Estimación de medidas de riesgo de mercado en series financieras mexicanas: un estudio comparativo entre procesos GARCH con innovaciones Gaussianas, t-Student y Pareto Generalizada

Alberto Saavedra Espinosa
Universidad Nacional Autónoma de México, Facultad de Ciencias

Resumen

Los objetivos de este trabajo son investigar si: i) un modelo GARCH con innovaciones modeladas mediante una Distribución Pareto Generalizada (DPG), complementado con un pronóstico EWMA de volatilidad para considerar problemas prácticos que pueden surgir en aplicaciones GARCH que comprenden largos periodos de tiempo, estima adecuadamente medidas de riesgo (VaR y *Expected Shortfall*) para series financieras mexicanas a altos niveles de confianza; ii) las estimaciones de dicho modelo son mejores que aquellas entregadas por un GARCH con innovaciones Gaussianas o t-Student. Nuestras evaluaciones de calidad y comparación entre modelos consisten de backtests de las medidas de riesgo de cada método utilizado en el presente artículo. Nuestros resultados muestran que: i) la metodología utilizada estima apropiadamente nuestras dos medidas de riesgo; ii) el modelo GARCH-DPG entrega mejores resultados que los modelos GARCH-Normal y GARCH-t-Student. Nuestros resultados se limitan a estimaciones de medidas de riesgo a un día. Hasta donde sabemos, nuestros resultados sobre el *Expected Shortfall* son los primeros de su clase para series mexicanas. Concluimos que el estudio alcanzó sus objetivos y existen importantes áreas de oportunidad para estudios posteriores.

Estimation of market risk measures in Mexican financial time series: a comparative study between GARCH processes with Gaussian, Student-t and Generalized Pareto innovations

Abstract

The objectives of this work are to investigate whether: i) a GARCH model with Generalized Pareto Distribution (GPD) innovations, complemented with an EWMA volatility forecast in order to consider practical problems that might arise in GARCH applications that comprise long periods of time, appropriately estimate risk measures (VaR and Expected Shortfall) for Mexican financial series, at high confidence levels; ii) the estimates yielded by such model are better than those given by a GARCH with Gaussian or Student-t innovations. Our quality assessment and comparison between models consist of backtests of the risk measures estimates yielded by each method used in this paper. Our results show that: i) the methodology used in this paper appropriately estimates our two risk measures; ii) the GARCH-GPD model yields better results than the GARCH-Gaussian and GARCH-t-Student models. Our results are limited to one-day risk measures estimates. As far as we know, our results on the Expected Shortfall are the first of its kind for Mexican series. We conclude that the study achieved its objectives and there are important areas of opportunity for further studies.

1. Introduction

Modern society relies on the proper functioning of the financial system and has a collective interest in its stability. Financial crises, such as the one of 2008, have pointed out the importance of ensuring that financial institutions properly measure and balance the risks they take and hold enough capital to withstand any foreseeable problems. Consequently, society views risk management positively and confers regulators the task of forging a framework that safeguards its interests.

Therefore, it is important to develop and test accurate methodologies to measure the risks to which a financial institution is exposed. In this regard, the main financial risk categories are market risk, credit risk, operative risk, and liquidity risk. This paper studies and assesses a technique to measure the best-known risk in the banking industry: market risk.

The studied technique uses daily information of the changes of market risk factors and has two important components. The first one is a stochastic volatility model to model and forecast the risk factors dynamics. The second one is the use of Extreme Value Theory (EVT) to complement such dynamic model.

Specifically, it uses a GARCH model with Generalized Pareto innovations to estimate two important risk measures: Value-at-Risk (VaR) and Expected Shortfall.[1] Alexander McNeil and Rüdiger Frey first suggested this kind of model in their extraordinary contribution Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach (McNeil, J. A. and Frey, R., 2000). Such technique takes advantage of the power of the Peaks Over Threshold (POT) method to describe tails of univariate distributions to estimate risk measures for financial time series at high confidence levels.

However, in contrast with McNeil, J. A. and Frey, R., 2000, the technique used in this paper investigates how suitable is to complement the GARCH volatility-modelling component of the mentioned methodology with an Exponential Weighted Moving Average (EWMA) approach, on the days when the GARCH specification presents a practical problem such as a non-significant parameter.[2] This feature might be an advantage in applications that cover periods of considerable length.

---

[1] From a mathematical point of view, VaR is a quantile of the loss distribution, X, of an asset whereas Expected Shortfall is given by $E[X \mid X > VaR]$. Regarding the interpretation of these risk measures, VaR can be thought as the loss that is not exceeded with a high probability (the so-called confidence level) while Expected Shortfall represents the expected loss given that the loss X exceeds VaR.

[2] EWMA is a more informal volatility forecasting technique popularized in J.P. Morgan's RiskMetrics. Specifically, EWMA is a volatility exponential smoothing (it weights the data from most recent to most distant with a sequence of exponentially decreasing weights that sum to almost one) of the form $P_{t+1}Y_{t+1}^2 = \lambda Y_t^2 + (1 - \lambda)P_{t-1}Y_t^2$. Thus, EWMA is essentially a recursive scheme where the prediction at time t is obtained from the prediction at time $t - 1$. The choice of the parameter λ is subjective; the smaller the value the less weight is put on the most recent information. In Section 2, we discuss an approach that suggest how to select this value based on a GARCH specification. In this regard, more information about this technique and its relationship with GARCH models is available in Sections 4.2.4 and 4.4.1 of McNeil, A. J. et al. (2005).

Thus, the objectives of the presented study are to investigate:

    i) If the risk measures estimates that the proposed technique yields for Mexican data, during both crisis and relative calm seasons in financial markets, and especially for high confidence levels, are well behaved.

    ii) How does the risk measures estimates yielded by the proposed technique compare with estimations with another GARCH modeling approach: to model GARCH innovations with a Gaussian or Student-t distribution.

For both points, our quality assessment and comparison between methods will be made using backtests of the risk measures estimates yielded by each method used in this paper.

The Mexican financial series used for our study are the USD/MXN exchange rate (FIX) and the main Mexican stock index, the Prices and Quotes Index (IPC for its acronym in Spanish).

When exploring relevant literature on this type of applications we find that Kourouma, L. et al. (2011) carried out an exercise similar to ours; however, such work focuses on US stock indices and asset prices such as oil and corn during a financial crisis. As far as we know, there has not been an application like ours in Mexican time series.

On the other hand, in Fernández, V. P. (2003) and Aguirre, A. I. et al. (2013) a GARCH model is used in conjunction with the POT Method to analyze, among other assets, a Mexican financial series: the IPC. However, such applications focus solely on the estimation and evaluation of VaR. That is, the Expected Shortfall analysis, which is an equally important risk measure, is left out.

In addition, in López, E. (2013) a GARCH model and another technique from the EVT, the Block Maxima Method, is used to analyze another Mexican financial series: the FIX exchange rate. However, this work also focuses solely on VaR; other Mexican financial series and the Expected Shortfall are left out of the analysis.

Consequently, our work is separated from those made previously in several senses. First, while some attention had been devoted to the study of the IPC using GARCH-GPD methods, these exercises focused solely on the estimation and evaluation of VaR. That is, they ignored the Expected Shortfall, which is an equally important risk measure that has the advantage of having better theoretical properties than VaR; specifically, to be a coherent risk measure.

Thus, as far as we know this is the first analysis of this nature that estimates and evaluates Expected Shortfall estimates at high confidence levels for Mexican series.

On the other hand, our application also contrasts with previous work in the fact that it analyzes the efficiency in Mexican data of the GARCH-GPD method in periods both of crisis (unlike Fernández, V. P. (2003)) and relative calm (in contrast to Kourouma, L. et al. (2011), López, E. (2013) and Aguirre, A. I. et al. (2013)).

Indeed, because the quality of conditional VaR and Expected Shortfall estimates was evaluated at high confidence levels, the length of the backtests used was long enough to evaluate the methodology used both in periods of crisis (as in 2008) and relative calm in the markets.

Another peculiarity of our application is related to a classic problem of applications with extremes: the choice of a threshold from which the tail of a distribution is considered to start. Specifically, our work explored how adequate for Mexican data is the methodology suggested by McNeil, A. J. and Frey, R. (2000) to choose an optimal threshold for the right tail of a GARCH innovation distribution. Investigating the adequacy of this method in Mexican series is of great importance: it shows the presence of an important empirical fact in national data, which may be relevant in subsequent analyzes.

