Bovespa Index volatility forecasting

Bruno Tebaldi

Computational Methods in Empirical Finance
prof. Pedro Valls Pereira

November 28, 2019

1. Introduction
2. Methodology
    - Volatility Models
    - Forecast Evaluation
3. Data
    - Bovespa index history
    - Bovespa index series
4. Results
    - Daily returns model
    - Out-of-sample forecast
    - Normal vs Skewed t-Student
    - Forecast Window
    - Parameter re-estimation
    - Realized volatility comparison
    - E-GARCH Parameter Estimates
5. Conclusion

Introduction

- We follow the footsteps of Brownlees, Engle & Kelly; *A Practical Guide to Volatility Forecasting through Calm and Storm*
- We use of similar models, to the Bovespa Index.
- The models take into account several stylized fact of financial time series.

Volatility Models

The following models were analized:

- ARCH(1) model
- GARCH(1,1) model
- E-GARCH(1,1) model
- GJR-Garch(1,1) model
- EWMA model
- APARCH(1,1) model

For each model we make a comparison between Normal and Skewed t-Student, as well as different forecasting windows.

Two loss functions are used to evaluate the forecasting error.

- The mean squared error loss depends solely on the additive forecast error, $\hat{\sigma}_t^2 - h_{t|t-k}$.

$$MSE : L(\hat{\sigma}_t^2, h_{t|t-k}) = \left(\hat{\sigma}_t^2 - h_{t|t-k}\right)^2 \tag{1}$$

- The quasi-likelihood-based loss, named for its close relation to the Gaussian likelihood, depends only on the standardized residual $\hat{\sigma}_t^2 / h_{t|t-k}$.

$$QL : L(\hat{\sigma}_t^2, h_{t|t-k}) = \frac{\hat{\sigma}_t^2}{h_{t|t-k}} - \ln\left(\frac{\hat{\sigma}_t^2}{h_{t|t-k}}\right) \tag{2}$$

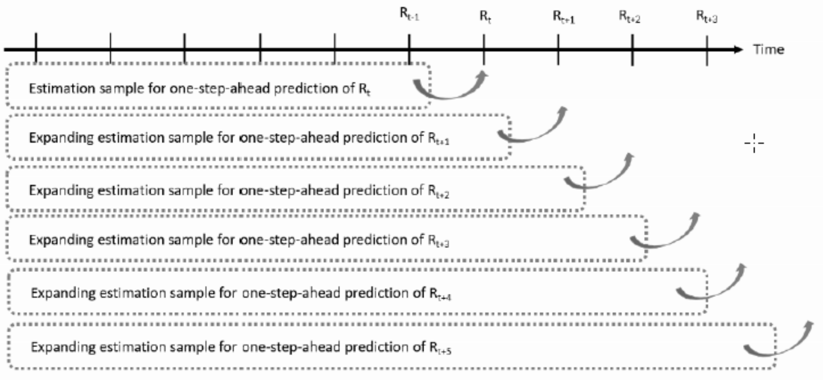
Forecast Window

Out-of-sample forecasts are typically computed using one of two methods:

1. Recursive (expanding window)
2. Rolling (moving window).

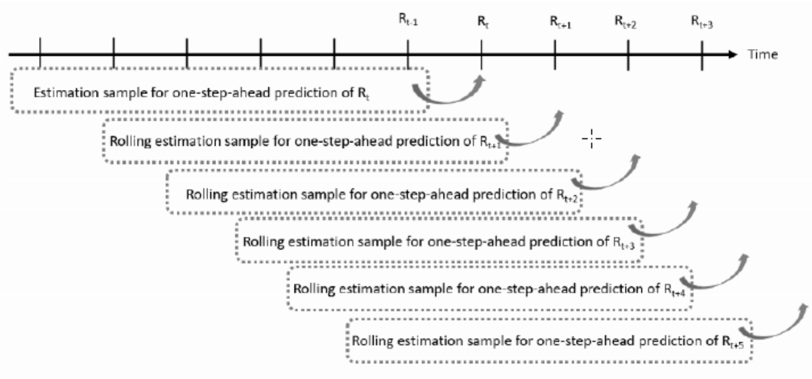
Forecast Window

Recursive (expanding window)



