

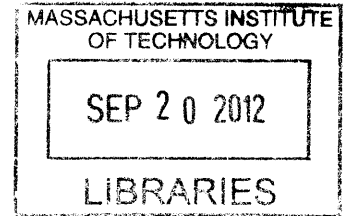
**Essays on Financial Economics**

by

Alberto Vargas Mendoza

B.S. Applied Mathematics  
Instituto Tecnológico Autónomo de México, 2005

**ARCHIVES**



Submitted to the Department of Economics  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

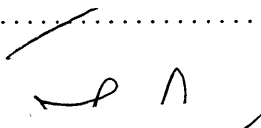
at the

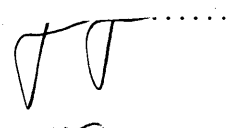
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2012

© 2012 Alberto Vargas Mendoza. All rights reserved

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Signature of Author .....  
  
Department of Economics  
11 July 2012

Certified by .....  
  
Victor Chernozhukov  
Professor of Economics  
Thesis Supervisor

Certified by .....  
  
Anna Mikusheva  
Castle-Krob Career Development Associate Professor of Economics  
Thesis Supervisor

Accepted by .....  
Michael Greenstone  
3M Professor of Environmental Economics  
Chairman, Departmental Committee on Graduate Studies



# Essays on Financial Economics

by

Alberto Vargas Mendoza

Submitted to the Department of Economics  
on 11 July 2012, in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy in Economics

## Abstract

This thesis consists of three independent essays on Financial Economics. In chapter one I investigate the possible mispricing of European-style options in the Mexican Stock Exchange. The source of this problem is that when the Mexican Stock Exchange introduced options over its main index (the IPC) in 2004, it chose Heston's (1993) "square root" stochastic volatility model to price them on days when there was no trading. I investigate whether Heston's model is a good specification for the IPC and whether more elaborate models produce significantly different option prices. To do so, I use an MCMC technique to estimate four different models within the stochastic volatility family. I then present both classical and Bayesian diagnostics for the different models. Finally, I use the transform analysis proposed by Duffie, Pan and Singleton (2000) to price the options and show that the prices implied by the models with jumps are significantly different from those implied by the model currently used by the exchange.

Next, I turn to a problem in behavioral portfolio choice: It has been shown that the portfolio choice problem faced by a behavioral agent that maximizes Choquet expected utility is equivalent to solving a quantile linear regression. However, if the agent faces a vast set of assets or when transaction fees are considerable, it becomes optimal for the agent to take non-zero positions on only a subset of the available assets. Thus, in chapter 2, I present a portfolio construction procedure for this context using L1 penalized quantile regression methods and explore the performance of these portfolios relative to their unrestricted counterparts.

In chapter three, co-authored with Victor Chernozhukov, we present the Extended Pareto Law as an alternative for modeling operational losses. Through graphical examination and formal goodness of fit tests we show that it outperforms the main parsimonious alternative, Extreme Value Theory, in terms of statistical fit. Finally, we show that using the Extended Pareto Law as a modeling technique also leads to reasonable capital requirements.

Thesis Supervisor: Victor Chernozhukov  
Title: Professor of Economics

Thesis Supervisor: Anna Mikusheva  
Title: Castle-Krob Career Development Associate Professor of Economics



# Acknowledgements

The completion of this thesis would not have been possible without the support and encouragement of many people. Among them, I must begin by thanking my advisors Victor Chernozhukov and Anna Mikusheva, whose guidance was invaluable throughout the writing of this work; I truly cannot thank them enough for their time, interest and effort. I am also greatly indebted to Jerry Hausman who agreed to be my third recommender; whose comments were crucial in producing this final version and whose great generosity I will always appreciate. I would also like to thank Whitney Newey, Isaiah Andrews and the participants of MIT's lunch time econometrics seminar for their useful comments.

Among the faculty at MIT I am especially grateful to Panle Jia Barwick, Sara Ellison, Peter Temin, Glenn Ellison, Drazen Prelec, Esther Duflo, Daron Acemoglu, Abhijit Banerjee, Olivier Blanchard, Ricardo Caballero, Ben Olken, Stephen Ross, Antoinette Schoar, Jerome Detemple and Jun Pan. Whether I was their student, teaching assistant or just having a casual conversation with them they were always generous with their time, and their insights have been invaluable in shaping my ideas on economics and teaching.

Outside the classroom the administrative staff at the department of economics has always been incredibly courteous and helpful. I would like to thank especially Emily Gallagher, Kim Scantlebury, Beata Shuster and Katie Lewis for all their hard work during the job market and Gary King for his help through all my time at MIT.

