

Facultad de Economía y Empresa

Departamento de Administración y Economía de la Empresa



**“FORECASTING FINANCIAL RISK IN HIGHLY
VOLATILE SCENARIOS: PARAMETRIC
DISTRIBUTIONS AND THE GRAM-CHARLIER
DENSITY”**

by

ANDRÉS MORA VALENCIA

Supervisors:

JAVIER PEROTE PEÑA

ESTHER B. DEL BRÍO GONZÁLEZ

RESUMEN EN INGLES

Salamanca, May 2014

CHAPTER II: Risk forecasting in the subprime and sovereign debt crisis

II.1. Introduction

Recent literature has extensively focused on studying the causes and consequences of the subprime and sovereign debt crisis – e.g. see Shiller (2008), Kolb (2010), Duca et al. (2010). The origin of the crisis was the sharp decline of the U.S. housing prices in 2006, triggered by the enormous amount of subprime mortgages contracted in a decade characterized by low interest rates and irrational expectations about the sustainability of real estate market prices. The crisis was amplified by different mechanisms such as the highly leveraged banks, the deregulation of the financial system, the growth of securitization, the misbehavior of rating agencies and the so-called ‘credit crunch’ (Hull, 2009), and it caused the bankruptcy of many investment banks (Bear Stearns and Lehman Brothers) and the bailout of insurance companies (AIG). This situation caused panic in the financial markets (Gorton, 2009), dramatically increased systemic risk (Harrington, 2009) and turned into a global financial crisis and a major recession (Mishkin, 2011).

From a risk management perspective, the crisis increased volatility in the financial markets (Schwert, 2011) and demanded new risk measures and methodologies capable of accurately estimating the regulatory capital of financial institutions. Risk measures are commonly obtained by quantile-based methods (Dowd and Blake, 2006), and among these the most widely used is the Value-at-Risk (hereafter, VaR). Different approaches have also been used to compute VaR (see Jorion, 2006; Alexander, 2009; Hubbert, 2012; and Hull, 2012, among others) but there is no consensus on the most appropriate methodology. Former VaR models have been criticized because the normality assumption involves risk underestimation, and thus skewed and heavy-tailed distributions (e.g., Bali and Theodossiou, 2008) have been proposed.

The current paper expands on this issue by comparing the performance of VaR forecasts obtained by the normal distribution (benchmark) to four natural candidates that account for the heavy tails of stock returns: the Student’s t , a skewed variant of the Student’s t distribution (Hansen, 1994), the extreme value theory (EVT) approach (Embrechts et al., 1997; Reiss and Thomas, 1997; Coles, 2001; and McNeil et al., 2005) and the semi-nonparametric approach based on the Gram-Charlier (GC) density, which is an expansion around the normal density allowing for skewness and excess kurtosis (Gram, 1879; Charlier, 1905; Edgeworth, 1907).

The VaR forecasting performance of the models is analyzed for the high volatility scenario of the recent subprime and sovereign debt crises. Furthermore, we compare how VaR measures are affected by the occurrence of extreme events in different

economic areas, e.g. United States, Europe and emerging markets. For this purpose three leading world stock indices are considered: MSCI Europe, MSCI USA and MSCI Emerging Markets. For these indices, historical daily losses are compared with the maximum loss forecasted for each method considering a one-day-ahead horizon. VaR forecasts are computed by assuming an ARMA-GARCH model for the conditional mean-variance and computing the quantile of the assumed distribution at 99% confidence level. This technique is known as backtesting (Zumbach, 2006). According to this procedure, it is expected that for 1% of the cases (days of the sample) the historical losses will fall outside the estimated VaR when VaR at 99% is calculated. This idea allows a straightforward implementation of VaR backtesting or forecasting performance tests (see, e.g., Christoffersen, 2003). We are interested in showing the impact of the recent crises on forecasts of the VaR methodology performance and thus the backtesting period is divided into two subperiods: pre-crisis and crisis, the latter including the subprime and the sovereign debt crises.

The results show that both the normal and Student's t are inadequate for high confidence levels and/or high volatility periods, although the skewed Student's t (hereafter, skewed- t) outperforms the Student's t . On the other hand, GC and EVT produce accurate VaR forecasts in these contexts. Therefore the optimal VaR model depends not only on the assumed confidence level but also on the market conditions observed.

The rest of the paper is organized as follows: Section 2 presents the models and VaR methodology, Section 3 analyzes the data and the empirical results on VaR forecasting, and section 4 summarizes the main results of the article.

II. 2. Models and Methodology

Since Mandelbrot (1963) the normality assumption of stock returns is deemed inappropriate, revealing the following stylized empirical regularities (Cont, 2001): (1) a sharp peak at the mean; (2) heavy tails; (2) skewness; (4) volatility clustering; (5) slow decay in the autocorrelation function of the absolute returns. To account for the leptokurtosis implied by the first two properties, the use of non-Gaussian distributions is proposed, of which the Student's t is the most widely used. Incorporating asymmetries requires, however, the use of other densities such as the skewed- t (Lambert and Laurent, 2001; and Giot and Laurent, 2003). Alternatively, for purposes of measuring risk, the EVT has directly focused on reproducing the behavior at the tails. Within the EVT framework, different approaches have been proposed, such as the generalized extreme value distribution or the generalized Pareto distribution (GPD). In this article, we compare the VaR performance of both the Student's t and skewed- t to the so-called peaks over threshold (POT) method, which is based on the GPD (Smith, 1989; Davison and Smith, 1990; and Leadbetter, 1991). Furthermore, we also incorporate the semi-nonparametric estimation that is based on the asymptotic properties of the GC type A series when approximating a frequency function (see Kendall and Stuart, 1977, p.

168–72). Most of the financial literature about semi-nonparametric methodologies is devoted to price derivatives following the seminal papers of Jarrow and Rudd (1982) and Corrado and Su (1996). However, only a few papers focus on the Gram-Charlier application to risk management (Mauleón and Perote, 2000; Mauleón, 2003; Marumo and Wolff, 2007; Puzanova et al. 2009; Níguez and Perote, 2012).

On the other hand, stock returns also seem to have a small predictable component in the conditional mean that has traditionally been modeled according to simple ARMA structures. Nevertheless squared returns exhibit particular dynamics (conditional heteroskedasticity, clusters of volatility and long memory) that have been extensively studied since Engle (1982) and Bollerslev (1986) introduced ARCH and GARCH models. As we focus on VaR performance due to the distributional hypotheses, the model implemented in this article incorporates an ARMA(1,1) and a GARCH(1,1) for modeling the conditional mean and variance, in accordance with the common use in risk management applications (see e.g. McNeil et al., 2005; or Jondeau et al., 2007). We define the complete model in equations (1) to (4) below.

$$r_t = \mu_t + \sigma_t Z_t, \quad (1)$$

$$\mu_t = \varphi + \phi\mu_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t, \quad (2)$$

$$z_t = \varepsilon_t/\sigma_t, \quad Z_t \sim G(0,1), \quad (3)$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \quad (4)$$

where, for the sake of comparison, different standardized (i.e. zero mean and unit variance) density specifications are considered for G . In particular, we consider the four following probability density functions (pdf).

(i) The normal pdf:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}. \quad (5)$$

(ii) The Student's t pdf:

$$t(z) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi(\nu-2)}\Gamma(\frac{\nu}{2})} \left(1 + \frac{z^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \quad (6)$$

where Γ is the gamma function and ν is the degrees of freedom parameter.

(iii) The skewed-t pdf by Fernández and Steel (1998):

$$g(z) = \begin{cases} -\frac{2}{\gamma+\frac{1}{\gamma}} t(\gamma z) & \text{for } x < 0, \\ \frac{2}{\gamma+\frac{1}{\gamma}} t\left(\frac{z}{\gamma}\right) & \text{for } x \geq 0, \end{cases} \quad (7)$$

where γ is the shape parameter, which incorporates the skewness, and $t(z)$ is the Student's t pdf in equation (6).

(iv) The GC Type A density is given by:

$$f(z, \mathbf{d}) = (1 + \sum_{s=1}^n d_s H_s(z)) \phi(z), \quad (8)$$

where $\phi(z)$ denotes the normal pdf in equation (5), $\mathbf{d}' = (d_1, \dots, d_n) \in \mathbb{R}^n$ and H_s is the Hermite polynomial (HP) of order s , which can be defined in terms of the derivatives of $\phi(z)$ as

$$\frac{d^s \phi(z)}{dz^s} = (-1)^s H_s(z) \phi(z) \quad (9)$$

In particular, the first eight HP are: $H_1(z) = z$, $H_2(z) = z^2 - 1$, $H_3(z) = z^3 - 3z$, $H_4(z) = z^4 - 6z^2 + 3$, $H_5(z) = z^5 - 10z^3 + 15z$, $H_6(z) = z^6 - 15z^4 + 45z^2 - 15$, $H_7(z) = z^7 - 21z^5 + 105z^3 - 105z$, $H_8(z) = z^8 - 28z^6 + 210z^4 - 420z^2 + 105$.

It is noteworthy that some authors use GC density to denote a positive transformation of the Gallant and Nychka (1987) type (e.g. Jondeau and Rockinger, 2001, or León et al. 2009) of the truncated GC series. We implement the original GC Type A expansion, which is simpler and more useful for VaR applications. Furthermore, most of the empirical studies truncate the expansion in $n = 4$ and employ only two terms of the expansion, d_3 and d_4 , since these terms account for skewness and excess kurtosis, respectively. Then we initially estimate VaR taking into account the GC expansion truncated at the fourth order, estimated via method of moments (GC-MM) and maximum likelihood (GC-ML1). We follow the procedure proposed by Del Brio and Perote (2012) to estimate the density parameters in two steps. First we estimate the conditional mean and variance using Quasi Maximum Likelihood (QML) and obtain the standardized residuals and, second, we estimate the d_s parameters for the standardized residuals. For the GC-ML1 and GC-MM models, the estimates are obtained by maximizing the log-likelihood function, Eq. (53), and applying directly the Eqs. (39)–(46) from Brio and Perote (2012, p. 534–5), respectively. Furthermore, we consider a second GC model expanded up to the eighth term and estimated by maximum likelihood (GC-ML2). For this model, we identify the “optimal” truncation order using the Akaike Information Criterion (AIC).

In this model, the estimated VaR with a confidence level α is computed as the estimated α -quantile, $\hat{q}_\alpha(z)$, of the assumed G distribution. Therefore, the predicted VaR for the variable r at the time horizon $t+1$ and with confidence level α is given in equation (10).

$$\text{VaR}_{t+1}^\alpha = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} \hat{q}_\alpha(z), \quad (10)$$

where $\hat{\mu}_{t+1}$ and $\hat{\sigma}_{t+1}$ are the predictions for the mean and standard deviation conditioned by the available information at time t , Ω_t , based on the ARMA-GARCH model described in equations (2) and (4).

Alternatively, VaR can also be computed by the EVT methodology through two different approximations: block maxima and POT. We implement the latter method following the two-step procedure proposed by McNeil and Frey (2000).

In the first step, the ARMA(1,1)-GARCH(1,1) model is fitted using QML and in the second step the so-obtained standardized residuals are used to implement the POT methodology using 10% of the tail of the distribution as the threshold. Thus, $\hat{q}_\alpha(Z)$ is given by

$$\hat{q}_\alpha(Z) = u + \frac{\beta}{\xi} \left(\left(\frac{1-\alpha}{N_u/n} \right)^{-\xi} - 1 \right) \quad (11)$$

where u is the estimated threshold, N_u is the number of exceedances over the threshold, n is the sample size (thus N_u/n is a non-parametric estimator of the empirical distribution tail) and β and ξ are the scale and shape parameters of the GPD. The cumulative distribution function (cdf) of the GPD distribution is given by

$$F(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi}, & \xi \neq 0, \\ 1 - e^{-x/\beta}, & \xi = 0. \end{cases} \quad (12)$$

The weakness of the EVT lies in the threshold selection, which involves a tradeoff between bias and variance in the estimation of the parameters, especially the shape parameter ξ . This parameter can be estimated by bootstrapping or graphical techniques, but there is no optimal method to choose the appropriate threshold. Some empirical studies have shown that a good approximation is to choose as threshold 5% or 10% of the data in the tail of the distribution. We adopt the latter, as in McNeil and Frey (2000).