Forecast Window

Rolling (moving window)



Realized Volatility

The realized volatility are calculated using the following methodology:

- Denote by $p_{d,m}$ the log-price of the asset in the period m ($m = 1, \ldots, M$) of day d ($d = 1, \ldots, D$), the corresponding intraday return in the period d and m

$$r_{d,m} = p_{d,m} - p_{d,m-1} \qquad m = 2, \ldots, M \qquad d = 1, \ldots, D \tag{3}$$

$$r_{d,1} = p_{d,1} - p_{d-1,M} \tag{4}$$

- The realized variance for an asset in day d (VR_d) is then given by

$$VR_d = \begin{cases} \sum_{m=2}^{M} r_{d,m}^2 & \text{if day d-1 is excluded} \\ r_{d,1}^2 + \sum_{m=2}^{M} r_{d,m}^2 & \text{otherwise} \end{cases} \tag{5}$$

1 Introduction

2 Methodology
- Volatility Models
- Forecast Evaluation

3 Data
- Bovespa index history
- Bovespa index series

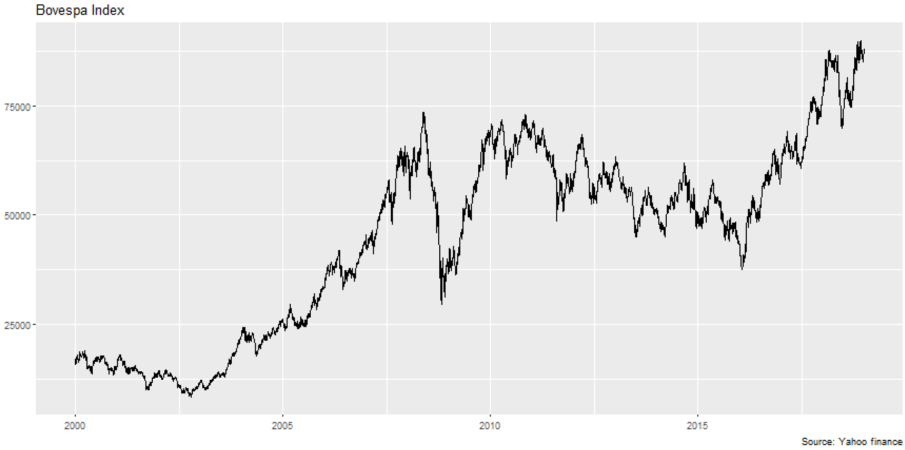
4 Results
- Daily returns model
- Out-of-sample forecast
- Normal vs Skewed t-Student
- Forecast Window
- Parameter re-estimation
- Realized volatility comparison
- E-GARCH Parameter Estimates

