MODELING THE DISTRIBUTION OF EXCHANGE RATE TIME SERIES AND MEASURING THE TAIL AREA: AN EMPIRICAL APPLICATION OF THE COLOMBIAN FLEXIBLE EXCHANGE RATE RETURNS

Héctor Manuel Zárate Solano*

Recibido: febrero 2004
Aprobado: mayo 2004

RESUMEN

El conocimiento de la distribución de probabilidad de los retornos de la tasa de cambio y la medición de las área extremas son tópicos en la literatura de finanzas que han sido analizados por procedimientos de estimación paramétricos y no paramétricos. Sin embargo, un conflicto de robustez surge debido a que estas series de tiempo son leptocurticas. Más aún, se ha observado que en varias economías en desarrollo la fase inicial del régimen flexible de tasa de cambio ha presentado volatilidad alta. En esta investigación se cubren dos objetivos: primero, parametrizar varias clases de distribuciones que permitan tener una nueva descripción del proceso generador de la tasa de cambio durante el régimen flexible. Segundo, cuantificar el área extrema a través del estimador de Hill. Está estrategia requiere que el número de observaciones extremas sea conocido. Así basado en la teoría de estadísticas de orden se implementa una regla de decisión encontrada por simulación de Monte Carlo bajo varias distribuciones. El modelo de decisión es formulado de tal manera que el error cuadrado medio es minimizado.

Palabras clave: Áreas extremas, distribuciones leptocúrticas, simulaciones de Monte Carlo.
Clasificación JEL: C14, C15, C16, C22, C44, E17, F31.

---

* The results and opinions expressed here are those of the author and do not compromise the Banco de la República nor its board of governors. I am thankful to Professor Roger Koenker at University of Illinois at Urbana-Champaign, for his orientation and suggestions in a previous draft. I would like to thank my colleagues Norberto Rodríguez for his comments, Juan Manuel Julio for helpful discussions, and an anonymous referee. Any remaining errors are my responsibility.

ABSTRACT

Modeling the unconditional distribution of returns on exchange rate and measuring its tails area are issues in the finance literature that have been studied extensively by parametric and non-parametric estimation procedures. However, a conflict of robustness is derived from them because the time series involved in this process are usually fat tailed and highly peaked around the center. Moreover, it has been an empirical fact that the initial phase of a freely floating exchange rate regime has experienced high volatility across many economies. The purpose of this paper is twofold. First, we try to capture the behavior of the Colombian exchange rate under the flexible system by fitting special types of distributions in order to obtain a new insight of the underlying distribution. Secondly, we measure the tail area through the Hill estimator. This strategy requires the number of extreme observations in the tails to be known. Therefore, the decision rule of choosing an optimal cutting observation based on the idea of spacing statistics is implemented by using a Monte Carlo simulation under different underlying distributions. The decision model is formulated in such a way that the mean squared error is minimized.

Key words: Tail area; Long-tailed distributions; Monte Carlo simulation.
JEL Classification: C14, C15, C16, C22, C44, E17, F31.

I. INTRODUCTION

The methods of modeling the behavior of the exchange rate and drawing inferences about the tail areas of the distribution have been an issue of great interest in econometric literature and applied statistics.[1] As with stock returns, there is a consensus on the stylized fact that the empirical distributions of exchange rate returns are fat-tailed and more peaked around the center than the normal distribution. Combination of these two facts is known in the distributional theory as leptokurtosis. Recent studies suggest that several empirical distributions that seem to capture the properties of exchange rate returns, work as appropriate data generating functions. These kinds of distributions can be classified into two broad types: stable Paretian and random summation stable distributions.[2]

On the other hand, the simplicity to obtain the tail area using the Hill estimator has been the source of a wide range of applications in many areas of economics such as monetary policy, insurance claims and income distribution. However, its implementation requires the choice of a number of extreme order statistics r from a sample of size n. There are different approaches in selecting this cutting observation such as: the Bayesian framework, the likelihood techniques and Hill's alternatives. Recently Hsieh (1999) developed a procedure of choosing an optimal r based on the idea of spacing statistics and on a decision rule under the squared error loss function.

The plan of this paper is as follows: First, we describe the data and usefulness of modeling and measuring the tail of exchange rate returns. Secondly, we study the main characteristics of the distributions involved in modeling the flexible exchange rate return. Then we model the unconditional distribution of returns on exchanges rates under these kinds of parametric distributions for the flexible system implemented in this country since August-1999. In section 3, we applied the decision model and the optimal decision rule to choose the cutting observation through a Monte Carlo simulation analysis. Therefore, we applied the optimal decision rule to currency exchange rate returns and estimated the upper tail. Finally, we present the conclusions and some recommendations.

---

1 Blackwell and Hodges started measuring the tail area of convolutions of distributions. Wallace used the normal distribution to approximate the tail area for t and chi-square distribution. Approximations were extended later to the exponential family and for discrete and continuous cases such as: normal, binomial, gamma, and beta. Lindsay (1989) provided a mathematical instrument based on moment spaces of unknown mixed distributions. Hill (1975), Weissman (1985) developed general approaches without assuming a specific global parametric form but based in fitting a suitable parametric model or a Zipf form to a few of the largest or smallest order statistics. Dumouchel (1983) evaluated its robustness based upon the stability property of the distribution and suggested a robust procedure to estimate and compare tail shapes. Jureèková, Koenker y Portnoy (2001) studied the tail behaviour of the least-squares estimator in the linear regression model.

2 Other alternatives used frequently in empirical work are: The Student t distribution and the mixture of normal distributions.

II. ¿WHY MEASURE THE TAIL AREA OF THE EXCHANGE RATE RETURNS?

Knowledge of the unconditional probability distribution of the exchange rate returns has important finance and economics applications in situations where uncertainty related to exchange rate movements must be measured. For example, analysis regarding the effects of exchange rate volatility on international trade and capital mobility, models of foreign exchange transactions costs, mean-variance analysis of international assets portfolios, studies of exchange rate efficiency, or pricing of options on foreign currencies. Furthermore, the inflation targeting framework follows the exchange rate dynamics. Finally, we must take account of recent evidence suggesting a strong correlation between nominal and real exchange rates under a floating rate system.

Summing up unpredictable changes in exchange rate has pervasive effects for macroeconomic stabilization, with consequences for prices, wages, interest rates, production levels, and employment opportunities. These facts have direct or indirect implications for the welfare of economics agents. Frequently a measure of dispersion such as the variance of exchange rate changes is used to account for the uncertainty, and the implementation of some parametric models based on normal assumptions is typically exploited. However, it is now generally accepted among economists in this field that short term foreign currencies are fat tailed behaved and more peaked around the origin.[3]

The α- area of the tail of a distribution is also analyzed in this context as a measure of the extreme(?) behavior of exchange rate movements and is related to the fourth raw moment.[4] By using a direct estimation of α we also test for regime switches affecting changes of the distribution of exchange rates over time. Thereby, the economic interpretation when α is increased is that extreme exchange rate changes have become less frequent over time. Thus, the instability of the exchange rate over time can be tested through the hypothesis $H_0 : \alpha < 2$. Hence one may wonder whether the change of regimes or international environment has led to changing the tail of the distribution. Thus, we could match the effect of the changes under the stability test.