After this introduction, our paper structure is as follows: section two describes the methodology used in our study, including the general development of our backtesting procedures; section three presents the main results of our application; section four presents our conclusions and a few ideas for future work on this topic.

2. Methodology

Let $\{X_t\}_{t=1}^{\infty}$ be a stationary time series that represents daily observations of log-returns of the price of a financial asset.[3] We will assume that the dynamics of X_t is given by a GARCH(1,1)[4] model:

$$X_t = \sigma_t Z_t, \ \ \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (1)$$

Where $\alpha_0 > 0, \ \alpha_1 \geq 0, \ \beta_1 \geq 0$ and the innovation process Z_t is a Strict White Noise (i.e. it is formed by independent and identically distributed random variables) with distribution function $F_Z(z)$ with zero mean and unitary variance. We will suppose that σ_t is measurable with respect to $\mathfrak{F}_{t-1}$, the available information of the log-returns process up to time $t-1$.

Given that our backtests will have a considerable length,[5] we will complement our volatility modelling with an EWMA approach. This is related with two typical scenarios that could affect our GARCH volatility forecasting:

1. The parameter $\widehat{\alpha_0}$ of the GARCH model fitted to the series is not statistically significant (which in our application is checked with the traditional t-statistic test of statistical significance).

---

[3] Given that risk management typically focuses in analyzing financial loses, in risk measuring studies is typical to associate a positive sign to losses and a negative one to gains. This is done for both practical (e.g. work with positive random variables) and theoretical (e.g. facilitate the application of certain statistical techniques to losses data) reasons. Our paper follows this practice and associates a positive sign to Mexican peso depreciations and IPC losses.

[4] There are many possible elections for the series dynamics (e.g. ARCH, HARCH processes), however, our election of the GARCH(1,1) is motivated by the capacity of such model to parsimoniously reproduce some of the main stylized facts (see Section 4.1.1 of McNeil, A. J. et al. (2005)) of financial returns series. Furthermore, studies like Hansen, P. R. and Lunde, A. (2005) have shown that such models are able to deliver accurate short-term volatility forecasts for series of financial returns. Evidence of the goodness of fit that this model delivers for our series is shown in Section 3.

[5] In fact, as our goal is to evaluate risk measures estimations at high confidence levels (e.g. $\alpha = 0.999$), our backtests will have to use several years of daily information in order to observe loses that occur with the corresponding low probabilities (e.g. 0.001).

2. The GARCH process fitted to the series is not stationary (or, equivalently, that $\widehat{\alpha_1} + \widehat{\beta_1} > 1$).

While such cases are rare (specially the second one), they could appear in some days in an application that covers an extended period of time (approximately 20 years) like ours. Thus, while the volatility of our series will always be estimated through a GARCH(1,1) model, its forecasts will instead be given by:

$$\widehat{\sigma_{t+1}^2} = \begin{cases} \widehat{\alpha_1}X_t^2 + (1 - \widehat{\alpha_1})\widehat{\sigma_t^2} & si \ \widehat{\alpha_1} + \widehat{\beta_1} > 1 \ o \ \mathfrak{P}(\widehat{\alpha_0}) > 0.05 \\ \widehat{\beta_1} + \widehat{\alpha_1}X_t^2 + \widehat{\beta_1}\widehat{\sigma_t^2} & in \ any \ other \ case \end{cases} \qquad (2)$$

Where $\widehat{\alpha_0}$, $\widehat{\alpha_1}$ and $\widehat{\beta_1}$ are the estimated values of the GARCH model fitted to our series on day t and $\mathfrak{P}(\alpha_0)$ is the p-value of the t-statistic test applied to $\widehat{\alpha_0}$.

This is, on the days when any of the cases previously mentioned occur, we use an EWMA approach to forecast the volatility of our series. On the rest of the days, we use the forecast given by our GARCH specification.

Note that, given the definition of (2), our EWMA specification almost replicates the volatility forecast suggested by the GARCH model.[6] This is a desirable feature in such quantity if the GARCH model achieves to estimate appropriately the volatility of Mexican financial time series.

Since our application will investigate the adequacy of different models for the innovation distribution, we estimate the parameters of our GARCH models using a Quasi-Maximum likelihood approach. This is in order to avoid making any prior assumptions on the innovation distribution.[7]

Given that,

$$F_{X_{t+1}|\mathfrak{F}_t}(x) = P(\sigma_{t+1}Z_{t+1} \leq x \mid \mathfrak{F}_t)$$
$$= F_Z\left(\frac{x}{\sigma_t}\right),$$

we have that VaR and Expected Shortfall formulas turn out to be:

$$VaR_\alpha^t = \sigma_{t+1}q_\alpha(Z), \qquad (3)$$
$$ES_\alpha^t = \sigma_{t+1}ES_\alpha(Z) \qquad (4)$$

Where σ_{t+1} follows the specification given in (2) and $q_\alpha(Z)$ and $ES_\alpha(Z)$ denote the risk measures associated to the distribution of Z_t, which, by hypothesis, do not depend on t.

When the innovation distribution is modeled by a Gaussian or Student-t distribution, such risk measures can be computed easily. In fact, they are given by:

---

[6] Indeed, works like Martínez, J. (2014) and ours show that for certain Mexican financial series, the value of the sum $\widehat{\alpha_1} + \widehat{\beta_1}$ is typically close to 1.

[7] For those interested readers who are not familiar with Quasi-Maximum likelihood estimation procedures, we recommend reading Chapter 4 of McNeil, A. J. et al. (2005).

Risk measures formulas for the Gaussian and Student-t models		
Model	$q_\alpha(Z)$	$ES_\alpha(Z)$
$N(\mu, \sigma^2)$	$\mu + \sigma\Phi^{-1}(\alpha)$	$\mu + \dfrac{\sigma\phi(\Phi^{-1}(\alpha))}{1-\alpha}$
$t(v)$	$t_v^{-1}(\alpha)$	$\dfrac{g_v(t_v^{-1}(\alpha))}{1-\alpha}\left(\dfrac{v + (t_v^{-1}(\alpha))^2}{v-1}\right)$

Table 1: Formulas for quantile (second column) and Expected Shortfall (third column) for the Gaussian and Student-t models.

On the other hand, we can use the POT method to model the innovation distribution. The POT method is based on the Gnedenko-Balkema-Pickands-de Haan theorem, which is a limit result from the Extreme Value Theory that, roughly speaking, says that the excess distribution, $F_u(x) = P(X - u \leq x \mid X > u)$, of an i.i.d. sequence of random variables can be approximated by a Generalized Pareto Distribution, for a big enough threshold u.

Specifically, as Aguirre, A. I. et al. (2013) summarizes, the Gnedenko-Balkema-Pickands-de Haan theorem tells us is that, for a large class of univariate distributions, the excess distribution F_u, for big values of u, is approximately equal to

$$F_u \approx G(x) = \begin{cases} 1-\left(1-\frac{x}{\beta}\right)^{-1/\xi} & si\ \xi \neq 0 \\ 1-e^{-\frac{x}{\beta}} & si\ \xi = 0 \end{cases}$$

Where $\xi \in \mathbb{R}, \beta = \sigma + \xi(u - \mu)$. ξ and β are known, respectively, as the form and scale parameters of the GPD.

In the former expression, three possible distributions can be obtained, depending on the value of the parameter ξ of the GPD. If $\xi > 0$, the GPD is a Pareto distribution with parameters $\alpha = 1/\xi$ and $\kappa = \beta/\xi$, for $x > 0$. For $\xi = 0$ the GPD corresponds to an exponential distribution with rate $1/\beta$ and $x > 0$. Finally, if $\xi < 0$, the GPD take the form of type II Pareto Distribution, which is defined in the range $0 < x < \beta/\xi$.

There are at least two ways to work with thresholds. One is through the value of the threshold itself; i.e. a relative large value in terms of the magnitude of our observations. The second one is to express the threshold in terms of the number of data points that are above it. An example of the second approach for a sample of 1,000 observations would be to fix the value of u equal to the $(k + 1)$ order statistic; thus, if $k = 100$ we have 100 observations above u and we consider that the tail of the distribution of our sample consists of the 10% biggest observations in it.

