Daily Volatility Forecasting with Off-the-Shelf Models: A Comparative Study Under Stress

Carlos Trucíos and Felipe Marques

Universidade Estadual de Campinas (UNICAMP).

26º Simpósio Nacional de Probabilidade e Estatística

ctruciosm.github.io
ctrucios@unicamp.br

Table of contents

Motivation

Motivation

1. **Benchmarks:** If we truly want practitioners to adopt the new procedure we are proposing, it must outperform strong benchmarks. It is very difficult to convince people from other fields to use our methods if they believe they offer no real advantage.

Motivation

1. **Benchmarks:** If we truly want practitioners to adopt the new procedure we are proposing, it must outperform strong benchmarks. It is very difficult to convince people from other fields to use our methods if they believe they offer no real advantage.
2. **High-Dimensional Settings:** Even in high-dimensional settings, we still need appropriate univariate volatility models in the first stage of the procedure.

Motivation

1. **Benchmarks:** If we truly want practitioners to adopt the new procedure we are proposing, it must outperform strong benchmarks. It is very difficult to convince people from other fields to use our methods if they believe they offer no real advantage.
2. **High-Dimensional Settings:** Even in high-dimensional settings, we still need appropriate univariate volatility models in the first stage of the procedure.
3. **ML in Social Media:** Many of these works focus mainly on "horse races" comparisons between models, sometimes with little or no understanding of how the procedures actually work.

Introduction

Introduction

- Forecasting volatility plays a crucial role in a wide range of applications across economics, finance, insurance, energy, climate, etc.

Introduction

- Forecasting volatility plays a crucial role in a wide range of applications across economics, finance, insurance, energy, climate, etc.
- Given its importance, numerous approaches have been developed for forecasting daily volatility.

Introduction

- Forecasting volatility plays a crucial role in a wide range of applications across economics, finance, insurance, energy, climate, etc.
- Given its importance, numerous approaches have been developed for forecasting daily volatility.
- While several options benefit researchers and experienced practitioners, they pose significant challenges for (untrained) practitioners, who must choose among these models for their daily tasks, often with limited or no information to guide their decisions.

Main goals and contribution

- Offer practical guidance for researchers and practitioners in choosing the most suitable volatility model for the problem at hand, using accessible, user-friendly code.

Introduction

Main goals and contribution

- Offer practical guidance for researchers and practitioners in choosing the most suitable volatility model for the problem at hand, using accessible, user-friendly code.
- Identify a strong benchmark model, which can also be used as a solid first step (univariate) for High-Dimensional volatility modelling.

Introduction

Main goals and contribution

- Offer practical guidance for researchers and practitioners in choosing the most suitable volatility model for the problem at hand, using accessible, user-friendly code.
- Identify a strong benchmark model, which can also be used as a solid first step (univariate) for High-Dimensional volatility modelling.

Main goals and contribution

- Offer practical guidance for researchers and practitioners in choosing the most suitable volatility model for the problem at hand, using accessible, user-friendly code.
- Identify a strong benchmark model, which can also be used as a solid first step (univariate) for High-Dimensional volatility modelling.

Scope of the Study

While we acknowledge alternative approaches for forecasting daily volatility, mainly those based on high-frequency data, this study focuses on off-the-shelf models that rely solely on daily data.

Models

Models

- **GARCH**: Generalized Autoregressive Conditional Heteroskedasticity, introduced by Bollerslev (1986).

Models

- **GARCH**: Generalized Autoregressive Conditional Heteroskedasticity, introduced by Bollerslev (1986).
- **FIGARCH**: Fractionally Integrated Generalised Autoregressive Conditional Heteroscedasticity, introduced by Baillie et al. (1996).

Models

- **GARCH**: Generalized Autoregressive Conditional Heteroskedasticity, introduced by Bollerslev (1986).
- **FIGARCH**: Fractionally Integrated Generalised Autoregressive Conditional Heteroscedasticity, introduced by Baillie et al. (1996).
- **SV**: (vanilla) Stochastic Volatility model, proposed by Taylor (1982).

Models

- **GARCH**: Generalized Autoregressive Conditional Heteroskedasticity, introduced by Bollerslev (1986).
- **FIGARCH**: Fractionally Integrated Generalised Autoregressive Conditional Heteroscedasticity, introduced by Baillie et al. (1996).
- **SV**: (vanilla) Stochastic Volatility model, proposed by Taylor (1982).
- **MSGARCH**: Markov-Switching GARCH proposed by Haas et al. (2004), which allows for multiple volatility regimes.

Models

- **GARCH**: Generalized Autoregressive Conditional Heteroskedasticity, introduced by Bollerslev (1986).
- **FIGARCH**: Fractionally Integrated Generalised Autoregressive Conditional Heteroscedasticity, introduced by Baillie et al. (1996).
- **SV**: (vanilla) Stochastic Volatility model, proposed by Taylor (1982).
- **MSGARCH**: Markov-Switching GARCH proposed by Haas et al. (2004), which allows for multiple volatility regimes.
- **GAS**: Generalized Autoregressive Score model, introduced independently by Creal et al. (2013) and Harvey (2013).

Why these models?

They combine widespread empirical adoption with ready availability across multiple software environments.

Models

Why these models?

They combine widespread empirical adoption with ready availability across multiple software environments.