2.1. Data Description

In applied work with exchanges rates, returns are preferred to levels when there is a high frequency data. The main reason is that investors compare returns rather than levels and capital movements are the prime cause of short run exchange fluctuations. The

---

3 It means that a distribution has fatter tails, relative to the normal distribution.

4 This moment gives information about kurtosis. While in the parametric framework the tail area can be calculated through $\alpha(x) = \int_x^\infty f(x)dx$, the non- parametric one estimates the heaviness of the tail without assuming a particular global parametric form.

stylized facts of exchange rates can be summarized as follows: the logarithm of exchange rates is usually non stationary (Fig 3). This fact may cause problems of spurious correlation unless proper care is taken. In contrast, returns are stationary (figure 4) and also exhibit little serial correlation (fig 5). Kurtosis is significantly different from zero. In general, sample moments of returns imply lack of skewness for freely floating currencies.

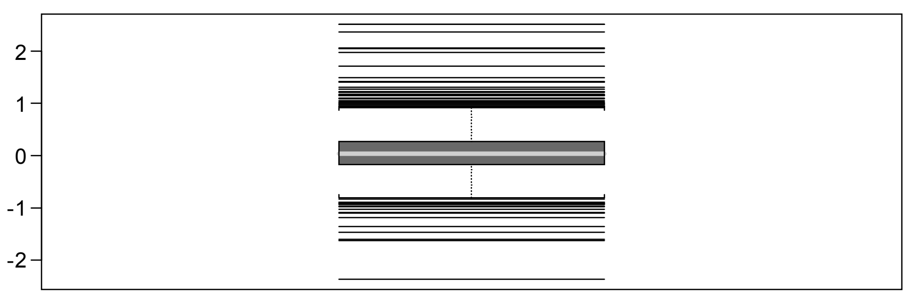
The basic data for our analysis consists of the daily exchange rates (in terms of the U.S. dollar), which cover the period since the floating exchange rate system started. Thus, the whole sample goes from September 29-1999 to October 10-2003 excluding holidays. We calculate the rate of change by taking the logarithm difference between the close of two successive trading days.[5]

For reference purposes, we record in the table and figures below some basic sample statistics. Shown in table 1 are: the sample standard deviation (volatility), the sample skewness, the sample normalized kurtosis and the quantiles of highest observations.

Table 1
Sample Statistics for Daily Returns

Std. Deviation	Skewness	Kurtosis	99%	95%
0.47	0.59	4.10	1.41	0.86

Figure 1
Box Plot for Returns



At a glance a boxplot shows the location, spread, and skewness of the data, along with observations that may be outliers. The bar across the box at the median summarizes location; the length of the box shows spread and the relative position of the median provide an indication of skewness.

[5] $R_t = 100*\ln(tc_t/tc_{t-1})$, where tc is the official exchange rate.

Figure 2
Upper Tail Percentile Window

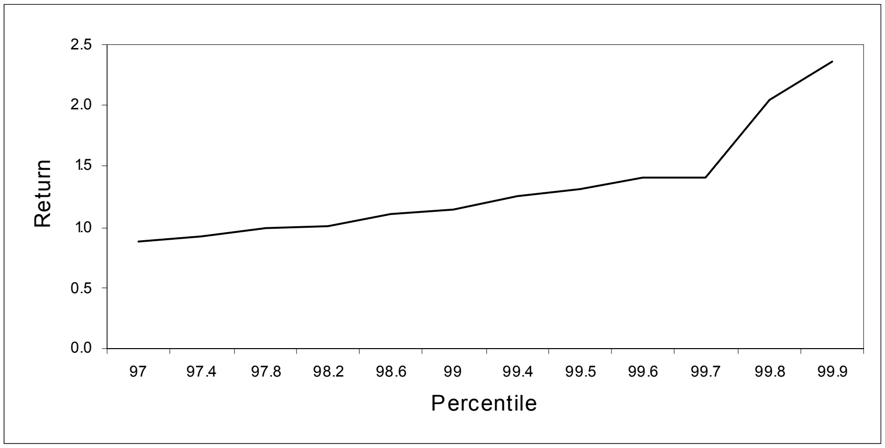


Figure 3
Daily Official Exchange Rate

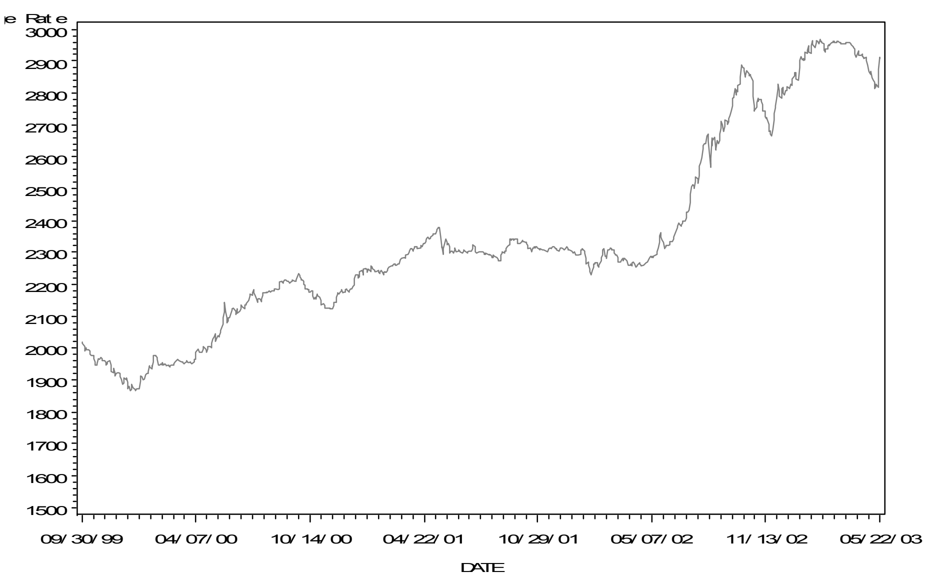


Figure 4

Daily Returns

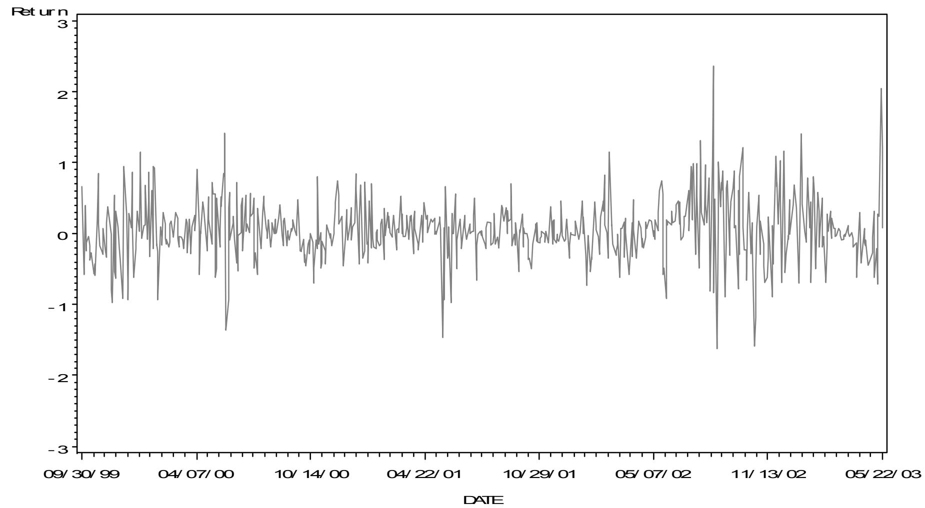
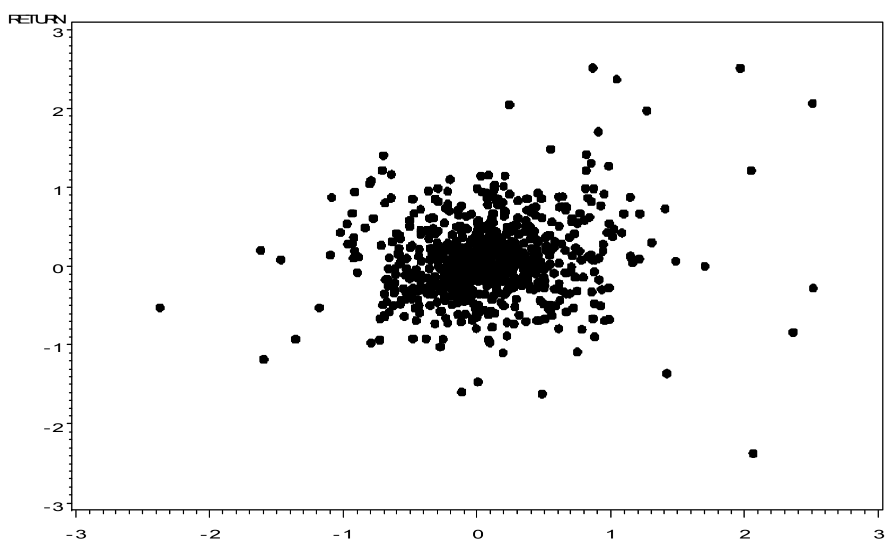