II. 3. Empirical Application

3.1. Data and in-sample results

In this section, we compare the performance of the above-mentioned models for computing VaR for different world stock indices: MSCI Europe, MSCI USA and MSCI Emerging Markets (EM) and in two different volatility scenarios, which we call pre-crisis and crisis period. All data were obtained from Datastream; for more details see the Appendix. The sample comprises almost 16 years of daily data from December 1997 to the first quarter of 2013. We split this sample into two sub-samples and choose as the crisis starting date July 2006, one year before the date when Bear Stearns hedge funds reported massive losses. Table 1 displays the descriptive statistics for continuously compounded returns of these series, defined as $r_t = 100 \log(P_t/P_{t-1})$, where P_t represents the corresponding price index.

Table 1. Descriptive statistics for European stock index returns

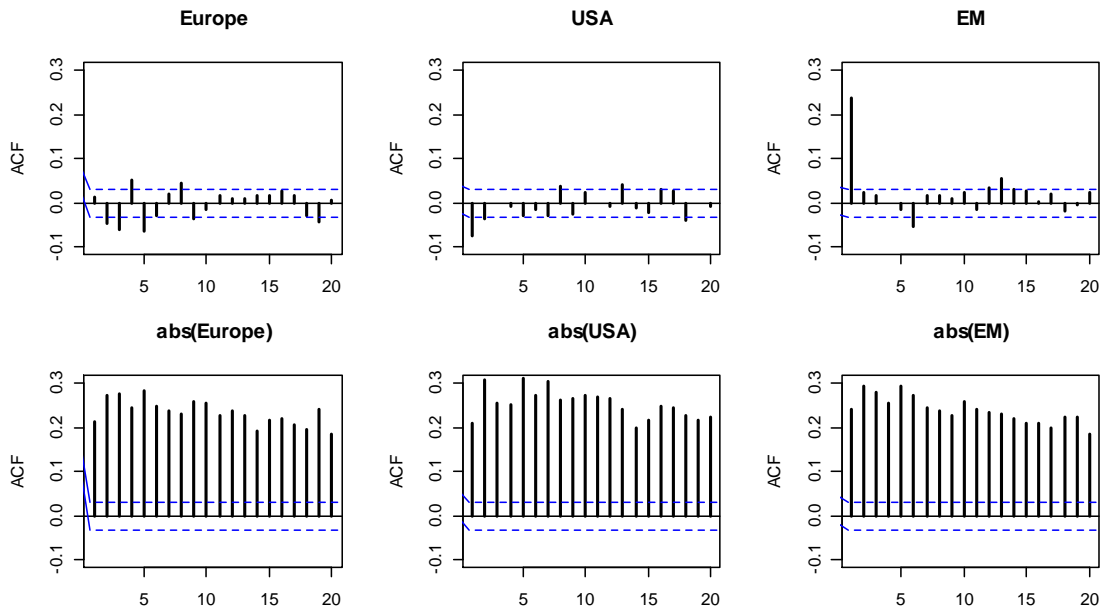
	Total period	Pre-crisis period	Crisis period
Panel A: MSCI Europe			
Mean	0.0025	0.0079	-0.0028
Median	0.0322	0.0311	0.0326
Standard deviation	1.4648	1.1567	1.7189
Variance	2.1457	1.3379	2.9546
Excess kurtosis	6.1617	2.3405	5.5684
Skewness	-0.0963	-0.2128	-0.0492
Range	20.8762	10.8752	20.8762
Minimum	-10.1783	-5.6814	-10.1783
Maximum	10.6979	5.1938	10.6979
Panel B: MSCI USA			
Mean	0.0036	-0.0071	0.0143
Median	0.0258	0.0000	0.0489
Standard deviation	1.3131	1.1375	1.4681
Variance	1.7241	1.2939	2.1552
Excess kurtosis	7.8484	2.7571	8.8208
Skewness	-0.1829	0.0964	-0.3154
Range	20.5563	11.7713	20.5563
Minimum	-9.5137	-6.1609	-9.5137
Maximum	11.0426	5.6104	11.0426
Panel C: MSCI EMERGING MARKETS			
Mean	0.0268	0.0323	0.0213
Median	0.1094	0.1128	0.0957
Standard deviation	1.2850	1.0174	1.5060
Variance	1.6511	1.0350	2.2681
Excess kurtosis	7.7059	2.3419	7.2308
Skewness	-0.5113	-0.6116	-0.4406
Range	20.0673	9.8876	20.0673
Minimum	-9.9944	-5.8312	-9.9944
Maximum	10.0729	4.0564	10.0729

Descriptive statistics show that the mean return for the European stock index is positive in the pre-crisis period, but becomes negative in the crisis period. However, the mean return for the other indices is positive in the crisis period. Moreover, volatility and kurtosis increase in the crisis period compared to the pre-crisis period. The variation range also reveals that the data are more disperse in the crisis period, almost double in contrast with the pre-crisis period. Skewness of the MSCI USA seems to be positive in the pre-crisis period and negative in the crisis period, which means that lower (higher) returns were more likely to be obtained in the pre-crisis (crisis) period. The MSCI Europe and MSCI Emerging Markets indices, however, exhibit negative skewness in both periods. The deviations of the median from the mean and the values of the excess kurtosis justify the use of VaR based on skewed and leptokurtic distributions such as the skewed-t.

Figure 1 displays the autocorrelation function (ACF) of the return series (upper graphs) and the absolute return series (lower graphs) using the total sample. The ACF of the

return series shows that there is a slightly autoregressive structure in the data and thus either an AR(1) or ARMA(1,1) structure might be identified. The ACF of the absolute return series reveals a strong presence of conditional heteroskedasticity in the data that can be adequately captured by a GARCH(1,1) process.

Figure 1. Autocorrelation functions for the selected stock index returns in levels and absolute values



Next we proceed to choose between the three plausible models for conditional mean - white noise, AR(1) and ARMA(1,1) - according to accuracy criteria. Table 2 shows the log-likelihood values of these three models combined with a GARCH(1,1) for modeling conditional variance and under different distributional hypotheses, either normal, Student's t or skewed-t. The results show clear evidence in favor of the autoregressive models but they do not strongly support the ARMA(1,1) versus the AR(1) model. We choose the ARMA(1,1) since it has slightly higher log-likelihood values and it nests the AR(1).

Table 2. Log-likelihood for different conditional mean-variance models and under different distributional assumptions.

	EUROPE	USA	EM
Panel A: Normal			
GARCH(1,1)	-6358.328	-5908.279	-6052.282
AR(1)-GARCH(1,1)	-6358.105	-5904.808	-5944.017
ARMA(1,1)-GARCH(1,1)	-6356.587	-5899.511	-5943.020
Panel B: Student's t			
GARCH(1,1)	-6337.145	-5833.96	-6002.947
AR(1)-GARCH(1,1)	-6336.956	-5829.498	-5899.354
ARMA(1,1)-GARCH(1,1)	-6336.815	-5822.885	-5897.602

Panel C: Skewed Student's t			
GARCH(1,1)	-6329.544	-5826.936	-5979.551
AR(1)-GARCH(1,1)	-6329.466	-5820.701	-5886.280
ARMA(1,1)- GARCH(1,1)	-6324.653	-5809.244	-5883.286

Table 3 presents the Maximum Likelihood (ML) estimates of the parameters of the ARMA(1,1)-GARCH(1,1) model under the three distributional hypotheses. P-values for testing the significance of each parameter are given in parentheses. These values show that the GARCH(1,1) parameters are statistically significant but not all the parameters of the ARMA(1,1) are statistically different from zero. This fact is in line with the 'small predictable component of the conditional mean' stylized fact featured by stock returns. Moreover GARCH(1,1) processes exhibit persistent behavior since they are close to the non-stationarity (i.e., $\alpha+\beta$ is close to one). This fact captures the 'long memory' or the 'persistence of conditional variance' usually found in this type of data. With all this information, we decided to use the ARMA(1,1)-GARCH(1,1) model when implementing the backtesting technique to investigate the performance of the different distributional hypotheses for VaR computation. Table 3 also includes the estimates for the shape parameter (degrees of freedom) and skew parameter of the Student's t distributions. Both parameters are significant, which shows that the distribution is leptokurtic and asymmetric.

Table 3. Parameters of the ARMA(1,1)-GARCH(1,1).

	EUROPE	USA	EM
Panel A: Normal			
φ	0.0141 (0.0374)	0.0111 (0.0134)	0.0798 (0.0000)
ϕ	0.7699 (0.0000)	0.7682 (0.0000)	0.1338 (0.0975)
θ	-0.7896 (0.0000)	-0.8124 (0.0000)	0.1159 (0.1549)
ω	0.0180 (0.0000)	0.0139 (0.0000)	0.0294 (0.0000)
α	0.0901(0.0000)	0.0783 (0.0000)	0.1005 (0.0000)
β	0.9017 (0.0000)	0.9130 (0.0000)	0.8793 (0.0000)
Panel B: Student's t			
φ	0.0836 (0.0415)	0.0179 (0.0019)	0.0913 (0.0000)
ϕ	-0.3100 (0.5733)	0.7198 (0.0000)	0.0870 (0.2712)
θ	0.3194 (0.5576)	-0.7711 (0.0000)	0.1506 (0.0569)
ω	0.0163 (0.0002)	0.0099 (0.0008)	0.0257 (0.0000)
α	0.0849 (0.0000)	0.0799 (0.0000)	0.0952 (0.0000)
β	0.9086 (0.0000)	0.9169 (0.0000)	0.8876 (0.0000)
ν	10 (0.0000)	6.4203 (0.0000)	8.7273 (0.0000)
Panel C: Skewed Student's t			
φ	0.0118 (0.0119)	0.0127 (0.0016)	0.0708 (0.0002)
ϕ	0.7628 (0.0000)	0.7160 (0.0000)	0.0262 (0.7447)
θ	-0.7962 (0.0000)	-0.7821 (0.0000)	0.2007 (0.0114)

ω	0.0165 (0.0001)	0.0094 (0.0008)	0.0239 (0.0000)
α	0.0808 (0.0000)	0.0774 (0.0000)	0.0885 (0.0000)
β	0.9115 (0.0000)	0.9183 (0.0000)	0.8941 (0.0000)
ν	10 (0.0000)	6.9753 (0.0000)	9.2225 (0.0000)
γ	0.8980 (0.0000)	0.8945 (0.0000)	0.8819 (0.0000)

P-values for the t-test in parentheses

The parameters of the EVT and the semi-nonparametric VaR methodologies implemented in the article are displayed in Table 4. In these cases two-step estimation was implemented following Del Brio et al. (2011), i.e. returns were filtered according to the ARMA(1,1)-GARCH(1,1) estimated in the first step by QML. The shape parameter for EVT, ξ in Panel A, is not significantly different from zero for Europe and Emerging Market indices but it is for the USA. This means that, after filtering the returns by an ARMA(1,1)-GARCH(1,1) model, the standardized residuals exhibit medium-tailed distributions (note that if $\xi = 0$ then GPD becomes the exponential distribution). Regarding the GC densities three alternative models are estimated: the GC expanded to the fourth term and estimated either by the method of moments (Panel B) or maximum likelihood (Panel C) and the GC expanded to the fourth order (Panel D). In all cases parameters d_3 and d_4 confirm the presence of (negative) skewness and leptokurtosis. Nevertheless, not all the parameters of the larger expansion (GC-ML2) are significantly different from zero. Despite this fact we maintain the polynomial structures in order to compare the effects not only of the method of estimation but also of the expansion length in the VaR forecasting performance implemented in the next section.

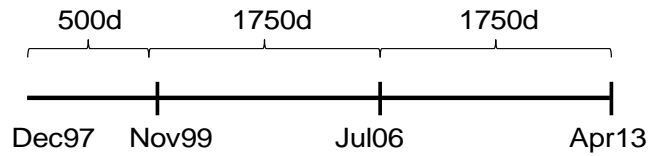
Table 4. Parameters of the standardized EVT and GC.

