ are the parameters, the transition probability for $i = 0, 1$ is given by:

$$P(S_t = i | S_{t-1} = i, D(S_{t-1}) = d) = \begin{cases} sin(\gamma_1(i) + \gamma_2(i)d)^2, & \text{if } d \leq \tau \\ \\ sin(\gamma_1(i) + \gamma_2(i)\tau)^2, & \text{if } d > \tau \end{cases} \qquad i = 0, 1. \tag{2}$$

Given the duration dependence and the transition probabilities in Equations (1)-(2), the model for the conditional volatility equation is given by:

$$r_t = \sigma(S_t, D_t) z_t, \quad z_t \sim N(0, 1), \quad S_t = 0, 1 \tag{3}$$

$$\sigma(S_t, D_t) = (\omega(S_t) + \zeta(S_t) D_t)^2, \tag{4}$$

where, the latent state variable affects the level of volatility, $\omega(S_t) = \omega_0(1 - S_t) + \omega_1 S_t$, while the duration of the states, D_t, affect the dynamics of volatility through $\zeta(S_t) D_t$, where, $\zeta(S_t) = \zeta_0(1 - S_t) + \zeta_1 S_t$. z_t is is assumed to follow an identically and independently normal distribution.

2.2. Forecast Combination

To tackle the problem of duration selection for volatility forecasting, we rely on a model combination procedure to aggregate N individual models forecasts (each one indexed by a fixed duration memory) into a pooled modeling approach. Let, $\hat{\sigma}^2_{t+1}$, be the weighted average of the individual volatility forecasting models $\left\{ \hat{\sigma}^2_{i,t+1} \right\}_{i=1}^N$:

$$\hat{\sigma}^2_{t+1} = \sum_{i=1}^{N} w_{i,t} \, \hat{\sigma}^2_{i,t+1}, \tag{5}$$

where $\{w_{i,t}\}_{i=1}^N$ are the combining weights at time t and N is taken over a set of DDMS models restricted by lower (upper) bound of τ.

In our empirical analysis, we set the lower and upper bounds of τ using the smoothed probabilities of the restricted model ($\gamma_2(i) = 0$, $\zeta(i) = 0$). More precisely, we estimate the classical Markov-switching model (without duration dependence) and take the mean value using the shortest (longest) period that each regime has a smooth probability greater than 0.5 (on average). Despite our duration selection bands' lack of statistical properties,

[1] Although is necessary to set a duration memory for the DDMS estimation, our paper analysis remain on the uncertainty regarding the choice value of τ.

this data-driven method follows in part the business cycle identification methodology. However, we are particularly interested in capturing the recurring low and high frequency of conditional volatility to set duration uncertainty.[2]

For the combining weights, we use two methods: (i) fixed weights, i.e., the simple average: $w_{i,t} = \frac{1}{N}$, and (ii) moving weights, i.e., the discount mean square prediction error (DMSPE). In particular, this method uses the functions of the historical forecasting performance of the individual models over a holdout out-of-sample period as weights:

$$w_{i,t} = \varphi_{i,t}^{-1} \Big/ \sum_{j=1}^{N} \varphi_{j,t}^{-1}, \tag{6}$$

where,

$$\varphi_{i,t} = \sum_{s=m+1}^{t} \theta^{t-s} \left(\sigma_s^2 - \hat{\sigma}_{i,s}^2\right)^2, \tag{7}$$

where θ is the discount factor. For $\theta = 1$, there is no discounting and the individual forecasts are uncorrelated. When $\theta < 1$, greater weight is attached to the recent forecast accuracy of the individual models (see Stock and Watson, 2004). Although much research has been done on model combination techniques, we focus on simpler methods since we want to marginalize the gains of using different duration's models.

2.3. Model Estimation

The DDMS model can be estimated using the maximum-likelihood approach (see Maheu and McCurdy, 2000a) or using MCMC methods (see Pelagatti, 2007). In any case, the DDMS model can be collapsed into a first-order Markov model with N states, if we define a new latent variable $\mathcal{S}_t$, which covers all possible paths from $S_t = i$ up to τ, for $i = 0, 1$. More specifically, the transition matrix for $\mathcal{S}_t$, is given by:

$$P = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{NN} \end{bmatrix}, \tag{8}$$

where $p_{ij} = P(\mathcal{S}_t = i | \mathcal{S}_{t-1} = j)$, with $i, j = 1, \ldots, N$ is constructed using the transition probability equation, and the number of states N is function of τ given by $N = 2 + 2(\tau - 1)$.[3] On this basis, the filter and the likelihhod can be performed following Hamilton (1989).