Figure 5

Scatter Diagram for Returns

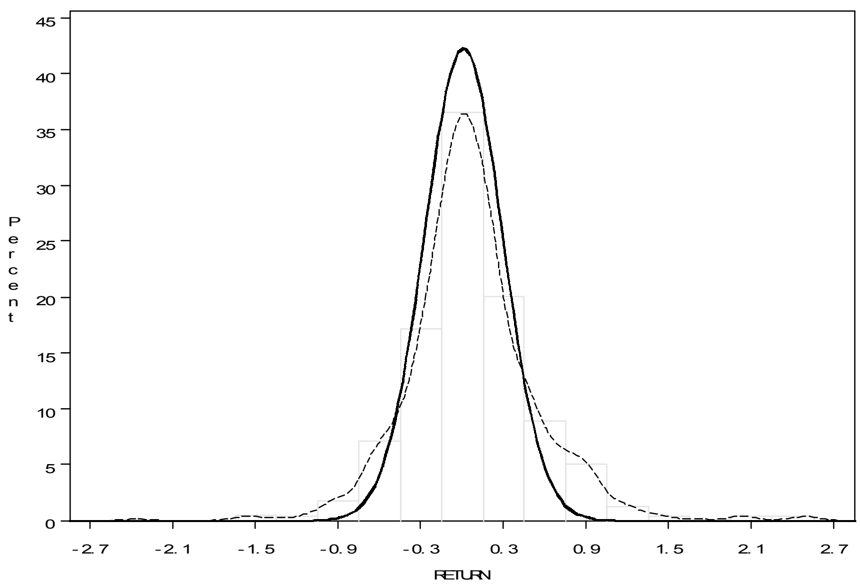


III. MODELING THE DISTRIBUTION OF EXCHANGE RATE TIME SERIES

It is believed as an empirical fact that the distribution of exchange rate returns is usually fat-tailed and more peaked around the origin than the normal one. Various statistical probability distributions seem to share these features: the student *t* distribution, the mixture of two normal distributions,[6] the stable Paretian distributions and the random geometric summation stable distributions.[7] In this section, we describe the main characteristics of these kinds of distributions and compare them with the empirical distribution built with non-parametric fitting that serves as a benchmark.

Figure 6

Estimated Normal and Empirical Density for Exchange Rate Returns



3.1. Stable Paretian distributions[8]

These kinds of distributions have several theoretical properties.[9] They are stable with respect to addition and scaling which implies the same shape over different periods. This property is relevant in finance: If weekly returns are the sum of daily

[6] This fact allows enjoying the nice and useful normal properties.
[7] This family covers distributions such as: Laplace and Weibull.
[8] From which the lower the exponent the fatter the tails are. These distributions fit nicely the ancient Law of Proportional Effect proposed by Robert Gibrat. Its main assumption is that the exponent does not change when observations are summed. Moreover, the implications of these kinds of distributions are related with the exact amount of tail-fatness that is involved. Summing up, the sum stable laws are also related with the additive property.
[9] "The only way the sum can get large is by one of the summands getting large".

returns, then weekly returns have a stable distribution with the same tail index as the daily returns. They are flexible in fitting exchange rates and almost all the studies take into account this kind of distributions.

More specifically, let $X_1, X_2, \ldots X_n$ be a random sample with the same distribution function *F*. *F* is Paretian stable if there exist constants $a_n > 0$ and $b_n \in R$ such that for any *n*

$a_n(X_1 + X_2 + \ldots + X_n) + b_n \overset{d}{=} X_1$. Where *d* means *distributed*.

The characteristic function[10] has the explicit representation:

$$\int e^{itx} dF(x) = \begin{cases} \exp\{-c^{\alpha} |t|^{\alpha} [1 - i\beta sign(t) \tan \frac{\pi\alpha}{2}] + i\delta t\}, \text{ if } \alpha \neq 1 \\ \exp\{-c|t|[1 + i\beta \frac{2}{\pi} sign(t) \ln|t|] + i\delta t\}, \text{ if } \alpha = 1 \end{cases}$$

Where $sign(t) = \begin{cases} 1 \text{ if } t > 0 \\ 0 \text{ if } t = 0 \\ -1 \text{ if } t < 0 \end{cases}$

The characteristic exponent $\alpha\left(0 < \alpha < 2\right)$ is the index of stability and can also be interpreted as a shape parameter; β $(-1 \leq \beta \leq 1)$ is the skewness parameter; δ $(\delta \in R)$ is a location parameter; and c $(c > 0)$ is the scale parameter.

The stability property of a distribution plays an important role in measuring the area of the tail. It is based on the characteristic that a quantity may have a stable distribution if it can be thought as the sum of independent effects. In other words, the *iid* variables X_i have a stable distribution with index α if $X_1 + \ldots + X_n$ has the same distribution as: $\delta_n + n^{1/\alpha} X_1,\ 0 < \alpha \leq 2$. If $\alpha = 2$ then *X* is distributed normal, whereas if $\alpha < 2$ the distribution is called Stable Paretian due to the fact that the tail probabilities are quite approximated to those of the Paretian probability distribution in which $P(X > x) = kx^{\alpha}, as \quad x \to \infty$.