Let

$$r_t = (P_t - P_{t-1})/P_{t-1} \approx \log(P_t/P_{t-1})$$

denote the return at time t, where P_t represents the closing price at time t. We assume $\mathbb{E}(r_t \mid \mathcal{F}_{t-1}) = 0$.

GARCH model

Assumes that the conditional variance at time t is fully determined by past squared returns and its own past values. In its simplest form, the model is specified as:

$$r_t = \sigma_t \epsilon_t, \tag{1}$$

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{2}$$

where $\omega > 0$ and $\alpha, \beta \geq 0$ are model parameters, σ_t^2 represents the conditional variance (or squared volatility) at time t, and the innovation term ϵ_t has zero-mean and unit-variance.

Models

FIGARCH model

Extends the GARCH framework by capturing long-memory behaviour in volatility. In the FIGARCH(1, d, 1) model, the volatility equation is given by:

$$\sigma_t^2 = \omega + (\phi - \beta) r_{t-1}^2 + \beta \sigma_{t-1}^2 + \sum_{j=1}^{\infty} \frac{d\Gamma(j-d)}{\Gamma(j+1)\Gamma(1-d)} (r_{t-j}^2 - \phi r_{t-j-1}^2), \tag{3}$$

where ω, ϕ, β, and d are model parameters, with $d \in [0, 1]$ governing the degree of long memory.

Models

SV model

Assumes that the log-conditional variance evolves stochastically following an AR(1) process. Its dynamics can be described as follows:

$$r_t = \exp(h_t/2)\epsilon_t, \tag{4}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sigma\eta_t, \tag{5}$$

where h_t is the log conditional variance at time t, μ, ϕ and σ are parameters to be estimated, $\eta_t \sim N(0,1)$.

MSGARCH model

Extends single-regime GARCH to K volatility regimes via an unobserved state variable S_t with transition probabilities following a first-order Markov chain: $\pi_{ij} = \mathbb{P}(S_t = j \mid S_{t-1} = i)$.

The Haas et al. (2004) specification with $K = 2$ regimes, avoids path dependence by tracking regime-specific conditional variances:

$$r_t = \sigma_t^{(k)} \epsilon_t, \tag{6}$$

$$\sigma_t^{2(k)} = \omega^{(k)} + \alpha^{(k)} r_{t-1}^2 + \beta^{(k)} \sigma_{t-1}^{2(k)}, \quad k \in \{1, 2\} \tag{7}$$

where $\omega^{(k)} > 0$ and $\alpha^{(k)}, \beta^{(k)} \geq 0$.

Models

GAS model

Time-varying parameters $\boldsymbol{\theta}_t$ are driven by the score of the conditional density function $p(r_t; \boldsymbol{\theta}_t)$. In its general form:

$$\boldsymbol{\theta}_{t+1} = \kappa + A s_t + B\boldsymbol{\theta}_t, \quad \text{where } s_t = \underbrace{S_t(\boldsymbol{\theta}_t)}_{\mathcal{I}_t(\boldsymbol{\theta}_t)^{-\gamma}} \underbrace{\nabla_t(r_t, \boldsymbol{\theta}_t)}_{\text{score}}$$

GAS model

Time-varying parameters $\boldsymbol{\theta}_t$ are driven by the score of the conditional density function $p(r_t; \boldsymbol{\theta}_t)$. In its general form:

$$\boldsymbol{\theta}_{t+1} = \kappa + As_t + B\boldsymbol{\theta}_t, \quad \text{where } s_t = \underbrace{S_t(\boldsymbol{\theta}_t)}_{\mathcal{I}_t(\boldsymbol{\theta}_t)^{-\gamma}} \underbrace{\nabla_t(r_t, \boldsymbol{\theta}_t)}_{\text{score}}$$

A particular case based on the Student-t distribution is given by:

$$\log(\dot{\sigma}_{t+1}) = \tilde{\kappa} + \tilde{\alpha}\left[\frac{(\nu+1)r_t^2}{\nu\dot{\sigma}_t^{\,2} + r_t^2} - 1\right] + \tilde{\beta}\log(\dot{\sigma}_t), \tag{8}$$

where ν stands for the degrees of freedom of the Student-t distribution. This model is known as Beta-t-EGARCH of Harvey and Chakravarty (2008).

Monte Carlo Simulations

Monte Carlo Simulations

Simulation setup

- The five models previously described are used both as true DGPs and for generating one-step-ahead volatility forecasts

Monte Carlo Simulations

Simulation setup

- The five models previously described are used both as true DGPs and for generating one-step-ahead volatility forecasts
- Sample sizes: $(n = 500, 1000, 2500)$

Monte Carlo Simulations

Simulation setup

- The five models previously described are used both as true DGPs and for generating one-step-ahead volatility forecasts
- Sample sizes: $(n = 500, 1000, 2500)$
- Innovation distributions: standardized Normal and Student-t_7

Simulation setup

- The five models previously described are used both as true DGPs and for generating one-step-ahead volatility forecasts
- Sample sizes: $(n = 500, 1000, 2500)$
- Innovation distributions: standardized Normal and Student-t_7
- Parameter values were selected to reflect characteristics of both emerging and developed markets as well as to reflect values commonly adopted in the literature.