[2]For more details, see the empirical results.

[3]This equation is adapted from Maheu and McCurdy (2000a) since we are not modeling the conditional mean equation.

3. Data and Results

3.1. Data

For our study's purposes, we use daily prices of the Bitcoin (in U.S dollar) traded on Bitstamp from January 2015 to January 2020. This data series is an interesting case of study since cryptocurrency is more volatile than traditional assets, and recently, it has gained increased attention in empirical studies (see Corbet and Katsiampa, 2020). Table 1 reports some descriptive statistics of the bitcoin log-returns. The mean is 0.0017 and is statistically close to zero, while the standard deviation is 0.0331 with an annualized value of 0.6330. The largest price increase is 0.1585, while the largest price decrease is -0.2175. The skewness is small and negative, and the kurtosis is higher than the normal distribution value. The Jarque-Bera (JB) statistics also indicated the departure from normality, while the Lagrange Multiplier Test and the Ljung-Box statistics are highly significant, suggesting ARCH effects in the data series.

Table 1: **Descriptive Statistics**

Obs.	Mean	Std. Dev.	Max.	Min.	Skewness	Kurtosis	JB	LM(8)	$Q^2(8)$
1824	0.0017	0.0331	0.1585	-0.2175	-0.3527	7.3730	1491.2^{***}	244.2^{***}	433.4^{***}

Note: This table presents the summary statistics of the Bitcoin daily log-returns from 01/01/2015 to 01/01/2020. JB refers to the Jarque-Bera test of normality. LM is the Lagrange Multiplier test for ARCH effects in the demeaned returns, while Q^2 is the corresponding Ljung-Box statistic on the squared demeaned returns, respectively. *** indicates the rejection of the null hypotheses at the 1% level.

The out-of-sample analysis is conducted through expanding window estimation from April 2018 to January 2020 (641 daily observations). This forecasting sample period is very volatile and was selected to test the forecasting ability of the DDMS model under challenging market conditions. Furthermore, it extends other empirical applications that use Markov-switching models to forecast bitcoin volatility (see, for example Ardia et al., 2019; Segnon and Bekiros, 2020). To evaluate the DDMS combination volatility forecast's, we use intraday 5-minute quotes as an alternative source to proxy volatility instead of squared daily returns (see Andersen and Bollerslev, 1998).[4] Following Trucíos (2019), we use the realized variance (RV) and few other realized measures robust to microstructure noise and jump, which are widely used nowadays. More specially, we use the bipower variation - BV (Barndorff-Nielsen and Shephard, 2004), MinRV and MedRV (Andersen et al., 2012) as proxies of true volatility. Details about realized measures, see Table 6 in Appendix.

[4]The data series were obtained from bitcoincharts website: https://bitcoincharts.com.

3.2. Empirical Motivation Methodology

We start the predictive analysis of the DDMS model combination, reporting our empirical strategy to set the lower and upper bounds of τ. Using the in-sample data set from January of 2015 to March 2018 (1183 observations), Table 2 presents the average of the Markov-switching model's smoothed probabilities. Using the shortest (*) and longest (**) periods which each state presents a chance greater than 0.5, we proxy the lower and upper bounds of τ taking the mean of these values, i.e., $\tau_{min} = 5 = (3^* + 7^*)/2$ and $\tau_{max} = 236 = (335^{**} + 138^{**})/2$, respectively. Once we define these values, the model combination scheme is constructed using equally spaced models over the duration index set. Particularly, we take five models with durations: 5, 60, 120, 180, and 236, to run the next forecasting exercises.

Table 2: **Smoothed Probability of the Restricted Model.**

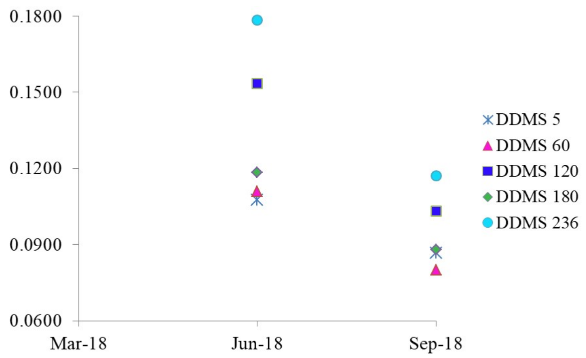
$S_t = 0$	Days	Avg. Prob.	$S_t = 1$	Days	Avg. Prob.
-	-	-	1-45	45	0.946
46-54	9	0.711	55-59	5	0.953
60-73	14	0.842	74-82	9	0.924
83-99	17	0.861	100-102	3^*	0.684
103-183	81	0.969	184-191	8	0.780
192-226	35	0.939	227-235	9	0.957
236-294	59	0.980	295-343	49	0.913
344-354	11	0.789	355-357	3	0.725
358-375	18	0.907	376-385	10	0.912
386-509	124	0.983	510-517	8	0.762
518-524	7^*	0.791	525-551	27	0.957
552-574	23	0.958	575-579	5	0.886
580-717	138^{**}	0.991	718-745	28	0.979
746-793	48	0.922	794-821	28	0.853
822-848	27	0.918	849-1183	335^{**}	0.976