	EUROPE	USA	EM
Panel A: EVT			
ξ	-0.0290 (0.2861)	-0.1963 (0.0000)	-0.0674 (0.0670)
β	0.4949 (0.0000)	0.5836 (0.0000)	0.5427 (0.0000)
Panel B: GC-MM			
d_3	-0.0338	-0.0759	-0.0499
d_4	0.0264	0.0675	0.0421
Panel C: GC-ML1			
d_3	-0.0345 (0.0000)	-0.0449 (0.0000)	-0.0370 (0.0000)
d_4	0.0235 (0.0000)	0.0347 (0.0000)	0.0305 (0.0000)
Panel D: GC-ML2			
d_3	-0.0242 (0.0035)	-0.0651 (0.0000)	-0.0405 (0.0000)
d_4	0.0270 (0.0000)	0.0285 (0.0012)	0.0417 (0.0000)
d_5	0.0056 (0.0418)	-0.0150 (0.0028)	-0.0031 (0.1944)
d_6	0.0021 (0.1232)	0.0000 (0.5130)	0.0049 (0.0095)
d_7	0.0000 (0.4540)	-0.0027 (0.0003)	-0.0008 (0.0598)
d_8	0.0004 (0.0186)	0.0008 (0.0032)	0.0005 (0.0212)

3.2. Backtesting

In order to test the validity of the distributional assumptions regarding stock returns (normal, Student's t, skewed-t, GC and EVT), the historical series, r_1, \dots, r_m , are compared to the VaR_t^α predicted for the day $t = \{n + 1, \dots, m\}$ by using a time window of the n previous days. In order to test the performance of these models on different volatility scenarios (pre-crisis and crisis period), we implement the backtesting technique as illustrated in Figure 2. We consider a time window of 500 days for computing every one-step-ahead VaR prediction and a total period of 3500 days as the backtesting or out-of-sample period. The backtesting period is divided into two identically sized sub-periods of 1750 days each: the pre-crisis period (November 1999 – July 2006) and the crisis period (July 2006 – 1st quarter of 2013).

Figure 2. Backtesting periods



The predicted VaR is compared to the observed return at 99% confidence level. Therefore, when calculating VaR at 99%, we expect that in 1% of the backtesting days the negative returns will exceed VaR predictions. These values are referred to as ‘violations’ or ‘exceptions’. More specifically, if I_t is the indicator defined in equation (21),

$$I_t := 1_{\{r_{t+1} > \text{VaR}_t^\alpha\}}, \quad (21)$$

a ‘violation’ occurs when $r_{t+1} > \text{VaR}_t^\alpha$, and then the indicator function takes on value 1. Otherwise, I_t takes on value 0 [whenever $r_{t+1} \leq \text{VaR}_t^\alpha$]. Therefore, if the VaR methodology is adequate, it is expected that the violation indicator function values will behave as realizations of independent and identically distributed (iid) Bernoulli experiments with success probability equal to $1 - \alpha$, i.e. $\sum_{t=1}^m I_t \sim \text{Bin}(m, 1 - \alpha)$. Thus the null hypothesis that ‘the model adequately estimates VaR’ can be tested by a straightforward one-sided binomial hypothesis test. The alternative hypothesis suggests that the method underestimates or overestimates the VaR calculation depending on the number of expected violations.

We apply this backtesting procedure and binomial test to investigate the performance of ARMA-GARCH models with different distributional assumptions: normal, Student's-t, skewed-t, GC and EVT-POT. It is noteworthy that every prediction is based on the estimated distribution conditioned by the available information set, which is updated as new information is released to the market. Therefore the performance of every

distribution depends on its capability of adaptation to the new volatility environment and the occurrence of extreme events. Figures 3 – 6 illustrate how distributional parameters adapt to the time-varying scenario for the MSCI Europe index (the corresponding figures for the USA and Emerging Markets indices present similar patterns and are available upon request). In particular, Figure 3.A and Figure 3.B display the changes of the degrees of freedom parameter of Student's t and skewed-t over time. Note that the Y-axis is truncated at 10 since for bigger values the distribution is not very different from the standard normal. It is clear that this parameter decreases (increases) as the volatility increases (decreases) but it remains above 4 so that kurtosis is still well-defined. Figure 3.C displays the evolution of the skew parameter, which is relatively stable throughout the pre-crisis period but has a decreasing pattern at the beginning of the crisis period.

Figure 3: Student's t distributions time-varying parameters for MSCI EUROPE index returns.

Figure 3.A. Student's t shape (degrees of freedom) parameter

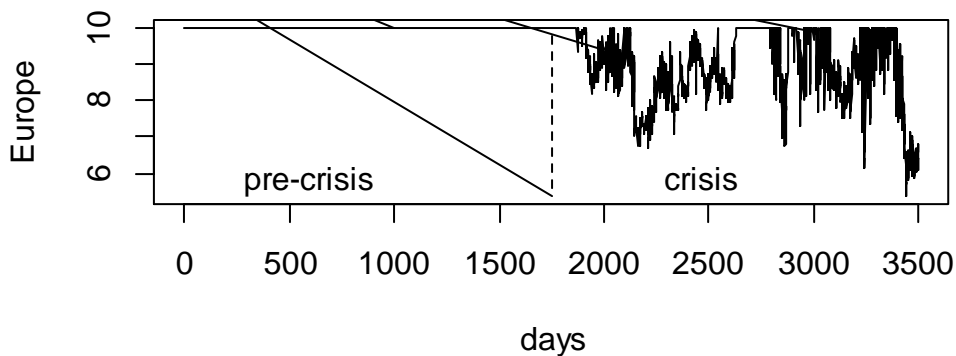


Figure 3.B. Skewed-t shape (degrees of freedom) parameter

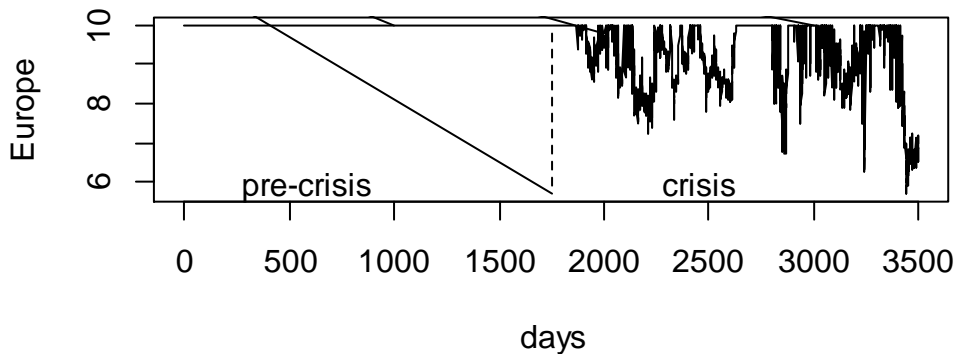


Figure 3.C. Skewed-t skew parameter

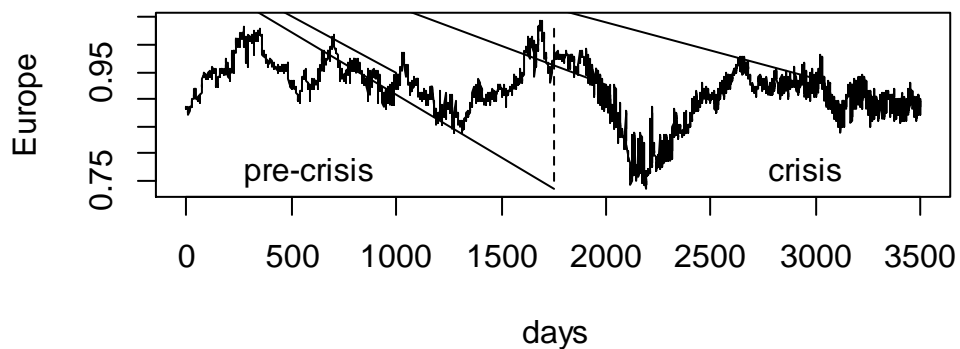


Figure 4 displays the shape parameters for the GPD implemented in the EVT model. Parameter ξ seems to be negative (evidencing shorter tails) in the pre-crisis period and positive (heavier tails) at the beginning of the crisis period and becomes negative again at the end, and parameter β is very volatile and seems to increase in the crisis period.

Figure 4: Generalized Pareto time-varying parameters for MSCI EUROPE index returns.

Figure 4.A. Generalized Pareto shape parameter (ξ)

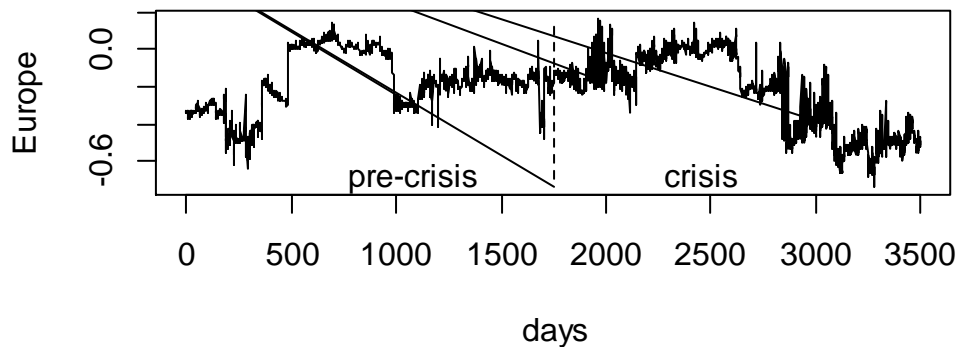
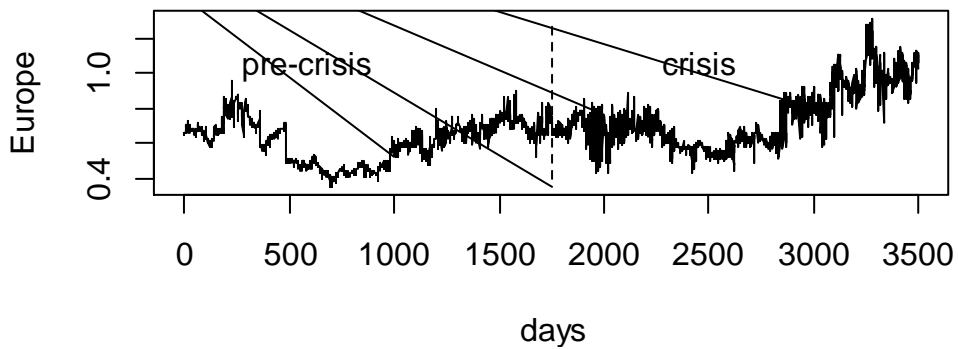


Figure 4.B. Generalized Pareto scale parameter (β)



Figures 5.A. and 5.B. illustrate the dynamics for GC-ML1 parameters, d_3 and d_4 , which account for skewness and excess kurtosis, respectively. Parameter d_3 estimates are positive within a specific time range in the pre-crisis period, whereas this parameter remains negative for the whole crisis period. Regarding parameter d_4 estimates, they are positive and increasing with extreme values occurrence.

Figure 5: GC-ML1 time-varying parameters for MSCI EUROPE index returns.

Figure 5.A. GC-ML1 d_3 parameter

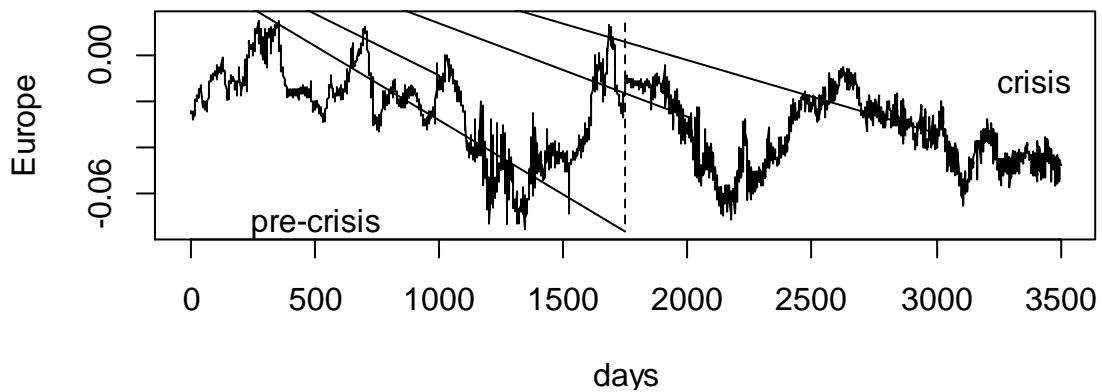


Figure 5.B. GC-ML1 d_4 parameter

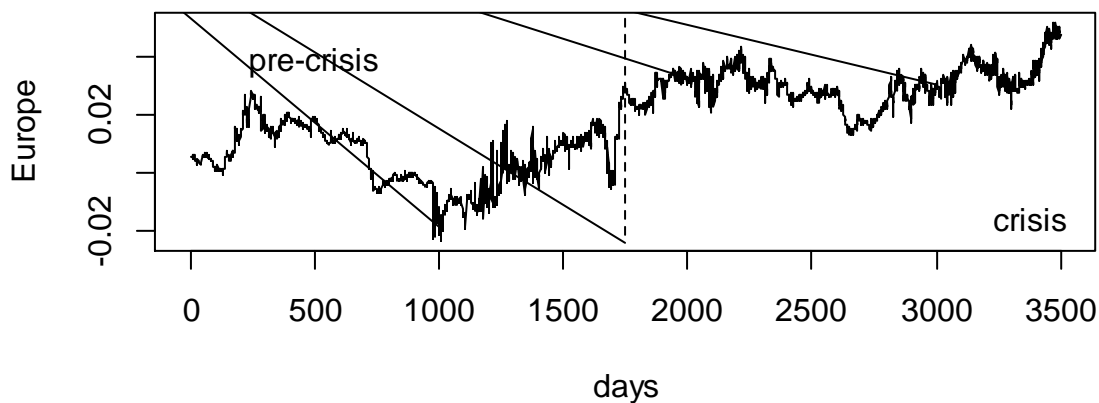


Figure 6 shows the behavior of the parameters of the GC expanded up to the eighth term (GC-ML2). Note that in the first part of the crisis period, only d_3 and d_4 are included according to the AIC criterion. In this context, the interpretation of these parameters is consistent with the expected values of skewness (negative) and excess kurtosis (positive). Nevertheless, when a larger expansion is chosen, skewness is captured by the interaction among the odd parameters and heavy-tailed patterns are featured by the interaction among the even parameters. Therefore the insights about the combinations that may incorporate a certain degree of skewness and kurtosis are not straightforward. This is what happens, for instance, in the last part of the crisis period, where odd (even) parameters seem to be significantly positive (negative), although they also present an extreme volatility. Furthermore, the high presence of extreme values leads to both asymmetries and heavy tails and thus the relation between even and odd parameters also plays an important role in accounting for them.

