Mitigating the choice of the duration in DDMS models through a parametric link

Fernando Henrique de Paula e Silva Mendes[a,*], **Douglas Eduardo Turatti**[b] **and Guilherme Pumi**[a]

Abstract

One of the most important hyper-parameters in duration-dependent Markov-switching (DDMS) models is the duration of the hidden states. Because there is currently no procedure for estimating this duration or testing whether a given duration is appropriate for a given data set, an ad hoc duration choice must be heuristically justified. This is typically a difficult task and is likely the most delicate point of the modeling procedure, allowing for criticism and ultimately hindering the use of DDMS models. In this paper, we propose and examine a methodology that mitigates the choice of duration in DDMS models when forecasting is the goal. The idea is to use a parametric link instead of the usual fixed link when calculating transition probabilities. As a result, the model becomes more flexible and any potentially incorrect duration choice (i.e., misspecification) is compensated by the parameter in the link, yielding a likelihood and transition probabilities very close to the true ones while, at the same time, improving forecasting accuracy under misspecification. We evaluate the proposed approach in Monte Carlo simulations and using real data applications. Results indicate that the parametric link model outperforms the benchmark logit model, both in terms of in-sample estimation and out-of-sample forecasting, for both well-specified and misspecified duration values.

1 Introduction

Many authors have suggested variants of the Markov-switching model since the seminal paper of Hamilton (1989). In this context, Durland and McCurdy (1994) proposed 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, such as if the continuation of expansions or recessions is dependent on how long the economy has been in those regimes, this modeling approach also includes bull and bear market stock identification (Maheu and McCurdy, 2000a) and foreign exchange volatility estimation (Maheu and McCurdy, 2000b) since duration can be included as a conditioning variable in the conditional mean/variance equations. Based on that, the empirical literature has also explored generalizations of the DDMS model, broadening the baseline duration dependence structure in different directions (see, for example, Lam, 2004; Pelagatti, 2007; Bejaoui and Karaa, 2016, among others.).

One of the key components in DDMS models is the duration value to be used in the application, which is a user-specified parameter that cannot be estimated. The duration selection in the existing literature is either arbitrary, under the argument that the impact of the duration dependence vanishes after selecting a "reasonably" large value or based on a grid search aimed at maximizing the likelihood. The difficulty in selecting and justifying the

[*]Corresponding author. This Version: May 10, 2023

[a]Graduate Program in Statistics - Federal University of Rio Grande do Sul.

[b]Aalborg University Business School.

E-mails: fernandohpsm@hotmail.com (F.H.P.S Mendes), guilherme.pumi@ufrgs.br (G. Pumi) and det@business.aau.dk (D.E. Turatti)

duration is the main criticism of DDMS models, often hindering their use in applications. In contrast to the existing literature, Isogai et al. (2008) showed, using a bayesian approach, that duration structures are not necessarily monotonic and, therefore, cannot be described by the conventional models. The authors restrict their approach to a bull and bear market analysis in the same spirit of Maheu and McCurdy (2000a) through a country-by-country stock data analysis; however, they do not extend their approach to other DDMS-type models and restrict the study to an in-sample design.

In this context, the main contribution of the present paper is to develop and examine a strategy for estimating DDMS models that attenuates the problem of duration selection and justification by allowing for a completely arbitrary choice. Our approach is based on the use of the asymmetric Aranda-Ordaz parametric link function (Aranda-Ordaz, 1981) instead of the (fixed) logit link in the transition probabilities in DDMS models. The idea behind this approach is that any potentially incorrect duration choice is compensated for by the parameter in the link, increasing model flexibility by letting the data "tell" which is the "best" link, allowing the model to capture the latency of the Markov chain endogenously, improving the fit and forecasting accuracy under misspecification of the duration.

To illustrate our approach empirically, we extend the research that evaluates alternative volatility modeling and forecasting methods for S&P500 daily log returns by broadening the traditional DDMS specifications to include the Aranda-Ordaz link. More precisely, we conduct a pairwise Diebold-Mariano-West statistical test for the one-day ahead forecast stock volatility from April 2018 to January 2020, comprising 443 out-of-sample observations. For robustness checking, we apply different volatility proxies and loss functions commonly found in the literature. Overall, our modeling approach outperforms the benchmark logit specification under certain reasonable conditions, mitigating the duration uncertainty choice. In addition, we also compare the proposed specification to Garch-type models. We did not find statistical support to reject the null hypotheses of equal predictive accuracy, while these results were not observed for the models with logit link.

To evaluate the performance of the proposed approach, we conduct a Monte Carlo simulation study inspired by two classical applications of DDMS models: (i) a bull and bear market identification for the in-sample stock market cycles probabilities and (ii) a point volatility forecasting exercise conducted for different out-of-sample horizons. In general, we observe that the proposed Aranda-Ordaz approach improves estimates in different directions; however, a typical result is the advantage (in proportion) of the Aranda-Ordaz over the logit for higher likelihood values. Over the stock market probabilities, we observe that the Aranda-Ordaz model also presents superior frequency with probability closer to the true values under duration misspecification. From the forecasting perspective, the apply of the Aranda-Ordaz is advantageous regardless of the true value of the duration. This indicates that even when the correct duration is used, the model based on the logit cannot provide the "best possible" forecasts.

The remainder of this paper is organized as follows: Section 2 presents the proposed model; Section 3 describes the optimization algorithm designed to conduct the maximum likelihood estimation. Section 4 explore Monte Carlo simulations. The data and our empirical findings are found in Section 5; Section 6 concludes the paper.

2 The proposed model

Based on Maheu and McCurdy (2000a,b), we start by considering a simple stochastic volatility model given by

$$Y_t = \mu(S_t) + \sigma\big(S_t, D(S_t)\big) Z_t,$$

where S_t denotes the state mixing variable, $D(S_t)$ is the duration of the state S_t, at time t, and $Z_t \sim N(0,1)$ is an i.i.d. error term. The duration $D(S_t)$ depicts the length of a run of realizations of a particular state and, in principle, could grow very large causing estimation problems and numerical instability. To avoid such problems, we define

$$D(S_t) := \min\big\{D(S_{t-1}) I(S_t = S_{t-1}) + 1, \tau\big\},$$

where $\tau \in \mathbb{N}$ is a user chosen threshold such that the duration is accounted for up to time τ, and I is the indicator function. The transition probabilities associated to the latent states S_t are parameterized using a similar approach as in generalized linear model by means of a link. The most commonly applied link is the logit, which yields

$$P\big(S_t = i | S_{t-1} = i, D(S_{t-1}) = d\big) = \frac{\exp\left(\gamma_1(i) + \gamma_2(i)(d \wedge \tau)\right)}{1 + \exp\left(\gamma_1(i) + \gamma_2(i)(d \wedge \tau)\right)}, \quad i = 0, 1, \tag{1}$$

where $d \wedge \tau = \min\{d, \tau\}$ and $\gamma_j(i)$, $i, j \in \{1, 2\}$, are parameters to be estimated.