5 Conclusion

The Bovespa Index

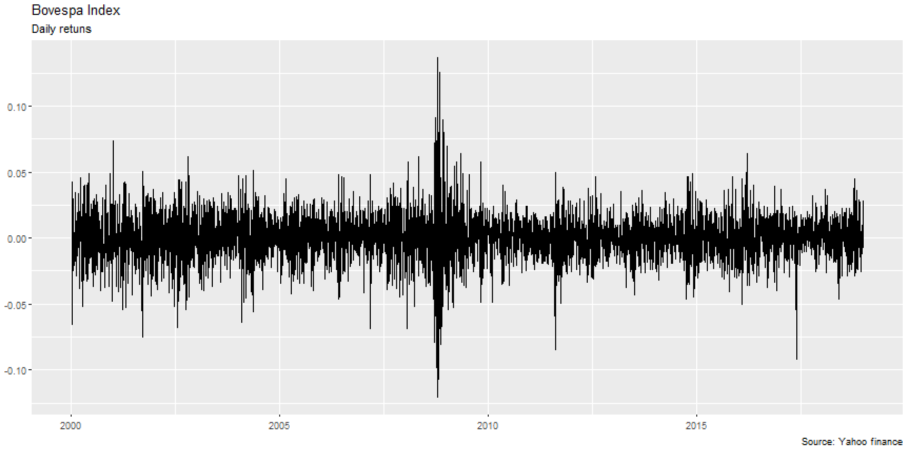
1. The Bovespa Index (Ibovespa) is the result of a theoretical portfolio of assets.
2. The Ibovespa's objective is to be the indicator of the average performance of quotations of the most tradable and representative assets of the Brazilian stock market.
3. The Index is composed of shares and units exclusively of shares of companies listed on the Bm&fBovespa that meet the inclusion
4. The Bovespa index suffered a methodology change on 11/09/2013

The Bovespa Index
Level



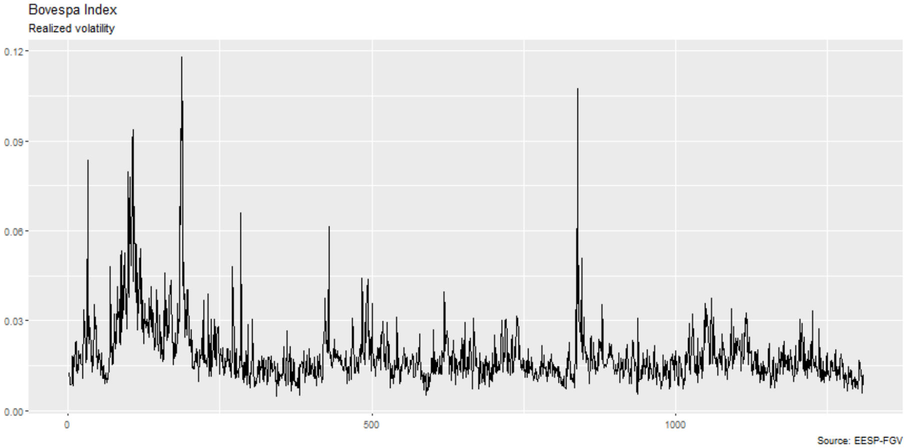
The Bovespa Index

Daily returns



The Bovespa Index High Frequency Data

The high frequency data is calculated for the period 1998-04-06 to 2003-08-03.



Holidays and other eventual gaps in the series were handdled by using a local level model with a Kalman filter.

	Date	Open	High	Low	Close	Adj. Close	Volume
Min.	2000-01-03	8.397	8.513	8.225	8.371	8.371	0
1st Qu.	2004-08-03	23.225	23.456	22.961	23.235	22.757	0
Median	2009-05-11	50.414	51.065	49.891	50.423	50.089	1.981.600
Mean	2009-05-27	44.812	45.287	44.342	44.826	44.465	7.266.875
3rd Qu.	2014-03-10	61.264	61.945	60.747	61.272	61.138	3.502.075
Max.	2018-12-28	89.820	91.242	89.429	89.820	89.820	232.265.300
NA's	84	84	84	84	84	0	84

1 Introduction
2 Methodology
- Volatility Models
- Forecast Evaluation

3 Data
- Bovespa index history
- Bovespa index series

4 Results
- Daily returns model
- Out-of-sample forecast
- Normal vs Skewed t-Student
- Forecast Window
- Parameter re-estimation
- Realized volatility comparison
- E-GARCH Parameter Estimates

5 Conclusion

Daily returns model

The following models are fitted to the daily returns. The selection of the model is based by the Akaike Information criteria.