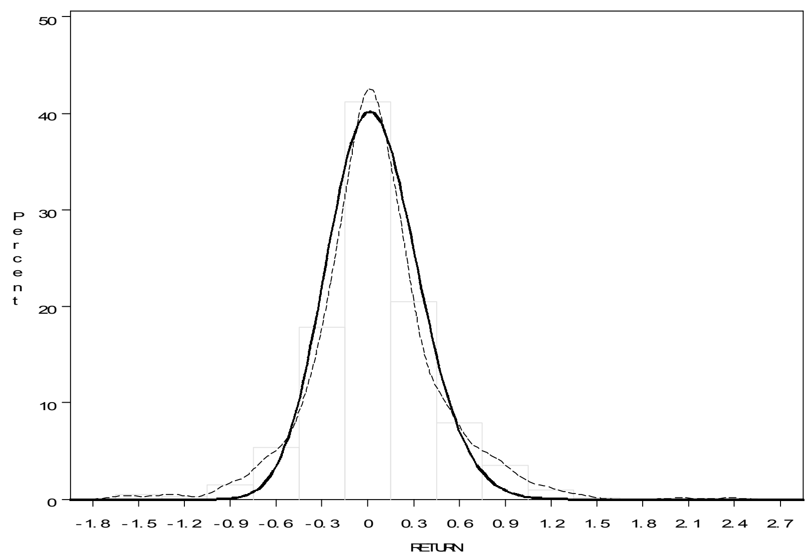
---

[10] The density is determined uniquely by the characteristic function. Regularly varying: $G(x) = e^{-x^{\beta}} = L(x)/x^{\alpha}$ where L is slowly varying, this is: $L(tx)/L(x) \to 1$ as $x \to \infty$ for all t; for example this covers Pareto and log-gamma distributions.

Stable distributions have nice properties and provide parameters of location, scale and skewness. If a stable distribution with $\alpha < 2$ is a good candidate to explain data, therefore, the property $P(X > x) = kx^{\alpha}$ is used to estimate the probability of extreme deviations and it is used as a good indicator of tail behavior. In this case the maximum likelihood estimator will have an asymptotically normal distribution with mean α and variance determined by the Fisher Information frontier.

Figure 7

Estimated Gamma and Empirical Density for Exchange Rate Returns



3.2. The geometric summation model and some of its properties

The idea underlying the geometric summation model is to allow for the fact that financial markets may with some small probability change their characteristics in any given period.

To state the geometric summation stable model formally, let X_i be the change of an exchange rate during the period $t = t_0 + i$. In each period we expect the occurrence of an event that significantly changes the characteristic of the return process. Let $T(p)$ be "the number of periods after which such an event is expected to occur". Therefore, we assume that $T(p)$ is an independent random variable geometrically distributed. So, its probability function is given by:

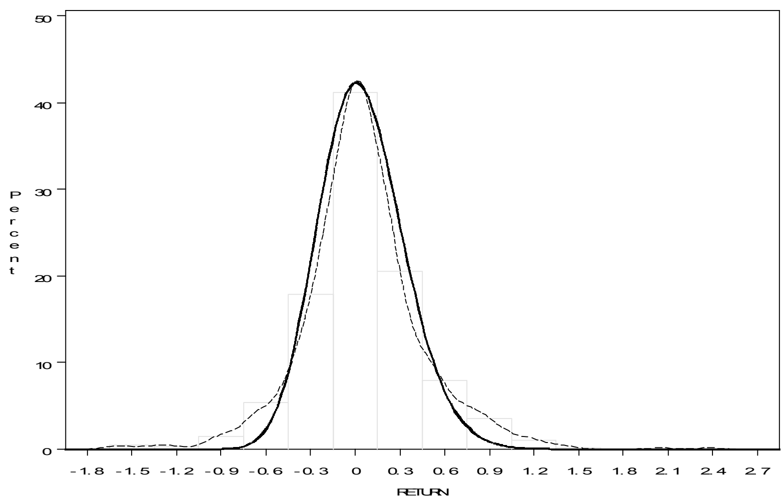
$$\Pr\{T(p) = k\} = (1-p)^{k-1} p, \text{ k=1, 2,}$$

The geometric sum $G_P = \sum_{I=1}^{T(P)} X_I$ represents the accumulation of the X_i's up to the event at time $t_0 + T(p)$, which means the total change of the exchange rate over that period of time. The distributions that belong to this family are: Lognormal, Weibull and Laplace.[11]

The use of random summation stable distributions has been proposed recently as an effective alternative in financial modeling. The main attribute is that they are stable under different underlying probability schemes.

Figure 8

Estimated Log-normal and Empirical Density for Exchange Rate Returns

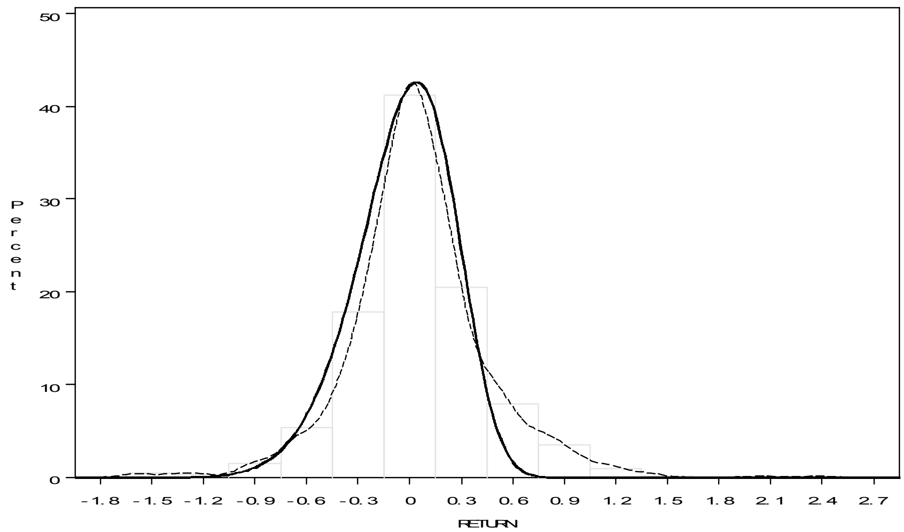


[11] Lognormal: the distribution of $e^{\mu+\sigma U}$ where U is standard normal. Weibull with decreasing rate, $G(x) = e^{-x^{\beta}}$ with $0 < \beta < 1$ where $G(x) = 1 - G(x)$ is the tail.

Weibull: $F(x;\alpha,\lambda) = \begin{cases} 1 - \exp\{-\lambda x^{\alpha}\}, & si\ x \geq 0 \\ 0 & si\ x < 0 \end{cases}$

Laplace: $G(x) = \frac{\lambda}{2}\int_{-\infty}^{x} e^{-\lambda|u|} du$

Figure 9
Estimated Weibull and Empirical Density for Exchange Rate Returns



3.3. Goodness of fit Tests

A goodness of fit test is a statistical hypothesis test, that is used to assess formally whether the observations are an independent sample from a particular distribution with distribution function $\hat{F}$. That is, a goodness of fit test can be used to test the following null hypothesis:

H_0 : *The exchange rate returns come from the distribution function* $\hat{F}$[12]

Kolmogorov-Smirnov Tests. Compare an empirical distribution function with the distribution function of the hypothesized distribution.

$D_n = \sup\{| F_n(x) - \hat{F}(x)|\}$. A large value of D_n indicates a poor fit.

Anderson-Darling Tests: This test is designed to detect discrepancies in the tails and has higher power that the K-S test. The statistics is defined by:

---

[12] As usual in the folklore of statistical science failure to reject H_0 should not be interpreted as "accepting H_0 as being true".

$A_n^2 = n\int_{-\infty}^{\infty}[F_n(x)-\hat{F}(x)]^2\Psi(x)\hat{f}(x)dx$. Where $\psi(x) = 1/\{\hat{F}(x)[1-\hat{F}(x)]\}$ is the weight function. A large value of A_n^2 indicates a poor fit.

Table 2
Evaluation of Candidates Models

Model	$S-K$	$A-D$
Normal	0.10552	47.4159
Weibull	0.93143	160.8622
Gamma*	0.07541	27.3441
Log-Normal	0.07946	35.3716

*The sum stable distributions are the best candidates.