In this work we propose to parameterize the transition probabilities upon applying a twice differentiable one-to-one parametric link function $g(\cdot, \lambda) : (0,1) \to \mathbb{R}$, in the same generalized linear model approach as before. That is, we consider

$$g\big(P\big(S_t = i | S_{t-1} = i, D(S_{t-1}) = d\big); \lambda\big) = \gamma_1(i) + \gamma_2(i)(d \wedge \tau), \quad i = 0, 1,$$

or, equivalently,

$$P\big(S_t = i | S_{t-1} = i, D(S_{t-1}) = d\big) = g^{-1}\big(\gamma_1(i) + \gamma_2(i)(d \wedge \tau); \lambda\big) \quad i = 0, 1, \tag{2}$$

where λ is a parameter to be estimated from the data. One of the most commonly applied parametric link function is the so-called asymmetric Aranda-Ordaz link function (Aranda-Ordaz, 1981), given by

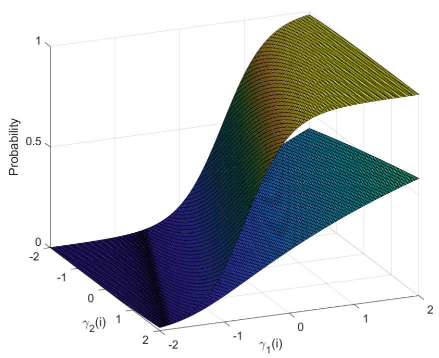
$$g(y; \lambda) = \log\left(\frac{(1-y)^{-\lambda} - 1}{\lambda}\right),$$

for $y \in (0,1)$ and $\lambda > 0$, whose inverse is given by

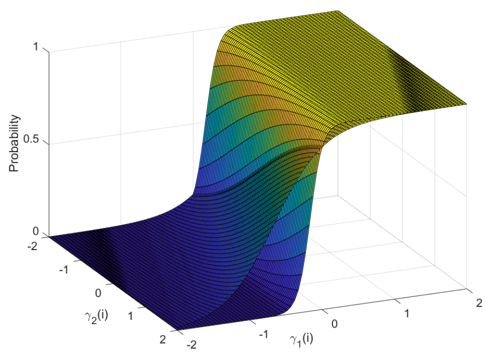
$$g^{-1}(x; \lambda) = 1 - \big(1 + \lambda e^x\big)^{-\frac{1}{\lambda}},$$

for $x \in \mathbb{R}$. Observe that $\lim\limits_{\lambda \to 0+} g^{-1}(x; \lambda) = 1 - e^{-e^x}$ and $\lim\limits_{\lambda \to 0+} g(x; \lambda) = \log\big(-\log(1-x)\big)$ which is the so-called cloglog link function.

To exemplify the effects of the parameters d and λ in the transition probabilities (2) as a function of $\gamma_1(0)$ and $\gamma_2(0)$. Figure 1(a) presents the case $d = 3$, and $\lambda \in \{1, 12\}$ and in Figure 1(b), we have $\lambda = 1$ and $d \in \{3, 12\}$. In both cases we have transition probabilities parameters $\gamma_j(0) \in [-2, 2]$, for $j \in \{1, 2\}$.



(a) $d = 3$, $\lambda \in \{1, 12\}$ and $i = 0$



(b) $\lambda = 1$, $d \in \{3, 12\}$ and $i = 0$

Figure 1: Transition probabilities shape for different values of λ and d.

3 Optimization Strategy

Parameter estimation of DDMS models can be challenging and prone to numerical issues. The likelihood function may have multiple local maxima, flat and spiky regions making numerical optimization difficult. Adding an extra parameter through the Aranda-Ordaz link may further complicate the estimation process, as the transition parameters and the link function are free to vary. This suggests that the likelihood function may have issues identifying the transition matrix and the link function parameters, implying it can be near flat on some directions. Additionally, the estimation process in DDMS models with a large duration parameter can also pose computational difficulties. A large duration parameter often leads to a large sparse transition matrix in the extended representation of the DDMS model. This results in a sparse transition matrix that may approach singularity for several combinations of parameters. Singularity means that some states are given probabilities numerically close to zero, causing the unconditional probabilities to not exist. This results in the likelihood function being undefined at multiple points in the parameter space.

To address these challenges, we developed an optimization algorithm especially tailored to maximize the log-likelihood function of DDMS models. The proposed approach applies a combination of random and grid search techniques for initial parameter values and constrained numerical optimization within defined bounds to tackle multimodality. We also impose non-linear restrictions to ensure the invertibility of the transition matrix. The details of the algorithm are outlined below.

1. Find the starting values using a combination of random and grid search: Let κ be the number of parameters in the DDMS model. Define a vector $\boldsymbol{b}$ consisting of 100 evenly spaced values between 0.1 and 10 for the parameter λ. For the remaining parameters, create a $(\kappa - 1) \times 100$ matrix C with random numbers, where each row is drawn from a continuous uniform distribution. The bounds of the uniform distributions are heuristically defined and may vary depending on the particular model and dataset. The resulting draws can then be represented as $\left[C' \quad \boldsymbol{b}_i \otimes \mathbf{1}_{100}\right]'$, for all $i = 1, \ldots, 100$, where $\otimes$ denotes the Hadamard (elementwise) product and $\mathbf{1}_k \in \mathbb{R}^k$ denotes a vector of ones.

2. Evaluate the likelihood function at each of the points defined by $\left[C' \quad \boldsymbol{b}_i \otimes \mathbf{1}_{100}\right]'$, for all $i = 1, \cdots, 100$. Sort the results in decreasing order and store the top s values. Then proceed with the first set of parameters.

3. Let $\boldsymbol{\theta}_0 := (\theta_{0,1}, \cdots, \theta_{0,\kappa})'$ be the current vector of starting values for optimization. We look for a local maxima in the domain $[\theta_{0,1} - r, \theta_{0,1} + r] \times \cdots \times [\theta_{0,\kappa} - r, \theta_{0,\kappa} + r]$, where r can either depend on the values of $\boldsymbol{\theta}_0$ or be fixed exogenously. For simplicity, we set $r = r_1$. However, it is important to note that some parameters may have restricted parameter spaces, such as λ, which must be positive. In such cases, the bounds must satisfy the restrictions on the parameter space.