Note: This table reports the shortest (*) and longest (**) periods which each state presents a smoothed probability greater than 0.5 on average. These values are used to characterize the lower (upper) bound in our DDMS combination; $\tau_{min} = 5 = (3^* + 7^*)/2$ and $\tau_{max} = 236 = (335^{**} + 138^{**})/2$, respectively.

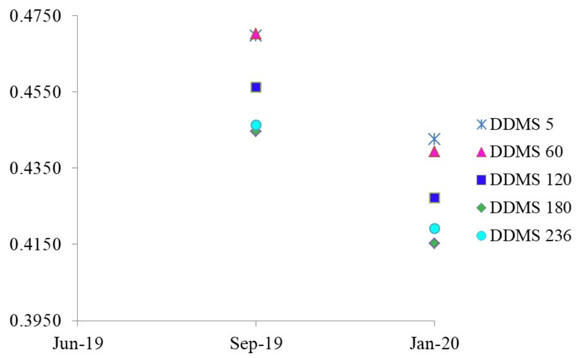
At this point, it is straightforward to notice that our modeling approach ultimately depends on the in-sample data size estimation since it affects the regime classification used to set the lower and upper bounds of τ. However, our empirical research focuses on the idea that the DDMS model can generate different conditional volatility forecasts for every duration parameter selected. This enables us to explore the model combination technique to incorporate duration uncertainty in the analysis.

Before we conduct a statistical test for the predictive ability of the DDMS combination, we first explore the empirical issues of fixing the duration. Figure 1 displays the accumulated loss function for the one-day-ahead forecasting as the out-of-sample sizes changes. More precisely, we use the previously selected duration models and the realized variance. Additionally to the mean squared error - MSE (upper side figures), we also compute the QLIKE robust loss function (bottom side figures). Based on Patton (2011), the use of the robust function leads to selecting true volatility forecasts even if proxies are imperfect due to noise presence. In general, for the beginning of the out-of-sample (first two quarters), the DDMS 60 is the best model. On the other hand, the DDMS 180 and 236 present the best results for the ending out-of-sample periods (last two quarters). Our empirical finding is more evident using the robust loss function, as we see by the vertical distance between the best model and other specifications.

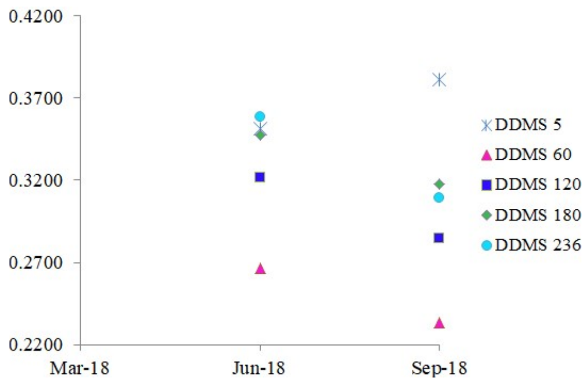
Figure 1: **Forecasting Sample Size and Duration Dependence**
Loss Functions & RV 5 min.



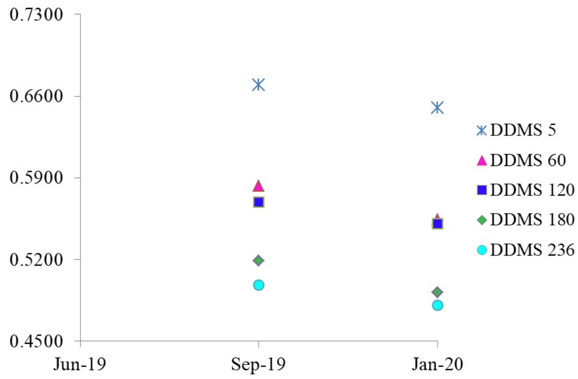
(a) MSE (10^5) x 1, 2 QTD.



(b) MSE (10^5) x 6, 7 QTD.



(c) QLIKE x 1, 2 QTD.



(d) QLIKE x 6, 7 QTD.