Financially, coming to MIT would have been impossible without external support from several sources. Thus, I would like to acknowledge financial support from the MIT Club of Mexico and from the Mexican Government through CONACyT's doctoral grant 205834.

Moreover, I would not have arrived to MIT without the support of many people. I must first of all thank Pedro Aspe, whose continued support, for over a decade now, has been invaluable academically, professionally and personally. Many other of my professors at ITAM also played a crucial part in my arriving at MIT, and I would particularly like to thank Todd Keister, Cesar Martinelli, Ricard Torres, Gonzalo Hernández, Alejandro Hernández, José Luis Morales, Karim Anaya, Rubén Hernández, Mercedes Gregorio, Guillermo Grabinsky and Rigoberto Vera.

I enjoyed immensely sharing this time at MIT with some extraordinary classmates, from the first semester doing many problem sets with Bruno Ferman to the last semesters sharing an office with Joaquín Blaum, I could not have asked for a better set of classmates than the entering class of 2006. During these years, having friends that came from Mexico to Cambridge at the same time as me was also

a great source of support, so I am grateful to Luis Zermeño, Arturo Aguilar and Maria Elena Ortega. As for the friends who were geographically far but were always there for me, I would like to thank Luis Serra, Elisa Ortega, Héctor Saenz, Mariana Pérez, María José Ferrer, Verónica Padilla, Samyra Uribe, Rocío Sandoval, Antonia Delgado, Rodrigo García de León, Armando Martínez, Rafael Gamboa and Sergio Sánchez.

Finally, but perhaps most important of all, I must thank my family for all their support. My grandmother Clara Vargas has always been an example of hard work and dedication. As for my parents, Olga and José Luis, I cannot begin to describe how thankful I am for all the opportunities they have given me, for all the support and for so many things I could never finish listing here. To my dear wife Mónica, I can only say that I simply cannot imagine the last decade without her, that none of this would have been possible without her and that I dedicate this work to her.

# Contents

<b>1</b>	<b>Bayesian Estimation and Option Mispricing</b>	<b>13</b>
1.1	Introduction . . . . .	13
1.2	Models of Price Dynamics . . . . .	16
1.2.1	General Stochastic Volatility with Jumps Model . . . . .	20
1.3	Estimation Methodology . . . . .	21
1.3.1	Estimated Model and Priors . . . . .	22
1.3.2	Estimation Algorithm . . . . .	24
1.4	Data . . . . .	28
1.4.1	Descriptive Statistics . . . . .	29
1.5	Estimation Results and Model Selection . . . . .	30
1.5.1	Estimation Results . . . . .	30
1.5.2	Implied Jump Process . . . . .	32
1.5.3	Model Selection: Test of Residuals . . . . .	37
1.5.4	Model Selection: Bayes Factors . . . . .	39
1.6	Option Pricing Application . . . . .	40
1.7	Conclusions and Direction of Future Research . . . . .	44
1.8	Appendix A. Drawing from the Conditional Posteriors . . . . .	45
<b>2</b>	<b>Pessimistic Portfolio Selection with Many Assets</b>	<b>61</b>
2.1	Introduction . . . . .	61
2.2	Choquet Expectation and Choquet Expected Utility in Portfolio Selection . . . . .	63

2.2.1	Choquet Expected Utility as an Alternative to Expected Utility . . . . .	63
2.2.2	Choquet Expectation and Coherent Risk Measures . . . . .	66
2.2.3	Pessimistic Portfolios . . . . .	68
2.2.4	Portfolio Selection when $p$ is Large . . . . .	70
2.3	Performance Analysis . . . . .	72
2.3.1	Simulation Exercise: Procedure . . . . .	72
2.3.2	Simulation Exercise: Results . . . . .	73
2.3.3	Fama-French 100 Portfolios . . . . .	77
2.4	Conclusions . . . . .	84
2.5	Appendix B. Portfolio Selection as Regression . . . . .	86
2.5.1	Markowitz Portfolios . . . . .	86
2.5.2	Portfolios that Minimize $\alpha$ -risk . . . . .	87
2.5.3	Portfolios that Maximize Expected Return with an $\alpha$ -risk Restriction . . . . .	89
2.6	Appendix C. Adding the $l_1$ Restriction . . . . .	91