4. To guarantee the existence of unconditional probabilities, which are defined by $\pi = (A'A)^{-1}A'\begin{bmatrix}\mathbf{0}_N \\ 1\end{bmatrix}$, where $A = \begin{bmatrix}I_N - \mathcal{P} \\ \mathbf{1}'_N\end{bmatrix}$, with $\mathcal{P}$ denoting the transition matrix for the extended states[4], and $\mathbf{0}_N$ the null vector in $\mathbb{R}^N$, ensure that the reciprocal condition number of the matrix $(A'A)$ is above the machine precision. In practice, a small value such as 10^{-9} is sufficient to virtually eliminate numerical issues caused by a near singular transition matrix.

5. Use a derivative based numerical optimization method to find the maximum of a nonlinear, multivariate function with bounds, as specified in 3 and nonlinear constraints as defined in 4.

6. The optimization in step 5 is considered successful if the first-order optimality measure is close to zero and the proposed solution does not approach the boundary defined in step 3. We evaluate the proximity of the proposed solution to the bounds by computing the absolute percentage difference. This percentage difference should be above a specified threshold. More specifically, let
$$\ell_i^- := \frac{|\theta_{1,i} - (\theta_{0,i} - r)|}{|\theta_{1,i}|}, \qquad \text{and} \qquad \ell_i^+ := \frac{|\theta_{1,i} - (\theta_{0,i} + r)|}{|\theta_{1,i}|},$$
for $i \in \{1, \cdots, \kappa\}$, where $\theta_{1,i}$ is the proposed value for the ith parameter. We say that the proximity criteria is met for a given threshold $\delta > 0$ if $\min\{\ell_1^-, \cdots, \ell_\kappa^-, \ell_1^+, \cdots, \ell_\kappa^+\} > \delta$. Note that some parameters have restricted spaces, such as $\lambda > 0$. In such cases, it is acceptable for the estimation to be close to the restrictions, and the percentage proximity criteria should not be calculated. We apply $\delta = 0.01$ in most cases.

7. If the first-order optimality measure or the proximity criteria are not satisfied:
   7.1. If the first-order optimality measure is not close to zero, repeat the optimization process by returning to step 2 and choosing the next starting value. Continue this process until the first-order optimality criterion is met.
   7.2. If the first-order optimality is close to zero but the proximity criteria is not satisfied, proceed to step 3 and adjust the value of r_1 to r_2, where, $r_2 > r_1$. If this second round of optimization still fails to satisfy the criteria, return to step 3 and use a much larger value of r, $r_3 > r_2$, only for those parameters that do not meet the criteria.
   7.3. Report the optimization as unsuccessful if all s stored initial values fail to simultaneously meet both the first-order optimality and the proximity criteria.

We use the algorithm described above to obtain maximum likelihood estimates for both the Aranda-Ordaz and Logit DDMS models. The optimization set-up is identical for both models, except in the logit case, where only the matrix of random numbers C is used for the search of starting points. It is worth noting that we employ the same matrix C for both models, ensuring that the starting points search is comparable across models.

[4]See Maheu and McCurdy (2000a) for details.

4 Monte Carlo Simulation

In this section, we perform a Monte Carlo simulation study to compare the proposed Aranda-Ordaz approach to the traditional logit case following the contributions of Maheu and McCurdy (2000a,b). More precisely, our analysis is divided into two model applications based on an empirically relevant set of parameters. In the first case, we analyze the in-sample state probabilities generated under duration uncertainty for both functions following a bull and bear market analysis as in Maheu and McCurdy (2000a). For the second case, our study structure is quite similar; however, we explore the out-of-sample context, considering a conditional variance specification following Maheu and McCurdy (2000b). All codes were written in Matlab by the authors, and are available upon request.

4.1 In-sample capabilities: the bull and bear market model

In the first Monte Carlo investigation, we focus on the in-sample predictive ability related to the transition probabilities of the proposed Aranda-Ordaz approach compared to the fixed link logit case. We consider the following model,

$$Y_t = \mu_0(1 - S_t) + \mu_1 S_t + \big((1 - S_t)\sigma_0 + S_t\sigma_1\big)Z_t, \quad Z_t \sim N(0, 1). \tag{3}$$

In choosing the simulation scenario, we consider parameters reflecting the bull and bear market stylized facts, similar to the real data application presented in Section 5.1. The bull (bear) market is characterized by positive (negative) mean returns and lower (higher) variance. For the transition probabilities parameters, we set $\gamma_2(i) > 0$ for $i = 0, 1$, and this reflects that the probability of staying in the bull (bear) market increases as the duration increases. This dependence structure can be interpreted as a momentum effect, as discussed in Maheu and McCurdy (2000a); Isogai et al. (2008); Shibata (2012), among others.

We consider model (3) with true duration $d_0 = 8$ and parameters $\mu_0 = -0.5$, $\mu_1 = 1.5$, $\sigma_0 = 6$ and $\sigma_1 = 2$. The transition probabilities are given by (1) with parameters $\gamma_1(0) = -1.8$, $\gamma_2(0) = 0.7$, $\gamma_1(1) = -0.8$ and $\gamma_2(1) = 0.6$.[5] We estimate the model with duration $d \in \{4, 6, 8, 10, 12\}$ using both, the proposed Aranda-Ordaz and the fixed logit link. We generate time series of length 1,000 and discarded the first 200 observations as burn-in, yielding a final sample size of $n = 800$. The experiment was replicated 1,000 times. For each time series we fit model (3) using the logit and Aranda-Ordaz links. For each approach, we obtain the predictive, filtered and smoothed probabilities, denoted respectively by $P(S_t = i|\varphi_{t-1})$, $P(S_t = i|\varphi_t)$ and $P(S_t = i|\varphi_T)$ as in Kim and Nelson (1999).