Model	**Const.**	**AR**	**MA**	**log likelihood**	**AIC**
ARMA(1,1)	0.0004 (0.0003)	-0.2944 (0.3830)	0.3075 (0.3798)	12576.58	-25147.16
ARMA(1,0)	0.0003 (0.0003)	0.0097 (0.0145)	-	12576.36	-25148.72
ARMA(0,1)	0.0003 (0.0003)	-	0.0103 (0.0149)	12576.37	-25148.74
ARMA(0,0)	0.0003 (0.0003)	-	-	12576.13	-25150.27

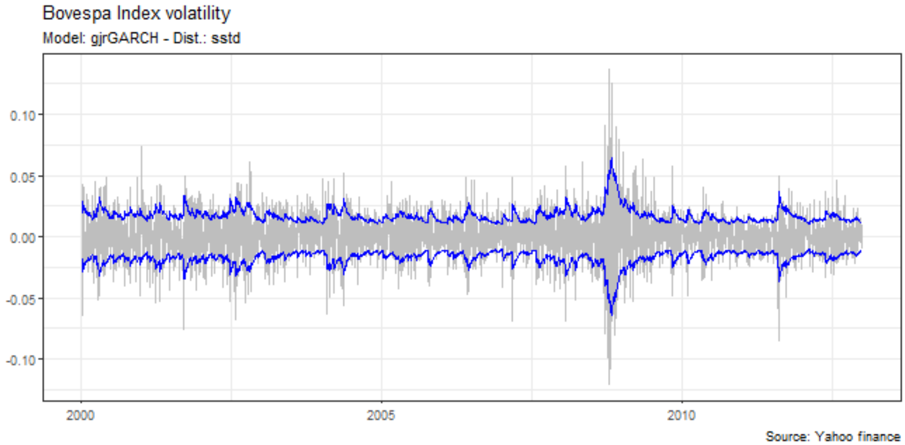
The ARMA(0,0) model will be used to model the mean of the returns.

In-Sample Fit

Model	Dist	Akaike	Bayes	Shibata	HannanQuinn
ARCH	Norm	-5.1607	-5.1552	-5.1607	-5.1588
ARCH	Sstd	-5.2325	-5.2233	-5.2325	-5.2292
GARCH	Norm	-5.3083	-5.3009	-5.3083	-5.3056
GARCH	Sstd	-5.3259	-5.3148	-5.3259	-5.3219
E-GARCH	Norm	-5.3252	-5.3159	-5.3252	-5.3219
E-GARCH	Sstd	-5.3413	-5.3283	-5.3413	-5.3366
GJR-Garch	Norm	-5.3295	-5.3202	-5.3295	-5.3262
GJR-Garch	Sstd	-5.3438^{*}	-5.3308^{*}	-5.3438^{*}	-5.3391^{*}
EWMA	Norm	-5.2947	-5.2910	-5.2947	-5.2934
EWMA	Sstd	-5.3185	-5.3111	-5.3185	-5.3159
APARCH	Norm	-5.3266	-5.3154	-5.3266	-5.3226
APARCH	Sstd	-5.3409	-5.3261	-5.3409	-5.3356

Table: Model Information Criteria

In-Sample Fit



Out-of-sample forecast

- All models are estimated using a Normal distribution as well as a Skewed t-Student distribution, using all 3 types of forecasting windows (expanding, moving 2 year window, moving 5 year window), with several parameters re-estimation strategies (5 days, 20 days, 60 days, 252 days).
- A total of 144 forecasting scenarios were analyzed.

Normal vs Skewed t-Student

- For the baseline comparison of between the Normal distribution and the Skewed t-Student we use the expanding window with 252 day period of parameter re-estimation.
- Since parameter re-estimation can be an important factor we will also present results for the expanding window with 5 day period of parameter re-estimation

Normal vs Skewed t-Student

252 day periodof parameter re-estimation

Table: MSE for a expanding window with 252 period of parameter re-estimation

	Norm	Sstd	DM p-value
ARCH	9.5927	1.7369+	0.0945
GARCH	1.5821*	1.5809+*	0.3517
E-GARCH	1.5837+	1.5866	0.3738
GJR-GARCH	1.6024+	1.6043	0.3777
EWMA	1.5846+	1.5851	0.3348
APARCH	1.6520+	1.6575	0.3783