Monte Carlo Simulations

Simulation setup

- The five models previously described are used both as true DGPs and for generating one-step-ahead volatility forecasts
- Sample sizes: $(n = 500, 1000, 2500)$
- Innovation distributions: standardized Normal and Student-t_7
- Parameter values were selected to reflect characteristics of both emerging and developed markets as well as to reflect values commonly adopted in the literature.
- Series are contaminated with a single additive outlier in the final month

Monte Carlo Simulations

Simulation setup

- The five models previously described are used both as true DGPs and for generating one-step-ahead volatility forecasts
- Sample sizes: $(n = 500, 1000, 2500)$
- Innovation distributions: standardized Normal and Student-t_7
- Parameter values were selected to reflect characteristics of both emerging and developed markets as well as to reflect values commonly adopted in the literature.
- Series are contaminated with a single additive outlier in the final month
- 1,000 Monte Carlo replications

Estimation:

- Parameters are estimated by Maximum Likelihood.

Estimation:

- Parameters are estimated by Maximum Likelihood.
- For SV, parameters are estimated by a Bayesian MCMC approach.

Estimation:

- Parameters are estimated by Maximum Likelihood.
- For SV, parameters are estimated by a Bayesian MCMC approach.
- In all cases, we are interested in $\sigma^2_{T+1} = \mathbb{V}(r_{T+1}|\mathcal{F}_T)$, where $\mathcal{F}_T$ is the information available up to time T

Monte Carlo Simulations

Observation:

Prior to our main simulation, we evaluated competing R packages in a preliminary Monte Carlo study using the Diebold–Mariano test (Diebold and Mariano, 1995).

- **GAS:** `GAS` (Ardia et al., 2019b) vs. `betategarch` (Sucarrat, 2013) shows equal predictive accuracy across all sample sizes.

Monte Carlo Simulations

Observation:

Prior to our main simulation, we evaluated competing R packages in a preliminary Monte Carlo study using the Diebold–Mariano test (Diebold and Mariano, 1995).

- **GAS:** `GAS` (Ardia et al., 2019b) vs. `betategarch` (Sucarrat, 2013) shows equal predictive accuracy across all sample sizes.
- **MSGARCH:** `FitML()` vs. `FitMCMC()` (Ardia et al., 2019a) yields equal accuracy for large samples, but favours `FitML()` in small samples.

Monte Carlo Simulations

Observation:

Prior to our main simulation, we evaluated competing R packages in a preliminary Monte Carlo study using the Diebold–Mariano test (Diebold and Mariano, 1995).

- **GAS:** `GAS` (Ardia et al., 2019b) vs. `betategarch` (Sucarrat, 2013) shows equal predictive accuracy across all sample sizes.
- **MSGARCH:** `FitML()` vs. `FitMCMC()` (Ardia et al., 2019a) yields equal accuracy for large samples, but favours `FitML()` in small samples.
- **SV:** `stochvol` (Kastner, 2016) vs. `stochvolTMB` (Wahl, 2025) performs equally in large samples, but `stochvolTMB` exhibits frequent convergence failures in smaller samples.

Monte Carlo Simulations

Model	Parameter values 1	Parameter values 2
GARCH	$\omega = 0.18, \alpha = 0.09, \beta = 0.89$	$\omega = 0.37, \alpha = 0.14, \beta = 0.77$
FIGARCH	$\omega = 0.08, \phi = 0.2, \beta = 0.5, d = 0.4$	$\omega = 0.21, \phi = 0.15, \beta = 0.3, d = 0.3$
GAS	$\tilde{\kappa} = 0.04, \tilde{\beta} = 0.95, \tilde{\alpha} = 0.05$	$\tilde{\kappa} = 0.01, \tilde{\beta} = 0.80, \tilde{\alpha} = 0.10$
SV	$\mu = 1, \phi = 0.95, \sigma_h = 0.15$	$\mu = 1.5, \phi = 0.9, \sigma_h = 0.20$
MSGARCH	$\omega_1 = 0.005, \alpha_1 = 0.025, \beta_1 = 0.95$ $\omega_2 = 0.1, \alpha_2 = 0.25, \beta_2 = 0.70$ $P = \begin{bmatrix} 0.75 & 0.30 \\ 0.25 & 0.70 \end{bmatrix}$	$\omega_1 = 0.02, \alpha_1 = 0.03, \beta_1 = 0.93$ $\omega_2 = 0.20, \alpha_2 = 0.10, \beta_2 = 0.80$ $P = \begin{bmatrix} 0.95 & 0.10 \\ 0.05 & 0.90 \end{bmatrix}$

Table 1: Parameter values used in the data-generating processes.

Monte Carlo Simulations

Table 2: One-step-ahead squared volatility forecasting performance measured by MSE. Shaded cells indicate best-performing models according to the 75% MCS. DGP: GARCH.