**3 Extended Pareto Law as a Parsimonious Alternative for Operational Risk Modeling 97**

3.1	Introduction . . . . .	97
3.2	The Extended Pareto Law . . . . .	100
3.2.1	The EPL and its Essential Properties . . . . .	100
3.2.2	Motivation for EPL . . . . .	101
3.2.3	Formal Diagnostics and Estimation of the EPL . . . . .	104
3.3	Empirical Results . . . . .	107
3.3.1	Model Validation . . . . .	109
3.3.2	Capital Reserve Estimation . . . . .	116
3.4	Conclusions . . . . .	117



# List of Figures

1-1	Daily log-returns of the IPC index 1987-2011. . . . .	30
1-2	Estimated volatility for two sub-periods under the SV, SVJ and SVIJ specifications. . . . .	31
1-3	Volatility estimates and probability of jump times under the SVJ specification. . . . .	32
1-4	Magnitude and probability of jumps under the SVJ specification. . .	33
1-5	Volatility estimates and probability of jump times under the SVJ specification for the 1994-95 period. . . . .	34
1-6	Returns, volatility and expected magnitude of jumps in both series under the SVCJ specification. . . . .	36
1-7	Returns, volatility, probability of jumps in returns and probability of jumps in volatility under the SVIJ specification. . . . .	36
1-8	Returns, magnitude of jumps in returns, volatility and expected size of jumps in volatility under the SVIJ specification. . . . .	37
1-9	QQ plots for the residuals under the four models. . . . .	38
1-10	Percentage difference in price implied by the SVJ model relative to those implied by Heston's model for out of the money options. . . .	43
1-11	Percentage difference in price implied by the SVJ model relative to those implied by Heston's model for in the money options. . . . .	43
2-1	Active assets as a function of $\ell_1$ constraint. . . . .	73
2-2	Expected and realized $\alpha$ -risk as a function of the constraint on $\ w\ _1$ . . . . .	74
2-3	Out-of-sample distribution for 5 portfolios with different $\ell_1$ restrictions. . . . .	75

2-4	Active assets as a function of the $\ell_1$ constraint when $\alpha$ -risk is restricted to $\alpha_0$ (Case 1) or $\alpha_E$ (Case 2). . . . .	76
2-5	Expected return as a function of $\ w\ _1$ . . . . .	76
2-6	Ex-ante and ex-post $\alpha$ -risk for both levels of risk restrictions. . . . .	77
2-7	Active assets vs $\ell_1$ constraint for the non-stress scenario. . . . .	78
2-8	Ex ante and ex-post $\alpha$ -risk for the non-stress scenario. . . . .	79
2-9	Ex-post cumulative returns for 5 portfolios: equally weighted, minimizing alpha risk (unrestricted and with no-short-sale restriction) and minimizing variance (unrestricted and with no-short-sale restriction). . . . .	79
2-10	Expected and realized return under the two levels of risk restriction. . . . .	81
2-11	Expected and realized $\alpha$ -risk under the two levels of risk restriction. . . . .	81
2-12	Out-of-sample cumulative returns for the stress scenarios. . . . .	82
2-13	Expected and realized return in the stress scenarios. . . . .	83
2-14	Expected and realized $\alpha$ -risk in the stress scenarios. . . . .	84
3-1	Hill plots for simulated Pareto and Extended Pareto samples. The bottom $x$ axis denotes $n$ (the number of upper order statistics used) and the top $x$ axis represents $X_{(n)}$ (the magnitude of the $n$ 'th largest loss). . . . .	103
3-2	$p$ -value of the KS test for the EPL (black line) and GPD (red dots) models as a function of the threshold level. Unit of measure: CPBP-2. . . . .	110
3-3	$p$ -value of the KS test for the EPL (black line) and GPD (red dots) models as a function of the threshold level. Unit of measure: EPWS-ALL. . . . .	111
3-4	PP and QQ plots for EPL and GPD models. Unit of measure: EPWS-ALL. . . . .	113
3-5	PP and QQ plots for EPL and GPD models. Unit of measure: CPBP-2. . . . .	114
3-6	Hill plots for the EPWS-ALL unit of measure. In each case the bottom $x$ axis denotes $n$ (the number of upper order statistics used) and the top $x$ axis represents $X_{(n)}$ (the magnitude of the $n$ 'th largest loss). . . . .	115

3-7 Hill plots for the CPBP-2 unit of measure. In each case the bottom  $x$  axis denotes  $n$  (the number of upper order statistics used) and the top  $x$  axis represents  $X_{(n)}$  (the magnitude of the  $n$ 'th largest loss). . 115



# Chapter 1

## Bayesian Estimation and Option Mispricing

### 1.1 Introduction

The Mexican derivatives market (MexDer) introduced options on the main index of the Mexican stock exchange (the IPC) in 2004. Banks and other financial institutions buy these options in order to hedge their positions on the IPC and often hold them until maturity. Thus, there is little trading in these products and in fact most days there is no trading for many of the options. In spite of this, by law, MexDer is required to publish a price for the options each day, even if there was no trading. These prices are then used by banks and investment firms to mark-to-market their positions.