Figure 6: GC-ML2 time-varying parameters for MSCI EUROPE index returns.

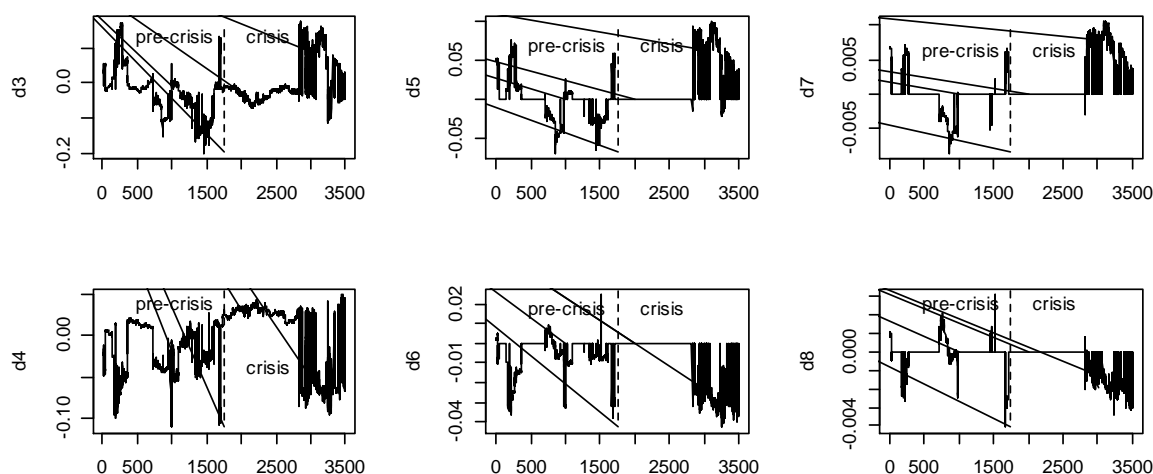


Table 5 displays the number of exceptions and the p-value for the binomial test (in parentheses) for the seven models (ARMA-GARCH-normal, ARMA-GARCH-t, ARMA-GARCH-skewed-t, ARMA-GARCH-EVT, ARMA-GARCH-GC-MM, ARMA-GARCH-GC-ML1, ARMA-GARCH-GC-ML2) at 99% confidence level and for the three stock indices. Table 4 shows two different panels, one for each backtesting period (pre-crisis and crisis periods).

For the pre-crisis period (Panel A), the normal distribution method underpredicts the tail behavior (i.e. the number of exceptions is significantly higher than the expected value) for all cases, while the skewed-t and GC-ML2 fail for the MSCI USA case. The t, EVT, GC-MM and GC-ML1 models cannot be rejected at 5 per cent confidence for any of the series. In the crisis period (Panel B), the normal significantly underpredicts the VaR for all series, while the method based on Student's t cannot be rejected on only one occasion (EM). The skewed-t performance, however, tends to overpredict risk (i.e. it involves overly conservative risk measures) although it cannot be rejected in any of the series. The EVT approach and GC models perform well for all cases since these methods focus on modeling the extreme values.

These results are consistent with the usual evidence found in stock return VaR performances, i.e. the normal is strongly rejected (especially for high confidence levels and volatile scenarios), the Student's t might be useful only if kurtosis and skewness are not severe and the skewed-t is an alternative to capture skewness although it might not be the best option for capturing kurtosis. The EVT and the GC densities involve accurate market risk measures since the former focuses on extreme values and the latter is very flexible to adapt to different scenarios with a variable number of parameters. Nevertheless we find that, for prediction purposes, the larger expansions do not seem to provide the best outcomes and also that the simpler MM estimation procedures involve accurate VaR forecasts. These results are consistent with Del Brio et al. (2011) and Níguez and Perote (2012), although the former article only focuses on in-sample fit and the latter analyzes VaR forecasting using positive transformations of GC series.

Table 5. VaR forecasting performance of different models

	EUROPE	USA	EM
Panel A: Pre-crisis period (November 1999 - July 2006)			
99% 1750 days	Expected number of exceptions = 17		
ARMA-GARCH-normal	26 (0.0331)	27 (0.0203)	38 (0.0000)
ARMA-GARCH-t	20 (0.3048)	13 (0.1685)	25 (0.0522)
ARMA-GARCH-skewed-t	16 (0.4197)	10 (0.0380)	17 (0.5157)
ARMA-GARCH-EVT	21 (0.2296)	22 (0.1670)	19 (0.3908)
ARMA-GARCH-GC-MM	21 (0.2296)	18 (0.4842)	19 (0.3908)
ARMA-GARCH-GC-ML1	21 (0.2296)	18 (0.4842)	20 (0.3048)
ARMA-GARCH-GC-ML2	22 (0.1670)	26 (0.0332)	20 (0.3048)
Panel B: Crisis period (July 2006 - first quarter of 2013)			
99% 1750 days	Expected number of exceptions = 17		
ARMA-GARCH-normal	41 (0.0000)	52 (0.0000)	31 (0.0021)
ARMA-GARCH-t	29 (0.0070)	29 (0.0070)	22 (0.1670)
ARMA-GARCH- skewed-t	12 (0.1104)	13 (0.1685)	12 (0.1104)
ARMA-GARCH-EVT	21 (0.2296)	21 (0.2296)	14 (0.2413)
ARMA-GARCH-GC-MM	20 (0.3048)	12 (0.1104)	16 (0.4197)
ARMA-GARCH-GC-ML1	20 (0.3048)	21 (0.2296)	17 (0.5157)
ARMA-GARCH-GC-ML2	25 (0.0522)	19 (0.3908)	15 (0.3265)

P-values for the binomial test are in parentheses. EVT considers a 10% threshold.

Finally, the stock index returns and their corresponding forecasted VaR at 99% confidence are plotted in Figure 7 for the three series and the whole backtesting period. It is clear that the normal (red line) is the distribution that produces less conservative VaR measures (lower values) and the EVT (dark blue line) and GC (light blue line) the ones that result in higher VaR predictions.