For the first approach, one can use the graph of the Sample Mean Excess Function to find appropriate threshold values to use in the POT Method. The Sample Mean Excess Function is an empirical estimator of the function $e(u) = E(Y - u \mid Y > u)$. For an i.i.d. sequence of n random variables, $\{Y_i\}_{i=1}^n$, it is given by:

$$e_n(u) = \frac{\sum_{i=1}^{n} max\{Y_i - u, 0\}}{\sum_{i=1}^{n} I_{\{Y_i > u\}}}.$$

That is, $e_n(u)$ represents the sum of the excesses above the threshold, u, divided by the number of observations that exceed u. The graph of the Sample Mean Excess Function is defined as:

$$\{(u, e_n(u)) \mid Y_{n,n} < u < Y_{1,n}\}$$

Where $Y_{1,n} > \cdots > Y_{n,n}$ represent the order statistics of the sample $\{Y_i\}_{i=1}^{n}$.

The interpretation of the graph of the Sample Mean Excess Function is: if from a certain threshold, the points show an linear upward trend, then there are signs of heavy tail behavior. In the same way, data with exponential distribution would give an approximately horizontal line. Lastly, data from a short tail distribution would show a linear downward trend. In all cases, the threshold from which any of these behaviors is noted might be a good candidate to apply the POT Method.

Regarding the second approach, McNeil, A. J. and Frey, R. (2000) suggest carrying out a simulation study. Specifically, because a pair of desirable characteristics in an estimator is that it is unbiased and with least Mean Squared Error (MSE), they suggest the following: identify a distribution that represents at least a good approximation to the true distribution of innovations (e.g. a Student-t distribution, because the series of financial returns are usually leptokurtic). If such a distribution can be seen as a particular case of the GPD (e.g. a Student-t distribution with v degrees of freedom can be considered as a GPD with a shape parameter $\xi = 1/v$), one can use this fact to estimate the bias and MSE of ξ, and some other amount of interest (e.g. a high quantile of the distribution) of the fitted GPD. Then, they suggest plotting the bias and MSE of these estimates as a function of k (the number of data that conform the tail of the distribution). Finally, one should explore the graphs generated by this method and find the value of k that minimizes both the bias and mean square error of the estimates of ξ and any other amount of interest.

Once the threshold for applying the POT Method has been selected, the parameters of the GPD can be estimated via maximum likelihood, based on the sample versions of the GARCH innovations; that is, the standardized residuals of the model. In our application, we estimate GPD parameters by Maximum likelihood.[8]

In this way, it is possible to use the approximation given by the Gnedenko-Balkema-Pickands-de Haan theorem and develop the following VaR and Expected Shortfall formulas for a distribution whose tails are modeled by a GPD:

$$q_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_u}(1-\alpha) \right)^{-\hat{\xi}} - 1 \right); \quad ES_\alpha = \frac{q_\alpha}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{1-\hat{\xi}} \qquad (5)$$

---

[8] For the reader interested in the details of the construction of the maximum likelihood estimators of the parameters of the GPD, we recommend reading Section 6.5.1 of Embrechts, P. et al. (1997).

Where N_u represents the amount of data above the threshold and n the total size of the sample used. Thus, by substituting these estimates in the equations given in (3) and (4), it is possible to construct the estimators of VaR_α^t and ES_α^t based on a GARCH-GPD model.

Backtesting

To evaluate the estimates of VaR_α^t and ES_α^t obtained through the GARCH-Normal, GARCH-Student-t and GARCH-GPD models it is possible to develop backtests.

We evaluate the performance of VaR_α^t and ES_α^t on a historical series $x_1, \ldots, x_m$, where $m \gg n$, based on a memory of n days on days $t \in T = \{n, \ldots, m-1\}$. This means that if we worked, for example, with a time window of $n = 1{,}000$ banking days, for each prediction of VaR_α^t and ES_α^t we would use, approximately, 4 years of daily information.

To perform our VaR_α^t backtest, on each day $t \in T$ we fit a new GARCH model to the corresponding loss observations, we calculate the standardized residuals of the model and determine a new estimator of the innovation distribution via The POT method (or any other model for innovations). Then, we use the formula given in (3) and, based on the model fitted to the log-returns, we estimate VaR_α^t. Next, we compare $\widehat{VaR_\alpha^t}$ with x_{t+1} for the values of α considered appropriate; in our case, we were interested in $\alpha \in \{0.95, 0.99, 0.995, 0.999\}$. We say that a violation occurred whenever $x_{t+1} > \widehat{VaR_\alpha^t}$.

It is possible to develop a binomial test based on the number of violations to evaluate the performance of our estimates of VaR_α^t. Assuming the dynamics of equations (2) and (3), the violation indicator at time $t \in T$ is a Bernoulli random variable:

$$I_t = I_{\{X_{t+1} > VaR_\alpha^t\}} = I_{\{Z_{t+1} > q_\alpha\}} \sim Bernoulli(1 - \alpha).$$

Moreover, I_t and I_s are independent for $s, t \in T$ and $s \neq t$, since Z_t and Z_s are independent. Thus,

$$\sum_{t \in T} I_t \sim Binomial(card(T), 1 - \alpha),$$

that is, the total number of violations has a binomial distribution under the proposed model.

In this way, under the null hypothesis that the model correctly estimates the conditional quantiles of the studied series, the empirical version of the statistic $\sum_{t \in T} I_{\{X_{t+1} > \widehat{VaR_\alpha^t}\}}$ comes from the $Binomial(card(T), 1 - \alpha)$ distribution.

Therefore, we can perform a two-tailed binomial test to test this null hypothesis against the alternative that the method entails a systematic error in the estimation of VaR_α^t that leads to very few or too many violations. A smaller p-value than our level of significance (typically 0.05) in this binomial test will be interpreted as evidence against the null hypothesis.

Now, it is also possible to develop a backtest to evaluate our estimates of ES_α^t. This backtest is similar to that developed in the case of VaR_α^t and will allow us to investigate whether the model proposed in the methodology discussed in this Section provides reasonable estimates of ES_α^t.

This time we are interested in the size of the difference between X_{t+1} and ES_α^t given the event of a violation to the VaR_α^t quantile. We define the residuals

$$R_{t+1} = \frac{X_{t+1} - ES_\alpha^t}{\sigma_{t+1}} = Z_{t+1} - E[Z \mid Z > q_\alpha(Z)].$$

It is clear that under our model (2) these residuals are i.i.d. and that, conditioned to the event $\{X_{t+1} > VaR_\alpha^t\}$ or equivalently $\{Z_{t+1} > q_\alpha(Z)\}$, have an expected value equal to zero.

Thus, suppose again that we perform our backtest over the days in the set T. We can form empirical versions of these residuals on the days when a violation to the corresponding quantiles occurs, that is, on the days when $\{x_{t+1} > \widehat{VaR}_\alpha^t\}$. We will refer to these residuals as excess residuals and will denote them by

$$\left\{ r_{t+1} : t \in T, x_{t+1} > \widehat{VaR}_\alpha^t \right\}, \quad where \quad r_{t+1} = \frac{x_{t+1} - \widehat{ES}_\alpha^t}{\sigma_{t+1}},$$

where $\widehat{ES}_\alpha^t$ is a conditional estimate of the Expected Shortfall.

In this way, under the null hypothesis that we correctly estimate the dynamics of the loss process (σ_{t+1}) and the first moment of the truncated innovation distribution ($E[Z \mid Z > q_\alpha(Z)]$), the excess residuals should behave as a sample of i.i.d. observations with mean zero.

To test the zero-mean hypothesis we use a bootstrap test that makes no assumption about the distribution underlying the excess residuals. Specifically, a two-tailed test that contrasts our null hypothesis that excess residuals have a mean statistically equal to zero against the alternative that such residuals have a mean different than zero; which would mean that the conditional Expected Shortfall was systematically underestimated or overestimated.

In essence, the bootstrap test we use consists of the following steps:

i)   take the excess residuals as data;
ii)   calculate the average of these data (this will be the initial value of the test statistic: the simple average);
iii)   resample various series satisfying the null hypothesis (i.e. series whose mean does equal zero) from the original data set (i.e. the excess residuals);
iv)   calculate and store the means of the resampled series;
v)   compare the values of the averages of the resampled series against the mean of the original series (the initial value of our test statistic);
vi)   estimate the p-value of the test by calculating the number of times that the resampled averages were further (either below or above) from zero than the average of our original series: if those events represent less than 5% of the tests, the hypothesis that the original data have average zero is rejected.

3. Results

In this section, we present the results of our application. First, we show how our methodology works on a particular day, for both the FIX and IPC series. Then, we present the results of the backtests used to evaluate the performance of such methodology in our data.

In total, our series are conformed by daily observations of FIX and IPC between January 2, 1996 and February 12, 2016. However, in the first part of this section we will use the following sets of information:

- FIX: from February 17, 2012 to February 12, 2016.
- IPC: from September 7, 2011 to August 31, 2015.