MexDer bases its pricing procedure on Heston's (1993) "square root" stochastic volatility model (MexDer (2008)). This specification offers more flexibility than the classical Black-Scholes (Black and Scholes (1973)) setting. While the latter assumes a constant level of volatility, Heston's model assumes that volatility follows a mean reverting diffusion process and thus allows for episodes of volatility-clustering. Further, Heston's model retains one of the most useful characteristics of the Black-Scholes setting: a closed-form option pricing formula. Thus, Heston's model becomes a natural candidate when modeling stocks or indices that exhibit

volatility-clustering.

Nonetheless, there is more than one way in which asset dynamics can depart from the geometric Brownian-motion assumption of Black-Scholes. In particular, stocks and stock indices tend to have days with extremely large returns (either positive or negative). This fact has led to a large literature including the work by Bakshi et al. (1997), Bates (2000), Pan (2002) and Eraker et al. (2003) that shows that models without jumps tend to be misspecified for stock indices. Although most of this research was conducted on data from US markets, the presence of days with crashes and large positive swings in the IPC suggests that models without jumps will also be misspecified for this index. Hence, a model that accurately describes the dynamics of the IPC will most likely have to include jumps in the return series or simultaneous jumps in the return and volatility dynamics.

In this paper I investigate whether the model used by the exchange is, in fact, misspecified for the IPC. Next, I test whether models with jumps represent a better alternative. Finally, I examine whether the more elaborate models imply significantly different option prices.

To formally show that jumps are necessary to fit the IPC data I estimate models with and without jumps for the IPC and test their relative performance. However, estimation in the context of stochastic volatility models is not trivial. The main complication arises from the fact that the volatility series is not observed. Including jumps complicates the estimation further in two directions. First, jumps increase the number of unobserved state variables in the model. Second, and perhaps more importantly, it is sometimes difficult for estimation algorithms to disentangle the variation arising from the diffusive element of the dynamics and that arising from the jump component (see e.g. Honore (1998)) .

In spite of these difficulties, several techniques have been developed to estimate and test this family of models. These have included calibration (Bates (1996)), Implied State GMM (Pan (2002)), Efficient Method of Moments (Andersen et al. (1999)) and Markov Chain Monte Carlo (Jacquier et al. (1994) and Eraker et al. (2003)).

When choosing an empirical technique for this problem, an important require-

ment was that I could use the output to price options. Hence, I need not only estimates of the parameters of the models, but also of the unobserved state variables. The MCMC technique of Eraker et al. (2003) does precisely this: it produces estimates of both the parameters and unobserved variables, and thus serves as the appropriate tool for the purpose of this paper.

Beyond the actual estimation, I am interested in the relative performance of the different models under consideration. To investigate this, I perform two kinds of tests. The first is a standard normality test of residuals, while the second is based on Bayes factors and is particularly useful for comparing nested models. Just like option valuation, this procedure depends crucially on estimating the latent variables as well as the parameters, and so, can only be conducted after certain estimation procedures.

From the estimation and model selection results, I conclude that the current stochastic volatility model used by MexDer is misspecified. Further, I find that the model with jumps only in returns is still misspecified, but is an improvement over Heston's model. Finally, I show that models with simultaneous jumps in volatility and returns represent a much better fit for the index dynamics.

Using the transform analysis introduced by Duffie et al. (2000), I then estimate option prices and show that those implied by the stochastic volatility and jumps models are significantly different than those implied by Heston's model. Hence, I find that the option prices implied by a better fitting model for IPC dynamics are significantly different than those currently used to mark-to-market positions.

The rest of the paper is organized as follows: section 2 introduces the family of stochastic volatility and jumps models that we will be working with, section 3 presents the estimation methodology, section 4 describes the data, section 5 presents the estimation and model selection results, section 6 presents the option pricing results, and section 7 concludes.