Values for the MSE are multiplied by 10^7
Diebold-Mariano Test alternative hypothesis: two sided
+ : lowest value across the line
* : lowest value across the column

Normal vs Skewed t-Student

252 day periodof parameter re-estimation

Table: QL for a expanding window with 252 day period of parameter re-estimation

	Norm	Sstd	DM p-value
ARCH	7.8714	2.5507+	0.02921
GARCH	2.4337+*	2.4375*	0.2159
E-GARCH	2.4453+	2.4537	0.2956
GJR-GARCH	2.4471	2.4463+	0.7158
EWMA	2.4603	2.4556+	0.2877
APARCH	2.4588+	2.4609	0.9981

Diebold-Mariano Test alternative hypothesis: two sided

+ : lowest value across the line

* : lowest value across the column

Normal vs Skewed t-Student

5 day periodof parameter re-estimation

Table: MSE for a expanding window with 5 day period of parameter re-estimation

	Norm	**Sstd**	**DM p-value**
ARCH	7.0980	1.7296+	0.1571
GARCH	1.5819*	1.5805+*	0.3486
E-GARCH	1.5823+	1.5848	0.3767
GJR-GARCH	1.6007+	1.6025	0.3620
EWMA	1.5841+	1.5845	0.3387
APARCH	1.6462	1.6381+	0.0957

Values for the MSE are multiplied by 10^7

Diebold-Mariano Test alternative hypothesis: two sided

+ : lowest value across the line

* : lowest value across the column

Expanding window vs Moving window

- For the baseline comparison of between the Expanding window vs Moving window we use consider a 5 day period of parameter re-estimation, as well as a 252 day period of parameter re-estimation.
- It's worth mentioning that the use of a Skewed t-Student distribution presented convergence problems for some models. We will focus on the Normal distribution results.
- We present a Diebold-Mariano test comparison between the models for the forecast scenario of the Normal distribution with a moving window of 5 years and re-estimation every 5 days.

Expanding window vs Moving window

MSE Results with Normal Dist.

Table: MSE comparison of forecast methods form Normal distribution

	Expanding	**Moving 2 years**	**Moving 5 years**
ARCH	7.09804	6.41130+	6.71682
GARCH	1.58187*	1.58828	1.57809+
E-GARCH	1.58228	1.60976	1.57625+*
GJR-GARCH	1.60071	1.58248+*	1.58965
EWMA	1.58413	1.59175	1.58364+
APARCH	1.64617	1.63488	1.63303+

Values for the MSE are multiplied by 10^7

Parameters re-estimated every 5 days

+ : lowest value across the line

* : lowest value across the column

Expanding window vs Moving window

MSE Results with Sstd. Dist.

Table: MSE comparison of forecast methods form Sstd. distribution

	Expanding	**Moving 2 years**	**Moving 5 years**
ARCH	1.72956	1.64733	1.67268
GARCH	1.58050	NA	NA
E-GARCH	1.58478	NA	1.57723
GJR-GARCH	1.60255	NA	NA
EWMA	1.58451	1.58882	1.58554
APARCH	1.63811	NA	NA

Values for the MSE are multiplied by 10^7
Parameters re-estimated every 5 days
NA: Not available
+ : lowest value across the line
* : lowest value across the column

Expanding window vs Moving window

QL Results with Normal Dist.

Table: QL comparison of forecast methods form Normal distribution

	Expanding	**Moving 2 years**	**Moving 5 years**
ARCH	6.81089+	29.99399	17.97508
GARCH	2.43581+*	2.45923	2.45747
E-GARCH	2.43971	2.47418	2.43676+*
GJR-GARCH	2.44566+	2.46250	2.44697
EWMA	2.45617+	2.46857	2.46430
APARCH	2.44892+	2.45127*	2.45905

Parameters re-estimated every 5 days

+ : lowest value across the line

* : lowest value across the column

Expanding window vs Moving window

DM-Test with Normal Dist.