DGP	Model	Normal									Student-t								
		Uncontaminated series			One additive outlier one month before			One additive outlier one week before			Uncontaminated series			One additive outlier one month before			One additive outlier one week before		
		500	1000	2500	500	1000	2500	500	1000	2500	500	1000	2500	500	1000	2500	500	1000	2500
GARCH	GARCH-N	2.344	1.178	0.373	13.015	8.970	6.785	319.274	324.534	318.670	3.343	2.281	0.588	14.024	11.762	8.009	361.461	369.344	362.271
	GARCH-T	2.421	1.222	0.379	9.673	7.951	6.567	278.359	307.652	316.499	3.352	1.968	0.542	10.322	8.912	7.253	313.738	337.345	350.449
	FIGARCH-N	3.339	2.897	1.943	19.240	17.063	13.584	384.186	426.769	449.714	6.711	4.516	3.969	22.653	19.595	17.003	428.713	473.458	499.663
	FIGARCH-T	3.358	2.897	1.998	17.514	16.362	13.561	373.596	430.544	451.341	5.070	4.135	3.173	18.639	17.842	15.653	412.454	470.124	506.026
	GAS-N	2.162	1.255	1.073	33.244	45.454	120.992	548.184	650.375	815.775	12.268	5.402	3.923	50.343	67.926	4845.224	671.572	752.936	6310.940
	GAS-T	2.231	1.143	0.805	5.787	3.550	3.494	79.420	89.359	148.213	7.972	6.590	6.402	10.573	7.065	6.484	53.138	35.673	31.520
	SV-N	2.106	1.766	1.732	3.289	3.379	3.648	107.579	148.655	160.356	8.322	8.475	8.125	8.167	7.862	7.605	69.457	100.918	140.882
	SV-T	2.511	2.049	1.945	3.139	2.935	3.051	48.927	69.377	97.142	8.269	8.301	8.539	7.885	8.056	7.020	29.685	33.006	38.779
	MS-N	3.764	2.369	1.280	11.388	11.449	11.669	238.699	341.235	397.056	10.687	5.938	2.884	14.300	8.720	5.572	153.560	188.893	204.737
	MS-T	3.602	2.285	1.143	9.923	11.173	11.642	159.322	260.841	349.349	11.112	7.117	3.481	14.571	12.157	9.351	151.144	161.214	209.294

Table 3: One-step-ahead squared volatility forecasting performance measured by MSE. Shaded cells indicate best-performing models according to the 75% MCS. DGP: FIGARCH.

DGP	Model	Normal									Student-t								
		Uncontaminated series			One additive outlier one month before			One additive outlier one week before			Uncontaminated series			One additive outlier one month before			One additive outlier one week before		
		500	1000	2500	500	1000	2500	500	1000	2500	500	1000	2500	500	1000	2500	500	1000	2500
FIGARCH	GARCH-N	0.159	0.127	0.094	0.666	0.501	0.485	22.463	22.331	24.091	0.380	0.227	0.198	0.948	0.843	0.995	28.295	30.046	37.482
	GARCH-T	0.163	0.130	0.095	0.476	0.446	0.481	16.779	20.362	23.907	0.332	0.192	0.183	0.871	0.783	1.049	24.919	30.117	37.259
	FIGARCH-N	0.346	0.229	0.142	1.477	1.263	0.997	26.750	28.074	27.572	0.533	0.382	0.258	1.811	1.536	1.389	34.819	36.350	37.841
	FIGARCH-T	0.356	0.268	0.157	1.422	1.372	1.066	23.397	26.874	28.036	0.540	0.382	0.232	1.756	1.534	1.257	30.227	34.676	36.151
	GAS-N	0.187	0.145	0.125	1.227	1.783	9.611	38.745	45.482	65.129	0.476	0.711	0.734	2.206	3.570	41.843	48.164	49.904	120.976
	GAS-T	0.174	0.125	0.114	0.252	0.224	0.269	4.908	5.880	10.582	0.630	0.392	0.322	0.829	0.518	0.377	5.602	4.489	3.958
	SV-N	0.182	0.156	0.157	0.191	0.213	0.287	5.364	8.521	11.514	0.344	0.276	0.393	0.427	0.289	0.328	5.165	7.109	13.872
	SV-T	0.215	0.172	0.169	0.212	0.202	0.261	2.872	4.354	6.995	0.441	0.308	0.399	0.454	0.284	0.306	3.643	3.705	4.291
	MS-N	0.262	0.201	0.146	0.656	0.719	0.845	14.871	21.616	23.430	0.791	0.711	0.444	1.213	1.011	0.861	13.453	16.514	16.949
	MS-T	0.273	0.208	0.149	0.622	0.680	0.754	10.686	15.160	19.314	0.835	0.874	0.804	1.212	1.399	1.216	11.240	13.295	17.701

Monte Carlo Simulations

Table 4: One-step-ahead squared volatility forecasting performance measured by MSE. Shaded cells indicate best-performing models according to the 75% MCS. DGP: GAS.