Specifically, we will use $1,000$ daily observations on the fitting of each GARCH model. This means that we will use, approximately, 4 years of daily information to estimate our risk measures.

The parameters of the GARCH models fitted to the data used in this section, as well as the p-values of the corresponding statistical significance tests, are presented in Table 2:

Parameter Significance, GARCH(1,1) model				
	FIX		IPC	
Parameter	Punctual estimation	P-value	Punctual estimation	P-value
α_0	2.220×10^{-7}	0.106	1.333×10^{-6}	0.0118
α_1	0.0581	7.150×10^{-12}	0.0671	4.460×10^{-7}
β_1	0.9396	$< 2 \times 10^{-16}$	0.9164	$< 2 \times 10^{-16}$

Table 2: Punctual estimation and p-value of the t-statistic test of statistical significance of GARCH(1,1) parameters for FIX and IPC series.

Table 2 shows that all the parameters of the GARCH model fitted to the IPC series are significant. In the FIX case, only the α_0 parameter is non-significant. According to (2), this means that we use an EWMA to forecast the volatility of the FIX series on this particular day, while that of the IPC series is estimated using the formula given by our GARCH specification:

$$
\begin{array}{llll}
& \text{FIX} & & \text{IPC} \\
\sigma_{t+1} & = & (\lambda X_t^2 + (1-\lambda)\sigma_t^2)^{1/2} & \sigma_{t+1} & = & (\alpha_1 X_t^2 + \beta_1 \sigma_t^2)^{1/2} \\
& = & \alpha_1 X_t^2 + (1-\alpha_1)\sigma_t^2 & & = & 0.0106 \quad (7) \\
& = & 0.009782 \quad (6) &
\end{array}
$$

In order to evaluate the goodness of fit of these models, we estimate their volatility and compute the sample versions of the innovations of our series. Specifically, if a GARCH model correctly estimates the volatility of a time series, we expect its innovations to behave as a Strict White Noise.
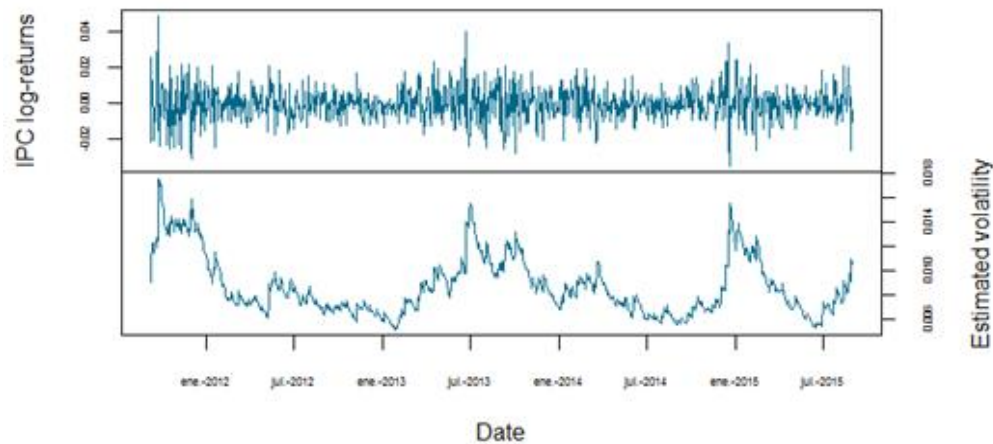
The volatility estimated by the fitted GARCH models is presented in Figure 1. Once this quantity has been estimated, we obtain the sample observations of the innovations of each model dividing each observation by the corresponding value of the estimated volatility.

Figure 1

GARCH estimation of FIX volatility
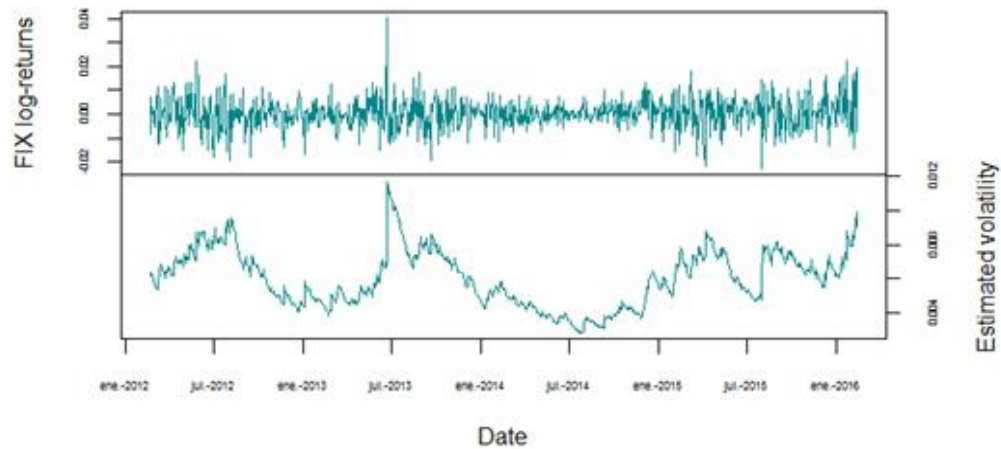

GARCH estimation of IPC volatility

For both series, we test the SWN hypothesis through graphic and numerical tests applied to their GARCH innovations, its squares, and its absolute values. For the first, we look at the graphs of the Autocorrelation function (ACF) and Partial Autocorrelation function (PACF). For the second, we apply a Ljung-Box test.

Figure 2 presents the graphs of the ACF and PACF for both FIX and IPC innovations.

Examining Figure 2 allows us to see that there is no evidence that FIX nor IPC innovations present a serial dependence of its lagged values. Thus, this Figure suggests that the GARCH model achieved to estimate correctly the volatility of both series.

Figure 2
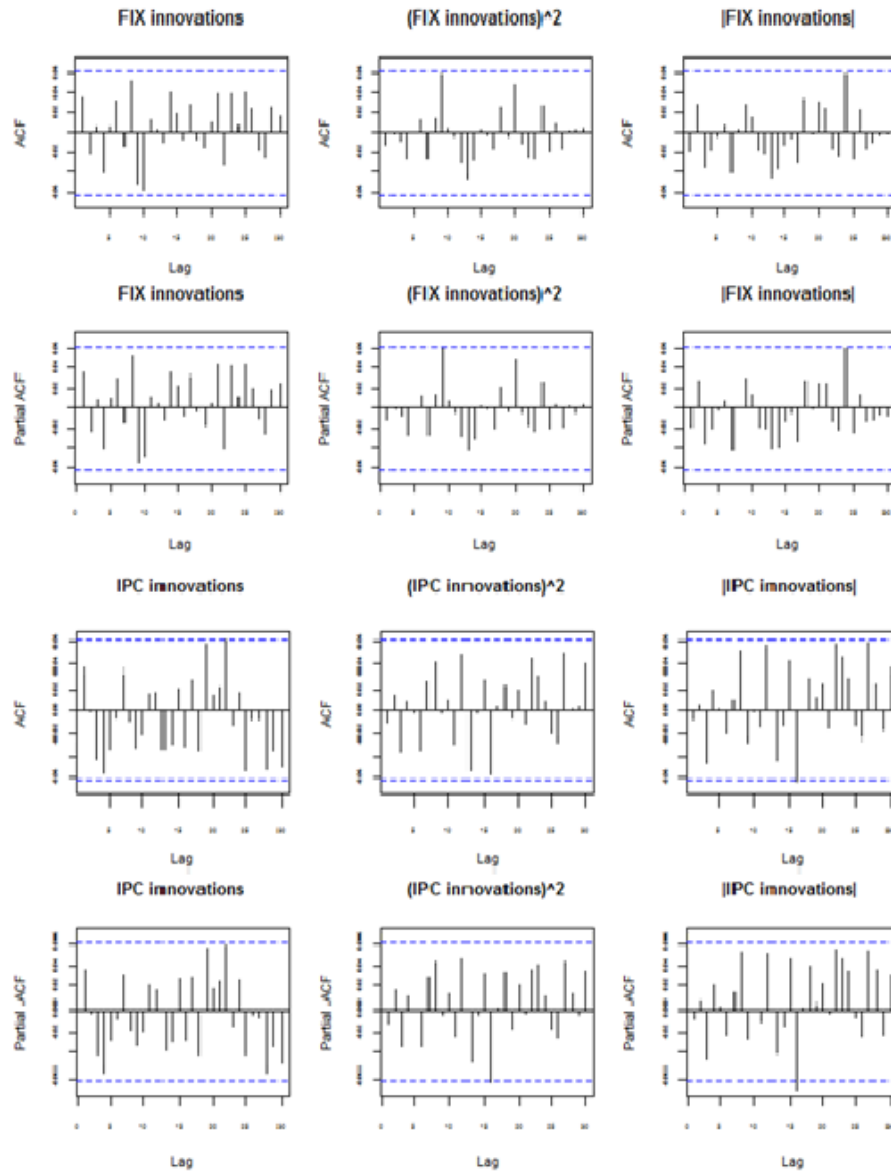
Regarding the Ljung-Box test, Table 3 summarizes the results of its application to the GARCH innovations of the FIX and IPC series.