Tables 1 and 2 show the simulation results. Table 1 presents the proportion of times the MAPE of the Aranda-Ordaz probability's MAPE is less than that of the logit's as well as the average MAPE difference for each probability type computed. Table 2 shows the proportion of times the likelihood obtained using the proposed Aranda-Ordaz approach is greater than the likelihood obtained with the logit. From Table 1, we observe that when the model is misspecified (i.e. $\tau \neq \tau_0$), the Aranda-Ordaz approach yields more precise probabilities more often than the logit. The more distant is τ from τ_0, the better the Aranda-Ordaz performs in comparison to the logit case. Under the correct specification, however, the logit link outperforms the Aranda-Ordaz by a narrow margin, presenting the smallest of all MAPE differences, and a slightly superior proportion for smaller MAPE. On the other hand, the results presented in Table 2 shows, on the other hand, that the Aranda-Ordaz produces a higher likelihood in the vast majority of cases, even in the correctly specified scenario.

[5]The parameters value scale refers to log returns multiplied by 100.

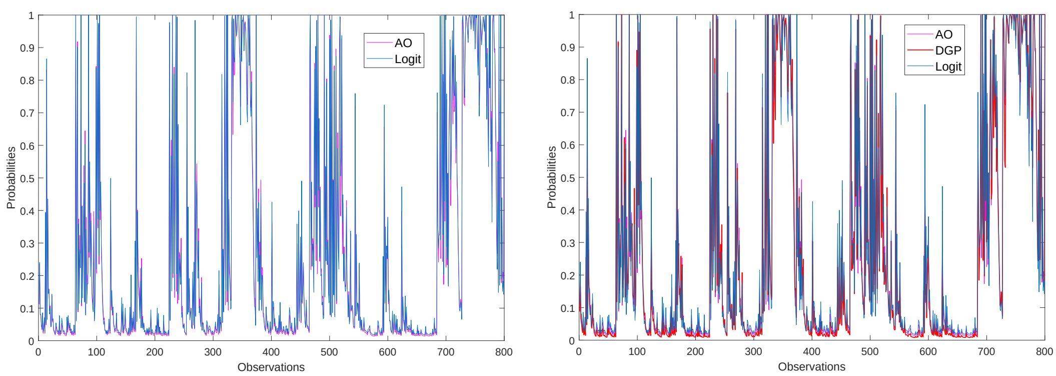
Table 1: Proportion of times in which the MAPE obtained with the Aranda-Ordaz is smaller than the logit's and the difference between the respective average MAPE.

Probabilities	Proportion					Difference				
τ	4	6	8	10	12	4	6	8	10	12
$P(S_t = i\|\varphi_{t-1})$	69%	52%	43%	60%	78%	0.0245	0.0057	-0.0042	0.0218	0.0334
$P(S_t = i\|\varphi_t)$	69%	53%	45%	60%	77%	0.0265	0.0067	-0.0030	0.0453	0.0561
$P(S_t = i\|\varphi_T)$	79%	55%	41%	55%	72%	0.0622	0.0138	-0.0087	0.1547	0.1700

Table 2: Frequency at which the likelihood values obtained with the Aranda-Ordaz is superior to the one obtained with the logit.

DGP	Likelihood Value Superiority Frequency				
$\tau_0 = 8$	94% ($\tau = 4$)	96% ($\tau = 6$)	91% ($\tau = 8$)	88% ($\tau = 10$)	91% ($\tau = 12$)

As an illustration, Figure 2 presents the bear market filtered probabilities. For the 200th path over 1,000 replications, and $\tau = 4$ (misspecified case), both links depicted similar market phases with different probabilities in some periods, as see Figure 2(a). We also compare these estimates to the DGP in Figure 2(b). Graphically speaking, the model's performance follows the true probability values, with the accumulated advantage of the Aranda-Ordaz specification period by period. The blue line (logit link) is much more evident over the red line (DGP), revealing some difference at this sing path. However, results confirmed over the battery of time series support the gain of the Aranda-Ordaz for misspecified duration cases.



(a) Logit and Aranda-Ordaz

(b) DGP, Logit, and Aranda-Ordaz

Figure 2: Bear market filtered probabilities, $\tau = 4$.

4.2 Out-of-Sample capabilities: Volatility Forecasting

In this Section, we compare the proposed Aranda-Ordaz approach to the traditional logit approach in terms of out-of-sample capabilities. To provide grounds for comparison, we conduct a series of Monte Carlo simulations considering

$$Y_t = \sigma\big(S_t, D(S_t)\big) Z_t, \quad \text{with} \quad \sigma\big(S_t, D(S_t)\big) = \big(\omega(S_t) + \zeta(S_t) D(S_t)\big)^2, \tag{4}$$

where the latent state affects the level of volatility, $\omega(S_t) = \omega_0(1 - S_t) + \omega_1 S_t$, while the duration of the states, $D(S_t)$, affect the dynamics of volatility through $\zeta(S_t)D(S_t)$, where, $\zeta(S_t) = \zeta_0(1 - S_t) + \zeta_1 S_t$. Z_t is assumed to follow an identically and independently normal distribution, and transition probabilities for the states S_t are given by (1).

We generated time series of size $n + 10$, say $y_1, \cdots, y_{n+10}$ for values of $\tau_0 \in \{15, 25, 35\}$. The last 10 values were reserved to create the conditional variances necessary for out-of-sample forecasting purposes. For each scenario, we estimate the model using $y_1, \cdots, y_n$ with the logit and the Aranda-Ordaz link functions for $\tau \in \{\tau_0 - 10, \tau_0 - 5, \tau_0, \tau_0 + 5, \tau_0 + 10\}$. Next, for each estimated model we obtain h-steps ahead forecast for forecasting horizons $h \in \{1, \cdots, 10\}$. Let $\hat{\sigma}^2_{n+1}, \cdots, \hat{\sigma}^2_{n+10}$ and $\tilde{\sigma}^2_{n+1}, \cdots, \tilde{\sigma}^2_{n+10}$ denote the forecasted values of $\sigma^2_{n+1}, \cdots, \sigma^2_{n+10}$ using the Aranda-Ordaz and the logit, respectively. For each horizon $h \in \{1, \cdots, 10\}$ and each method, we calculate the forecasting mean absolute percentage error (MAPE), given by

$$\mathrm{MAPE}_{\mathrm{AO}}(h) = \frac{1}{h}\sum_{k=1}^{h}\left|\frac{\hat{\sigma}^2_{n+k} - \sigma^2_{n+k}}{\sigma^2_{n+k}}\right|, \qquad \text{and} \qquad \mathrm{MAPE}_{\mathrm{logit}}(h) = \frac{1}{h}\sum_{k=1}^{h}\left|\frac{\tilde{\sigma}^2_{n+k} - \sigma^2_{n+k}}{\sigma^2_{n+k}}\right|.$$

We replicate the experiment 1,000 times. To simplify the exposition, in Table 3 we present the difference between the average MAPE of the proposed Aranda-Ordaz approach and the logit, for each forecast horizon, that is, $D(h) := \mathrm{MAPE}_{\mathrm{logit}}(h) - \mathrm{MAPE}_{\mathrm{AO}}(h)$, for $h \in \{1, \cdots, 10\}$.