IV. LET THE TAILS SPEAK FOR THEMSELVES.

Given the dispute over the specific distribution and the fact that returns are fat tailed we use a robust procedure based on Zipf's distributions to measure the upper tail of the distribution.[13] This distribution is the discrete version of the symmetric stable distribution class.

A natural and effective way of modeling the tail behavior is looking before at the data and letting the tails speak for themselves. Its main characteristic is that inferences about tails do not depend on the center of the distribution. Thus, there is not a global distribution that governs the behavior of the tails. Let $X^{(1)}$, $X^{(2)}$ ... $X^{(n)}$, from the distribution F with algebraic tail form $1-F(x)\sim Cx^{-\alpha}$ as $x \rightarrow 0$. Where $X^{(1)}$ is the reversed order statistic. α is called the upper tail index. Hill introduced an estimator (Hill, 1975) derived from conditional maximum likelihood considerations on the descending order statistic defined as:

$$\alpha_H^{(r)} = \frac{r+1}{\sum_{i=1}^{r} i\ln(\frac{x^{(i)}}{x^{(i+1)}})}$$

[13] The law, introduced by the linguist George Kingsley Zipf, describes the relation between the data value and its order in a time series accordingly ordered. Zipf´s law can be stated as: $rx_{(r)} =$ constant, where r is the rank of the observation and $x_{(r)}$ is the magnitude or the frequency of the occurrence. The main characteristics of this are a few elements that score very high, a medium number of elements with middle scores and a huge number of elements that score very low.

Where $r + 1$ is the number of observations above the threshold D. This formula requires a choice of r.

4.1. Procedure to choose the cutting observation *r*

A variety of analytical techniques can be used in finding the number of extreme order statistics. A general methodology proposed by Hill is based mainly on the Renyi representation theorem, which states that for $i = 1, 2, ..., n$.

$$x^{(i)} = F^{-1}[Exp - (\frac{e_1}{n} + \frac{e_2}{n-1} + ... + \frac{e_i}{n-i+1})]$$

Where e_i are independent exponentially distributed random variables, each one with expectation 1. Thus, using the fundamental result in the theory of rank statistics that $F(X) \approx U$ and $\log U \approx e$. Inverting and solving for e_i

$$e_j = (n - j + 1)(\ln F(X^{j-1} - \ln F(x^{(j)})$$

By definition $F(x^{(0)}) = 1$ and j=1, 2,…, n.

Assuming $F(x) \sim 1 - Cx^{-\alpha}$ for $x \geq D$ with D unknown, Hill proposes to construct random variables $V_i = \log X^{(i)} - \log X^{(i+1)}$ where $X^{(i)}$ is the i^{th} reverse-order statistic and (??) the tail area can be calculated as: $\alpha_r = (r^{-1} \sum_{i=1}^{r} iV_i)^{-1}$. Summing by parts we can express $V_i = \ln(X^i / X^{i+1})$ for $i = 1, 2, ..., n-1$.

It follows that conditional upon $X^{(r+1)} \geq D$, $\alpha\, iV_i = e_{n-i+1}$ for $i = 1, 2, ..., r$. The choice of $r + 1$ is crucial to know the behavior of iV_i . If it is small then iV_i follows an exponential distribution with parameter α for $i = 1, 2, ..., r$.

On the other hand, iV_i show a systematic difference from the exponential distribution. Hill proposes some methodologies of estimating the cut observation, r . First, a general frequentist approach, which tests the hypothesis that iV_i comes from an exponential distribution for $i = 1, 2, ..., r$. The chi-square distribution is used as the appropriate theoretical distribution which decides if the hypothesis is accepted for a particular r. Thus, we can increase r step by step until we do not find enough evidence to accept the hypothesis that iV_i follows an exponential distribution. Furthermore, Hill proposes other test statistics as measures of discrepancy:

$$H^{(r)} = \alpha_0^2 \sum_{i=1}^{r} (iV_i - \alpha_o^{-1})^2$$

And $K^{(r)} = \sum_{i=1}^{r} (\ln(iV_i) - r\sum_{i=1}^{r} \ln(iV_i)]^2$

Where:

$$\alpha_0 = \alpha_0(r) = r\left[\sum_{i=1}^{r} iV_i\right]^{-1} = r\left[\sum_{i=1}^{r} \ln Y^{(i)} - r\ln Y^{(r+1)}\right]^{-1}.$$

The distributions of the statistics, conditional upon $Y^{(r+1)} \geq D$ can be derived. Hsieh similarly looked at a variant of the above formulation and proposed a robust procedure to find r. The tests statistic is:

$$H^{(r)} = \alpha_0^2 \sum_{i=1}^{r} (iV_i - \alpha_0^{-1})^2 = \frac{\sum_{i=1}^{r} (e_i - e\ )^2}{e^2}$$

which measures the discrepancy between the underlying distribution and the exponential distribution. The idea of this procedure for choosing the optimal r is to test the hypothesis that iV_i has an exponential distribution for $i = 1, 2, ..., r$ by using the statistic $H^{(r)}$. If the hypothesis is accepted at a particular r, then r can be increased until the evidence to accept the hypothesis is not enough. Hsieh developed a systematic methodology to implement this test statistic (Hsieh 1999) through a Monte Carlo simulation study. We adapted the optimal rule for determining r under specific distributions and sample size.

4.2. The optimal decision rule

Decisions problems are well-defined by a set that contains: an action space A, a parameter space Θ, and a loss function $L: Ax\Theta \to R$. Under this framework the tail index represents the true state of nature, thus $\alpha \in \theta$, the action space set, formed by all possible actions that a statistician can take for a given sample $X = x$ is given by $A = \left\{\alpha_H^r : 2 \leq r+1 \leq n\right\}$ which is a set of all Hill estimates for a known sample of size n.

The decision function (rule) is determined by $\delta_s(X) = \alpha_H^{r'}$ where $r' = \max\left\{2, \min\left\{r : \phi_H > s\right\} - 1\right\}$. s is a predetermined decision parameter and its choice is based on the tradeoff between the variation and bias in the estimates. The statistician must choose the parameter *S* taking into account that the risk function is minimized.[14]

4.3. Monte Carlo analysis

We perform a Monte Carlo simulation experiment suggested by Hsieh in order to find the optimal *S*.[15] Through this experiment we obtain the value of α under the underlying Pareto, log normal gamma and Weibull distributions and for our specific sample size. Then, we calculate the risk function that is minimized, which in turn allows us to formulate s as a function of the sample size. Finally we obtain the decision rule. The simulation process is described below:

Parameter space

$\Theta = \{\alpha : \alpha = 1.0, 2.0(.1)\}$, which means that α varies from 1 to 2 with a bean of 0.1

Choice of *S*:

$\Psi = \{s : s = .1.0, 3.0(.1)\}$, which means that s varies from 0.1 to 3 with a bean of 0.1.

For a fixed α we generated *k* random samples of size *n* from the chosen distributions F_1. Given a fixed *S* and the *ith* random sample, the decision rule $\partial_s(x_i) = \alpha_H^{(r_i')} = \alpha^{(r_i')}$ is applied and $r_i' = \min\{r : \phi_H > s\}$. Thus, the Hill estimate $\alpha_i, i = 1, 2, \ldots k$ is calculated and evaluated to respect to different loss functions. The process is repeated for each value of s and the optimal decision *S* is chosen to one that minimizes the risk of ∂_s at α.