Table: DM Test for Normal distribution, moving window of 5 years and re-estimation every 5 days

	ARCH	GARCH	E-GARCH	GJR-GARCH	EWMA	APARCH
ARCH	NA	0.9939	0.9940	0.9940	0.9939	0.9940
GARCH	0.0061*	NA	0.7080	0.7048	0.2764	0.6476
EGARCH	0.0060*	0.2920	NA	0.2778	0.2867	0.2282
GJRGARCH	0.0060*	0.2952	0.7222	NA	0.2883	0.2065
EWMA	0.0061*	0.7236	0.7133	0.7117	NA	0.6829
APARCH	0.0060*	0.3524	0.7718	0.7935	0.3171	NA

DM-Test: The alternative hypothesis is that the model on the column is less accurate than model on the row.
* : less than 5%

Expanding window vs Moving window

MSE Results with Normal Dist.

Results for the 252 day period of parameter re-estimation are similar. Its worth mentioning that the E-GARCH model presented convergence problems when using a 2 year moving window.

Table: MSE comparison of forecast methods form Normal distribution

	Expanding	**Moving 2 years**	**Moving 5 years**
ARCH	9.59273	7.18291+	12.63411
GARCH	1.58209*	1.58932*	1.57497+*
E-GARCH	1.58371	Inf	1.57816+
GJR-GARCH	1.60242	1.62359	1.59712+
EWMA	1.58464	1.59499	1.58393+
APARCH	1.65203	1.63419+	1.64973

Values for the MSE are multiplied by 10^7

Parameters re-estimated every 252 days

+ : lowest value across the line

* : lowest value across the column

Parameter re-estimation

- For the parameter re-estimation comparison we consider a Normal distribution with a 5 year moving window.

Parameter re-estimation

Table: MSE comparison of re-estimation form Normal distribution in a 5 year moving window

	5 days	**20 days**	**60 days**	**252 days**
ARCH	6.7168+	7.1792	9.1130	12.6341
GARCH	1.5781	1.5782	1.5789*	1.5750+*
E-GARCH	1.5762+*	1.5781*	1.5796	1.5782
GJR-GARCH	1.5897+	1.5928	1.5964	1.5971
EWMA	1.5836+	1.5852	1.5851	1.5839
APARCH	1.6330+	1.6392	1.6426	1.6497

Values for the MSE are multiplied by 10^7

+ : lowest value across the line

* : lowest value across the column

QL Results

Realized volatility comparison

- The intraday data allowed for a 3 year (almost 4 year) comparison of the realized volatility.
- The forecast were performed considering a 3 year expanding window, and a 3 year moving window. Both with 5 day period of parameter re-estimation.

Table: MSE comparison in a 3 year expanding window

	Norm	**Sstd**	**DM p-value**
ARCH	12.5964	0.5145+	0.2188
GARCH	0.3111	0.2964+	0.4948
E-GARCH	0.2295+	0.2310	0.1303
GJR-GARCH	0.2600	NA	NA
EWMA	0.2262*	0.2231+	0.2671
APARCH	0.2654	NA	NA

Values for the MSE are multiplied by 10^7
Diebold-Mariano Test alternative hypothesis: two sided
Parameters re-estimated every 5 days
+ : lowest value across the line
* : lowest value across the column