## 1.2 Models of Price Dynamics

The Black-Scholes model introduced in 1973 is one of the most widely used tools in finance. Not only does Black-Scholes provide a set of reasonable assumptions for the dynamics of asset prices, it also has the great advantage of implying a closed-form formula for the pricing of European-style options. Formally, the Black-Scholes model assumes that the prices ( $S_t$ ) of the underlying asset follow a Geometric Brownian Motion, which implies

$$dS_t = mS_t dt + \sigma S_t dW_t,$$

where  $W$  is a standard Brownian motion process and  $m$  and  $\sigma$  are fixed parameters. By a simple application of Ito's lemma it implies that  $Y_t = \log(S_t)$  behaves according to

$$dY_t = \mu dt + \sigma dW_t.$$

This formulation provides an immediate interpretation of the model:  $dY_t$  can be thought of as an instantaneous measure of log-returns. Thus, the model tells us that log-prices grow according to a deterministic trend ( $\mu dt$ ) that is perturbed by a Brownian motion with volatility  $\sigma$ . Putting together these two assumptions we have a very reasonable starting point for modeling asset dynamics. Further, under no arbitrage assumptions, the Black-Scholes model implies that the price of a European call that matures at time  $T$  with strike price  $K$  is given by

$$S_t N(d_1) - e^{-r(T-t)} K N(d_2), \tag{1.1}$$

where  $N$  is the standard normal cumulative distribution function,  $r$  is the risk free rate and  $d_1$  and  $d_2$  are functions of  $K, S_t, (T - t), \sigma$  and  $r$ .

However, the dynamics of stocks and stock indices exhibit phenomena that are not consistent with the assumptions of the Black-Scholes model. In particular, stock indices tend to experience volatility-clustering; that is, there are periods when the volatility of returns is very high and periods when it is low.

Further, all the variables in the option pricing formula are observable, except  $\sigma$ . Thus, if option prices are observed,  $\sigma$  can be calculated from the option pricing



formula, and if the Black-Scholes model were true, the estimated  $\sigma$  should be the same for every option regardless of its strike price or time to maturity. However, a well-known result is that the implied volatility is highest when the strike price is very high or very low. This result, known as "volatility smiles" or "volatility smirks", is extensively documented in Stein (1989), Aït-Sahalia and Lo (1998) and Bakshi et al. (2000).

Thus, stochastic volatility models were introduced as generalizations of the Black-Scholes framework that allows for volatility-clustering and potentially provides a solution to the "volatility smiles" issue. Some of the first models to include a volatility term that could vary with time were those of Scott (1987), Hull (1987), and Wiggins (1987). However, it was Heston's 1993 model that became widely used because it shared a crucial characteristic with the Black-Scholes model: it offered a closed-form solution to pricing European options. Formally, Heston's model assumes the following asset dynamics:

$$\begin{aligned} dS_t &= mS_t dt + \sqrt{V_t} S_t dW_t^1 \\ dV_t &= \kappa (\theta - V_t) dt + \sigma \sqrt{V_t} dW_t^2, \end{aligned}$$

where  $W_t^1$  and  $W_t^2$  are Brownian motion processes with correlation  $\rho$ . Thus, Heston's model is a generalization of Black-Scholes, where the volatility of the price diffusion is allowed to vary with time. In particular, the volatility is allowed to follow a diffusion similar to an Ornstein–Uhlenbeck process, but where the volatility of the diffusive element is proportional to  $\sqrt{V_t}$ . In spite of this difference, it is important to note that a feature that the volatility process shares with Ornstein–Uhlenbeck processes is the mean reversion of the deterministic drift. This gives an immediate interpretation to  $\theta$  as a long-term value the process tends to return to and  $\kappa$  as the speed with which the process will tend to return to  $\theta$ .

Precisely because volatility is allowed to vary, this specification can accommodate periods when returns will be very volatile, but the model also implies that volatility will tend to return to its long term value  $\theta$  and thus periods of high volatility will be transient.

This added flexibility, together with the availability of a closed-form pricing for-

mula for European-style options, makes the Heston model a considerable improvement over Black-Scholes. However, the added flexibility provided by the introduction of stochastic volatility does not seem to be enough.

Episodes of "crashes" cannot be captured by these dynamics. For instance, under Heston's model, a crash like that of 1987 when the S&P 500 lost more than 20% in a single day would imply an observation of more than 8 standard deviations away from the mean. In order to capture these crashes, models with jumps in the return series were introduced. These jumps take the form of occasional shocks to the return series which will explain large (either positive or negative) returns in a single day which could not be explained by a continuous diffusion such as those in Black-Scholes' or Heston's specification.

The use of jumps in returns goes back to, at least, Merton's 1976 model which augments the Black-Scholes model by adding normally distributed jumps that occur according to a Poisson process. So, following the notation I have been using, Merton's model implies

$$dY_t = \mu dt + \sigma dW_t + \xi^y dN_t^y,$$

where  $N$  is a Poisson process, and  $\xi^y$  is a Gaussian random variable.