DGP	Model	Normal									Student-t								
		Uncontaminated series			One additive outlier one month before			One additive outlier one week before			Uncontaminated series			One additive outlier one month before			One additive outlier one week before		
		500	1000	2500	500	1000	2500	500	1000	2500	500	1000	2500	500	1000	2500	500	1000	2500
GAS	GARCH-N	0.051	0.032	0.021	0.163	0.085	0.048	5.400	5.196	4.889	0.138	0.091	0.052	0.399	0.245	0.138	7.449	7.063	6.698
	GARCH-T	0.052	0.033	0.021	0.109	0.070	0.046	4.322	4.641	4.705	0.159	0.075	0.046	0.307	0.175	0.122	5.613	5.971	6.258
	FIGARCH-N	0.099	0.073	0.049	0.335	0.247	0.166	6.728	7.109	6.495	0.283	0.212	0.134	0.767	0.682	0.558	10.358	10.744	10.185
	FIGARCH-T	0.098	0.080	0.054	0.298	0.226	0.164	6.292	6.722	6.122	0.257	0.192	0.134	0.721	0.668	0.579	9.315	9.973	10.072
	GAS-N	0.030	0.015	0.007	0.248	0.212	0.228	10.076	11.830	13.472	0.138	0.127	0.146	0.804	0.414	0.351	11.008	11.641	11.438
	GAS-T	0.032	0.018	0.007	0.067	0.042	0.024	0.925	1.201	2.307	0.079	0.038	0.014	0.119	0.055	0.023	0.783	0.621	0.538
	SV-N	0.065	0.052	0.044	0.063	0.051	0.043	1.251	1.688	1.850	0.106	0.074	0.055	0.117	0.077	0.052	0.594	0.800	1.051
	SV-T	0.078	0.062	0.050	0.079	0.061	0.048	0.701	0.958	1.265	0.098	0.061	0.044	0.088	0.046	0.027	0.362	0.410	0.458
	MS-N	0.067	0.050	0.025	0.168	0.128	0.089	3.717	5.165	5.793	0.166	0.106	0.057	0.214	0.144	0.097	2.206	2.287	2.767
	MS-T	0.069	0.051	0.028	0.160	0.147	0.099	2.601	4.103	5.124	0.144	0.085	0.052	0.215	0.125	0.096	2.313	2.049	2.128

Monte Carlo Simulations

Table 5: One-step-ahead squared volatility forecasting performance measured by MSE. Shaded cells indicate best-performing models according to the 75% MCS. DGP: SV.

DGP	Model	Normal									Student-t								
		Uncontaminated series			One additive outlier one month before			One additive outlier one week before			Uncontaminated series			One additive outlier one month before			One additive outlier one week before		
		500	1000	2500	500	1000	2500	500	1000	2500	500	1000	2500	500	1000	2500	500	1000	2500
SV	GARCH-N	1.420	1.308	1.226	2.430	1.963	1.717	24.271	22.149	20.415	2.372	2.004	1.797	3.297	2.698	2.332	20.477	17.130	15.515
	GARCH-T	1.408	1.307	1.222	2.057	1.835	1.667	16.582	18.484	19.177	2.266	2.004	1.804	3.002	2.563	2.306	14.283	14.134	14.552
	FIGARCH-N	1.664	1.569	1.508	3.863	3.772	3.976	34.900	34.728	32.740	3.287	2.587	2.321	5.412	4.667	4.615	31.486	29.313	25.552
	FIGARCH-T	1.678	1.587	1.499	3.713	3.711	4.201	27.727	29.787	32.480	3.258	2.598	2.508	5.308	4.566	5.146	25.305	24.686	26.047
	GAS-N	1.404	1.299	1.221	2.541	2.406	2.354	33.335	35.001	36.294	2.208	1.981	1.841	3.030	2.605	2.435	23.172	21.353	19.369
	GAS-T	1.431	1.275	1.195	1.581	1.386	1.296	4.194	4.143	4.478	1.895	1.648	1.489	2.094	1.775	1.564	3.982	3.172	2.633
	SV-N	1.376	1.245	1.178	1.431	1.302	1.247	3.946	5.856	7.650	1.905	1.847	1.789	1.953	1.882	1.796	2.538	2.268	2.010
	SV-T	1.443	1.266	1.191	1.952	1.292	1.236	2.844	3.451	4.487	1.842	1.666	1.553	1.812	1.649	1.506	2.317	2.033	1.834
	MS-N	1.391	1.329	1.233	1.923	1.667	1.557	11.538	12.202	11.942	2.085	1.769	1.662	2.437	1.971	1.848	6.257	6.033	6.813
	MS-T	1.370	1.368	1.249	1.811	1.732	1.625	9.536	9.402	9.768	2.133	1.805	1.607	2.629	2.113	1.833	10.402	8.374	6.931

Table 6: One-step-ahead squared volatility forecasting performance measured by MSE. Shaded cells indicate best-performing models according to the 75% MCS. DGP: MSGARCH.

DGP	Model	Normal									Student-t								
		Uncontaminated series			One additive outlier one month before			One additive outlier one week before			Uncontaminated series			One additive outlier one month before			One additive outlier one week before		
		500	1000	2500	500	1000	2500	500	1000	2500	500	1000	2500	500	1000	2500	500	1000	2500
MSGARCH	GARCH-N	0.019	0.010	0.007	0.055	0.020	0.009	1.136	0.991	0.851	0.039	0.021	0.013	0.068	0.037	0.016	1.310	1.158	1.022
	GARCH-T	0.021	0.011	0.007	0.041	0.018	0.009	0.755	0.773	0.778	0.038	0.022	0.012	0.059	0.037	0.018	0.990	1.002	0.996
	FIGARCH-N	0.038	0.030	0.023	0.110	0.095	0.090	1.279	1.317	1.384	0.079	0.051	0.035	0.129	0.124	0.105	1.449	1.482	1.456
	FIGARCH-T	0.039	0.033	0.025	0.104	0.104	0.099	1.046	1.198	1.351	0.078	0.059	0.043	0.160	0.150	0.132	1.340	1.441	1.582
	GAS-N	0.020	0.018	0.015	0.060	0.042	0.043	2.165	2.648	3.627	0.109	0.060	0.143	0.122	0.108	0.203	4.509	4.048	5.973
	GAS-T	0.016	0.012	0.010	0.022	0.015	0.011	0.145	0.137	0.159	0.041	0.029	0.025	0.046	0.032	0.025	0.138	0.094	0.075
	SV-N	0.027	0.024	0.026	0.025	0.021	0.024	0.156	0.220	0.274	0.047	0.044	0.041	0.046	0.041	0.040	0.097	0.094	0.104
	SV-T	0.033	0.029	0.030	0.032	0.027	0.027	0.117	0.148	0.196	0.049	0.048	0.046	0.048	0.047	0.043	0.075	0.069	0.072
	MS-N	0.027	0.016	0.007	0.055	0.041	0.018	0.546	0.591	0.602	0.057	0.029	0.012	0.060	0.036	0.018	0.425	0.259	0.288
	MS-T	0.026	0.022	0.012	0.058	0.046	0.031	0.438	0.543	0.605	0.047	0.042	0.013	0.053	0.062	0.020	0.457	0.438	0.401