Table 3: Predictive ability of the logit and the Aranda-Ordaz links in a simulated variance forecasting exercise. Presented are the difference between the MAPE of the proposed Aranda-Ordaz approach and the logit. Positive values favor the Aranda-Ordaz specification while negative values stand for the logit case.

DGP	Models	Forecasting horizon (h)									
		1	2	3	4	5	6	7	8	9	10
$\tau_0 = 15$	$\tau = 5$	0.0232	0.0173	0.0151	0.0153	0.0146	0.0136	0.0128	0.0122	0.0115	0.0104
	$\tau = 10$	0.0749	0.0614	0.0533	0.0463	0.0407	0.0357	0.0323	0.0293	0.0264	0.0235
	$\tau = 15$	0.1295	0.1001	0.0840	0.0720	0.0637	0.0574	0.0516	0.0467	0.0430	0.0399
	$\tau = 20$	0.0328	0.0278	0.0236	0.0197	0.0163	0.0141	0.0121	0.0102	0.0093	0.0091
	$\tau = 25$	0.0482	0.0399	0.0319	0.0262	0.0217	0.0182	0.0155	0.0133	0.0114	0.0096
$\tau_0 = 25$	$\tau = 15$	0.4406	0.3994	0.3713	0.3505	0.3312	0.3129	0.2972	0.2833	0.2712	0.2608
	$\tau = 20$	0.0959	0.0909	0.0860	0.0829	0.0807	0.0788	0.0754	0.0724	0.0701	0.0683
	$\tau = 25$	0.1750	0.1578	0.1477	0.1399	0.1350	0.1293	0.1236	0.1188	0.1144	0.1106
	$\tau = 30$	0.4244	0.3862	0.3602	0.3375	0.3180	0.3020	0.2874	0.2728	0.2600	0.2481
	$\tau = 35$	0.2614	0.2354	0.2168	0.2056	0.1960	0.1876	0.1798	0.1726	0.1661	0.1594
$\tau_0 = 35$	$\tau = 25$	0.6613	0.6415	0.6260	0.6135	0.6018	0.5902	0.5806	0.5723	0.5666	0.5623
	$\tau = 30$	1.3123	1.2725	1.2461	1.2205	1.1912	1.1630	1.1386	1.1173	1.0966	1.0771
	$\tau = 35$	1.1140	1.0849	1.0577	1.0303	1.0026	0.9816	0.9653	0.9517	0.9371	0.9227
	$\tau = 40$	1.1296	1.0663	1.0188	0.9727	0.9341	0.9062	0.8848	0.8684	0.8528	0.8353
	$\tau = 45$	1.0551	1.0083	0.9653	0.9238	0.8887	0.8585	0.8373	0.8161	0.7945	0.7707

Table 3 only contains positive values indicating that, on average, forecasting using the Aranda-Ordaz is advantageous regardless the true value of τ_0 and the duration τ used in the estimation procedure. It is interesting to notice that the smallest values of the difference $D(h)$ are usually not obtained when $\tau = \tau_0$, which may be a consequence of model complexity, indicating that even when the correct duration is used, the model based on the logit is not capable to provide the "best possible" forecasts. Regarding the magnitude of τ_0, the higher the duration, the bigger the difference $D(h)$ for all forecasting horizons. The smallest differences $D(h)$ are obtained for $\tau_0 = 15$, $\tau = 5$, ranging between about 10% and 23%, while the overall higher are obtained for $\tau_0 = 35$ and $\tau = 30$, ranging from over 100% to a bit over 130%.

The results on Table 3 show that, on average, applying the Aranda-Ordaz link is advantageous over the logit even when the duration and the link function are correctly specified. One question that remains is how often (if at all) the likelihood of the proposed Aranda-Ordaz approach is higher than the logit. To shed some light into this issue, under the same DGP as in Section 4.2, we compare the likelihood obtained using the proposed Aranda-Ordaz approach and the logit.

Table 4 presents the frequency at which the log-likelihood obtained using the Aranda-Ordaz link is superior to the one obtained with the logit. The first column presents the true value τ_0 applied in the data generating process, while the value of τ used in the estimation procedure is displayed in parenthesis. We observe that in all scenarios considered in the simulation, using the Aranda-Ordaz link yields superior likelihood in average, including the case where the link and duration are correctly specified. This result is expected since the Aranda-Ordaz link is more flexible than the logit. The frequency at which the likelihood is greater for the Aranda-Ordaz ranges from 62% to 91%. Interestingly, the higher the true value of τ, the greater this frequency is.

Table 4: Frequency at which the Aranda-Ordaz link likelihood value is superior to the one obtained with the logit link in 1,000 simulated trials.

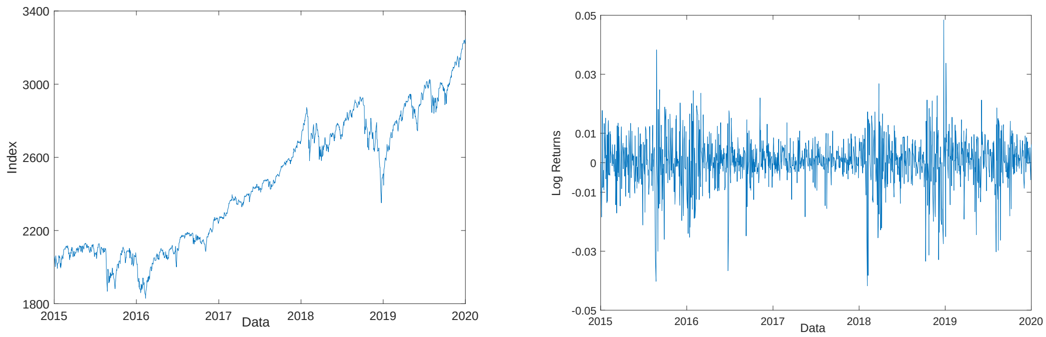
DGP	Log-Likelihood Value Superiority Frequency				
$\tau_0 = 15$	62% ($\tau = 5$)	70% ($\tau = 10$)	63% ($\tau = 15$)	62% ($\tau = 20$)	68% ($\tau = 25$)
$\tau_0 = 25$	89% ($\tau = 15$)	76% ($\tau = 20$)	78% ($\tau = 25$)	87% ($\tau = 30$)	68% ($\tau = 35$)
$\tau_0 = 35$	87% ($\tau = 25$)	91% ($\tau = 30$)	88% ($\tau = 35$)	82% ($\tau = 40$)	84% ($\tau = 45$)

5 Empirical Exercise

5.1 S&P500

In this section, we present a real data application of the proposed methodology considering the log returns of daily closing S&P500 index from January 2nd 2015 to January 2nd 2020, yielding a sample size $n = 1{,}259$, as seen in Figure 3. The descriptive statistics is presented in Table 5. We observe a mean very close to zero and a small standard deviation, with an annualized value of 0.1344 ($0.0085 \times \sqrt{250}$). The Jarque-Bera (JB) normality test rejects the null hypothesis of an underlying normal distribution, which is further corroborated by the kurtosis, considerably higher than the normal distribution's. The Lagrange Multiplier and Ljung-Box tests are highly significant, suggesting ARCH effects in the log returns.