3.3. Point Forecasting

We conduct the point forecasting analysis comparing the duration dependence model combination with the GARCH-type specifications. More precisely, we use the GARCH, EGARCH, and GJR-GARCH models, with (without) switching process as benchmark models.[5] Although the literature about bitcoin volatility has widely recognized the GARCH-type specifications, we broad these studies using the duration dependence model. We also expand our point forecasting exercise in two directions. First, we also combine the GARCH-type specifications in line with the duration models. Secondly, we compare the combination approach with the fixed duration model. In this case, we use the intermediate duration model, DDMS 120, since its raking remains relatively stable as the

[5]See the model specifications in Table 7 in Appendix.

out-of-sample sizes changes, as seen in Figure 1. We compare the forecast results using the Diebold-Mariano-West test (Diebold and Mariano, 1995; West, 1996) through the proxies volatiles (RV, BV, MinRv, and MedRV) for the MSE and QLIKE loss functions.

Table 3 shows the results for the fixed weighted model combination. In most cases, the *t*-statistic is positive, indicating that the benchmark model forecasts generated a larger average loss than the DDMS combination. Notably, exceptions are obtained exclusively using the MSE loss function for the models GARCH, GJR-GARCH, and EGARCH (not for all proxies). For the robust QLIKE loss function, the t-statistic is positive for all models, and the null hypothesis of equal predictive accuracy is rejected at most of the models at 0.10 (or lower) level. Only 5 of 32 pair-wise results is significate greater than 0.10 level. Despite using a simple combination strategy, the overall results indicate that the duration dependence combination surpasses some of the empirical literature's established models.

Table 3: **Diebold–Mariano–West test**
Fixed Weighted - Mean

Models	Loss Functions	MedRV (5 Min.)	MinRV (5 Min.)	BV (5 Min.)	RV (5 Min.)
DDMSC x GARCH	MSE	-0.50	-0.45	-0.71	-0.88
	QLIKE	**1.70**	**1.67**	**1.79**	**2.11**
DDMSC x MS-GARCH	MSE	1.11	1.11	1.42	**2.08**
	QLIKE	**1.97**	**1.83**	**2.40**	**3.43**
DDMSC x EGARCH	MSE	0.04	0.15	-0.03	-0.19
	QLIKE	1.56	1.51	1.60	**1.83**
DDMSC x MS-EGARCH	MSE	0.69	0.72	0.82	0.97
	QLIKE	**1.87**	**1.77**	**2.25**	**3.01**
DDMSC x GJR-GARCH	MSE	-0.54	-0.47	-0.75	-0.93
	QLIKE	**1.69**	**1.66**	**1.78**	**2.09**
DDMSC x MS-GJR-GARCH	MSE	**1.92**	**1.81**	**1.96**	**2.16**
	QLIKE	**1.72**	1.60	**1.76**	**1.97**
DDMSC x Mean Garch-Types	MSE	0.07	0.20	0.11	0.08
	QLIKE	1.58	1.51	**1.78**	**2.23**
DDMSC x DDMS 120	MSE	0.80	0.75	0.58	0.58
	QLIKE	**1.75**	**1.71**	**1.82**	**2.14**

Note: The table above presents the *t*-statistics from Diebold–Mariano–West tests of equal predictive accuracy for the benchmark models and the duration-dependent Markov-switching Combination (DDMSC) approach. In bold, we have the *t*-statistic greater than 1.65 (1.96) in absolute value, which indicates a rejection of the null of equal predictive accuracy at the 0.10 (0.05) level. A positive *t*-statistic indicates that the benchmark model forecast produced larger average loss than the DDMSC approach, while a negative sign indicates the opposite. In our forecast exercise, we use five equally spaced duration values, 5, 60, 120, 180 and 236, respectively

Tables 4 and 5 show the moving weighted model combination's results using $\theta = 0.9$ and $\theta = 1.0$ as discount factor values.[6] Although the same combination method is used, results for $\theta = 1.0$ share similar characteristics to using fixed weights. Moreover, Table 4 displays a small frequency of null rejections for the QLIKE loss function. This suggests that using $\theta = 1.0$ would be more appropriate for the DMSFE combination since it does not assign weights to recent observations. The rationale is that a large number of observations and consequently forecasting errors can be necessary to define a duration process for the volatility where they should induce similar effects on its memory. Thus, attaching higher weights only to closer observations is not enough to fully capture the memory process of the volatility.