Monte Carlo Simulations

- Results using QLIKE lead to the same conclusions.
- Results across alternative parameter settings remain consistent.
- Over 20 parameter settings were evaluated (excluding the FIGARCH DGP), yielding to similar conclusions.

Monte Carlo Simulations

- In general, for the single-regime and no-long-memory DGPs, in the absence of outliers, the forecasting model that matches the DGP tends to yield the best (or at least among the best according to the MCS) forecasting performance, being the performance better as the sample size increases.

Monte Carlo Simulations

- In general, for the single-regime and no-long-memory DGPs, in the absence of outliers, the forecasting model that matches the DGP tends to yield the best (or at least among the best according to the MCS) forecasting performance, being the performance better as the sample size increases.
- An exception arises in some scenarios involving the SV model with Student-t innovations, where the GAS model with Student-t distribution performs equal or even better.

Monte Carlo Simulations

- In general, for the single-regime and no-long-memory DGPs, in the absence of outliers, the forecasting model that matches the DGP tends to yield the best (or at least among the best according to the MCS) forecasting performance, being the performance better as the sample size increases.
- An exception arises in some scenarios involving the SV model with Student-t innovations, where the GAS model with Student-*t* distribution performs equal or even better.
- For the FIGARCH and MSGARCH DGPs, the forecasting procedure matching the DGP does not necessarily provide the best forecasting performance. To perform well, we need a large number of observations in the time series.

Monte Carlo Simulations

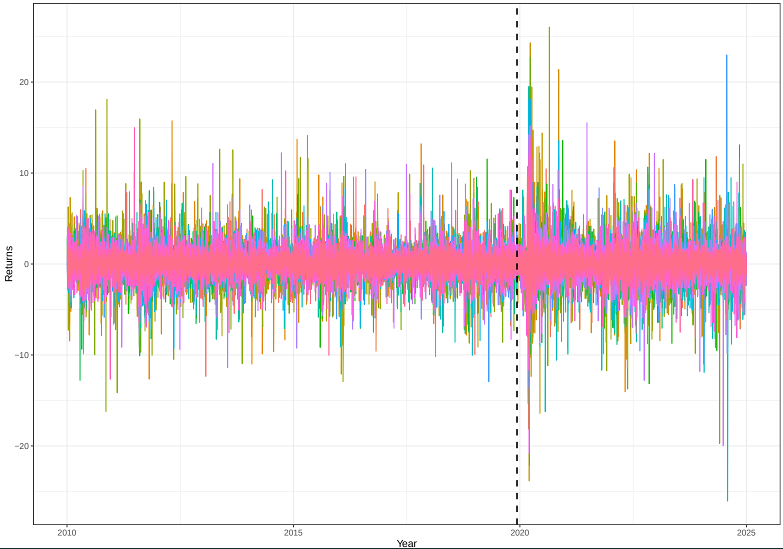
- In general, for the single-regime and no-long-memory DGPs, in the absence of outliers, the forecasting model that matches the DGP tends to yield the best (or at least among the best according to the MCS) forecasting performance, being the performance better as the sample size increases.
- An exception arises in some scenarios involving the SV model with Student-t innovations, where the GAS model with Student-t distribution performs equal or even better.
- For the FIGARCH and MSGARCH DGPs, the forecasting procedure matching the DGP does not necessarily provide the best forecasting performance. To perform well, we need a large number of observations in the time series.
- In the presence of extreme observations, the SV model (and sometimes also the GAS) under the Student-t innovation assumption emerge as best-performing model

Empirical Application

Empirical Application

- Daily return series of the constituents of the Dow Jones Industrial Average Index.
- The data, obtained from Economatica, span the period from January 4, 2010, to December 31, 2024, totalling 3,774 observations.
- Only stocks fully traded in the period of analysis were considered, ending up in 29 stocks.
- Windos size equal to 2500 days
- One-step-ahead forecast squared volatility is compared with the 5-min realized variance and bi-power-variation (obtained from the CaPiRe database)