Tables in the Appendix A shows the risk, $R_s(\alpha)$ at s=.1, 2.0 (.1) for Pareto random samples of different sizes. The value of s is calculated fitting the OLS estimator on the optimal value of *S* as a function of the sample size. We obtain for this specific

---

[14] The Risk function is given by: $R(\alpha, \delta(X)) = E_X L(\alpha, \delta(X)) = \int L(\alpha, \delta(x)) f(x/\alpha) dx$. In this case the loss function $L(\alpha, \delta_s(X))$ is defined on the product space $\Theta x A$.

[15] We would like to point out that the pXrice we have to pay for using estimators that come from simulation techniques is that the limiting distribution and the rate of convergence of the estimator are unknown.

sample size the value of the parameter that should be chosen by the statistician under the rule:

$$S^* = 1.57$$

4.4. Measuring the tail area of the flexible exchange rate return

In this section we calculate the amount of tail-fatness of the daily exchange rates returns based on the decision rule obtained previously. Table 1 shows the tail index estimation results from the rule. We reported: the number of observations, $\hat{r}+1$, the tail estimate $\hat{\alpha}_H^r$, the estimated cutoff point $\hat{D}$ and the standard deviation. The rule chooses $r + 1 = 72$ as the cut observation. Thus, the estimation of the tail is $\alpha_H^{72} = 2.56$ and the threshold above which the algebraic tail is valid is estimated above the return $\hat{D} = 0.62$.

Table 3

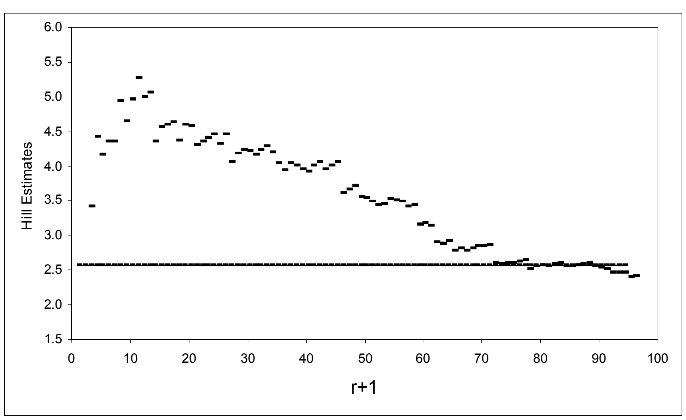
Tail index estimation
The standard deviation is calculated with the formula
$\alpha(r+1)/(r\sqrt{r-1})$

n	$r+1$	$\hat{D}$	$\hat{\alpha}$	std
498	72	0.62	2.56	0.31

We can test for $H_0 : \alpha < 2$ against the alternative $H_1 : \alpha > 2$ on the basis of the asymptotic normality. At the 5 percent significance level the asymptotic confidence interval is given by [1.92, 3.17]. Thus, H_0 is not rejected. Therefore, returns are clearly fat-tailed and the currency is highly volatile.

Figure 10 displays the estimation of the tail area as a function of the of the extreme order statistics r.

Figure 10
Optimal Decision Rule and Hill Estimates of Exchange Rate Returns



V. CONCLUSIONS

The paper focuses on two main issues related to the behavior of the Colombian flexible exchange rate returns: modeling the underlying distribution and measuring the amount of tail-fatness. First, we have fitted various candidate distributions to model the daily returns for the period in which the currency has floated freely. The change from a controlled exchange regime to a flexible one has always caused a period of high volatility during the beginning of the new period according with the international experience.

In this research we found that some distributions nicely fit the center of the distribution. However, the family of symmetric stable Paretian distribution has the best fit and dominates the other alternatives. The distributions of the geometric family under consideration are thin tailed and not capable of capturing the characteristics of the returns. The goodness of fit, measured in terms of Kolmogorov and Anderling tests shows the results. The normal distribution which is the model extensively used in finance appears to be very inappropriate for modeling return of exchange rates.

Future work in this topic will examine the fitting of other members of these families. It is also possible to explore other views which presume that the data come from distributions that vary over time. Thus, more information will lead us to test the *iid* hypothesis and to find distributions across days of the week in order to set appropriate statistical properties of the exchange rate time series. The results also suggest the modeling of conditional distribution of returns as residuals of an *ARMA* model.

Based on the optimal decision rule a practical estimation of the number of extreme order statistics required in the computation of the Hill estimator is extended to some special distributions. A statistical test of discrepancy and the moment space are used according to Hsieh to find the optimal decision parameter, which in turn sets the extreme cut observation and the threshold value. It is noted that the statistician should choose the decision parameter. The estimation is achieved from a Monte Carlo simulation strategy. The results indicate that the currency at the beginning of the flexible regime is highly volatile.

APPENDIX A: SIMULATION RESULTS

Choosing the optimal s from 500 Pareto samples of size 100 with shape Parameter α under the mean squares error (Minimum values in bold)

s	Tail index										
	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0.1	0.04150	0.02779	0.02470	0.02293	0.02020	0.01822	0.01677	0.01361	0.01173	0.00688	0.00761
0.2	0.03934	0.03384	0.02105	0.02283	0.01838	0.01388	0.01511	0.00888	0.01017	0.00794	0.00787
0.3	0.03210	0.02267	0.02743	0.02067	0.01658	0.01812	0.01499	0.01171	0.01066	0.00892	0.00778
0.4	0.03129	0.02970	0.01874	0.01875	0.01408	0.01677	0.01215	0.01054	0.01024	0.00967	0.00752
0.5	0.03362	0.02337	0.02080	0.02167	0.01540	0.01407	0.01364	0.01122	0.00921	0.00733	0.00814
0.6	0.02870	0.01962	0.02092	0.01406	0.01377	0.01242	0.01177	0.00944	0.00771	0.00712	0.00574
0.7	0.02803	0.02167	0.01974	0.02179	0.01426	0.01016	0.00910	0.00865	0.00707	0.00711	0.00740
0.8	0.02561	0.02769	0.02028	0.01876	0.01259	0.00938	0.01091	0.00986	0.00868	0.00737	0.00714
0.9	0.02369	**0.01398**	0.02032	0.01746	0.01602	0.01108	0.00965	0.00915	0.00930	0.00762	0.00544
1.0	0.02619	0.02024	0.02126	0.01344	0.01076	0.01259	**0.00675**	0.00893	0.00782	0.00586	0.00543
1.1	0.02240	0.01719	**0.01282**	0.01355	0.01160	0.00914	0.00765	0.00975	0.00868	0.00782	0.00624
1.2	0.01993	0.02417	0.01527	0.01353	0.01033	0.00979	0.00875	0.00948	0.00831	0.00764	0.00656
1.3	0.02247	0.02050	0.01538	0.01124	**0.01019**	0.00927	0.00946	0.00720	**0.00514**	0.00616	0.00706
1.4	0.02764	0.01912	0.01344	0.01101	0.01045	0.01140	0.00738	0.00656	0.00654	0.00496	0.00627
1.5	**0.01764**	0.01845	0.01592	**0.01044**	0.01166	0.01012	0.00762	0.00739	0.00581	0.00616	0.00572
1.6	0.02130	0.01634	0.01337	0.01356	0.01084	0.01044	0.00960	**0.00630**	0.00780	0.00582	0.00531
1.7	0.02676	0.01973	0.01415	0.01179	0.01358	**0.00725**	0.00898	0.00859	0.00524	**0.00465**	**0.00445**
1.8	0.02107	0.01786	0.01437	0.01170	0.01177	0.01014	0.00801	0.00858	0.00683	0.00765	0.00469
1.9	0.02158	0.01641	0.01604	0.01105	0.01023	0.00828	0.00889	0.00862	0.00657	0.00621	0.00581
Min	**0.01764**	**0.01398**	**0.01282**	**0.01044**	**0.01019**	**0.00725**	**0.00675**	**0.00630**	**0.00514**	**0.00465**	**0.00445**