Table 4: **Diebold–Mariano–West test**
Moving Weighted - DMSFE $\theta = 0.9$

Models	Loss Functions	MedRV (5 Min.)	MinRV (5 Min.)	BV (5 Min.)	RV (5 Min.)
DDMSC x GARCH	MSE	-0.47	-0.41	-0.69	-0.86
	QLIKE	**1.74**	**1.66**	**1.83**	**2.23**
DDMSC x MS-GARCH	MSE	1.13	1.13	1.43	**2.10**
	QLIKE	1.58	1.55	**2.11**	**3.35**
DDMSC x EGARCH	MSE	0.09	0.19	0.01	-0.13
	QLIKE	1.52	1.44	1.56	**1.86**
DDMSC x MS-EGARCH	MSE	0.72	0.75	0.85	1.02
	QLIKE	1.55	1.51	**2.05**	**3.05**
DDMSC x GJR-GARCH	MSE	-0.51	-0.43	-0.72	-0.91
	QLIKE	**1.73**	**1.65**	**1.81**	**2.21**
DDMSC x MS-GJR-GARCH	MSE	**1.93**	**1.82**	**1.97**	**2.18**
	QLIKE	1.31	1.32	1.36	1.53
DDMSC x DMSFE Garch-Types	MSE	0.02	0.17	0.08	0.05
	QLIKE	1.45	1.37	**1.67**	**2.26**
DDMSC x DDMS 120	MSE	0.91	0.88	0.68	0.68
	QLIKE	1.49	1.45	1.48	**1.88**

Note: The table above presents the t-statistics from Diebold–Mariano–West tests of equal predictive accuracy for the benchmark models and the duration-dependent Markov-switching Combination (DDMSC) approach. In bold, we have the t-statistic greater than 1.65 (1.96) in absolute value, which indicates a rejection of the null of equal predictive accuracy at the 0.10 (0.05) level. A positive t-statistic indicates that the benchmark models forecast produced larger average loss than the DDMSC approach, while a negative sign indicates the opposite. In our forecast exercise, we use five equally spaced duration values, 5, 60, 120, 180 and 236, respectively.

[6] We use 30 days for the holdout out-of-sample period.

Table 5: **Diebold–Mariano–West test**
Moving Weighted - DMSFE $\theta = 1.0$

Models	Loss Functions	MedRV (5 Min.)	MinRV (5 Min.)	BV (5 Min.)	RV (5 Min.)
DDMSC x GARCH	MSE	-0.52	-0.47	-0.73	-0.89
	QLIKE	**1.70**	**1.66**	**1.79**	**2.12**
DDMSC x MS-GARCH	MSE	1.12	1.12	1.44	**2.10**
	QLIKE	**1.98**	**1.84**	**2.40**	**3.43**
DDMSC x EGARCH	MSE	0.04	0.14	-0.04	-0.18
	QLIKE	1.56	1.51	1.60	**1.83**
DDMSC x MS-EGARCH	MSE	0.70	0.72	0.83	0.99
	QLIKE	**1.88**	**1.78**	**2.24**	**3.01**
DDMSC x GJR-GARCH	MSE	-0.56	-0.50	-0.76	-0.94
	QLIKE	**1.69**	**1.65**	**1.78**	**2.10**
DDMSC x MS-GJR-GARCH	MSE	**1.94**	**1.83**	**1.98**	**2.18**
	QLIKE	**1.74**	1.61	**1.77**	**1.97**
DDMSC x DMSFE Garch-Types	MSE	0.02	0.15	0.06	0.04
	QLIKE	1.59	1.52	**1.77**	**2.21**
DDMSC x DDMS 120	MSE	0.78	0.70	0.58	0.60
	QLIKE	**1.74**	**1.70**	**1.82**	**2.14**

Note: The table above presents the t-statistics from Diebold–Mariano–West tests of equal predictive accuracy for the benchmark models and the duration-dependent Markov-switching Combination (DDMSC) approach. In bold, we have the t-statistic greater than 1.65 (1.96) in absolute value, which indicates a rejection of the null of equal predictive accuracy at the 0.10 (0.05) level. A positive t-statistic indicates that the benchmark models forecast produced larger average loss than the DDMSC approach, while a negative sign indicates the opposite. In our forecast exercise, we use five equally spaced duration values, 5, 60, 120, 180 and 236, respectively.

3.4. Risk Prediction Evaluation

Apart from computing statistical measures of predictive accuracy as the point forecast-based ones reported in the previous section, it is also important to evaluate the performance of the volatility forecasts in terms of the out-of-sample VaR. VaR is the most used measure of financial risk, it is easy and intuitive for non-specialists to understand and offers a more sensible measure of the risk than variance since it focuses on losses (see Nieto and Ruiz, 2016, for a comprehensive survey on VaR forecasting). Assuming that the returns are zero mean and serially uncorrelated, the one-step-ahead VaR conditional on information available at time t is computed as

$$VaR^{\alpha}_{t+1|t} = Q_{\alpha}\sigma_{t+1|t}, \tag{9}$$

where Q_{α} is the α quantile of the standardized error distribution (assumed normal in our case[7]) and $\sigma_{t+1|t}$ is the one-step-ahead volatility forecast.