Table 3 shows that the hypothesis of non-correlation was not rejected for neither of our series, in any of the nine cases considered. Therefore, Table 3 provides further evidence that the GARCH model adequately estimated the volatility of our series.

Ljung-Box test on FIX and IPC innovations: P-values			
Data	$h = \ln(n)$	$h = 20$	$h = 30$
FIX	0.6967	0.6721	0.7181
FIX squared	0.9752	0.8967	0.9871
FIX on absolute value	0.7131	0.8834	0.9272
IPC	0.2196	0.4325	0.2474
IPC squared	0.9075	0.8241	0.7588
IPC on absolute value	0.9674	0.6353	0.4096

Table 3: P-values of the Ljung-Box tests applied to the innovations of the GARCH(1,1) models fitted to the FIX (upper half of the table) and IPC (bottom half of the table) series. Ljung-Box tests were based on three different lags (h), which are shown by column.

Given that the results contained in Figure 2 and Table 3 give us good certainty that the GARCH model correctly estimates the volatility dynamics of our series, we now continue with the analysis of their innovations.

Our first objective is to find a first appropriate threshold to use in the POT Method. Thus, we use the graph of the Sample Mean Excess Function. Figure 3 shows this graph for both the IPC and FIX innovations.
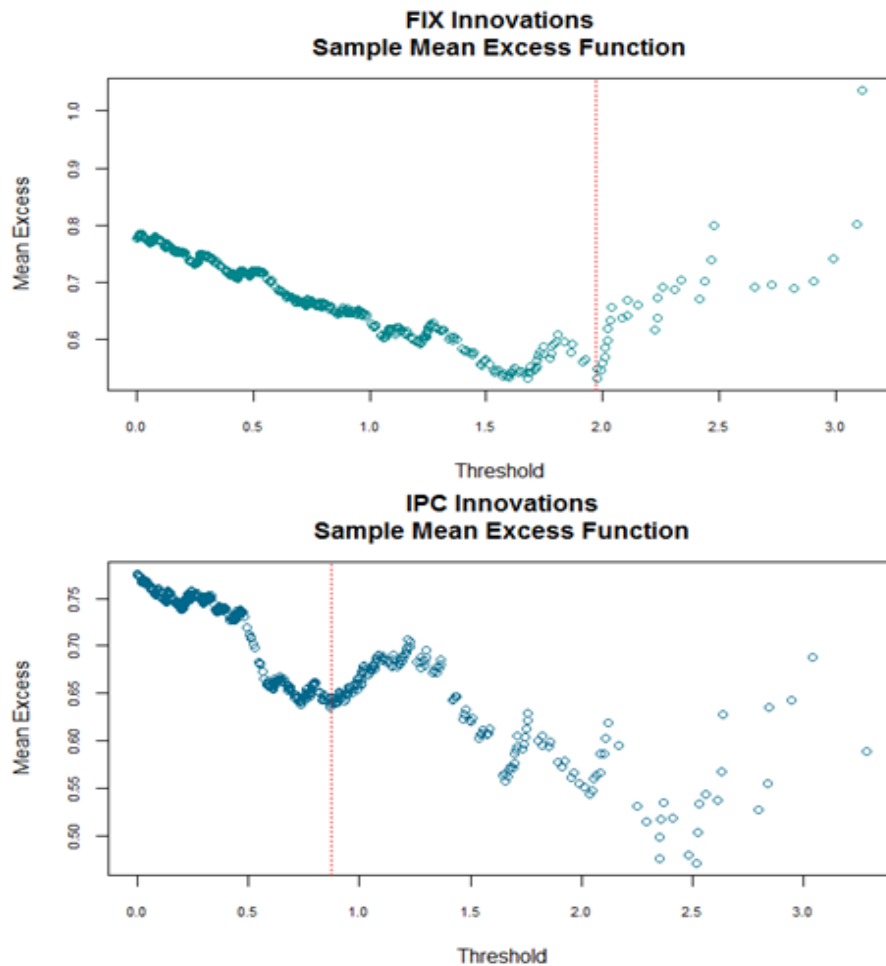
Figure 3

Figure 3 shows a couple of potential thresholds to make a first GPD fit to the GARCH innovations of our series. In particular, it suggests using a threshold around 2 for the FIX innovations and one around 1 for IPC innovations. Therefore, we fit a GPD model to the innovations series using the suggested thresholds.[9] A visual assessment of the quality of such fit is shown in Figure 4.

Explicitly, Figure 4 contains the curves of the excess distribution (on the left) and the tail (on the right) of the GPD model, for the FIX and IPC innovations, with the empirical versions of such curves superimposed respectively. Hence, this Figure let us see that the thresholds suggested by Figure 3 allow the GPD model to achieve an excellent goodness of fit in both series of innovations.

Figure 4
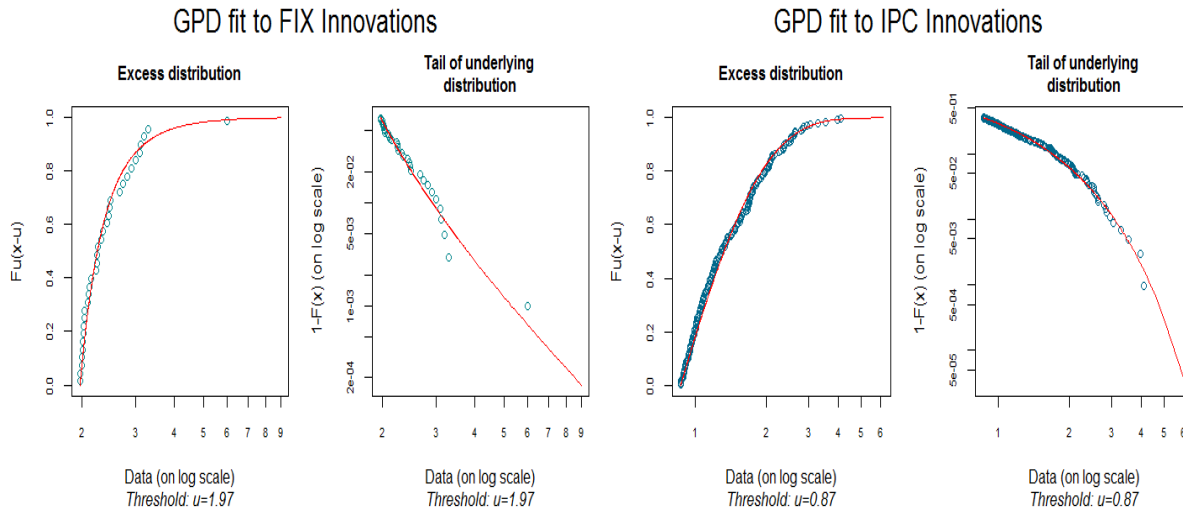
Parameters of the GPD fit: FIX innovations				Parameters of the GPD fit: IPC innovations			
n	N_u	$\hat{\xi}$	$\hat{\beta}$	n	N_u	$\hat{\xi}$	$\hat{\beta}$
511	34	0.3467	0.3497	484	182	-0.0370	0.6610

Once we have seen that our first GPD models had a good fit to the innovation series, we compare such fit with the one that we would get with a Gaussian and Student-t models. The estimated parameters of the three models are summarized in Table 4.

Models fitted to the FIX and IPC innovations: estimated parameters

Model	FIX parameters		IPC parameters	
Gaussian	$\mu = 0.0367$	$\sigma = 0.9998$	$\mu = 0.0244$	$\sigma = 1.005$
GPD	$\xi = 0.1515$	$\beta = 0.4465$	$\xi = -0.0353$	$\beta = 0.6619$
Student-t	$\nu = 6.9818$		$\nu = 9.9583$	

Table 4: Punctual estimations of the parameters of the Gaussian (first row), GPD (middle row), and Student-t (third row) models fitted to the FIX and IPC innovations.