Our goal is to conduct an out-of-sample forecasting exercise considering a 1-day step ahead volatility forecast through an expanding window using 443 daily observations, starting on April 1st, 2018. During this period, the S&P500 presented evidence of intra-state fluctuations, such as bear rallies (positive sub-trend) and bull corrections (negative sub-trend), encompassing a



(a) S&P500 Index

(b) S&P500 Log Returns

Figure 3: Time series plot of the S&P500 (a) Index and (b) Log Returns

Table 5: Summary statistics of the S&P500 daily log-returns. 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 1% significance level.

n	Mean	Std. Dev.	Max.	Min.	Skewness	Kurtosis	JB	LM(8)	Q_8^2
1259	0.0004	0.0085	0.0484	-0.0418	-0.5260	6.8360	829.9*	1899.8*	373.8*

primary bull market state as described by the empirical literature (see, for example, Maheu et al., 2021, for stock market cycles). We use this period of uncertainty and volatility to run a point forecasting exercise to compare the performance of a single Aranda-Ordaz DDMS to benchmark models.

5.2 Realised measures

A well-known problem in financial econometrics is that volatility is a latent variable and, therefore, cannot be directly observed, turning the evaluation of the predictive power of different approaches problematic. In this section, we consider the 5-minute intra-day quotes as an alternative source to proxy volatility. More specifically, besides the traditional realised variance (RV; see Andersen et al., 2003), we also consider alternative realised measures robust to jump and microstructure noise to proxy the true volatility, namely the bipower variation (BV; see Barndorff-Nielsen and Shephard, 2004), MinRV and MedRV (Andersen et al., 2012).

In general, these measures are regarded as better alternatives to proxy the true volatility than the squared daily return (see, for example, McAleer and Medeiros, 2008; Alizadeh et al., 2002, among others.), which are defined as:

$$\begin{aligned}
\mathrm{RV}_t &:= \sum_{i=1}^{m} r_{i,t}^2, \\
\mathrm{BV}_t &:= \frac{\pi}{2}\left[\frac{m}{m-1}\right]\sum_{i=1}^{m-1} |r_{i,t} r_{i+1,t}|, \\
\mathrm{MinRV}_t &:= \frac{\pi}{\pi-2}\left[\frac{m}{m-1}\right]\sum_{i=1}^{m-1} \min\big\{|r_{i,t}|, |r_{i+1,t}|\big\}^2, \\
\mathrm{MedRV}_t &:= \frac{\pi}{6-4\sqrt{3}+\pi}\left[\frac{m}{m-2}\right]\sum_{i=2}^{m-1} \mathrm{med}\big\{|r_{i-1,t}|, |r_{i,t}|, |r_{i+1,t}|\big\}^2,
\end{aligned}$$

where, $r_{i,t}$ is the ith high-frequency return of day t with $m = 78$ cases. The 5 minutes intraday quotes range from 9.30 AM to 4.00 PM, and the time series were obtained from First Rate Data website https://firstratedata.com.

5.3 Robust loss function

Typically, volatility proxies are used to evaluate volatility forecasts. However, these proxies are estimates of the integrated variance and, as such, they are imperfect. Patton (2011) defines a sense of robustness for loss functions in ranking volatility forecasts, and based on this concept, derived a general class of loss functions that are robust in that sense. Letting $\bar{\sigma}^2$ the volatility proxy and $\hat{\sigma}^2$ the volatility forecasts, we consider three loss functions to evaluate volatility forecasts that are members of Patton (2011)'s class, namely, the MSE, the QLIKE, and a measure denote hereafter by RLF, given by

$$\begin{aligned}
&\mathrm{MSE}(\bar{\sigma}^2, \hat{\sigma}^2) := \frac{(\bar{\sigma}^2 - \hat{\sigma}^2)^2}{2}, \quad \mathrm{QLIKE}(\bar{\sigma}^2, \hat{\sigma}^2) := \frac{\bar{\sigma}^2}{\hat{\sigma}^2} + \log\left(\frac{\bar{\sigma}^2}{\hat{\sigma}^2}\right) - 1, \\
&\mathrm{RLF}(\bar{\sigma}^2, \hat{\sigma}^2) := \hat{\sigma}^2 - \bar{\sigma}^2\left[\log\left(\frac{\bar{\sigma}^2}{\hat{\sigma}^2}\right) - 1\right].
\end{aligned}$$

Observe that Patton (2011)'s MSE and QLIKE losses differ from the usually applied loss functions of the same name.

5.4 Results

We consider the volatility model given by (4), which was already considered in the Monte Carlo simulation study. To conduct reliable evaluation between Logit and Aranda-Ordaz links, we also take the same setup for starting point parameters to optimize the model's likelihood functions. The study is constructed using pairwise Diebold-Mariano-West test.

5.4.1 Single Models

Table 6 presents the t-statistics from Diebold-Mariano-West tests of equal predictive accuracy for the benchmark logit model for $\tau \in \{5, 10, 15, 20, 25\}$ and the Aranda-Ordaz approach with $\tau = 15$ considering three loss functions, MSE, RLF and QLIKE, and four realised variances, MedRV, MinRV, BV and RV. A positive t-statistic indicates that the benchmark model forecast produced larger average loss than the Aranda-Ordaz link. A t-statistic greater than 1.65 and 1.96 in absolute value indicates a rejection of the null of equal predictive accuracy at the 0.10 and 0.05 levels. These statistics are marked with one and two asterisk, respectively.