The VaR forecast adequacy of the models is assessed through backtesting procedures. VaR backtesting procedures usually check the assumptions of correct coverage of the unconditional and conditional left–tail of log–returns distribution and that the VaR failures

[7] See Fameliti and Skintzi (2020) for the same error distribution assumption for the model combination case and Ardia et al. (2018) for the parametric construction of the VaR for the Markov-switching model.

occur independently. As common practice in the VaR forecasting literature, we apply the independence test (IND) of Christoffersen (1998), the unconditional coverage (UC) test of Kupiec (1995), the conditional coverage (CC) test of Christoffersen (1998), the dynamic quantile (DQ) test of Engle and Manganelli (2004), and for model comparison we also compute the average quantile loss function (AQL) of González-Rivera et al. (2004). An explanation of each of the statistical tests is provided by Nieto and Ruiz (2016) and Troster et al. (2019).

Table 6 reports the hit rate (returns smaller than VaR), the p-values of the IND, UC, CC and DQ tests, and the average loss (AQL) for the 1% and 2.5% VaR levels obtained with each of the models considered in the paper. DDMSC(M) denotes the fixed weighted model combination while DDMSC(0.9) and DDMSC(1.0) are the moving weighted model combinations using θ=0.9 and 1.0, respectively, as discount factor values.[8] Gray cells in the columns for the tests indicate p-values greater than 0.05 (non-rejection at the 5% signifcance level), while bold values in the column of AQL indicate the model with the best performance (lowest average loss). For both 1% and 2.5% VaR levels we observe that, although all the models exceed the proportion of VaR failures, the DDMSC specifications report hit rates closer to the confidence VaR levels than those of the Garch-type models. We can also observe that, for both 1% and 2.5% VaR levels, all the models passed the independence test of VaR failures.

When looking at the coverage tests UC, CC and DQ for the 1% VaR level, we find that all the models are rejected at the usual 5% significance level. Considering the 2.5% VaR level, the DDMS specifications show p-values much higher than the GARCH-type models. For the UC test, only the GARCH combinations are rejected. However, as we move to more robust tests (CC and DQ), only the DDMS 120, DDSMC and MS-EGARCH models are not rejected. Comparing these models, the DDMSC(M) report the lowest average loss. It is worth noting, that all the DDMSC combining strategies showed average losses lower than the DDMS 120's loss, indicating that taking into account the uncertainty about the duration dependence parameter is important to improve VaR forecasts of DDMS models.

[8]We report only the results for DDMSC(0.9) and DDMSC(1.0) procedures estimated using the RV proxy. Results using the other realized variance proxies for the moving weighted model combination strategy were similar, which are, therefore, available upon request.

Table 6: **Backtesting results**

Models	1% VaR						2.5 % VaR					
	Hit rate	IND	UC	CC	DQ	AQL	Hit rate	IND	UC	CC	DQ	AQL
DDMS 120	2.6%	0.431	0.001	0.003	0.000	1.262	3.1%	0.131	0.349	0.206	0.174	2.280
DDMSC(M)	2.1%	0.272	0.015	0.028	0.019	**1.213**	3.3%	0.161	0.239	0.187	0.190	**2.251**
DDMSC(0.9)	2.1%	0.272	0.015	0.028	0.018	1.219	3.3%	0.161	0.239	0.187	0.187	2.258
DDMSC(1.0)	2.1%	0.272	0.015	0.028	0.019	1.214	3.3%	0.161	0.239	0.187	0.190	2.252
GARCH	2.9%	0.105	0.000	0.000	0.000	1.298	3.8%	0.057	0.061	0.028	0.004	2.345
MS-GARCH	2.8%	0.490	0.000	0.001	0.000	1.292	3.8%	0.057	0.061	0.028	0.005	2.317
EGARCH	2.5%	0.375	0.002	0.006	0.003	1.220	3.8%	0.057	0.061	0.028	0.005	2.271
MS-EGARCH	2.5%	0.384	0.002	0.006	0.008	1.219	3.8%	0.274	0.061	0.095	0.092	2.253
GJR-GARCH	3.1%	0.131	0.000	0.000	0.000	1.294	3.8%	0.057	0.061	0.028	0.004	2.342
MS-GJR-GARCH	2.8%	0.490	0.000	0.001	0.000	1.311	3.8%	0.057	0.061	0.028	0.001	2.370
Mean Garch-Types	2.6%	0.431	0.001	0.003	0.001	1.265	3.9%	0.073	0.036	0.022	0.003	2.312
DMSFE 0.9 Garch-Types	2.6%	0.431	0.001	0.003	0.001	1.265	3.9%	0.073	0.036	0.022	0.003	2.309
DMSFE 1.0 Garch-Types	2.6%	0.431	0.001	0.003	0.001	1.265	3.9%	0.073	0.036	0.022	0.003	2.311