APPENDIX A (continued): SIMULATION RESULTS

Choosing the optimal s from 500 Pareto samples of size 300 with shape Parameter α under the mean squares error (Minimum values in bold)

s	Tail index										
	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0.1	0.04195	0.02845	0.03105	0.01882	0.01117	0.02239	0.01236	0.01400	0.01138	0.00938	0.00920
0.2	0.03021	0.02669	0.02170	0.01722	0.01855	0.01558	0.01185	0.01284	0.01369	0.01155	0.00750
0.3	0.03012	0.03094	0.02312	0.01717	0.01528	0.01400	0.01151	0.01357	0.00796	0.00837	0.00850
0.4	0.02138	0.02357	0.01604	0.01791	0.01486	0.00981	0.01543	0.01270	0.01242	0.00638	0.00837
0.5	0.03669	0.02567	0.03062	0.02002	0.01644	0.00989	0.00933	0.01055	0.00951	0.01227	0.00625
0.6	0.03646	0.02707	0.02280	0.01170	0.01718	0.01083	0.01418	0.00777	0.00887	0.01025	0.00569
0.7	0.02789	0.02523	0.02109	0.01767	0.01350	0.01176	0.01343	0.00904	0.00916	0.00919	0.00636
0.8	0.03748	0.02469	0.01822	0.02076	0.01186	0.01021	0.01174	0.00885	0.00802	0.00934	0.00722
0.9	0.01822	0.02564	0.02032	0.01740	0.01289	0.01176	0.00822	0.01106	0.00722	0.00609	0.00639
1.0	0.02505	0.02112	0.02496	0.01200	0.01234	0.00960	0.01004	0.00923	0.00545	0.00770	0.00838
1.1	0.02305	0.01468	0.01809	0.01474	0.01100	0.01140	0.00909	0.01097	0.00702	0.00738	0.00708
1.2	0.01801	**0.01259**	0.01600	0.01171	0.01310	0.00853	0.00797	0.00794	0.00826	0.00721	0.00730
1.3	0.02488	0.02621	0.01702	0.01196	0.01060	0.00993	0.00899	0.00872	0.00717	0.00572	0.00703
1.4	0.02299	0.01451	0.01678	0.02082	0.00914	0.01231	0.00688	0.00679	0.00543	0.00668	0.00586
1.5	0.02685	0.01636	0.01805	0.01429	**0.00687**	**0.00641**	0.01082	0.00846	**0.00481**	0.00659	**0.00493**
1.6	0.01734	0.01624	0.02363	**0.01105**	0.01119	0.00890	0.00802	0.00725	0.00555	0.00694	0.00509
1.7	**0.01383**	0.01608	**0.01595**	0.01662	0.01314	0.01030	0.00781	0.01241	0.00966	0.00662	0.00658
1.8	0.03113	0.01822	0.02188	0.01632	0.01009	0.00995	**0.00462**	0.00486	0.00709	0.00459	0.00793
1.9	0.03182	0.01970	0.02577	0.01859	0.00800	0.02259	0.00634	**0.00470**	0.00553	**0.00292**	0.00768
Min	**0.01383**	**0.01259**	**0.01595**	**0.01105**	**0.00687**	**0.00641**	**0.00462**	**0.00470**	**0.00481**	**0.00292**	**0.00493**

APPENDIX A (continued): SIMULATION RESULTS

Choosing the optimal s from 500 Pareto samples of size 400 with shape Parameter α under the mean squares error (Minimum values in bold)

s	Tail index										
	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0.1	0.03643	0.02817	0.02651	0.02225	0.02098	0.01505	0.01332	0.01238	0.01182	0.00976	0.01024
0.2	0.03411	0.02911	0.02834	0.02118	0.01776	0.01771	0.01278	0.01086	0.01042	0.01125	0.00869
0.3	0.03506	0.02777	0.02621	0.01977	0.01718	0.01437	0.01236	0.01270	0.01194	0.01143	0.00775
0.4	0.03512	0.02619	0.02323	0.01543	0.01836	0.01391	0.01142	0.01031	0.01008	0.00782	0.00862
0.5	0.03252	0.02549	0.02286	0.01960	0.01671	0.01420	0.01269	0.00941	0.00956	0.00913	0.00782
0.6	0.02724	0.02371	0.02011	0.01783	0.01499	0.01206	0.01108	0.01019	0.01035	0.00907	0.00740
0.7	0.02686	0.02513	0.01679	0.01716	0.01308	0.01215	0.01040	0.00966	0.01043	0.00797	0.00733
0.8	0.02703	0.02331	0.01906	0.01766	0.01311	0.01369	0.01053	0.00864	0.00854	0.00709	0.00751
0.9	0.02693	0.01970	0.01546	0.01698	0.01356	0.01198	0.00972	0.01005	0.00783	0.00758	0.00651
1.0	0.02737	0.02327	0.01677	0.01792	0.01225	0.00996	0.00882	0.00849	0.00695	0.00677	0.00583
1.1	0.02283	0.01967	0.01739	0.01563	0.01172	0.01048	0.00880	0.00864	0.00906	0.00645	0.00647
1.2	0.02433	0.01930	0.01817	0.01551	0.01117	0.00999	0.00886	0.00817	0.00743	0.00693	0.00551
1.3	0.02565	0.01812	0.01506	0.01364	**0.01044**	0.00991	0.00794	0.00784	0.00743	0.00679	0.00512
1.4	0.02324	0.01896	0.01746	0.01299	0.01209	0.01021	0.00954	0.00735	0.00689	0.00635	0.00541
1.5	0.02164	0.02115	0.01398	**0.01210**	0.01078	0.00996	0.00963	0.00805	0.00774	0.00699	0.00526
1.6	**0.02082**	0.01986	0.01745	0.01372	0.01126	**0.00889**	0.00973	0.00666	0.00695	**0.00572**	0.00617
1.7	0.02358	0.01795	0.01845	0.01351	0.01226	0.00907	0.01004	0.00713	0.00641	0.00596	0.00512
1.8	0.02129	0.02405	0.01674	0.01361	0.01073	0.01023	0.00809	**0.00661**	**0.00634**	0.00697	**0.00508**
1.9	0.02254	**0.01427**	**0.01266**	0.01307	0.01385	0.01097	**0.00577**	0.00776	0.00645	0.00643	0.00526
Min	**0.02082**	**0.01427**	**0.01266**	**0.01210**	**0.01044**	**0.00889**	**0.00577**	**0.00661**	**0.00634**	**0.00572**	**0.00508**

APPENDIX B

This section contains a brief summary of some of the most popular techniques used until now to draw inferences about the area of tail distributions. Even though there are variations and improvement on some of them, the purpose is to summarize the basic ideas employed in the task of estimating these kinds of probabilities.