Empirical Application



	RV										BPV									
	GARCH-N	GARCH-T	FIGARCH-N	FIGARCH-T	GAS-N	GAS-T	MS-N	MS-T	SV-N	SV-T	GARCH-N	GARCH-T	FIGARCH-N	FIGARCH-T	GAS-N	GAS-T	MS-N	MS-T	SV-N	SV-T
MMM	0.340	0.347	0.372	0.353	0.374	0.285	0.291	0.297	0.285	0.261	0.363	0.370	0.401	0.379	0.397	0.301	0.274	0.321	0.278	0.260
AMZN	0.274	0.272	0.304	0.285	0.284	0.272	0.294	0.268	0.210	0.210	0.302	0.299	0.333	0.314	0.309	0.301	0.275	0.297	0.229	0.230
AXP	0.319	0.308	0.340	0.327	0.336	0.308	0.301	0.307	0.256	0.252	0.340	0.326	0.366	0.350	0.356	0.330	0.278	0.328	0.260	0.262
AMGN	0.251	0.266	0.262	0.290	0.277	0.248	0.338	0.261	0.248	0.254	0.255	0.270	0.270	0.304	0.281	0.250	0.292	0.260	0.235	0.236
AAPL	0.285	0.296	0.308	0.314	0.290	0.308	0.305	0.292	0.247	0.252	0.321	0.333	0.346	0.352	0.326	0.346	0.302	0.329	0.273	0.280
BA	0.235	0.257	0.281	0.282	0.221	0.257	0.521	0.261	0.228	0.224	0.256	0.282	0.312	0.312	0.241	0.284	0.465	0.287	0.244	0.242
CAT	0.233	0.254	0.277	0.284	0.240	0.252	0.258	0.250	0.203	0.208	0.259	0.283	0.307	0.315	0.267	0.279	0.241	0.278	0.217	0.224
CVX	0.237	0.246	0.257	0.265	0.237	0.239	0.302	0.223	0.219	0.215	0.255	0.266	0.280	0.288	0.255	0.259	0.283	0.241	0.232	0.228
CSCO	0.339	0.315	0.350	0.333	0.385	0.289	0.306	0.315	0.262	0.244	0.374	0.350	0.392	0.372	0.422	0.321	0.310	0.348	0.276	0.259
KO	0.189	0.183	0.229	0.222	0.214	0.195	0.268	0.184	0.209	0.191	0.200	0.194	0.248	0.242	0.227	0.206	0.279	0.194	0.203	0.187
HD	0.194	0.200	0.217	0.223	0.203	0.203	0.261	0.197	0.199	0.189	0.207	0.215	0.237	0.244	0.216	0.218	0.238	0.210	0.202	0.193
HON	0.222	0.233	0.246	0.244	0.227	0.229	0.263	0.223	0.210	0.204	0.234	0.246	0.263	0.262	0.241	0.243	0.254	0.234	0.212	0.208
INTC	0.339	0.353	0.382	0.384	0.359	0.308	0.390	0.306	0.263	0.256	0.372	0.385	0.421	0.421	0.392	0.337	0.348	0.334	0.268	0.271
IBM	0.282	0.297	0.340	0.354	0.309	0.269	0.274	0.288	0.249	0.223	0.314	0.329	0.375	0.393	0.343	0.298	0.285	0.320	0.260	0.238
JNJ	0.193	0.194	0.198	0.195	0.202	0.194	0.235	0.192	0.219	0.206	0.201	0.201	0.208	0.208	0.209	0.202	0.239	0.197	0.210	0.198
JPM	0.255	0.272	0.301	0.307	0.303	0.260	0.254	0.257	0.227	0.222	0.278	0.296	0.330	0.337	0.328	0.285	0.248	0.279	0.242	0.236
MCD	0.247	0.250	0.281	0.300	0.271	0.258	0.337	0.258	0.289	0.268	0.231	0.234	0.265	0.294	0.246	0.247	0.298	0.242	0.254	0.234
MRK	0.222	0.226	0.317	0.264	0.260	0.219	0.270	0.220	0.240	0.225	0.229	0.233	0.331	0.284	0.265	0.226	0.253	0.226	0.226	0.213
MSFT	0.264	0.277	0.322	0.289	0.279	0.285	0.274	0.263	0.234	0.232	0.294	0.308	0.358	0.321	0.309	0.317	0.275	0.292	0.255	0.255
NKE	0.326	0.332	0.442	0.444	0.382	0.305	0.324	0.335	0.281	0.258	0.352	0.357	0.482	0.485	0.407	0.334	0.315	0.368	0.290	0.273
PG	0.181	0.180	0.201	0.191	0.195	0.188	0.253	0.180	0.219	0.198	0.180	0.181	0.212	0.203	0.195	0.190	0.248	0.182	0.199	0.183
GS	0.227	0.230	0.265	0.270	0.251	0.219	0.251	0.228	0.195	0.195	0.246	0.249	0.288	0.294	0.272	0.238	0.239	0.250	0.210	0.208
TRV	0.209	0.230	0.233	0.239	0.230	0.211	0.250	0.221	0.224	0.198	0.216	0.240	0.252	0.256	0.241	0.223	0.232	0.233	0.217	0.196
UNH	0.234	0.251	0.267	0.273	0.243	0.230	0.294	0.238	0.239	0.216	0.248	0.266	0.287	0.293	0.260	0.240	0.266	0.251	0.228	0.209
VZ	0.320	0.315	0.329	0.341	0.354	0.302	0.385	0.320	0.354	0.321	0.318	0.316	0.343	0.358	0.353	0.301	0.368	0.322	0.327	0.301
V	0.304	0.308	0.309	0.305	0.307	0.305	0.350	0.296	0.296	0.282	0.316	0.320	0.326	0.321	0.322	0.316	0.342	0.308	0.291	0.279
WMT	0.312	0.307	0.324	0.362	0.356	0.262	0.328	0.277	0.291	0.255	0.313	0.300	0.337	0.382	0.350	0.262	0.301	0.280	0.268	0.232
DIS	0.307	0.312	0.344	0.348	0.318	0.275	0.324	0.289	0.261	0.239	0.333	0.337	0.375	0.380	0.345	0.299	0.314	0.315	0.268	0.251
CRM	0.311	0.329	0.315	0.348	0.401	0.288	0.318	0.289	0.238	0.227	0.342	0.362	0.349	0.384	0.435	0.320	0.302	0.321	0.259	0.249