Note: This table reports the backtesting results for the VaR forecasts at the risk levels of 1% and 2.5% obtained with the duration-dependent Markov-switching Combination (DDMSC) and the benchmark models. The hit rate and the *p*-values of the independence (IND), unconditional coverage (UC), conditional coverage (CC) and dynamic quantile (DQ) tests are also reported. We also report the average quantile loss function (AQL). Gray cells denote *p*-values greater than 0.05, indicating the non-rejection of the null hypothesis of the test, and we highlight in bold the model with lowest average loss for each VaR level.

4. Conclusion

This paper investigates the out-of-sample properties of the duration-dependent Markov-switching model to forecast the conditional volatility while allowing for uncertainty in its duration process. We rely on the model combination approach and aggregate N individual model forecasts, each indexed by a fixed duration memory, into a pooled modeling approach to tackle duration selection. More precisely, our empirical strategy set the lower and upper bounds of the duration index set using the smooth probabilities of the classical Markov-switching model. Our empirical analysis is conducted through a pairwise statistical test for the one-day ahead predicted bitcoin volatility from April 2018 to January 2020, 641 out-of-sample observations. We apply different volatility proxies and loss functions commonly found in the literature for robustness checking. Overall results indicated that our modeling approach outperforms the Markov-switching GARCH-types models when considering point forecasts.

Regarding the out-of-sample VaR performance of the volatility forecasts, backtesting procedures were conducted. For the 2.5% VaR forecasts, ours proposals deliver satisfactory results, outperforming the Garch-type models and the DDMS 120 model. It is worth noting that, improvements over the benchmark models were achieved even by simple combining strategies as a simple model averaging. These results suggest the importance of taking into account the uncertainty about the duration dependence parameter when evaluating risk measures. For the 1% VaR forecasts, the results could be improved by considering heavy-tailed distributions in the estimation of our proposals. Such extensions are left for future research.

References

Andersen, T.G., Bollerslev, T., 1998. Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. International Economic Review 39, 885–905.

Andersen, T.G., Dobrev, D., Schaumburg, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. Journal of Econometrics 169, 75–93.

Ardia, D., Bluteau, K., Boudt, K., Catania, L., 2018. Forecasting risk with Markov-switching GARCH models:A large-scale performance study. International Journal of Forecasting 34, 733–747.

Ardia, D., Bluteau, K., Rüede, M., 2019. Regime changes in Bitcoin GARCH volatility dynamics. Finance Research Letters 29, 266–271.

Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and Bipower Variation with Stochastic Volatility and Jumps. Journal of Financial Econometrics 2, 1–37.

Bejaoui, A., Karaa, A., 2016. Revisiting the bull and bear markets notions in the Tunisian stock market: New evidence from multi-state duration-dependence Markov-switching models. Economic Modelling 59, 529–545.

Cai, J., 1994. A Markov Model of Switching-Regime ARCH. Journal of Business & Economic Statistics 12, 309–316.

Caldeira, J.F., Moura, G.V., Santos, A.A., 2015. Measuring risk in fixed income portfolios using yield curve models. Computational Economics 46, 65–82.

Christoffersen, P.F., 1998. Evaluating interval forecasts. International economic review , 841–862.

Corbet, S., Katsiampa, P., 2020. Asymmetric mean reversion of Bitcoin price returns. International Review of Financial Analysis 71, 101267.

Diebold, F.X., Mariano, R.S., 1995. Comparing Predictive Accuracy. Journal of Business & Economic Statistics 13, 253–263.

Durland, J.M., McCurdy, T.H., 1994. Duration-Dependent Transitions in a Markov Model of U.S. GNP Growth. Journal of Business & Economic Statistics 12, 279–288.

Engle, R.F., Manganelli, S., 2004. Caviar: Conditional autoregressive value at risk by regression quantiles. Journal of business & economic statistics 22, 367–381.

Fameliti, S.P., Skintzi, V.D., 2020. Predictive ability and economic gains from volatility forecast combinations. Journal of Forecasting 39, 200–219.