This family of models will be able to explain days with sudden large returns. However, since it assumes  $\sigma$  to be fixed throughout time, it will not be able to describe processes where there are periods of high volatility and periods of low volatility. Thus, a natural idea is to combine stochastic volatility and jumps in the return series. Following this direction, Bates (1996) introduces the following model:

$$\begin{aligned} dY_t &= \mu dt + \sqrt{V_t} dW_t + \xi^y dN_t^y \\ dV_t &= \kappa (\theta - V_t) dt + \sigma \sqrt{V_t} dW_t^2. \end{aligned}$$

This gives us a specification that can accommodate both volatility-clustering via the stochastic volatility element and sudden large returns through the jump component. So far, we have a model built around Black-Scholes, with added characteristics tailored to explain the phenomena that Black-Scholes cannot. Hence, we would ex-

pect this new model to provide a good fit to the data of stock indices. In spite of this, Bakshi et al. (1997), Bates (2000) and Pan (2002) find that models with jumps only in the return series are still misspecified for stock index data. The rationale is the following: having jumps in the return series can explain an isolated episode of a very large return. However, data shows that frequently, when there is a day with a very large loss, the next day tends to have a very large gain. In contrast, under Bates' specification, jumps arrive according to a Poisson process with arrivals a few times per year. Thus, observing consecutive days with jumps on several occasions would once more imply that we are observing very low probability events.

A natural extension of Bates' model is to add a jump process to the dynamics of the volatility series. This will allow for a sudden increase in the volatility of returns, therefore allowing for periods when there are very large returns on consecutive days. Although this is achieved partially by introducing stochastic volatility, by introducing jumps in  $V$  we allow for a sudden "burst" in the returns instead of a gradual increase in their dispersion. Thus, Duffie et al. (2000) introduce a family of models that generalize Bates' as follows:

$$\begin{aligned} dY_t &= \left(m - \frac{1}{2}V_t\right) dt + \sqrt{V_t}dW_t + \xi_t^y dN_t^y \\ dV_t &= \kappa(\theta - V_t) dt + \sigma\sqrt{V_t}dW_t^2 + \xi_t^v dN_t^v, \end{aligned} \tag{1.2}$$

where  $N^v$  is a Poisson process, and  $\xi^v$  represents the magnitude of jumps in volatility at time  $t$ . Although Duffie et al. (2000) allow for a variety of distributions for  $\xi^v$ , we will be interested in the case where  $\xi_t^v$  is exponentially distributed. Clearly, the model described by system (1.2) is a very general specification that allows for jumps in both returns and volatility. Thus, sudden "bursts" in the volatility of returns can be described in this framework. Further, since it retains jumps in returns, it can potentially encompass all the phenomena that we have described above and that simpler models are not capable of capturing.

For the rest of this paper, we will work with a variation on model (1.2). We begin by presenting this general stochastic volatility with jumps model and later describe restrictions that will allow us to test whether jumps are necessary for the particular index we are modeling.

### 1.2.1 General Stochastic Volatility with Jumps Model

Let  $S_t$  denote the price of the index at time  $t$  and let  $Y_t = \log(S_t)$ . The most general form of stochastic volatility model with jumps which we are interested in can be expressed as

$$\begin{aligned} dY_t &= \mu dt + \sqrt{V_{t-}} dW_t^1 + \xi^y dN_t^y \\ dV_t &= \kappa(\theta - V_{t-}) dt + \sqrt{V_{t-}} \sigma_v \left( \rho dW_t^1 + \sqrt{1 - \rho^2} dW_t^2 \right) + \xi^v dN_t^v, \end{aligned} \quad (1.3)$$

where  $V_{t-} = \lim_{s \uparrow t} V_s$ ,  $(W^1, W^2)$  is a standard Brownian Motion in  $\mathbb{R}^2$ ,  $N_t^y$  and  $N_t^v$  are Poisson processes with constant intensities  $\lambda_y$  and  $\lambda_v$  respectively and  $\xi_t^v \sim \exp(\mu_v)$  and  $\xi_t^y \sim N(\mu_y + \rho_J \xi_t^v, \sigma_y^2)$  are the jump sizes in volatility and returns respectively.

Thus, we are assuming that log-prices behave according to a Brownian motion with drift  $\mu$  and volatility  $\sqrt{V}$ . The departure from standard Brownian motion with drift is that this process is hit by normally distributed shocks with arrival times that follow a Poisson process. Furthermore, the volatility  $\sqrt{V}$  itself behaves according to a mean-reverting diffusion with long term value  $\theta$  that is hit by exponentially distributed shocks. These shocks also occur according to a Poisson process. Finally, the Brownian motions driving both diffusions are assumed to have a correlation of  $\rho$ .