For the MSE and RLF loss functions, the t-statistic is positive for all pairwise models evaluation. The null hypothesis of equal predictive accuracy is rejected for most cases. The only exception occurs for $\tau = 15$ and partially for $\tau = 20$. Considering the latter, the null is rejected only for the RLF loss case. The t-statistic for the QLIKE is negative, meaning the Aranda-Ordaz link produces a larger average loss than the logit, and this result is observed for all cases. However, there is no statistical evidence of difference in predictive ability between the links. In general, we observe the flexibility of the Aranda-Ordaz over different logit duration setups attenuating at some point, the choice of the duration uncertainty nature.

Table 6: Results from the Diebold-Mariano-West test applied to single models. Presented are the t-statistics from Diebold-Mariano-West test along with their p-value in parentheses.

Models		Loss	MedRV		MinRV		BV		RV	
A-O $\tau = 15$	Logit $\tau = 5$	MSE	2.34	(0.019)	2.36	(0.018)	2.35	(0.019)	2.32	(0.020)
		RLF	2.19	(0.029)	2.64	(0.008)	2.29	(0.022)	1.93	(0.054)
		QLIKE	-0.87	(0.384)	-0.81	(0.418)	-0.86	(0.390)	-0.90	(0.368)
	Logit $\tau = 10$	MSE	2.02	(0.043)	2.02	(0.043)	2.02	(0.043)	2.02	(0.043)
		RLF	2.62	(0.009)	2.67	(0.008)	2.64	(0.008)	2.60	(0.009)
		QLIKE	-0.77	(0.441)	-0.68	(0.497)	-0.76	(0.447)	0.81	(0.418)
	Logit $\tau = 15$	MSE	1.31	(0.190)	1.32	(0.187)	1.31	(0.190)	1.30	(0.194)
		RLF	1.17	(0.242)	1.58	(0.114)	1.25	(0.211)	0.95	(0.342)
		QLIKE	-0.87	(0.384)	-0.81	(0.418)	-0.86	(0.390)	-0.89	(0.373)
	Logit $\tau = 20$	MSE	1.07	(0.285)	1.07	(0.285)	1.08	(0.280)	1.08	(0.280)
		RLF	1.77	(0.077)	1.69	(0.091)	1.75	(0.080)	1.81	(0.070)
		QLIKE	-0.79	(0.430)	-0.70	(0.484)	-0.77	(0.441)	-0.83	(0.407)
	Logit $\tau = 25$	MSE	2.41	(0.016)	2.46	(0.014)	2.43	(0.015)	2.42	(0.016)
		RLF	2.72	(0.007)	3.46	(< 0.001)	2.89	(0.004)	2.36	(0.018)
		QLIKE	-0.73	(0.465)	-0.62	(0.535)	-0.71	(0.478)	-0.78	(0.435)

5.4.2 Combination Models

In this section, we consider a model combination approach to aggregate N individual model forecasts, each indexed by a fixed duration, into a pooled model scheme to incorporate duration choice uncertainty. Let, $\hat{\sigma}^2_{t+1}$ be the weighted average of the N individual volatility forecasts models $\{\hat{\sigma}^2_{i,t+1}\}_{i=1}^N$, that is

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

where $\{w_{i,t}\}_{i=1}^N$ are the combining ex-ante weights at time t and N is taken over a set of DDMS models restricted by lower (upper) bound of τ. We consider the naïve method of taking uniform weights $w_{i,t} := \frac{1}{N}$ and the discount mean square prediction error (DMSPE) combining method (Stock and Watson, 2004), whose weight are given by

$$w_{i,t} = \left[\varphi_{i,t}\sum_{j=1}^{N}\varphi_{j,t}^{-1}\right]^{-1}, \qquad \varphi_{i,t} = \sum_{s=m+1}^{t}\theta^{t-s}(\hat{\sigma}^2_s - \hat{\sigma}^2_{i,s})^2,$$

where θ is the discount factor and m is the observations contained in the subsample to estimate the models[6]. If $\theta = 1$, 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 our goal is to marginalize the gains of different duration's models. Table 7 displays the results obtained by evaluating the

Table 7: Results from the Diebold-Mariano-West test applied to the combination approach. Presented are the t-statistics from Diebold-Mariano-West test along with their p-value in parentheses.

Models		Loss	MedRV		MinRV		BV		RV	
A-O $\tau = 15$	Logit Fixed Weight	MSE	2.36	(0.018)	2.30	(0.021)	2.34	(0.019)	2.39	(0.017)
		RLF	4.04	(< 0.001)	4.14	(< 0.001)	4.08	(< 0.001)	3.69	(< 0.001)
		QLIKE	-0.76	(0.447)	-0.67	(0.503)	-0.75	(0.453)	-0.81	(0.418)
	Logit Moving Weight $\theta = 0.9$	MSE	2.30	(0.021)	2.37	(0.018)	2.32	(0.020)	2.28	(0.023)
		RLF	2.23	(0.026)	3.05	(0.002)	2.42	(0.016)	1.87	(0.062)
		QLIKE	-0.79	(0.430)	-0.71	(0.478)	-0.78	(0.435)	-0.84	(0.401)
	Logit Moving Weight $\theta = 1.0$	MSE	2.69	(0.007)	2.61	(0.009)	2.70	(0.007)	2.74	(0.006)
		RLF	3.22	(0.001)	4.64	(< 0.001)	3.52	(< 0.001)	2.48	(0.013)
		QLIKE	-0.78	(0.435)	-0.69	(0.490)	-0.77	(0.441)	-0.83	(0.407)

logit model combination approach to assess the effectiveness of the Aranda-Ordaz link. Similar to the findings in Table 6, a positive t-statistic is observed for all pairwise model evaluations in the cases of MSE and RFL loss functions. For both the moving and fixed weight approaches, the null hypothesis of equal predictive accuracy is rejected, implying that the flexible link outperforms the pooling approach. Moreover, the Aranda-Ordaz link yields a greater average loss than the logit combination, as indicated by the negative t-statistic for QLIKE. However, no statistical evidence was found to support a difference in predictive performance.

5.4.3 Garch-type models

Our analysis also include the traditional Garch-type models in a pairwise evaluation. Based on Haas et al. (2004), we use a plain vanilla specification, which is given by

$$Y_t = \sigma_{k,t} Z_t,$$
$$\sigma^2_{k,t} = \omega_k + \alpha_k Y^2_{t-1} + \beta_k \sigma^2_{k,t-1}$$

where k is the number of regimes. We considered $k = 1$, the traditional model, and $k = 2$, the two regime model. In both models Z_t's are i.i.d. $N(0, 1)$. Despite many variations of this modeling approach, our objective is to compare the performance of the links relative to a simple and useful model.