---

[9] Threshold selection is a problem that typically involves the analysis of several numerical, graphical, and model adequacy tests (e.g. investigate the stability of the form parameter of the GPD for along a certain spectral of thresholds) that are not discussed here for briefness reasons. We just show the results of the thresholds that best satisfied our criteria. Notwithstanding, the interested reader can find more information about threshold selection analysis for Mexican financial time series in Campa, M. A. (2001).

Figure 5 contains a graph that shows a comparison of the goodness of fit of the three models at the (right) tail of the distributions of both innovation series. Such Figure let us see that, in both cases, the GPD model has the best fit; moreover, it shows us that the Student-t distribution seems to overestimate the heaviness of the right tail of our two innovations series, whereas the Gaussian clearly underestimates it.

Figure 5
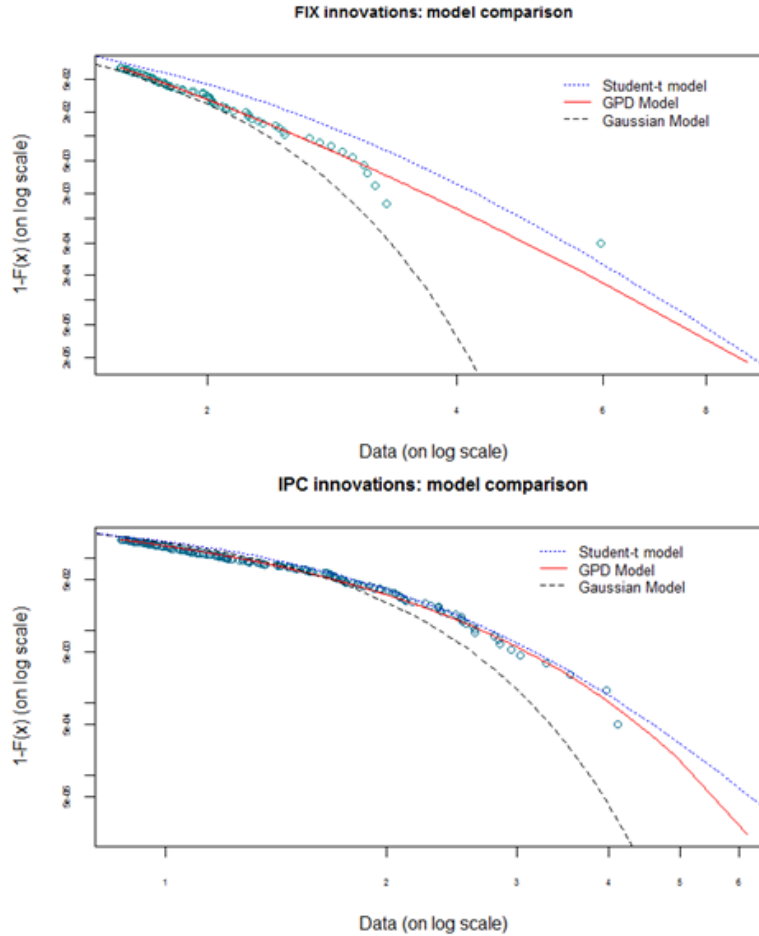
Using the formulas given in Table 1, equation (5), the parameters in Table 3 and the selected thresholds, we can estimate VaR and Expected Shortfall for the innovations of our series. Table 5 contains such estimations.

The quantities in Table 5 show that the monotony relationship between the tails estimated by our three models for innovations is also present in our risk measures estimations: The Student-t model yields the biggest estimations, the Gaussian model the smallest ones, and those of the GPD model lie between the former two.

Now, once we have estimated our risk measures for the innovation distribution and forecasted the volatility of our series, we can estimate the corresponding risk measures for our Mexican financial series. To do so, we use the formulas given in (3) and (4).

Risk measures for the innovations series

	FIX innovations							
α	0.95		0.99		0.995		0.999	
Model	VaR_α	ES_α	VaR_α	ES_α	VaR_α	ES_α	VaR_α	ES_α
Gaussian	1.681	2.099	2.363	2.702	2.612	2.929	3.126	3.404
GPD	1.717	2.270	2.572	3.277	3.009	3.792	4.220	5.219
Student-t	1.895	2.597	3.000	3.774	3.503	4.327	4.792	5.773
	IPC innovations							
α	0.95		0.99		0.995		0.999	
Model	VaR_α	ES_α	VaR_α	ES_α	VaR_α	ES_α	VaR_α	ES_α
Gaussian	1.678	2.087	2.363	2.690	2.613	2.916	3.130	3.391
GPD	1.706	2.317	2.695	3.272	3.104	3.667	4.016	4.548
Student-t	1.813	2.410	2.766	3.367	3.172	3.788	4.149	4.821

Table 5: Punctual estimations of VaR_α and ES_α obtained through the Gaussian, Student-t, and GPD models fitted to the FIX and IPC innovations, for several confidence levels (α).

Table 6 contains the estimations of our conditional risk measures. Note that, because such quantities are just an escalated version of the estimates contained in Table 5, they present the same monotony relationship.

Conditional Risk Measures for Mexican financial series

	FIX							
α	0.95		0.99		0.995		0.999	
Model	VaR_α^t	ES_α^t	VaR_α^t	ES_α^t	VaR_α^t	ES_α^t	VaR_α^t	ES_α^t
Gaussian	0.0164	0.0205	0.0231	0.264	0.0256	0.0286	0.0306	0.0333
GPD	0.0165	0.0226	0.0264	0.0325	0.0306	0.0367	0.0405	0.0466
Student-t	0.0185	0.0254	0.0293	0.0369	0.0343	0.0423	0.0469	0.0565
	IPC							
α	0.95		0.99		0.995		0.999	
Model	VaR_α^t	ES_α^t	VaR_α^t	ES_α^t	VaR_α^t	ES_α^t	VaR_α^t	ES_α^t
Gaussian	0.0181	0.0225	0.0255	0.0290	0.0282	0.0314	0.0337	0.0366
GPD	0.0188	0.0254	0.0296	0.0346	0.0335	0.0379	0.0407	0.0440
Student-t	0.0195	0.0258	0.0297	0.0359	0.0339	0.0403	0.0441	0.0511

Table 6: Punctual estimations of VaR_α^t and ES_α^t obtained through the GARCH(1,1) model fitted to the FIX and IPC series varying the models (Gaussian, Student-t, and GPD) fitted to the FIX and IPC innovations, for several confidence levels (α).

With the former results ends the first part of our application. This first approximation suggest that, among our three alternatives, the GARCH-GPD model seems to have the best fit to the right tail of our series. Moreover, it signals that the GARCH-Student-t model might overestimate such tails, whereas the GARCH-Gaussian model may underestimate them. Nevertheless, in order to validate such results, we must perform a formal evaluation, which is the purpose of the second part of this section.

Notice that in order to implement a backtest of our conditional risk measures estimates we need to automate all the steps followed in the previous section. This can be done easily for all such steps, except for one: the selection of the threshold to use in the POT Method.

Indeed, in an automatized routine, it would be impossible to analyze in each step the information given by the graph of the Sample Mean Excess Function or any other graphical tool. Thus, for our backtest purposes, we need an automatable algorithm to select an appropriate threshold to fit a GPD model to the GARCH innovations of our series.

To do so, we use the second way to work with thresholds: express them as a function of the number of observations above them. Now, it is important to notice that Figure 5 shows that the Studet-t distribution represents, at least, an approximation to the observed distribution of the innovations of our two series. Thus, we can apply the results of the simulation study of McNeil, A. J. and Frey, R. (2000) in our data.

Indeed, McNeil and Frey simulation study involved innovations series whose distributions was well approximated by a Studet-t distribution. Therefore, given that our data satisfy such condition, we will apply their result and assume that the right tail of our both innovations distributions is formed by the 10% of the larger innovations.

To check if this approach to fit a GPD model yields good results in our data, we make a graphical assessment. Figure 6 shows a comparison of the fit, to the right tail of both FIX and IPC innovations distribution, of a Gaussian model, a Student-t model and GPD model fitted following the results of McNeil and Frey's simulation study.

Examining Figure 6 we can notice that the goodness of fit achieved applying the result of McNeil and Frey is comparable to the one we achieved analyzing the graph of the Sample Mean Excess Function. In other words, the methodology yields excellent results in our Mexican financial data.