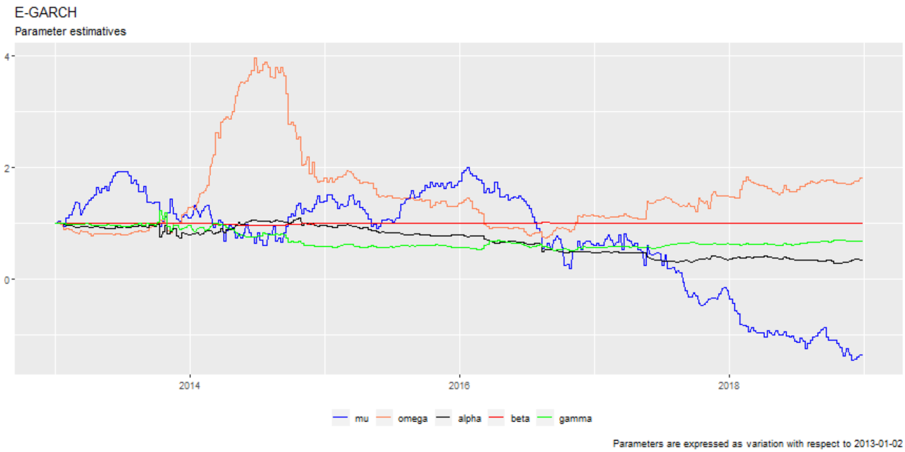
Table: MSE comparison between 3 year expanding window vs moving window

	Mov3(5)	Mov3(20)	Mov3(60)	Roll(5)	Roll(20)	Roll(60)
ARCH	4.2825	8.8842	0.5979	12.5964	0.5106	0.4952+
GARCH	0.3030+	0.3083	0.3148	0.3111	0.3156	0.3225
E-GARCH	0.2359	0.2282+	0.2447	0.2295	0.2340	0.2345
GJR-GARCH	0.2551+	0.2576	0.2588	0.2600	0.2625	0.2653
EWMA	0.2258*	0.2266*	0.2250+*	0.2262*	0.2266*	0.2258*
APARCH	0.2526	0.2574	0.2496+	0.2654	0.2667	0.2747

Values for the MSE are multiplied by 10^7
Mov3(x): moving window of 3 years with re-estimation every x days
Roll(x): expanding window with re-estimation every x days
+ : lowest value across the line
* : lowest value across the column

E-GARCH Parameter Estimates

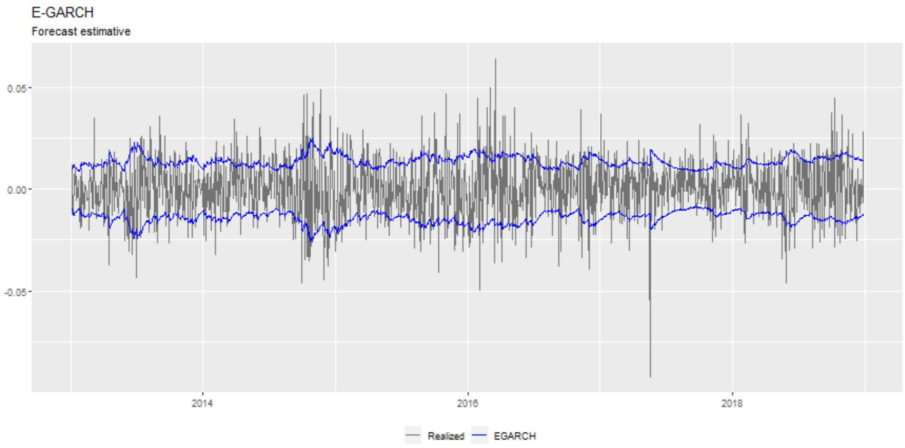
- The E-GARCH model presented a convergence problem when using a moving window of 2 years, therefore we investigate how is the evolution of the parameters across the time.
- For this task we use the 5 year moving window with re-estimation of parameters every 5 days.