González-Rivera, G., Lee, T.H., Mishra, S., 2004. Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. International Journal of forecasting 20, 629–645.

Gray, S.F., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42, 27–62.

Haas, M., Mittnik, S., Paolella, M.S., 2004. A New Approach to Markov-Switching GARCH Models. Journal of Financial Econometrics 2, 493–530.

Hamilton, J.D., 1989. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. Econometrica 57, 357–384.

Kupiec, P., 1995. Techniques for verifying the accuracy of risk measurement models. The J. of Derivatives 3.

Layton, A.P., Smith, D.R., 2007. Business cycle dynamics with duration dependence and leading indicators. Journal of Macroeconomics 29, 855–875.

Maheu, J.M., McCurdy, T.H., 2000a. Identifying Bull and Bear Markets in Stock Returns. Journal of Business & Economic Statistics 18, 100–112.

Maheu, J.M., McCurdy, T.H., 2000b. Volatility dynamics under duration-dependent mixing. Journal of Empirical Finance 7, 345–372.

Mikosch, T., Stărică, C., 2004. Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects. The Review of Economics and Statistics 86, 378–390.

Nieto, M.R., Ruiz, E., 2016. Frontiers in var forecasting and backtesting. International Journal of Forecasting 32, 475–501.

Patton, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160, 246–256.

Pelagatti, M., 2007. Duration Dependent Markov-Switching Vector Autoregression: Properties, Bayesian Inference, Software and Application. In: Business Fluctuations and Cycles, Science Publishers, Hauppauge NY.

Segnon, M., Bekiros, S., 2020. Forecasting volatility in bitcoin market. Ann Finance 16, 435–462.

Stock, J.H., Watson, M.W., 2004. Combination forecasts of output growth in a seven-country data set. Journal of Forecasting 23, 405–430.

Troster, V., Tiwari, A.K., Shahbaz, M., Macedo, D.N., 2019. Bitcoin returns and risk: A general garch and gas analysis. Finance Research Letters 30, 187–193.

Trucíos, C., 2019. Forecasting Bitcoin risk measures: A robust approach. International Journal of Forecasting 35, 836–847.

West, K.D., 1996. Asymptotic Inference about Predictive Ability. Econometrica 64, 1067–1084.

Appendix

Table 7: **Realized Measures**

RV	$\sum_{i=1}^{m} r_{i,t}^2$
BV	$\frac{\pi}{2}\left(\frac{m}{m-1}\right)\sum_{i=1}^{m-1}\lvert r_{i,t}\rvert\lvert r_{i+1,t}\rvert$
MinRV	$\frac{\pi}{\pi-2}\left(\frac{m}{m-1}\right)\sum_{i=1}^{m-1}\min(\lvert r_{i,t}\rvert\lvert r_{i+1,t}\rvert)^2$
MedRV	$\frac{\pi}{6-4\sqrt{3}+\pi}\left(\frac{m}{m-2}\right)\sum_{i=2}^{m-1} median(\lvert r_{i-1,t}\rvert\lvert r_{i,t}\rvert\lvert r_{i+1,t}\rvert)^2$

Note: The table above presents the realized measures used as volatilities proxies in our empirical research. Considering a trading day t divided into m equally sized sub-periods with return $r_{i,t}$ in sub-period i, we have $m = 288$ for 5 minutes intra-day Bitcoin prices quotes.

Table 8: **GARCH-type Specifications**

	$r_t = u_t, \quad u_t = \sigma_{k,t} z_t, \quad z_t \sim N(0,1)$
GARCH	$\sigma_{k,t}^2 = \omega_k + \alpha_{1,k} u_{t-1}^2 + \beta_k \sigma_{k,t-1}^2$
EGARCH	$\ln(\sigma_{k,t}^2) = \omega_k + \alpha_{1,k}(\lvert u_{t-1}/\sigma_{k,t-1}\rvert - E[\lvert u_{t-1}/\sigma_{k,t-1}\rvert]) u_{t-1}^2 + \alpha_{2,k} u_{t-1}/\sigma_{k,t-1} + \beta_k \ln(\sigma_{k,t-1}^2)$
GJR-GARCH	$\sigma_{k,t}^2 = \omega_k + (\alpha_{1,k} + \alpha_{2,k} I\{u_{t-1} < 0\}) u_{t-1}^2 + \beta_k \sigma_{k,t-1}^2$

Note: The table above presents the GARCH-type specifications. The number of regimes is defined by k=1,2 and $I\{\cdot\}$ is the indicator function. We follow the modelling approach of Haas et al. (2004).