Table 5: Out-of-sample average QLIKE forecasting performance for assets in the Dow Jones Industrial Average Index. The 5-minute realised variance (left panel) and realised bipower variation (right panel) are used as proxies for the latent variance. Shaded cells indicate models belonging to the 75% MCS.

Empirical Application

We extend our empirical analysis to forecast $\mathrm{VaR}^{2.5\%}_{T+1}$ and $\mathrm{ES}^{2.5\%}_{T+1}$:

- Out-of-sample evaluation relies on backtesting via calibration tests alongside scoring functions.
- Calibration tests used: Kupiec (1995), Christoffersen (1998), and Nolde and Ziegel (2017) and Bayer and Dimitriadis (2022) (following the selection guidelines of Hotta et al. (2024) and Yu et al. (2026)).
- Scoring functions: González-Rivera et al. (2004), Fissler and Ziegel (2016), Nolde and Ziegel (2017) and Taylor (2019)

Superior forecasting procedures are expected to pass the four calibration tests while remaining inside the Model Confidence Set (MCS) across all scoring functions.

Number of series passing all calibration tests and remaining within the MCS:

Model	No. of Series
GARCH-N	0/29
GARCH-T	10/29
FIGARCH-N	2/29
FIGARCH-T	18/29
GAS-N	0/29
GAS-T	18/29
SV-N	11/29
SV-T	20/29
MSGARCH-N	12/29
MSGARCH-T	21/29

Key Takeaways

- In small sample sizes, several models may deliver the same forecasting performance, even under misspecification.
- Matching the forecasting procedure to the true DGP does not always yield the best performance. In particular, caution is required with Markov-switching (MS) and long-memory models, which require large sample sizes.
- In the presence of extreme observations, SV-T outperforms its competitors, making it a strong benchmark when proposing outlier-robust procedures (surpassing standard GARCH alone is insufficient).

Acknowledgments

The authors gratefully acknowledge financial support from the São Paulo Research Foundation (FAPESP), under grants 2023/02538-0, 2023/01728-0, 2025/06131-7, and 2025/08502-2, as well as from the National Council for Scientific and Technological Development (CNPq), under grant 409630/2024-1. Computational resources were provided by the Centre for Applied Research on Econometrics, Finance and Statistics (CAREFS).

References

Ardia, D., Bluteau, K., Boudt, K., Catania, L., and Trottier, D.-A. (2019a). Markov-switching GARCH models in R: The MSGARCH package. *Journal of Statistical Software*, 91(4):1–38.

Ardia, D., Boudt, K., and Catania, L. (2019b). Generalized autoregressive score models in R: The GAS package. *Journal of Statistical Software*, 88(6):1–28.

Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30.

Bayer, S. and Dimitriadis, T. (2022). Regression-based expected shortfall backtesting. *Journal of Financial Econometrics*, 20(3):437–471.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):134–144.

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and osband's principle. *The Annals of Statistics*, pages 1680–1707.

González-Rivera, G., Lee, T.-H., and 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(4):629–645.

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.

Harvey, A. and Chakravarty, T. (2008). Beta-t-(E)GARCH. Cambridge Working Papers in Economics 08340, Faculty of Economics, Cambridge University.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: with Applications to Financial and Economic Time Series*, volume 52. Cambridge University Press.

Hotta, L., Trucíos, C., and Zevallos, M. (2024). A note about calibration tests for var and es. In Chiann, C., Pinheiro, A., and Toloi, C., editors, *Festschrift in Honor of Pedro A. Morettin: Time Series and Wavelets*. Springer, forthcoming.

Kastner, G. (2016). Dealing with stochastic volatility in time series using the R package stochvol. *Journal of Statistical Software*, 69(5):1–30.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2):73–84.

Nolde, N. and Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *The Annals of Applied Statistics*, 11(4):1833–1874.

Sucarrat, G. (2013). betategarch: Simulation, estimation and forecasting of Beta-Skew-t-EGARCH models. *The R Journal*, 5(2):137–147.

Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric laplace distribution. *Journal of Business & Economic Statistics*, 37(1):121–133.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75. *Time Series Analysis: Theory and Practice*, 1:203–226.

Wahl, J. (2025). *stochvolTMB: Likelihood Estimation of Stochastic Volatility Models*. R package version 0.3.0.

Yu, M. L., Trucíos, C., and Hotta, L. (2026). Outliers and the backtesting of risk measure forecasts: Calibration tests. Working paper.