B.1. Ruin probabilities

This topic is of importance in areas such as reliability, telecommunications systems and insurance risk. In the estimation process the Monte Carlo simulation is widely used to estimate probabilities, expectations or distributions that are not analytically available. Also some statistical methods involving order statistics are extensively used. Distributions under the risk theory can be classified into two groups. Light tails distributions which means the tail $B(x) = O(e^{-sx})$ for some s>0. Thus, the moment generating function is finite. In contrast $B(x)$ is heavy tail if $B(x) = \infty$ for all s>0. Light tails include Exponential, Gamma, Hyper-exponential. Heavy tails distributions include Weibull, Pareto, Loggama, mixtures of exponentials, the sub-exponential class of distributions and distributions with regularly varying tails.[1] The main objective of this approach is to determine the probability that the Value R_t drops below zero given its time evolution. Thus, $\psi(u,T) = P(\inf R_t < 0)$

B.2. Exponential Framework

Considering the method for tails of distributions that look exponential[1] ·, a general approach is derived for continuous distributions from the estimation of the function $k(x)$ and its correspondent derivative (Gross J and Hosmer D): If *X* is an exponential random variable then:

$$\int_x^{\infty} f(y)dy = -f^2(x)/f'(x)$$

Where $f(x)$ and $f'(x)$ are the density and its first derivative respectively. Then there exists a function $k(x)$ such that the tail area can be described by:

$$\int_x^{\infty} f(y)dy = k(x)f^2(x)/f'(x)$$

---

[1], The tail of a distribution is said to be regular varying with exponent α if $B(x) \sim \frac{L(x)}{x^{\alpha}}$, $x \rightarrow \alpha$ and L(x) is slowly varying, i.e. satisfies $\frac{L(xt)}{L(x)} \rightarrow 1, x \rightarrow \infty$. Examples: Pareto, Loggama, and pareto mixture of exponentials.

[1], Include standard normal, t, Poisson, binomial, Chi-squared

Where $k(x) = S(x) f^{'}(x) / f^2(x)$ and $S(x) = \int_x^\infty f(y)dy$

The investigation of $k(x)$ and its derivatives are the key element of approximating tail areas of continuous distributions. For the case of discrete distributions, one presumes the geometric distribution in which

$$\sum_{m=x}^{\infty} {}_x p_m = p_x^2 / \Delta p_x$$

Where $\Delta p_x = (p_x - p_{x+1})$. The sum of the tail probabilities $\sum_{m=x}^{\infty} {}_x p_m$ follows a geometric pattern in the sense that

$$\sum_{m=x}^{\infty} {}_x p_m = d(x) p_x^2 / \Delta p_x$$

Where:

$$d(x) = (\sum_{m=x}^{\infty} {}_x p_m) \Delta p_x / p_x^2$$

and its differences are the basis of approximating tail areas for discrete distributions.

B.3. Moments only for tails

Despite the bad reputation of the method of moments when comparing relative efficiency with maximum likelihood estimation, it offers a powerful set of mathematical tools in determining the tail of an unknown mixing distribution. Some special families of distributions have been catalogued as "quadratic variance property" given that the variance is a quadratic exponential function of the mean (normal, gamma, Poisson, Binomial, binomial, negative binomial, hyperbolic secant). In the set up of this problem a random variable has a mixture distribution relative to a parametric family of distributions $\{F_\theta : \theta \in \Omega\}$

$$F_Q(x) = \int F_\theta(x) dq(\theta) .$$

Where Q is the mixing distribution. The pth moment matrix

$$M_p = \begin{bmatrix} 1 & m_1 & . & . & m_p \\ m_1 & m_2 & . & . & m_{p+1} \\ m_2 & m_3 & . & . & m_{p+2} \\ . & . & . & . & . \\ m_p & m_{p+1} & . & . & m_{2p} \end{bmatrix}$$

The window function $\omega_p(x) = \{V_p^{'}(x)M_p^{-1}V_p(x)\}^{-1}$. Where $V_p(x)$ is the power vector $V_p(x) = (1, x, x^2, ..., x^p)$. The main result comes from the fact that given two arbitrary distributions F(.) and G(.), which is the target, they have the same first $2p$ moments: $m_i(F) = m_i(G) = m_i$ for $i = 0,1,2,...,2p$. With $m_o = 1$. Then, for all values of x, $|F(x) - G(x)| <= \omega_p(x)$.

B.4. The parametric solution inside the tails

The setup of the typical parametric solution is based on the shape of the likelihood of some appropriate functions. Thus, a distribution function that satisfies $F(x) \sim \{(x-\mu)/\beta\}^{\alpha}, x \downarrow \mu$ where μ is the threshold parameter and $(\alpha > 0, \beta > 0)$ are the scale and shape parameters respectively. The log-likelihood function having the first k order statistics if $X_1 < X_2 < ... < X_k$ is given by:

$$l(\alpha, \beta, \mu) = k \ln \alpha - k \ln \beta + (\alpha - 1)\sum_{i=1}^{k} \ln((X_i - \mu)/\beta) - n((X_k - \mu)/\beta)^{\alpha}$$

For α, μ fixed the maximum likelihood estimator of β is given by:

$$\beta = (n/k)^{1/\alpha}(X_k - \mu)$$

With corresponding maximized likelihood:

$$l(\alpha, \beta, \mu) = k \ln \alpha - k \ln n + k \ln k - k\alpha \ln(X_k - \mu) + (\alpha - 1)\sum_{i=1}^{k} \ln(X_i - \mu) - k$$

Setting $\partial l / \partial \alpha = 0$ and $\partial l / \partial \mu = 0$ respectively to obtain a system of simultaneous equation that can be solved iteratively to estimate α and μ. But when $\alpha < 1$ there is no consistent solution of the likelihood. If no local maximum exists the likelihood inference fails so we must use other estimators.

B.5. Log-spline Density Estimation

This method captures the tail of a density based on the idea of modeling the logarithm of a density by a spline function and then estimating the parameters of the model using maximum likelihood (Kooperberg and Stone (1991)). The logarithm of the density function is modeled by a restricted spline function

$$\log f(x;\theta) = c(\theta) + \theta_1 B_1(x) + ... + \theta_{J-1} B_{J-1}(x).$$

Where

$$\theta = (\theta_1, ..., \theta_{J-1})^T \text{. Such that } \theta_1 > 0, \theta_{J-1} > 0$$

$$c(\theta) = -\log[\int \exp\{\theta_1 B_1(x) + ... + \theta_{J-1} B_{j-1}(x)\} dx]$$

The log-likelihood function corresponding to the log-spline family is:

$$l(\theta) = \sum_{i=1}^{n} \log f(X_i; \theta)$$

Which is strictly concave and therefore can find the unique global maximizer. The automatic determination of knots can be done by an automatic knot selection technique.

REFERENCES

DuMouchel, W.H. (1983). "Estimating the Stable Index XXX in Order to Measure Tail Thickness: a Critique". *The Annals of Statistics* 11, 1019-1031.

Hill, B.M. (1975). "A simple General Approach to Inference about the Tail of a Distribution". *The Annals of Statistics* 3, 1063-1174.

Hsieh D. A. (1999). "Robustness of Tail Index Estimation". *Journal of Computational and Graphical Statistics* 8, 338-332.

Jureèková J., Koenker R. y Portnoy S. (2001). "Tail Behaviour at the Least Squares Estimator". *Statistics & Probability Letters* 55, 377-384.

Kooperberg, C., and Stone, C. (1991), "A Study of Logspline Density Estimation". *Computational Statistics and Data Analysis* 12, 327–347.

Lindsay B. (1989). "Moment Matrices: Applications in Mixtures". *The Annals of Statistics* 17 No. 2, 722-740.

Weissman I, (1985). "Maximum Likelihood Estimation of the Lower Tail of a Probability Distribution". *J.R. Statistics.Soc.* B 47 No 2, 285-298.