Figure 7. VaR at 99% in the backtesting period under different specifications

Figure 7.A. MSCI EUROPE

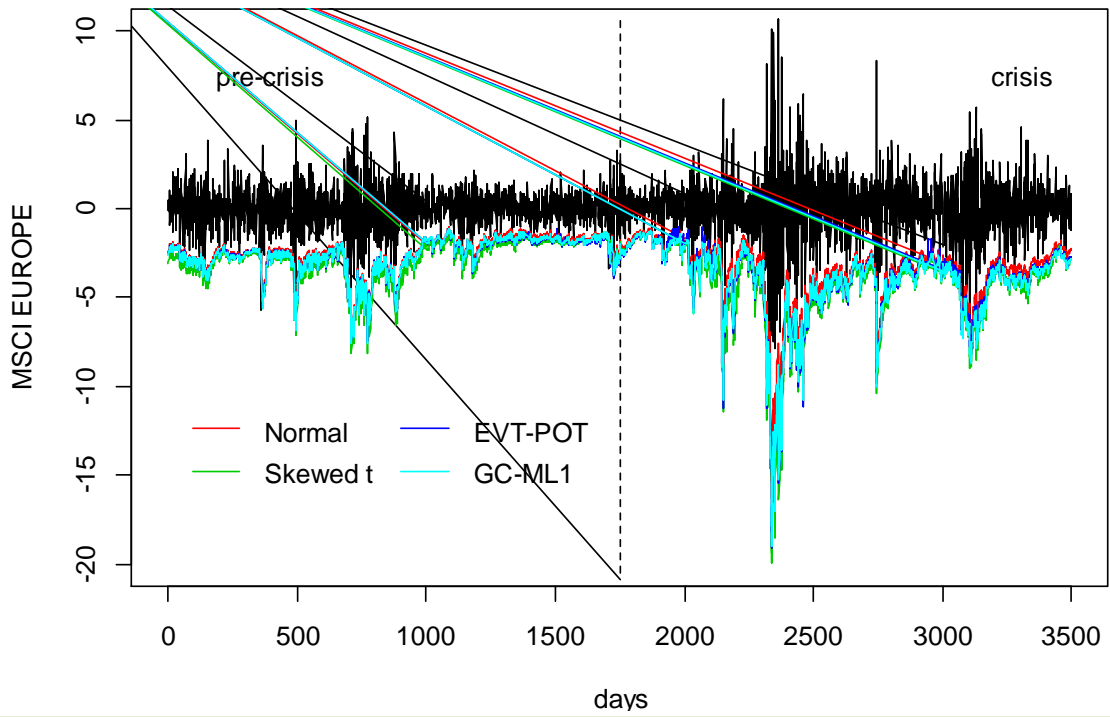


Figure 7.B. MSCI USA

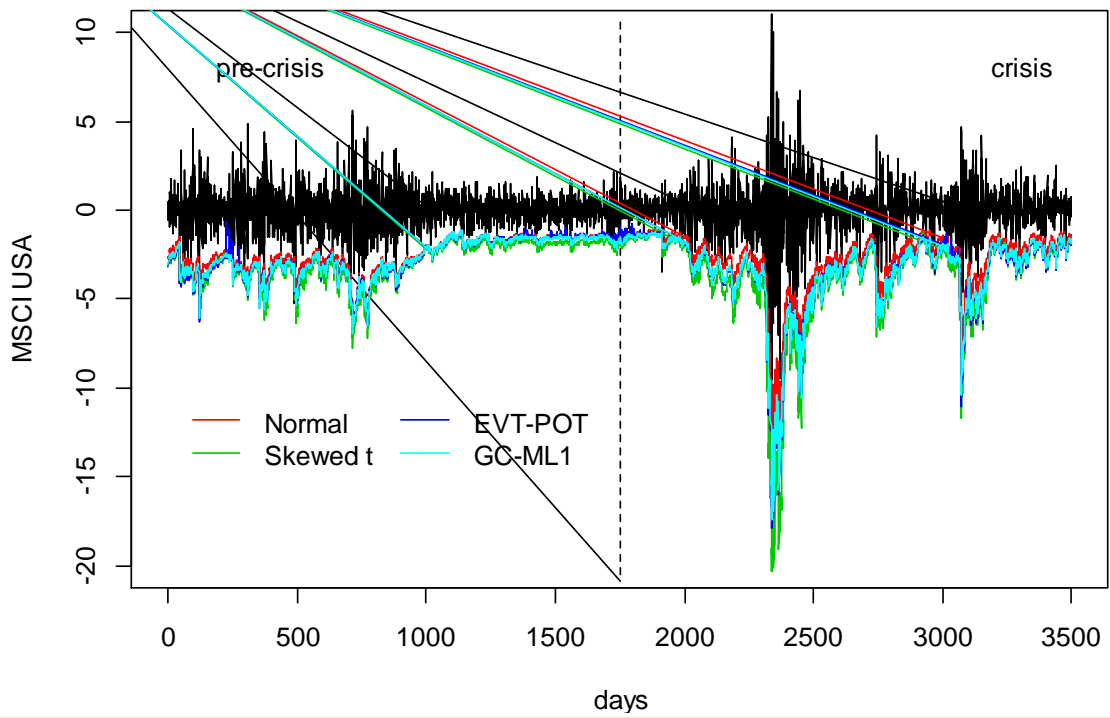
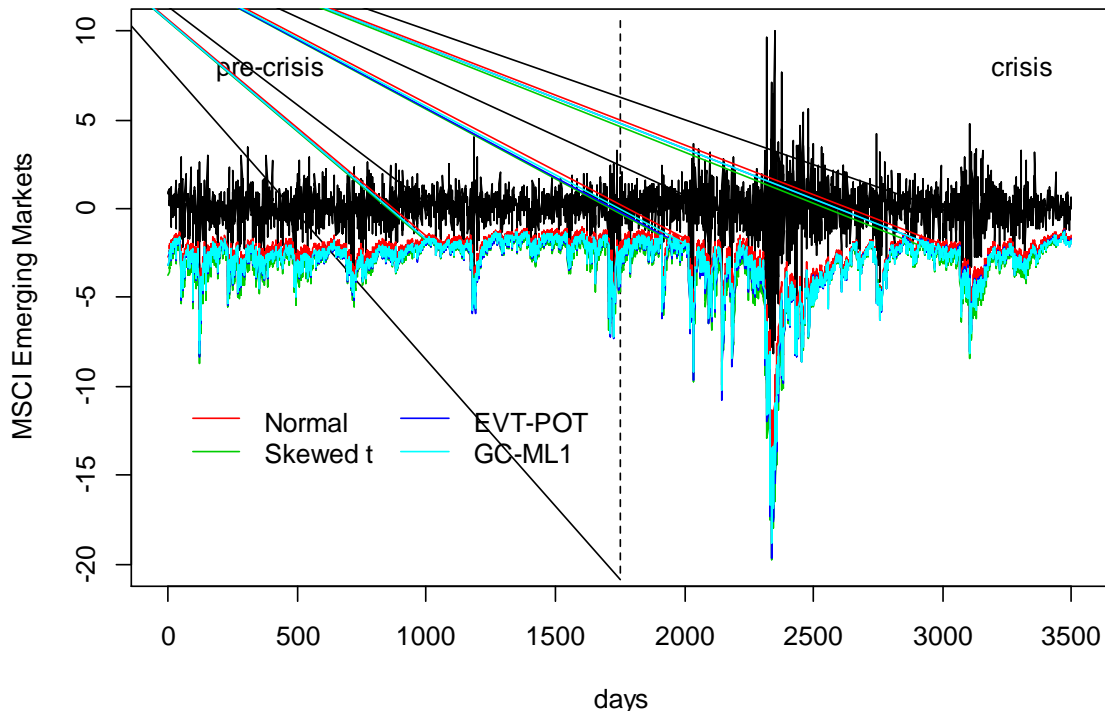


Figure 7.C. MSCI EMERGING MARKETS



II. 4. Conclusions

The recent financial crises have caused high volatility in financial markets and big losses for many investors. To quantify the potential losses and comply with regulatory capital requirements, financial institutions implement VaR methodologies. However, traditional VaR measures, based on the normal distribution, have been criticized because of their inability to adequately capture market risk, particularly at high confidence levels. For this purpose, the use of alternative thick-tailed and skewed distributions has been proposed, but there is still no consensus about the most appropriate methodology for VaR forecasting. In this article, we investigate this issue by comparing the relative performance of three parametric models (normal, Student's t and skewed- t), the EVT-POT approach and a semi-nonparametric model (GC) for stock indices of major economic areas: USA, Europe and emerging markets. We argue that the model ranking depends on the period under analysis, and thus we compare the sensitivity of VaR measures to the increase in volatility by studying VaR measures before and after the recent subprime and sovereign debt crisis.

Five main conclusions may be drawn from our study: (i) the normal underestimates risk even in low volatility scenarios; (ii) Student's t seems to be adequate for capturing VaR only for "relatively calm" periods and its skewed counterpart seems to be a better model for high volatility periods, although it tends to provide conservative risk measures; (iii) both EVT and GC are accurate methods for computing VaR at high confidence levels

(99%); (iv) the larger GC expansions (GC-ML2) do not necessarily improve VaR measures compared to the simpler two-parameter (skewness and excess kurtosis) formulations; and (v) the MM estimation method of GC densities seems to provide straightforward and accurate risk measures.

Result (i) is standard in the literature (e.g. see McNeil et al., 2005, p. 46-7 and 58, for a VaR performance comparison of the normal and Student's t at different confidence levels). Result (ii) emphasizes the poor VaR forecasting performance of Student's t in highly volatile scenarios and the better performance of skewed-t in this case (although likely at the expense of overly conservative VaR measures). This finding might be explained by the fact that positive skewed distributions capture the left tail of the empirical distribution in bear market periods but not in bull market periods, where the distribution exhibits negative skewness. Result (iii) is a consequence of the EVT (POT) approach, developed to capture extreme events, and the fact that this methodology is very sensitive to the threshold selection. This evidence is consistent with other EVT studies, e.g. Rachev et al. (2010). The accurate performance of the GC density lies in the asymptotic properties of the Hermite polynomial expansion and the flexibility of its formulation, which is capable of capturing not only leptokurtosis and skewness but also other features such as jumps in the probabilistic mass. Conclusion (iv) seems to contradict our former assessment but it supports the well-known fact that a good in-sample fit does not guarantee a good out-of-sample performance (see e.g. Hansen, 2009) and thus simpler models usually provide better forecasting outcomes. Finally, result (v) is in line with Del Brio and Perote (2011) and implies that accurate VaR forecasts according to the GC specification can be straightforwardly obtained by implementing MM techniques.

All these results highlight the fact that the optimal VaR model depends not only on the assumed confidence level (risk aversion) but also on the observed market conditions (volatility). Therefore our results could be summarized in a straightforward recommendation to risk managers: risk forecasting methodologies should accommodate the scenario in which forecasts are computed. Only by combining different methods or by using very flexible techniques can the regulatory capital and the provisions of financial institutions be accurately estimated. For this reason and according to our findings, we recommend implementation of the GC density to forecast VaR.

References

- Alexander, C., 2009. *Market Risk Analysis, Value at Risk Models*. The Wiley Finance Series (Volume IV).
- Bali, T.B, Theodossiou, P., 2008. Risk Measurement Performance of Alternative Distribution Functions. *Journal of Risk and Insurance* 75, 411-437.

- Bollerslev, T., 1986. Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics* 31, 307-327.
- Charlier, C.V., 1905. Über Die Darstellung Willkürlicher Funktionen. *Arkiv för Matematik, Astronomi och Fysik* 9, 1-13.
- Christoffersen, P., 2003. *Elements of Financial Risk Management*, San Diego, CA: Academic Press.
- Coles, S., 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, London.
- Cont, R., 2001. Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance* 1, 223-236.
- Corrado, C. J., Su T., 1996. S&P 500 Index Option Tests of Jarrow and Rudd's Approximate Option Valuation Formula. *Journal of Futures Markets* 16, 611-629.
- Del Brio, E. B., Perote J., 2012. Gram-Charlier Densities: Maximum Likelihood versus the Method of Moments. *Insurance: Mathematics and Economics* 51, 531–537.
- Del Brio, E.B., Níguez, T.M., J. Perote, 2011. Multivariate Semi-nonparametric Distributions with Dynamic Conditional Correlations. *International Journal of Forecasting* 27, 347-364.
- Davison, A.C., Smith R.L., 1990. Models for Exceedances over High Thresholds. *Journal of the Royal Statistical Society Series B* 52, 393-442.
- Dowd, K., Blake D., 2006. After VaR: The Theory, Estimation, and Insurance Applications of Quantile-Based Risk Measures. *Journal of Risk and Insurance* 73, 193-229.
- Duca, J. V., Muellbauer, J., Murphy, A., 2010. Housing Markets and the Financial Crisis of 2007–2009: Lessons for the Future. *Journal of Financial Stability* 6 (4) 203-217.
- Edgeworth, F.Y., 1907. On the Representation of Statistical Frequency by a Series. *Journal of the Royal Statistical Society (Series A)* 70, 102–106.
- Embrechts, P., Klüppelberg, C., T. Mikosch, 1997. *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.
- Engle, R.F., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United KingdomI. *Econometrica*, 50, 987-1007.
- Fernández, C., Steel M.F.J., 1998. On Bayesian Modeling of Fat Tails and Skewness. *Journal of the American Statistical Association* 93, 359–371.

- Gallant, A. R., Nychka, D. W., 1987. Semi-Nonparametric Maximum Likelihood Estimation. *Econometrica* 55, 363-390.
- Giot, P., Laurent S., 2003. Value-at-Risk for Long and Short Trading Positions. *Journal of Applied Econometrics* 18, 641-663.
- Gorton, G., 2009. The Subprime Panic. *European Financial Management* 15, 10-46.
- Gram, J.P., 1879. Om Raekkeudviklinger Bestemte Ved Hjaelp av de Mindste Kvadraters Methode (Kobenhavn).
- Hansen, B.E., 1994. Autoregressive Conditional Density Estimation. *International Economic Review* 35, 705-730.
- Hansen, P.R., 2009. In-Sample Fit and Out-of-Sample Fit: Their Joint Distribution and its Implications for Model Selection. Working Paper Stanford University.
- Harrington, S.E., 2009. The Financial Crisis, Systemic Risk, and the Future of Insurance Regulation. *Journal of Risk and Insurance* 76, 785-819.
- Hubbert, S., 2012. *Essential Mathematics for Market Risk Management*, Second Edition The Wiley Finance Series.
- Hull, J., 2009. The Credit Crunch of 2007: What Went Wrong? What Lessons Can Be Learned? *Journal of Credit Risk* 5, 3-18.
- Hull, J., 2012. *Risk Management for Financial Institutions*, Third Edition, Wiley Finance.
- Jarrow, R., Rudd A., 1982. Approximate Option Valuation for Arbitrary Stochastic Processes. *Journal of Financial Economics* 10, 347-369.
- Jorion, P., 2006. *Value at Risk: The New Benchmark for Managing Financial Risk*, Third Edition, McGraw-Hill.
- Kendall, M., Stuart, A., 1977. *The Advanced Theory of Statistics*. Griffin&Co, London.
- Kolb, R. E., 2010. *Lessons from the Financial Crisis: Causes, Consequences, and Our Economic Future*, Robert E. Kolb Series, Wiley.
- Jondeau, E., Rockinger, M. 2001. Gram-Charlier Densities. *Journal of Economic Dynamics and Control* 25, 1457-1483.
- Jondeau, E., Poon, S.-H., Rockinger M., 2007. *Financial Modeling Under Non-Gaussian Distributions*, Springer Verlag.
- Lambert, P., S. Laurent, 2001. Modelling Financial Time Series using GARCH-Type Models and a Skewed Student Density, Mimeo, Universite de Liège.

- Leadbetter, M. R., 1991. On a Basis for 'Peaks over Thresholds' Modeling. *Statistics & Probability Letters* 4, 357-362.
- León, A., Mencía, J., Sentana, E., 2009. Parametric Properties of Semi-nonparametric Distributions. *Journal of Business and Economic Statistics* 27, 176-192.
- Mandelbrot, B. 1963. The Variation of Certain Speculative Prices. *Journal of Business* 36, 394-419.
- Marumo, K., Wolff, R., 2007. Expansion Methods Applied to Asset Return Distributions. *Journal of Risk* 10, 3-24.
- Mauleón, I., 2003. Financial Densities in Emerging Markets: An Application of the Multivariate ES Density. *Emerging Markets Review* 4, 197-223.
- Mauleón, I., Perote, J. 2000. Testing Densities with Financial Data: An Empirical Comparison of the Edgeworth-Sargan Density to the Student's t. *European Journal of Finance* 6, 225-239.
- McNeil, A.J., Frey, R., 2000. Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach. *Journal of Empirical Finance* 7, 271-300.
- McNeil, A. J., Frey, R., Embrechts, P., 2005. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton.
- Mishkin, F.S., 2011. Over the Cliff: From the Subprime to the Global Financial Crisis. *Journal of Economic Perspectives* 25, 49-70.
- Ñíguez, T.M., Perote, J., 2012. Forecasting Heavy-Tailed Densities with Positive Edgeworth and Gram-Charlier Expansions. *Oxford Bulletin of Economics and Statistics* 74, 600-627.
- Puzanova, N., Siddiqui, S., Trede, M., 2009. Approximate Value-at-Risk Calculation for Heterogeneous Loan Portfolios: Possible Enhancements of the Basel II Methodology. *Journal of Financial Stability* 5, 374-392.
- Rachev, Z., Racheva-Iotova, B., S. Stoyanov, 2010. Capturing Fat Tails. *Risk Magazine*, May 2010, 72-77.
- Reiss, R.-D., Thomas, M., 1997. *Statistical Analysis of Extreme Values*, Basel: Birkhäuser.
- Schwert, G. W., 2011. Stock Volatility during the Recent Financial Crisis. *European Financial Management* 17, 789-805.
- Shiller, R. J., 2008. *The Subprime Solution: How Today's Global Financial Crisis Happened, and What to Do about It*, Princeton University Press.