This general specification includes the Black and Scholes (1973) and the Heston (1993) specification as particular cases: if  $\sigma_v = 0$  and  $N_t^y \equiv N_t^v \equiv 0$  for all  $t$ ,  $Y$  would follow a geometric Brownian motion, as described by the Black-Scholes model. If instead we only restrict  $N_t^y \equiv N_t^v \equiv 0$  for all  $t$  we have Heston's "square root" stochastic volatility model (SV model from now on). In addition, we will also be interested in the cases where:

1.  $N_t^v \equiv 0$  the case where there are jumps only in the returns and not in the volatility (SVJ model).
2.  $N_t^y \equiv N_t^v$  the case where there are correlated jumps in both returns and volatility, but they are restricted to occurring simultaneously (SVCJ model).

3. The case where there are jumps in both returns and volatility and these jumps occur at independent times (SVIJ model).

### 1.3 Estimation Methodology

With all the models presented above, we face a common problem: all of these models describe continuous time dynamics, but the data will invariably be discretely sampled. Thus, the estimation of the parameters will always be for a discrete time analog of the continuous time model we are interested in. Nonetheless, evidence that goes back at least as far as Merton (1980) shows that estimates of parameters based on discretization of continuous time models can be biased. Further, Melino (1994) offers a broad overview of the possible problems that may arise when estimating discretized versions of continuous-time models. However, through simulation exercises, Eraker et al. (2003) show that, for the range of parameters we are interested in, these biases become negligible when the discretization is at the daily level.

The structure of the Black-Scholes model or Merton's Jump-Diffusion lead to moment conditions and likelihood functions which are straightforward to construct, thus allowing for GMM estimation as in Fisher et al. (1996). The estimation of stochastic volatility models is more challenging, however, since the volatility series cannot be observed. Thus, several authors have estimated the unobserved parameters by using data contained in the prices of options. For instance, Ledoit et al. (2002) use the Black-Scholes implied volatilities of at-the money short-maturity options as a proxy for volatility. A more elaborate method presented in Pan (2002) calculates the implicit volatility from option prices, but using the option pricing model that corresponds to the asset dynamics she is assuming instead of the "naive" Black-Scholes implied volatilities. This then allows her to construct moment conditions, and form a variation of GMM, known as implied state GMM.

Another problem of having only discretely-observed data for a continuous-time model, is that the likelihood function of the observations is usually not explicitly computable. However, Aït-Sahalia (2002) proposes a maximum-likelihood method that approximates the likelihood function of the observations. To do so, he uses

Hermite polynomials to construct a sequence of closed-form functions that converge to the unknown likelihood function. This method was generalized to stochastic volatility models in Ait-Sahalia and Kimmel (2007). However, a clear drawback of this generalization is that it requires the prices of options or some proxy for the unobserved volatility.

Since the purpose of this paper is to show that the prices published by MexDer come from a misspecified model, we can not use the time series of option prices as part of the data used for inference. Thus, the methods described so far will not be useful in our context. Instead, we must explore the options available for estimation when the volatility series is not observable and option data is not available either.

These methods fall broadly into two categories. The first category comprises variations of the simulated method of moments proposed by Pakes (1989) and McFadden (1989). In particular, Gallant and Tauchen (1996) constructed a variation known as the efficient method of moments that has the significant advantage of being as efficient as maximum likelihood under certain conditions. This method has been successfully used by Andersen et al. (1997) and Andersen et al. (1999) to estimate a wide variety of models for the dynamics of stock indices.

The second approach, which we follow, uses Bayesian MCMC estimators. Their use dates back to Jacquier et al. (1994) but also includes Kim et al. (1999), Eraker (2001) and Eraker et al. (2003). A fundamental advantage of these methods is that they can produce estimates of the unobserved state variables and not only of the parameters. This is crucial for our purposes, since we are interested in using the results of the estimation to price options, and the (unobserved) volatility is an input required by the option-pricing formulas.

The rest of this section will be devoted to describing the Bayesian technique used for the estimation of the models of interest. This technique is a Markov Chain Monte Carlo procedure based on Eraker et al. (2003).