Notice that the threshold selection methodology used to fit the GPD models shown in Figure 6 can be easily automatized. Consequently, we can now automatize all the steps needed to apply the GARCH-GPD methodology and implement the backtests procedures described in Section 3.

To be consistent with the analysis performed in the first section, the VaR_α^t and ES_α^t backtest of our application will be based on a time mobile window of $n = 1,000$ banking days. This means that the set of days on which we will perform our backtests will be conformed by the daily observations of FIX and IPC log-returns from December 29, 1999 to February 12, 2016; the information of the days between January 2, 1996 and December 28, 1999 will be used to fit the first GARCH(1,1) model to each series.

In this way, the backtests of our series will be carried out on a set of $m = 4,060$ observations and each estimation of VaR_α^t and ES_α^t will be done based on approximately 4 years of daily information. Notice that the length of this period was extended as much as possible because we

will carry out evaluations on estimates of high quantiles and expected shortfalls at considerably high confidence levels.[10]

Figure 6

FIX innovations: model comparison

IPC innovations: model comparison
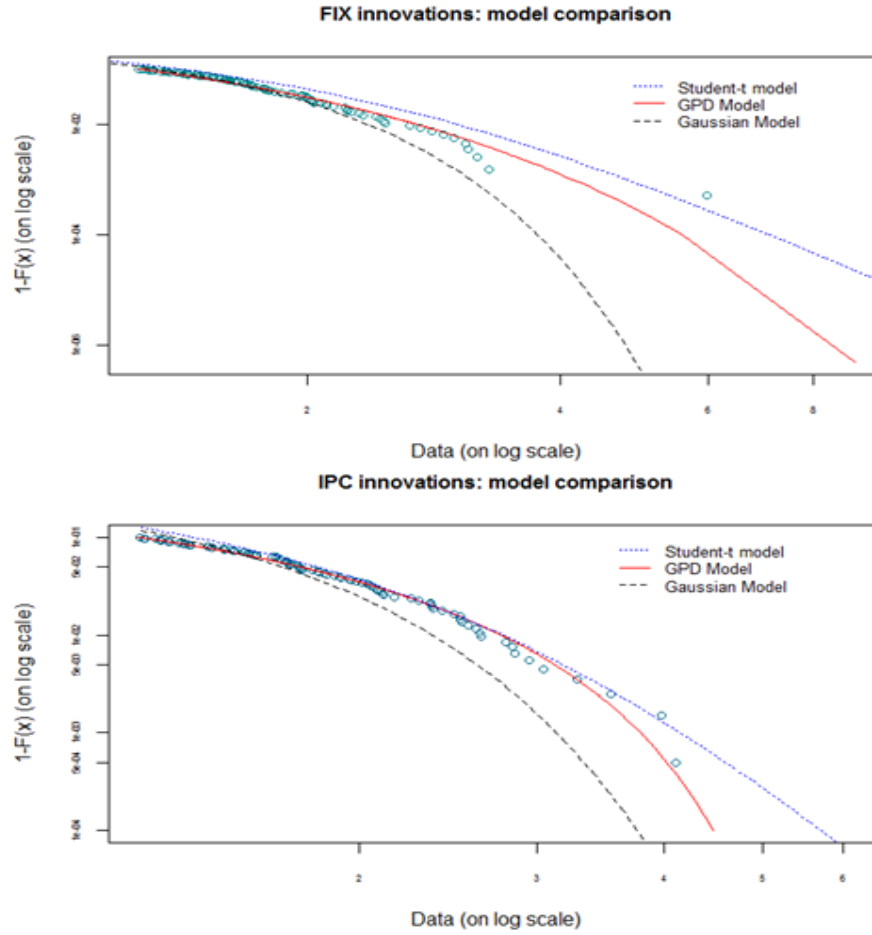
For FIX and IPC series, Table 7 shows the results of the VaR_α^t backtests. Specifically, Table 7 shows the p-values of the binomial tests applied to the samples of VaR violation indicators, the number of historical and (in parentheses) expected violations to each VaR_α^t estimate. The information is separated by confidence level and model fitted to our Mexican data.

On the other hand, Table 8 contains the results of the ES_α^t backtests. Such table shows the p-values of the zero-mean bootstrap tests applied to the excess residuals samples. The information is classified by model fitted to the series and the tested confidence levels.

---

[10] For instance, for the $VaR_{0.995}^t$ we expect to see, on average, $0.005 \times 4{,}060 \approx 20$ violations to such quantile. This means that we would have, approximately, just 20 observations to perform the corresponding binomial and bootstrap tests to evaluate the quality of the estimations of our risk measures.

<div align="center">VaR_α^t backtesting results</div>

	FIX			
α	0.95		0.99	
Model	Hist. Viol. (*203*)	P-value	Hist. Viol. (*41*)	P-value
GARCH-Gaussian	217	0.3132	77	3.25×10^{-7}
GARCH-GPD	205	0.8854	37	0.6359
GARCH-Student-t	169	0.0143	34	0.3432
α	0.995		0.999	
Model	Hist. Viol. (*20*)	P-value	Hist. Viol. (*4*)	P-value
GARCH-Gaussian	45	1.77×10^{-6}	19	6.30×10^{-8}
GARCH-GPD	19	0.9111	3	0.8046
GARCH-Student-t	15	0.2665	4	1

	IPC			
α	0.95		0.99	
Model	Hist. Viol. (*203*)	P-value	Hist. Viol. (*41*)	P-value
GARCH-Gaussian	201	0.9426	72	5.94×10^{-6}
GARCH-GPD	187	0.2642	34	0.3431
GARCH-Student-t	154	0.0003	27	0.0326
α	0.995		0.999	
Model	Hist. Viol. (*20*)	P-value	Hist. Viol. (*4*)	P-value
GARCH-Gaussian	43	9.50×10^{-6}	21	2.36×10^{-9}
GARCH-GPD	18	0.7377	3	0.8046
GARCH-Student-t	14	0.1813	2	04535

<u>Table 7</u>: VaR_α^t estimates backtest results. The table shows, by model fitted to the GARCH innovations, the p-values of the binomial tests applied to the samples of VaR_α^t violation indicators, historical violations to the VaR_α^t estimates and, in parentheses, the expected VaR_α^t violations by confidence level.

The results contained in Tables 7 and 8 show that the estimates of VaR_α^t and ES_α^t obtained through GARCH-Gaussian and GARCH-Student-t models present a similar behavior for the FIX and IPC series.

The GARCH-Gaussian model adequately estimates the VaR_α^t only for relatively low confidence levels ($\alpha = 0.95$). This is reflected in two aspects. The first one is a p-value greater than our significance level in the corresponding binomial test. The second corresponds to the fact that the observed violations to these quantiles are approximately equal to those expected.

For larger confidence levels, the VaR_α^t estimates obtained through the GARCH-Gaussian model are very poor: they strongly underestimate the true value of both time series quantiles. This can be seen in the p-value of the binomial test and in the high number of violations to the corresponding quantiles.

In contrast, the GARCH-Student-t model showed a clear tendency to overestimate VaR_α^t for relatively low confidence levels ($\alpha \leq 0.99$). This can be seen in the low number of violations to these quantiles. However, for higher confidence levels, this model provided quantiles estimates consistent with the history of both time series.

ES_α^t backtesting results			
FIX			
α	0.95	0.99	0.995
Model	P-value	P-value	P-value
GARCH-Gaussian	0.0048	0.0095	0.0429
GARCH-GPD	0.5941	0.7734	0.8051
GARCH-Student-t	0.5948	0.9959	0.6775
IPC			
α	0.95	0.99	0.995
Model	P-value	P-value	P-value
GARCH-Gaussian	0	0.0075	0.0154
GARCH-GPD	0.7949	0.7980	0.7332
GARCH-Student-t	0.9197	0.9536	0.7554

Table 8: Results of the ES_α^t estimates backtest. The table shows the p-values of the bootstrap tests applied to the residual samples of the difference between X_{t+1} and ES_α^t on the days when the VaR was exceeded.

In the case of ES_α^t, the GARCH-Gaussian model provides very poor estimates for all the confidence levels considered. This is shown in the result of our zero-mean bootstrap test: at any confidence level considered, the hypothesis that excess residual samples have a mean statistically equal to zero is strongly rejected.

These results, in conjunction with those of Table 7, reflect that the tail of the conditional distribution of our Mexican financial time series is considerably heavier than that of GARCH type models with Gaussian innovations. Consequently, quantifying the market risk of this type of series using processes with normal innovations (conditional normality) is a poor alternative: the probability that they assign to extreme events turns out to be far below of what is observed in reality.