Smith, R. L., 1989. Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone. *Statistical Science* 4, 367-393.

Zumbach, G., 2006. Backtesting Risk Methodologies from One Day to One Year. *Journal of Risk* 9, 55-91.

Appendix

Dataset description:

Index	Description
MSCI Europe	The MSCI Europe Index is a free float-adjusted market capitalization weighted index that is designed to measure the equity market performance of the developed markets in Europe. The MSCI Europe Index consists of the following 16 developed market country indices: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom.
MSCI USA	The MSCI USA Index is a free float-adjusted market capitalization index that is designed to measure large and mid cap US equity market performance. The MSCI USA Index is member of the MSCI Global Equity Indices and represents the US equity portion of the global benchmark MSCI ACWI Index.
MSCI Emerging Markets	The MSCI Emerging Markets Index is a free float-adjusted market capitalization index that is designed to measure equity market performance of emerging markets. The MSCI Emerging Markets Index consists of the following 21 emerging market country indices: Brazil, Chile, China, Colombia, Czech Republic, Egypt, Hungary, India, Indonesia, Korea, Malaysia, Mexico, Morocco, Peru, Philippines, Poland, Russia, South Africa, Taiwan, Thailand, and Turkey.

Source: Datastream, Thomson Financial.

CHAPTER III: A multivariate approximation to portfolio distribution

III.1. Introduction

During the last decades, the literature related to the search of statistical models to explain and forecast financial risk has undergone huge developments. The interest derives from the needs of risk managers of financial institutions who must decide on the most appropriate model for portfolio and risk management. For these purposes many perspectives have been proposed mainly concerning either the modelling of the conditional moment structure under normality or the underlying distribution of the asset returns.

Among the latter approach one of the most interesting and fruitful alternatives has been the semi-nonparametric (SNP hereafter) methodology developed by authors such as, Sargan (1975), Jarrow and Rudd (1982), Gallant and Nychka (1987), Gallant and Tauchen (1989), Corrado and Su (1997), Mauleón and Perote (2000), Nishiyama and Robinson (2000), Jondeau and Rockinger (2001), Velasco and Robinson (2001), Jurczenko et al. (2002), Verhoeven and McAleer (2004), Tanaka et al. (2005), León et al. (2005), Bao et al. (2006), Rompolis and Tzavalis (2006), León et al. (2009), Polanski and Stoja (2010), Níguez and Perote (2012) and Níguez et al. (2012). All these articles proposed the use of polynomial expansions of the Gaussian distribution to define density functions capable of capturing the stylized features of financial asset returns, besides of providing applications to the resulting densities to different contexts, e.g. hypotheses testing, density forecasting, Value-at-Risk (VaR hereafter), asset pricing or option valuation. The greater goodness-of-fit of this family of densities and the more accurate risk measures obtained, as shown in these papers, is a consequence of its more general and flexible representation, which admits as much parameters as necessary to capture the sharply-peaked, thick-tailed or skewed shapes of the underlying asset returns density.

These empirical findings result from the validity of the Gram-Charlier (GC hereafter) and Edgeworth series as asymptotic approximations – Charlier (1905) and Edgeworth (1907). As a matter of fact, under regularity conditions any frequency function can be expressed in terms of an infinite weighted sum of the derivatives of the standard Gaussian distribution or their corresponding Hermite polynomials (HP hereafter). The main shortcoming of these expansions is the fact that positivity of the finite (truncated)

expansions does not hold positivity in the entire domain of the parameter set – Barton and Dennis (1952). This problem, has been tackled in the literature by means of parametric restrictions (Jondeau and Rockinger, 2001), or through density reformulations based on the methodology of Gallant and Nychka (1987). These solutions are not always the best option since imposing positivity constraints may lead to sub-optimisation and positivity regions are not easy to be defined beyond the simpler cases (i.e. for expansions defined in terms of a couple of moments, usually skewness and kurtosis). Furthermore, positive transformations induce non-linearities among the distribution moments and the density parameters and in some cases lead to symmetric distributions. The former problem affects the straightforward interpretation of the parameters of the raw GC density and thus seriously restricts the implementation of the method of moments (MM hereafter). Alternatively, Maximum Likelihood (ML hereafter) techniques are usually employed although optimization algorithms usually fail to converge or do it to local optima. In addition, ML estimation only provides consistent estimates either under the normal or under the true density.

The extensions of GC densities to other continuous and differentiable non-normal densities have also been investigated. Particularly, the Poisson, Gamma, and Beta have been proposed as basis (GC Type B, Laguerre and Jacobi expansions, respectively). Nevertheless the validity of these series as asymptotic expansions and their empirical applicability are still to be proved (see Wallace, 1958, for a discussion on the validity of asymptotic expansions using non-normal densities as generating distributions). Generalizations of GC densities to the multivariate framework have also been proposed as alternatives to copula methods. In particular, Perote (2004) introduced a first definition and Del Brio et al. (2009; 2011) proposed more general formulations accounting for the positivity and the ‘curse of the dimensionality’ problems, in the same spirit as the DCC model by Engle (2002).

In this article we revise the aforementioned multivariate models focusing on the implementation of a straightforward MM estimation as alternative to traditionally used ML or Quasi ML (QML hereafter) techniques. This proposal is extremely simple for the GC densities since the even (odd) parameters are just linear combinations of the even (odd) density moments and the moment of order n depends only on the first n density parameters. Even more, the MM estimation involves consistent estimates, which is only guaranteed for ML under the true density and for QML under density misspecification and provided that first and second moments are correctly specified (Bollerslev and Wooldridge, 1992). We show that all these techniques, however, produce similar results. Furthermore, we implement a three-step estimation method which eases estimation of the portfolio density in relation to likelihood optimisation. We proceed as follows: Firstly we estimate the conditional variances under the normal distribution for every variable by QML; secondly, we estimate the rest of the GC density parameters for

every variable by MM; thirdly, we estimate the correlation among the portfolio variables by MM.

The remainder of the paper reviews the multivariate formulations for the GC expansions and explains the MM estimation (Section 2). Section 3 provides an application of our MM technique for the estimation and VaR computation of the multivariate density of a portfolio composed of three European stock indices. The last section (4) summarizes the main conclusions of the paper.

III. 2. Approximations to portfolio distribution

2.1. The univariate case

Let $\phi(x_i) = e^{-x_i^2/2} / \sqrt{2\pi}$ be the normal probability density function (pdf hereafter) and $H_s(x_i)$ the Hemite polynomial based on its s -th order derivative, which can be defined as in equation (1).

$$H_s(x_i) = (-1)^s \phi(x_i)^{-1} \frac{d^s \phi(x_i)}{dx_i^s} = s! \sum_{k=0}^{\lfloor s/2 \rfloor} \frac{(-1)^k}{k!(s-2k)! 2^k} x_i^{s-2k}. \quad (1)$$

These HP polynomials form an orthonormal basis. Thus, if $\mathbf{H}(x_i) = [H_1(x_i) \ H_2(x_i) \ \dots \ H_q(x_i)] \in \mathfrak{R}^q$ is the vector containing the first q HP then

$$\int \mathbf{H}(x_i) \mathbf{H}(x_i)' \phi(x_i) dx_i = \mathbf{S} = \text{diag} \{1!, 2!, \dots, q!\}. \quad (2)$$

Furthermore, the vector of HP can be written as

$$\mathbf{H}(x_i) = \mathbf{B} \mathbf{Z}_i + \boldsymbol{\mu}, \quad (3)$$

where $Z_i = [x_i \ x_i^2 \ \dots \ x_i^q]$, $\mu = [\mu_1 \ \mu_2 \ \dots \ \mu_q]$ is the vector containing the first q central moments of the normal distribution (i.e. $\mu_s = \frac{s!}{2(s-2)!}$ for s even, and zero otherwise) and

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ -3 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & -6 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ q! & q! & q! & q! & \dots & q! & 1 \\ \frac{q-2}{2} \frac{q-2}{2}! & \frac{q-4}{2} \frac{q-4}{2}! & \frac{q-6}{2} \frac{q-6}{2}! & \frac{q-8}{2} \frac{q-8}{2}! & \dots & 2(q-2)! & 1 \end{bmatrix}. \quad (4)$$

One of the main advantages of this sequence of HP is the fact that under certain regularity conditions (Cramér, 1926) a frequency function, $f(x_i)$, can be expanded formally in terms of GC Type A series, i.e.

$$f(x_i) = \sum_{s=0}^{\infty} \delta_s \frac{d^s \phi(x_i)}{dx_i^s} = \sum_{s=0}^{\infty} (-1)^s \delta_s H_s(x_i) \phi(x_i), \quad (5)$$

where the δ_s coefficients,

$$\delta_s = \frac{(-1)^s}{s!} \int_{-\infty}^{\infty} f(x_i) H_s(x_i) dx_i, \quad (6)$$

measure the deviations of $f(x_i)$ from $\phi(x_i)$ and can be also expressed in terms of the (non-central) moments of the random variable x_i with pdf $f(x_i)$.

Nevertheless for empirical purposes the asymptotic expansion needs to be truncated at a degree (q) and then the univariate GC density is defined as follows,

$$f_q(x_i, d_i) = [1 + H(x_i)' d_i] \phi(x_i), \quad (7)$$

where $\mathbf{d}_i = [d_{i1} \ d_{i2} \ \dots \ d_{iq}] \in \mathfrak{R}^q$ is a vector of parameters and, by convention, we consider $H_0(x_i)=1$ and $d_0=1$. This distribution in equation (7) satisfies interesting properties (see e.g. Mauleon and Perote, 2000). Among them, we enunciate in Proposition 1 the one in which the MM estimation method proposed in the present article is based.

Proposition 1: The first q moments of the GC distribution in equation (7) can be expressed as a linear function of the vector $\mathbf{d}_i' \in \mathfrak{R}^q$,

$$E[Z_i] = \mathbf{B}^{-1}(\mathbf{S}\mathbf{d}_i + \boldsymbol{\mu}), \quad (8)$$

where $Z_i = [x_i \ x_i^2 \ \dots \ x_i^q]$, and \mathbf{S} and \mathbf{B} are the matrices described in equations (2) and (4), respectively, and $\boldsymbol{\mu} \in \mathfrak{R}^q$ is the vector containing the first q central moments of the normal distribution.

Proof:

$$\begin{aligned} E[Z_i] &= \int Z [1 + \mathbf{H}(x_i)' \mathbf{d}_i] \phi(x_i) dx_i = \int \mathbf{B}^{-1}(\mathbf{H}(x_i) + \boldsymbol{\mu}) [1 + \mathbf{H}(x_i)' \mathbf{d}_i] \phi(x_i) dx_i \\ &= \mathbf{B}^{-1} \int \mathbf{H}(x_i) \phi(x_i) dx_i + \mathbf{B}^{-1} \int \mathbf{H}(x_i) \mathbf{H}(x_i)' \mathbf{d}_i \phi(x_i) dx_i + \mathbf{B}^{-1} \boldsymbol{\mu} \int \phi(x_i) dx_i + \mathbf{B}^{-1} \boldsymbol{\mu} \int \mathbf{H}(x_i)' \mathbf{d}_i \phi(x_i) dx_i \\ &= \mathbf{0} + \mathbf{B}^{-1} \mathbf{S} \mathbf{d}_i + \mathbf{B}^{-1} \boldsymbol{\mu} + \mathbf{0} = \mathbf{B}^{-1}(\mathbf{S} \mathbf{d}_i + \boldsymbol{\mu}). \end{aligned}$$

These relations among the density moments and parameters establish a straightforward way of estimating the density by the MM, given by

$$\hat{\mathbf{d}}_i = \mathbf{S}^{-1}(\mathbf{B} \hat{E}[Z_i] - \boldsymbol{\mu}), \quad (9)$$

where $\hat{E}[Z_i]$ is the vector containing the first q sample moments of variable x_i with pdf $f_q(x_i, \mathbf{d}_i)$.