Upon examination of the results presented in Table 8, it is observed that a positive t-statistic implies that the DDMS models yield larger average losses in comparison to Garch-type models. However, the null hypotheses of equal predictivity between the Aranda-Ordaz and Garch-type models cannot be rejected based on the MSE and QLIKE losses, indicating that these models are equally applicable. It is important to note that this evidence does not hold for the logit cases. For the single model, the null hypothesis is rejected for QLIKE, with

[6]We apply $m = 30$ days.

Table 8: Results from the Diebold-Mariano-West test applied to the Garch-type models. Presented are the t-statistics from Diebold-Mariano-West test along with their p-value in parentheses.

Models		Loss	MedRV		MinRV		BV		RV	
Garch	A-O $\tau = 15$	MSE	0.16	(0.873)	-0.24	(0.810)	0.08	(0.936)	0.51	(0.610)
		RLF	1.68	(0.093)	1.89	(0.059)	1.75	(0.080)	1.62	(0.105)
		QLIKE	1.06	(0.289)	1.08	(0.280)	1.06	(0.289)	1.05	(0.294)
	Logit $\tau = 15$	MSE	1.31	(0.190)	1.28	(0.201)	1.30	(0.194)	1.33	(0.184)
		RFL	3.25	(0.001)	3.04	(0.002)	3.18	(0.002)	3.35	(0.001)
		QLIKE	6.40	(< 0.001)	6.09	(< 0.001)	6.25	(< 0.001)	6.65	(< 0.001)
	Logit Fixed Weight	MSE	2.38	(0.017)	2.34	(0.019)	2.37	(0.018)	2.41	(0.016)
		RFL	4.47	(< 0.001)	4.37	(< 0.001)	4.39	(< 0.001)	4.38	(< 0.001)
		QLIKE	10.23	(< 0.001)	10.07	(< 0.001)	10.10	(< 0.001)	10.13	(< 0.001)
MS-Garch	A-O $\tau = 15$	MSE	0.82	(0.412)	0.16	(0.873)	0.65	(0.516)	1.23	(0.219)
		RFL	1.85	(0.064)	2.21	(0.027)	1.94	(0.052)	1.74	(0.082)
		QLIKE	1.06	(0.289)	1.08	(0.280)	1.06	(0.289)	1.06	(0.289)
	Logit $\tau = 15$	MSE	1.36	(0.173)	1.32	(0.187)	1.35	(0.177)	1.37	(0.171)
		RFL	3.49	(< 0.001)	3.29	(< 0.001)	3.42	(< 0.001)	3.56	(< 0.001)
		QLIKE	6.38	(< 0.001)	6.09	(< 0.001)	6.22	(< 0.001)	6.57	(< 0.001)
	Logit Fixed Weight	MSE	2.40	(0.016)	2.36	(0.018)	2.39	(0.017)	2.42	(0.016)
		RFL	4.45	(< 0.001)	4.36	(< 0.001)	4.37	(< 0.001)	4.37	(< 0.001)
		QLIKE	10.22	(< 0.001)	10.01	(< 0.001)	10.05	(< 0.001)	10.25	(< 0.001)

higher values for the MSE t-statistics. In the case of the logit model combination, this result is even more pronounced, as evidenced by the null rejection by all three losses, highlighting the advantages of the Aranda-Ordaz link in the context of volatility forecasting.

6 Conclusion

This paper proposes a methodology to address the issue of duration choice in DDMS models. The methodology involves the application of a parametric link function in place of the typical fixed link function to calculate transition probabilities. This parametric link function results in likelihood values and transition probabilities that are more accurate. The proposed approach is capable of significantly improving forecasting accuracy, especially in cases of duration misspecification.

Two Monte Carlo simulations, based on classical applications of DDMS models, are employed to evaluate the methodology. The results demonstrate that using the Aranda-Ordaz link function leads to more precise forecasts and transition probabilities, not only in cases of duration misspecification but also when the model is correctly specified. Furthermore, an empirical study is conducted to forecast the volatility of the S&P500, which illustrates the effectiveness of the proposed methodology. The results indicate that the Aranda-Ordaz DDMS outperforms fixed logit transitions in terms of forecast precision for most duration parameters. Moreover, the Aranda-Ordaz link function improves the forecasting performance to such an extent that it is equivalent to MS-Garch models. Notably, such an improvement in forecasting accuracy is not observed when using fixed link functions.

References

Alizadeh, S., Brandt, M. W., and Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, 57(3):1047–1091.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

Andersen, T. G., Dobrev, D., and Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, 169(1):75–93.

Aranda-Ordaz, F. J. (1981). On two families of transformations to additivity for binary response data. *Biometrika*, 68(2):357–363.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1):1–37.

Bejaoui, A. and 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.

Durland, J. M. and McCurdy, T. H. (1994). Duration-dependent transitions in a Markov model of U.S. GNP growth. *Journal of Business & Economic Statistics*, 12(3):279–288.

Haas, M., Mittnik, S., and Paolella, M. S. (2004). A new approach to Markov-Switching GARCH models. *Journal of Financial Econometrics*, 2(4):493–530.

Hamilton, J. (1994). *Time series analysis*. Princeton University Press, Princeton, New Jersey.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.

Isogai, A., Kanoh, S., and Tokunaga, T. (2008). A further extension of duration-dependent models. *The European Journal of Finance*, 14(5):427–449.

Kim, C.-J. and Nelson, C. R. (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. The MIT Press.

Lam, P.-s. (2004). A Markov-switching model of GNP growth with duration dependence. *International Economic Review*, 45(1):175–204.

Maheu, J. M. and McCurdy, T. H. (2000a). Identifying Bull and Bear markets in stock returns. *Journal of Business & Economic Statistics*, 18(1):100–112.

Maheu, J. M. and McCurdy, T. H. (2000b). Volatility dynamics under duration-dependent mixing. *Journal of Empirical Finance*, 7(3):345–372.

Maheu, J. M., McCurdy, T. H., and Song, Y. (2021). Bull and bear markets during the COVID-19 pandemic. *Finance Research Letters*, 42:102091.

McAleer, M. and Medeiros, M. C. (2008). Realized volatility: A review. *Econometric Reviews*, 27(1-3):10–45.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):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.

Shibata, M. (2012). Identifying bull and bear markets in Japan. *Asia-Pacific Financial Markets*, 19(2):99–117.

Stock, J. H. and Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6):405–430.