On the other hand, the GARCH-Student-t models provided good ES_α^t estimates at high confidence levels, ($\alpha > 0.99$). This is reflected in the results of the corresponding bootstrap tests: in these cases, the hypothesis that the excess residual samples have a mean statistically equal to zero is not rejected.

However, because these models significantly overestimated the true value of VaR_α^t for low confidence levels, the ES_α^t estimates for such confidence levels are poor, regardless of what is observed in the backtests of the ES_α^t. To see this, remember that the definition of ES_α^t is "the expected loss when VaR_α^t is exceeded"; therefore, an estimation error in VaR_α^t immediately affects any estimate of ES_α^t.

Finally, our backtests reveal that the GARCH-GPD model was the only one capable of delivering VaR_α^t and ES_α^t estimates consistent with what was historically observed in both time series, for all the confidence levels considered. This can be seen in the results of the binomial and bootstrap test: in all cases, the binomial behavior and zero-mean hypothesis were not rejected. In consequence, this model was the only one able to properly estimate the heaviness of the right tail of the conditional loss distribution of both Mexican time series.

4. Conclusions

Our work shows that the GARCH-GPD model was the only one capable of delivering VaR_α^t and ES_α^t estimates consistent with the history of both time series, for all confidence levels considered. This means that this model was the only one that properly estimated the heaviness of the right tail of the conditional loss distribution of both Mexican time series.

This is particularly relevant in the case of Expected Shortfall: through the applied zero mean bootstrap tests, our paper presents the first piece of evidence of this kind that points that the GARCH-GPD model satisfactorily estimates such risk measure, at high confidence levels, for financial Mexican financial time series. Moreover, our analysis also revealed that the GARCH-Normal (for any confidence level) and GARCH-Student-t (for confidence levels below 99%) models perform poorly when estimating such risk measure in Mexican series.

In addition, our two case studies showed that the GARCH-GPD hybrid model is able to provide reasonable estimates of VaR_α^t and ES_α^t both in times of crisis, where large-scale losses occur in clusters over a considerable period of time, and in times of relative calm in the markets.

Indeed, our backtests' length was large enough to test the GARCH-GPD model in both scenarios. Specifically, they contained in full the 2008 financial crisis and the high volatility period in the FIX Exchange Rate presented in late 2015 and early 2016 (which was strongly related to the worldwide drop in oil prices in that period and to the uncertainty about the monetary policy decision of USA's central bank).

Regarding similar studies on the subject, our results on VaR_α^t estimates obtained from a GARCH-GPD model are in accordance with what Kourouma, L. et al. (2011) (for $\alpha \in \{0.95, 0.99, 0.995\}$), Fernández, V. P. (2003) (for $\alpha \in \{0.99, 0.995\}$), and Aguirre, A. I. et al. (2013) (for $\alpha = 0.95$) found in their respective analyses. However, focusing on relative small confidence levels ($\alpha = 0.95$), they contrast with the particular finding of Fernández, V. P. (2003), who found that the GARCH-GPD model might yield poor VaR_α^t estimates.

For Expected Shortfall, our findings contrast with those of Kourouma, L. et al. (2011) in the international landscape, whose tests suggest that estimates of ES_α^t based on a GARCH-GPD might overestimate the value of such risk measure.

In general, the success of the GARCH-GPD model both in times of high volatility and in those of relative calm may be attributed to the flexibility of the GPD to model the tail of a distribution: through the sign and magnitude of its shape parameter, it can appropriately model heavy, lightweight and even short tails.

In this way, the above characteristic allows the GARCH-GPD model to adjust properly and timely its estimate of the heaviness of the tail of the innovation distribution. Moreover, this feature in the model seems to end up having the same effect on the tail of the conditional loss distribution. In other words, the GPD gives the model sufficient flexibility to adapt to periods where both large

(associated with higher magnitude innovations) and moderate (coupled with rather modest innovations) losses occur.

On the other hand, our work also confirmed that the method to select an optimal backtests threshold given by the simulation study proposed by McNeil, A. J. and Frey, R. (2000) is applicable to Mexican financial time series. This reveals an interesting empirical fact present in Mexican financial series: when filtering their volatility through a GARCH model, it is possible to consider that the tail of the innovation distribution associated with such model is constituted by, approximately, the largest 10% observations of the sample of such innovations.

The facts that: i) the GARCH-GPD model makes up a good tool to quantify the market risk in Mexican financial series; and ii) that such model has shown a better performance than the GARCH-Normal and GARCH-Student-t; have important implications for financial institutions with exposure to Mexican series: it yields a methodology that could make their market risk estimates more accurate. That is, a methodology that could make its regulatory capital closer to its economic capital.

However, the results presented in this paper are only a first step to verify that such an objective is achievable through this methodology. Indeed, further analysis of the effectiveness of this model to describe of Mexican series' returns should be made in the future. In particular, one could investigate if the results found in our application hold for Mexican financial returns of lower frequencies, such as 5 or 10 days. This mainly stems from the fact that the Basel Committee requires banks to calculate market risk capital requirements based on their 10-Day VaR, at 99% confidence.

Thus, an analysis as the one proposed would make possible to know whether Mexican banks would have enough incentives to follow to best practices in market risk management: one could calculate directly the capital requirements that institutions must satisfy and investigate their behavior under a risk-sensitive methodology, such as the one used in this paper.

Nevertheless, when performing an exercise as the described above, one should not overlook the Expected Shortfall. In the future, banking institutions' capital requirements may become more dependent on such risk measure, as set out in the document Minimum Capital Requirements for Market Risk, published by the Basel Committee in January 2016.[11]

Additionally, future studies should explore the behavior of market risk measures for Mexican series in a multivariate context. That is, they must take into account the dependence between several returns of Mexican assets: in both the short and long terms.

In this regard, it is our believe that such an analysis should also be based on EVT: traditional dependence measures, such as the Pearson correlation coefficient, are based on deviations from the mean and give the same weight to extreme observations than to the rest of them. Moreover, most of such dependence measures only consider linear dependence between random variables.

---

[11] Such document is available at http://www.bis.org/bcbs/publ/d352.pdf.

Instead, the dependence analysis of this class of data could be based on the use of copulas. In particular, one could investigate which copula allows a better estimation of conditional risk measures for a portfolio composed by the series used in this paper (or other Mexican series), when its marginal distributions are modeled, in each period, by a GARCH-GPD model (or another more general model that adequately models such distributions).

An analysis like the previous one should also be executed in Mexican series of returns of different frequencies, like 1, 5 or 10 banking days. The latter case could investigate the behavior of the market risk capital requirements of a financial institution in a more realistic context.

5. Acknowledgements

6. Bibliography

Aguirre, A. I. et al. (2013). *Estimación del valor en riesgo en la bolsa mexicana de valores usando modelos de heteroscedasticidad condicional y teoría de valores extremos*. Economía Mexicana NUEVA EPOCA, 22(1):177-205.

Campa, M. A. (2001). *Teoría de Valores Extremos con aplicaciones a medidas de riesgo*. (Bachelor Thesis). Universidad Nacional Autónoma de Mexico.

Embrechts, P. et al. (1997). Modelling extremal events for insurance and finance. Springer-Verlag.

Fernández, V. P. (2003). Extreme value theory: value at risk and returns dependence around the world. Recovered from: http://www.dii.uchile.cl/%7Eceges/publicaciones/ceges51.pdf.

Hansen, P. R. and Lunde, A. (2005). *A forecast comparison of volatility models: does anything beat a GARCH(1,1)?*. Journal of Applied Econometrics, 20: 873–889.

Kourouma, L. et al. (2011). Extreme value at risk and expected shortfall during financial crisis. CERAG. Recovered from: https://halshs.archives-ouvertes.fr/halshs-00658495/document.

López, E. (2013). *Valor en Riesgo del Peso Mexicano y otras divisas latinoamericanas ante eventos de crisis utilizando la Teoría de Valores Extremos*. (Bachelor Thesis). Universidad Nacional Autónoma de Mexico.

Martínez, J. (2014). *Modelos de volatilidad estocástica y su aplicación a la administración de riesgos financieros*. (Bachelor Thesis). Universidad Nacional Autónoma de Mexico.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Journal of empirical finance, 7(3):271-300.

McNeil, A. J. et al. (2005). Quantitative Risk Managements: Concepts, techniques and tools. Princeton University Press.