The truncated function in (7), however, does not guarantee positivity for all values of d_i and thus a positive (squared) transformation of the Gallant and Nychka's (1987) type is usually implemented. Next we explain the family of multivariate GC densities including these positive transformations.

2.2. The multivariate case

A random vector $X = [x_1 \ x_2 \ \dots \ x_n] \in \mathfrak{R}^n$ belongs to the multivariate GC (MGC hereafter) family of distributions if it is distributed according to the following pdf,

$$F(\mathbf{X}) = \frac{1}{n+1} \left[G(\mathbf{X}) + \left\{ \sum_{i=1}^n \frac{1}{c_i} \mathbf{h}(x_i)' \mathbf{A}_i \mathbf{h}(x_i) \right\} \left\{ \prod_{i=1}^n \phi(x_i) \right\} \right] \quad (9)$$

where

$$G(\mathbf{X}) = (2\pi)^{-n/2} |\Sigma^{-1/2}| \exp \left\{ -\frac{1}{2} \mathbf{X}' \Sigma^{-1} \mathbf{X} \right\} \quad (10)$$

is the multivariate normal pdf - with univariate marginals $\phi(x_i)$, \mathbf{A}_i is a positive definite matrix of order $(q+1)$, $\mathbf{h}(x_i) = [1 \ \mathbf{H}] \in \mathfrak{R}^{q+1}$ and $c_i = \int \mathbf{h}(x_i)' \mathbf{A}_i \mathbf{h}(x_i) \phi(x_i) dx_i$ ($\forall i=1,2,\dots, n$) are the constants that make the density integrating up to one - Del Brio et al. (2009). Without loss of generality, we are assuming the same truncation order q for every dimension i .

This MGC family encompasses many different distributions, such as the multivariate extensions of the GC density in León et al. (2009) or the Positive Edgeworth-Sargan in Níguez and Perote (2012). These two types of distributions are obtained by considering $\mathbf{A}_i = \mathbf{D}_i \mathbf{D}_i'$ and $\mathbf{A}_i = \text{diag} \{1, d_{i1}^2, \dots, d_{iq}^2\}$, respectively, where $\mathbf{D}_i = [1 \ \mathbf{d}] \in \mathfrak{R}^{q+1}$ (note that in both cases $c_i = 1 + \sum_{s=1}^q d_{is}^2$). However, in this paper we implement a related family of densities proposed in Perote (2004) which does not formally impose positive definiteness but presents other interesting advantages from an empirical perspective. Hereafter, we will refer to the pdf defined in eq. (12) below as MGC density.

$$F(\mathbf{X}) = G(\mathbf{X}) + \left\{ \sum_{i=1}^n H(x_i)' d_i \right\} \left\{ \prod_{i=1}^n \phi(x_i) \right\} \quad (12)$$

It is clear that for the MGC the marginal density of x_i is that of equation (7) and thus the MM estimation can be trivially implemented through the relation in equation (9). Even more, Del Brio et al. (2011) proved that an equivalent MGC density can be estimated by ML in two steps: In the first step, the conditional mean and variance of every variable are estimated by QML independently, and in the second step, the rest of the density parameters are jointly estimated in the standardised distribution. This paper proposes a similar three-step procedure based on the MM: First, QML estimates for conditional mean and variance of every variable are obtained independently by assuming a normal distribution. Second, the parameters for the univariate GC density of every standardised variable are estimated independently by the MM. Third, correlation parameters are approximated by the sample correlations.

III. 3. Empirical application

We illustrate the estimation procedure of the portfolio return distribution described in the previous section for a portfolio composed of three European stock indices: EUROSTOXX50, Ibex35 and Dax30. The sample comprises almost 10 years of daily data ($T=2,861$ observations) spanning from September 30th, 2002, to November 19th, 2013. We model continuously compounded returns, defined as $r_{it}=100\log(P_{it}/P_{i,t-1})$. Table 1 displays descriptive statistics for the series. These data feature the main empirical regularities of high-frequency financial returns: a small predictable component in the conditional mean, volatility clustering, skewness, leptokurtosis and, likely, multimodality (jumps) in the tails.

Table 1. Descriptive statistics for EUROSTOXX50, Ibex35 and Dax30.

	EUROSTOXX50	Ibex35	Dax30
Mean	0.01134	0.02025	0.04228
Variance	2.21881	2.26692	2.26598
Minimum	-8.20788	-9.58586	-7.43346
Maximum	10.43765	13.48364	10.79747
Skewness	0.07902	0.13892	0.08702

Kurtosis	5.45580	6.52309	5.38219
----------	---------	---------	---------

We specify a multivariate AR(1)-GARCH(1,1) structure for modelling conditional first and second moments and the MCG density in equation (12) for capturing the rest of the salient empirical regularities of the data. Thus, the multivariate model for the portfolio returns $\mathbf{r}_t = [r_{1t} \ r_{2t} \ r_{3t}] \in \mathcal{R}^3$ is:

$$\mathbf{r}_t = \Phi_0 + \Phi_1 \circ \mathbf{r}_{t-1} + \mathbf{u}_t, \quad (13)$$

$$\mathbf{u}_t | \Omega_{t-1} \approx MGC(0, \Sigma_t(\alpha, \rho)), \quad (14)$$

$$\Sigma_t(\alpha, \rho) = \sigma_t(\alpha) \mathbf{R}_t(\rho) \sigma_t(\alpha), \quad (15)$$

$$\mathbf{D}_t(\alpha)^2 = \text{diag}\{\alpha_{i0}\} + \text{diag}\{\alpha_{i1}\} \circ \mathbf{u}_{t-1} \mathbf{u}'_{t-1} + \text{diag}\{\alpha_{i2}\} \circ \mathbf{D}_{t-1}(\alpha)^2, \quad (16)$$

where Φ_0 and Φ_1 are 3×1 vectors containing the parameters of the AR(1) processes; $\text{diag}\{\alpha_{i0}\}$, $\text{diag}\{\alpha_{i1}\}$ and $\text{diag}\{\alpha_{i2}\}$ are diagonal matrices containing the parameters of the GARCH(1,1) processes (hereafter we refer to these parameters as Φ and α , respectively). Therefore, the variance and covariance matrix is decomposed in the diagonal matrix of conditional deviations, $\sigma_t(\alpha)$ and the symmetric correlation matrix, $\mathbf{R}_t(\rho)$, with general element $\{\rho_{ij}\}$ (hereafter we refer to the parameters in $\mathbf{R}_t(\rho)$ as ρ). Finally, \circ is the Hadamard product of two identical sized matrices (computed by element-by-element multiplication).

The estimation of the model in equations (13)-(16) through our proposed three-step MM is carried out in the following three stages:

Stage 1: Φ and α are estimated by QML as the values that maximise the log-likelihood of every variable under the Gaussian distribution, i.e.,

$$\{\hat{\phi}_{i0}, \hat{\phi}_{i1}, \hat{\alpha}_{i0}, \hat{\alpha}_{i1}, \hat{\alpha}_{i2}\} = \arg \max \left\{ \sum_{t=1}^T \left[\ln(\sigma_{it}^2) + \frac{(r_{it} - \phi_{i0} - \phi_{i1} r_{i,t-1})^2}{\sigma_{it}^2} \right] \right\}, \quad (17) \text{ s.t.}$$

$$\sigma_{it}^2 = \alpha_{i0} + \alpha_{i1}(r_{it} - \phi_{i0} - \phi_{i1}r_{it-1})^2 + \alpha_{i2}\sigma_{it-1}^2; \forall i=1,2,3.$$

Stage 2: The parameters (d) of the GC expansion are estimated independently for every dimension i by using the following correspondences,

$$\hat{d}_{i1} = \hat{m}_{i1} \quad (18)$$

$$\hat{d}_{i2} = (\hat{m}_{i2} - 1)/2 \quad (19)$$

$$\hat{d}_{i3} = (\hat{m}_{i3} - 3\hat{m}_{i1})/6 \quad (20)$$

$$\hat{d}_{i4} = (\hat{m}_{i4} - 6\hat{m}_{i2} + 3)/24, \quad (21)$$

$$\hat{d}_{i5} = (\hat{m}_{i5} - 10\hat{m}_{i3} + 15\hat{m}_{i1})/120 \quad (22)$$

$$\hat{d}_{i6} = (\hat{m}_{i6} - 15\hat{m}_{i4} + 45\hat{m}_{i2} - 15)/720 \quad (23)$$

$$\hat{d}_{i7} = (\hat{m}_{i7} - 21\hat{m}_{i5} + 105\hat{m}_{i3} - 150\hat{m}_{i1})/5040 \quad (24)$$

$$\hat{d}_{i8} = (\hat{m}_{i8} - 28\hat{m}_{i6} + 210\hat{m}_{i4} - 420\hat{m}_{i2} + 150)/40320 \quad (25)$$

$\forall i=1,2,3$, where $\hat{m}_{si} = \frac{1}{T} \sum_{t=1}^T \frac{(r_{it} - \bar{r}_i)^s}{S_i}$, $\forall s=1,2,\dots,8$, is the s -th order sample moment of the standardised series (\bar{r}_i and S_i denoting the average and standard deviation of r_{it} , respectively), which is a consistent estimate of the s -th order moment of the true distribution.

Stage 3: The Correlation matrix, $R(\rho)$, is estimated by computing the sample correlations among the portfolio variables:

$$\hat{\rho}_{ij} = \frac{1}{TS_i S_j} \sum_{t=1}^T (r_{it} - \bar{r}_i)(r_{jt} - \bar{r}_j), \forall i, j=1,2,3,4 \quad (i \neq j) \quad (26)$$

This method has the following advantages with respect to the ML: (i) it provides consistent estimates, i.e. the first step gives consistent (QML) estimates for conditional mean and variance parameters and the second step is also consistent since both log-

likelihood function is separable (see Del Brio et al., 2011) and the MM always yields consistent estimates. (ii) It is much simpler than the ML method with regards to convergence problems that may arise in optimization. (iii) It solves the curse of dimensionality of multivariate modelling, since it is not affected by the number of assets considered in the portfolio. (iv) Parameter estimates are the same regardless the expansion length and, as we show empirically, the procedure leads to very similar outcomes for the estimated density than those obtained by ML.

Table 2 provides two-step ML estimates (t-ratios are displayed in parentheses) for the parameters of the GC density of a portfolio composed of EUROSTOXX50, Ibex35 and Dax30 indices. We consider expansions up to the eighth term but d_{i1} and d_{i2} are constrained to zero since conditional means and variances are captured by the AR(1) and the GARCH(1,1) models, respectively. The AR(1)-GARCH(1,1) parameters are estimated in the first step by QML. These QML estimates confirm the presence of a small predictable component in conditional mean and the persistence and clustering in volatility ($\alpha_{i1} + \alpha_{i2}$ is estimated close to one). In the second step, the rest of the parameters of the density are estimated by either MM or ML applied to the series standardised by the estimated mean and variance of the previous step. The estimates of the GC densities exhibit the traditional behaviour of stock returns: (i) (negative) skewness is captured by parameter d_{i3} and the rest of the odd parameters are not significant; (ii) leptokurtosis is patent since d_{i4} is positive and significant; and (iii) presence of extreme values as high order moments (parameters d_{i6} and d_{i8}) are also significant. Note that the truncation order is chosen according to accuracy criteria (see the Akaike information criteria for the MGC model with 2 and 6 parameters), although the best model should eliminate the insignificant parameters that we still display in Table 2 for the sake of comparison. Finally, the third stage presents the estimate for the correlation matrix, which exhibits a positive correlation between EUROSTOXX50 and both Ibex35 and Dax30, but absence of correlation between the latter two indices.

Table 2. MGC density of stock indices: EUROSTOXX50, Ibex35 and Dax30.