E-GARCH
Parameter estimatives
4
2
0
2014
2016
2018
mu
omega
alpha
beta
gamma
Parameters are expressed as variation with respect to 2013-01-02

E-GARCH

Forecast Estimates



1 Introduction

2 Methodology
- Volatility Models
- Forecast Evaluation

3 Data
- Bovespa index history
- Bovespa index series

4 Results
- Daily returns model
- Out-of-sample forecast
- Normal vs Skewed t-Student
- Forecast Window
- Parameter re-estimation
- Realized volatility comparison
- E-GARCH Parameter Estimates

5 Conclusion

Conclusion

- Asymmetric models perform relatively well across all forecasting methods.
- The E-GARCH has convergence problems in the 2 year sample.
- Updating parameter estimates as frequently as possible provides the best forecasting performance.
- No evidence that the Student-t likelihood improves forecasting ability in comparison to the Normal.
- The use of Student-t may cause some convergence problems.
- The E-GARCH model reveals that the parameters have a great degree of variability without exhibiting any signs of breaks.

References

The Comprehensive R Archive Network
https://cran.r-project.org/
codes availiable in: github.com/btebaldi/EmpiricalFinance_1

Pereira, Pedro V.; Morrettin, Pedro A. (2019)
Econometria Financeira - Um Curso em Séries Temporais Financeiras
4th ed.

Pereira, Pedro V.
Lecture notes - Empirical finance I and II

A generic model for estimating the conditional variance of returns is the ARCH model.

$$h_t = \omega + \alpha(L)h_{t-1} + \nu_t \tag{6}$$

where $\alpha(L)$ is a lag-polynomial operator of the type $\alpha(L) = \alpha_1 L + \alpha_1 L^2 + \cdots + \alpha_m L^m$ and to ensure no negativity of conditional variance: $w, \alpha_i > 0$ for $i = 1, \ldots, m$. Also ν_t is and i.i.d. process with zero mean and unit variance.

back

The GARCH(1,1) model assumes the following form:

$$h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1} \tag{7}$$

Key features of this process are its mean reversion (imposed by the restriction $\alpha + \beta < 1$) and its symmetry - future variance responds as much to past positive returns as it does to negative returns.

back

The E-GARCH models the log of variance. The formulation for a E-GARCH(1,1) is as follows:

$$\ln(h_t) = \omega + \alpha(|\epsilon_{t-1}| - \mathbb{E}\left[|\epsilon_{t-1}|\right]) + \gamma\epsilon_{t-1} + \beta\ln(h_{t-1}) \tag{8}$$

where $\epsilon_t = r_t/\sqrt{h_t}$

back

The GJR-GARCH model also models asymmetry in the ARCH process. The GJR-GARCH(1,1) model assumes the following form:

$$h_t = \omega + (\alpha + \gamma \mathbb{I}_{r_{t-1}<c}) r_{t-1}^2 + \beta h_{t-1} \tag{9}$$

where $\mathbb{I}_{r_{t-1}}$ is an indicator equaling one when the previous period's return is below some threshold c (most commonly, $c = 0$).

back

The exponential smoothing is an IGARCH model with $\omega = 0$. In this case, the EWMA model assumes the following form:

$$h_t = \lambda h_{t-1} + (1 - \lambda) r\nu_t \tag{10}$$

back

The APARCH model assumes a specific parametric form for powers of this conditional heteroskedasticity.
In this case, the APARCH(1,1) model assumes the following form:

$$h_t^{\delta/2} = \omega + \alpha(|r_t| - \gamma r_t)^{\delta} + \beta h_{t-1}^{\delta/2} \tag{11}$$

back

Parameter re-estimation

Table: QL comparison of re-estimation form Normal distribution in a 5 year moving window

	5 days	**20 days**	**60 days**	**252 days**
ARCH	17.9751+	49.7862	31.2580	26.3828
GARCH	2.4575	2.4555	2.4437	2.4353+*
E-GARCH	2.4368*	2.4415*	2.4355+*	2.4451
GJR-GARCH	2.4470+	2.4484	2.4510	2.4473
EWMA	2.4643	2.4679	2.4558+	2.4563
APARCH	2.4591+	2.4650	2.4667	2.4598

+ : lowest value across the line
* : lowest value across the column

back