	EUROSTOXX50	Ibex35	Dax30
Stage 1			
ϕ_{i1}	0.06776 (3.290)	0.08020 (3.865)	0.09300 (4.409)
ϕ_{i2}	-0.05768 (-2.938)	-0.00669 (-0.323)	-0.03015 (-1.569)

α_{i0}	0.02389 (3.060)	0.02018 (2.825)	0.02276 (3.352)
α_{i1}	0.09721 (5.265)	0.09982 (5.078)	0.08949 (6.452)
α_{i2}	0.89217 (48.341)	0.89359 (48.676)	0.89925 (65.130)
Stage 2 (MM)			
d_{i3}	-0.04495	-0.04790	-0.05916
d_{i4}	0.05733	0.06572	0.04999
d_{i5}	-0.02341	-0.02207	-0.02759
d_{i6}	0.02333	0.02506	0.02251
d_{i7}	-0.01651	-0.01170	-0.01848
d_{i8}	0.01181	0.00959	0.01316
Stage 2 (ML)			
d_{i3}	-0.02260 (-1.89038)	-0.03181 (-2.54818)	-0.04127 (-4.05200)
d_{i4}	0.04212 (5.27076)	0.04937 (5.66917)	0.02997 (3.76514)
d_{i5}	0.00401 (0.89059)	-0.00176 (-0.36192)	-0.00188 (-0.46320)
d_{i6}	0.00531 (2.39301)	0.00808 (3.25338)	0.00065 (0.24688)
d_{i7}	0.00000 (0.14397)	-0.00011 (-0.15316)	-0.00080 (-1.19259)
d_{i8}	0.00101 (4.02314)	0.00102 (3.79608)	0.00068 (2.35141)

<i>Loglikelihood</i>	-1676.898	-1677.097	-1667.209
<i>AIC</i>	3365.796	3366.194	3346.418
<i>(6 parameters)</i>			
<i>AIC</i>	3380.726	3380.912	3363.116
<i>(2 parameters)</i>			
Stage 3			
EUROSTOXX50	1	0.04727	0.03616
Ibex35	0.04727	1	0.00000
Dax30	0.03616	0.00000	1

t-statistics in parentheses

For the sake of comparison, Table 3 shows the joint estimation for the multivariate Student's t distribution. For this distribution departures from normality are only captured by the degrees of freedom parameter (ν) and thus it is a much less flexible method of estimation.

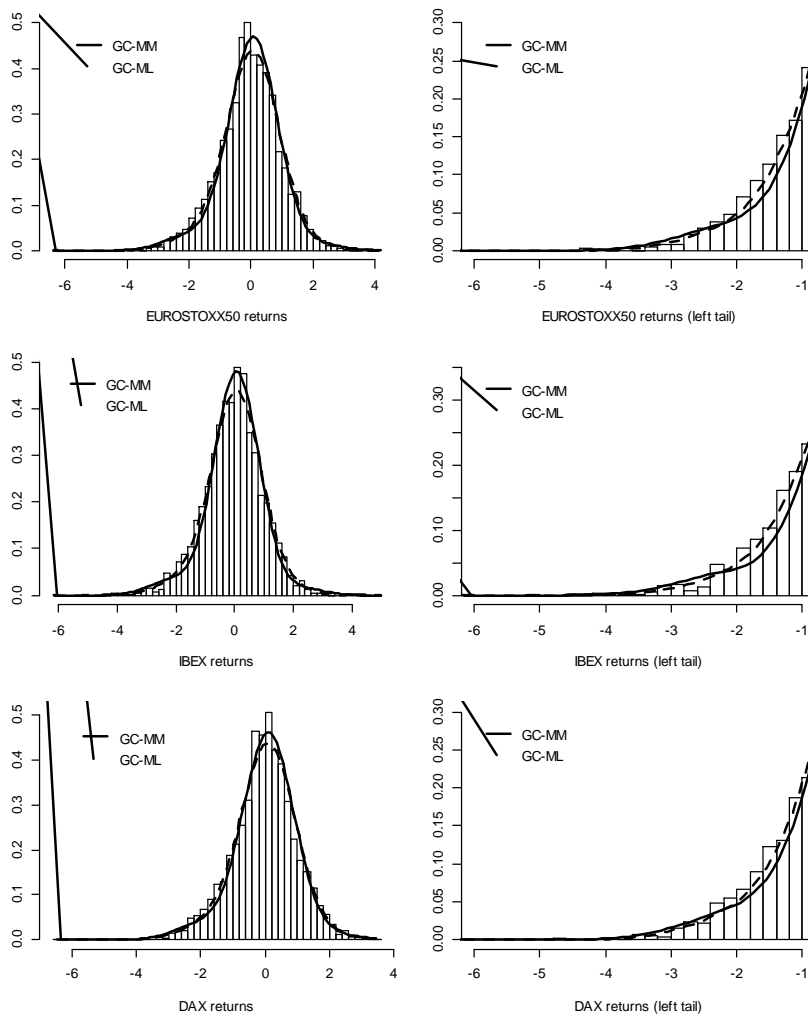
Table 3. Multivariate t density for EUROSTOXX50, Ibex35 and Dax30.

	EUROSTOXX50	Ibex35	Dax30
ϕ_{i1}	0.06776	0.08020	0.09300
ϕ_{i2}	-0.05768	-0.00669	-0.03015
α_{i0}	0.02389	0.02018	0.02276
α_{i1}	0.09721	0.09982	0.08949
α_{i2}	0.89217	0.89359	0.89925
ν	10.48116		
<i>Loglikelihood</i>	-11941.49		
Correlation Matrix			

EUROSTOXX50	1	0.04284	0.04419
Ibex35	0.04284	1	-0.00065
Dax30	0.04419	-0.00065	1

Figure 1 depicts the fitted GC marginal distributions of the returns of the EUROSTOXX50, Ibex35 and Dax30 indices compared to the histogram of the data (non-parametric estimation). Figures on the right column represent the distributions for the whole range and figures on the right the left tails (extreme values) of the corresponding distribution. The plots illustrate that both MM and ML methods (GC-MM and GC-ML, respectively) lead to very similar outcomes and that they approximate very accurately the empirical distribution of the portfolio. This evidence is even clearer in the tails of the distribution, which is the main focus of risk management.

Figure 1. Fitted GC distributions compared to the data histogram



Finally, we calculate VaR at 1% and 5% for an equally weighted portfolio formed with the three indices. For this purpose, 1,000 datasets of length 2,861 are simulated. Table 4 shows the average VaR and its standard error for the multivariate t, MGC estimated by MM and ML and the corresponding empirical VaR. The results illustrate how the MGC-MM model adequately captures portfolio's VaR and thus represents a very straightforward a useful method for risk management.

Table 4. Estimated VaR for an equally weighted portfolio of EUROSTOXX50, Ibex35 and Dax30.

	Multivariate t	MGC-MM	MGC-ML
Empirical VaR – 1%	-1.54395		
Mean VaR – 1%	-1.62432	-1.50203	-1.45894
Standard error	0.06022	0.05312	0.04867
Empirical VaR – 5%		-1.051891	
Mean VaR – 5%	-1.06971	-1.00847	-0.99632
Standard error	0.02996	0.02827	0.02711

III. 4. Conclusions

The GC density has revealed as a powerful tool to account for asset returns distribution because it asymptotically captures the true distribution and it represents a general and flexible approximation. Nevertheless this distribution has scarcely been used for capturing the multivariate behaviour of portfolio distributions due to the so-called ‘curse of dimensionality’ that particularly affects this type of distributions that depend on a large number of parameters. Furthermore the traditional ML estimation techniques usually fail to converge and, more importantly, do not guarantee consistency under possibly density misspecification. In order to solve these problems we propose a very simple three-step estimation method that combines QML estimation for conditional means and variances (Stage 1), MM estimation of the rest of the density parameters considering the univariate standardised marginal GC distributions and, finally, MM estimation of correlation coefficients. The validity of this proposal is based on three main properties of the MGC distribution: (i) Its marginals behave as univariate GC distributions; (ii) It admits an independent estimation of the first and second moments under the Gaussian hypothesis (QML); (iii) It exists a direct linear relation among density moments and parameters, which simplifies the implementation of the MM

techniques. Furthermore, this method is always consistent and may be straightforwardly implemented even for large portfolios.

An application of such procedure is performed for a portfolio composed of three European stock indices as an illustration of the method. The results are not very different from those obtained from QML estimation and thus it seems to be a straightforward method for estimating portfolio return distributions. The simplicity of the method as well as the asymptotic properties of the GC expansion makes this approach a very good approximation to portfolio distribution and thus a useful methodology for risk managers.

References

- Bao Y, Lee T-H, Saltoglu B (2006) Comparing density forecast models. *Journal of Forecasting* 26:203-225
- Barton DE, Dennis KER (1952) The Conditions under which Gram-Charlier and Edgeworth curves are positive definite and unimodal. *Biometrika* 39:425-427
- Bollerslev T, Wooldridge JM (1992) Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews* 11:143-172.
- Charlier CV (1905) Uber die darstellung willkurlicher funktionen. *Arvik fur Mathematik Astronomi och fysik* 9:1-13
- Corrado C, Su T (1997) Implied volatility skews and stock return Skewness and kurtosis implied by stock option prices. *European Journal of Finance* 3:73-85
- Cramér H (1926) On some classes of series used in mathematical statistics. Skandinaviske Matematikercongres, Copenhagen
- Del Brio EB, Níguez TM, Perote J (2009) Gram-Charlier densities: A multivariate approach. *Quantitative Finance* 9:855-868
- Del Brio EB, Níguez TM, Perote J (2011) Multivariate semi-nonparametric distributions with dynamic conditional correlations. *International Journal of Forecasting* 27:347-364
- Edgeworth FY (1907) On the representation of statistical frequency by series. *Journal of the Royal Statistical Society series A*, 80
- Engle RF (2002) Dynamic Conditional Correlation - A Simple Class of Multivariate GARCH Models. *Journal of Business and Economic Statistics* 20:339-350
- Gallant AR, Nychka DW (1987) Semiparametric maximum likelihood estimation. *Econometrica* 55:363-390
- Gallant AR, Tauchen G (1989) Semiparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications. *Econometrica* 57:1091-1120
- Jarrow R, Rudd A (1982) Approximate option valuation for arbitrary stochastic processes. *Journal of Financial Economics* 10:347-369

- Jondeau E, Rockinger M (2001) Gram-Charlier densities. *Journal of Economic Dynamics and Control* 25:1457-1483
- Jurczenko E, Maillat B, Negrea B (2002) Multi-moment option pricing models: a general comparison (part 1). Working Paper, University of Paris I Panthéon-Sorbonne.
- León A, Mencía J, Sentana E (2009) Parametric properties of semi-nonparametric distributions, with applications to option valuation. *Journal of Business and Economic Statistics* 27:176-192
- León A, Rubio G, Serna G (2005) Autoregressive conditional volatility, skewness and kurtosis. *Quarterly Review of Economics and Finance* 45:599-618
- Mauleon I, Perote J (2000) Testing densities with financial data: An empirical comparison of the Edgeworth-Sargan density to the Student's t. *European Journal of Finance* 6:225-239
- Nishiyama Y, Robinson PM (2000) Edgeworth expansions for semi-parametric averaged derivatives. *Econometrica* 68:931-980.
- Ñíguez TM, Perote J (2011) Forecasting heavy-tailed densities with positive Edgeworth and Gram-Charlier expansions. *Oxford Bulletin of Economics and Statistics* 74:600-627
- Ñíguez TM, Paya I, Peel D, Perote J (2012) On the stability of the CRRA utility under high degrees of uncertainty. *Economics Letters* 115:244-248
- Perote, J (2004) The multivariate Edgeworth-Sargan density. *Spanish Economic Review* 6:77-96
- Polanski A, Stoja E (2010) Incorporating higher moments into value at risk forecasting. *Journal of Forecasting* 29:523-535.
- Rompolis L, Tzavalis E (2006) Retrieving risk-neutral densities based on risk neutral moments through a Gram-Charlier series expansion. *Mathematic Modelling and Computation* 46:225-234
- Sargan JD (1975) Gram-Charlier approximations applied to t ratios of k-class estimators. *Econometrica* 43:327-346
- Tanaka K, Yamada T, Watanabe T (2005) Approximation of interest rate derivatives by Gram-Charlier expansion and bond moments. Working Paper IMES, Bank of Japan
- Velasco C, Robinson PM (2001) Edgeworth expansions for spectral density estimates and studentized simple mean. *Econometric Theory* 17:497-539
- Verhoeven P, McAleer M (2004) Fat tails in financial volatility models. *Mathematics and Computers in Simulation* 64:351-362
- Wallace DL (1958) Asymptotic approximations to distributions. *Annals of Mathematical Statistics* 29:635