### **1.3.1 Estimated Model and Priors**

We begin by constructing the Euler discretization of model (1.3) and we take it as a true model in the sense that the discretized model will have the same likelihood

function as the continuous time model sampled at discrete time intervals. Thus,

$$\begin{aligned} Y_t - Y_{t-1} &= \mu + \sqrt{V_{t-1}}\varepsilon_t^y + \xi_t^y J_t^y \\ V_t - V_{t-1} &= \kappa(\theta - V_{t-1}) + \sigma_v \sqrt{V_{t-1}}\varepsilon_t^v + \xi_t^v J_t^v, \end{aligned} \tag{1.4}$$

where  $J_t^y, J_t^v = 1$  indicates a jump arrival and  $\varepsilon_t^y, \varepsilon_t^v$  are standard normal random variables with correlation  $\rho$ .

Now, let  $\Theta$  denote the collection of parameters we wish to estimate so,

$$\Theta \subseteq \{\mu, \kappa, \sigma_v, \theta, \rho, \mu_y, \sigma_y, \rho_J, \mu_v, \lambda_y, \lambda_v\}.$$

In the context of Bayesian estimation, in addition to the model specification described by model (1.4) we will also need a set of priors for the parameters. Priors are usually thought of as containing some information about the parameters that is not contained in the data, but a statistician believes should be included in the estimation. However, for most of the parameters we wish to impose no such prior information. Hence, "uninformative" priors are used.

So, let  $p(\Theta)$  denote the prior distributions for the parameters. These prior distributions will be

$$\begin{aligned} \mu &\sim N(1, 25), & \mu_y &\sim N(0, 100), \\ \kappa &\sim N(0, 1), & \sigma_y^2 &\sim IG(5, 20), \\ \kappa\theta &\sim N(0, 1), & \mu_v &\sim \Gamma(20, 10), \\ \sigma_v^2 &\sim IG(2.5, 0.1), & \rho_J &\sim N(0, 4), \\ \rho &\sim U(-1, 1), & \lambda_v = \lambda_y &\sim \beta(2, 40). \end{aligned}$$

Notice, for instance, that the prior for  $\mu$  has a standard deviation of 5. Later, we will see that our data has an average daily return of around 0.1%. Therefore, by imposing a prior with a standard deviation several orders of magnitude larger than the initial guess for the parameter, we are essentially imposing no prior information. Analogous arguments can be constructed for most of the parameters except for the case of  $\lambda$  and  $\sigma_y^2$ . These priors restrict the models with jumps so that jumps are infrequent and of magnitudes larger than what could be explained by the stochastic

element of the diffusions. We could think that large returns observed infrequently could be part of a jump process with a much higher rate of occurrence and the possibility of either large or small jump magnitudes. However, by introducing frequent jumps the estimation procedure might have problems distinguishing between variation attributable to the diffusive element of the process and the jump element. These difficulties are described in Honore (1998) and in Aït-Sahalia (2009) and the references therein. Thus, it is reasonable to impose an informative prior for these two parameters.

Finally, the choice of the parametric forms of the priors is standard in the literature. Later we will see that the reason that these distributions have become standard is that the priors and their posteriors are conjugate distributions.

### 1.3.2 Estimation Algorithm

Let  $\Omega$  be the set of parameters together with the latent variables. Thus,  $\Omega$  includes the all the values we wish to estimate. Further, if we denote  $V, J, \xi^y$  and  $\xi^v$  as the entire vectors of  $V_t, J_t, \xi_t^v$  and  $\xi_t^y$  sampled at the same discrete intervals as  $Y_t$ , then

$$\Omega = \{\Theta, J, \xi^y, \xi^v, V\}.$$

Thus, the objective of our Bayesian estimation is to find the posterior distribution of  $\Omega$  given the observed log prices  $Y$ . A simple application of Bayes' theorem tells us that

$$p(\Omega|Y) \propto p(Y|\Omega) p(\Omega).$$

However,  $p(\Omega|Y)$  is not a standard distribution so drawing from it is not a trivial task. Therefore, to find  $p(\Omega|Y)$  we use a hybrid - MCMC algorithm from Eraker et al. (2003).

### MCMC Procedure

The general idea of MCMC procedures is to construct a Markov Chain that has as its stationary distribution the distribution we are interested in i.e.  $p(\Omega|Y)$ . We begin by dividing  $\Omega$  into subsets (which will be discussed later). Each element of  $\Omega$



is then assigned a starting value  $\Omega^0 = (\omega_1^0, \dots, \omega_r^0)$ . Then, the  $c$ 'th iteration of the algorithm consists of obtaining a draw for each  $\omega_i \in \Omega$ . Specifically, we draw the  $i$ 'th component  $\omega_i^c$  from the distribution of  $\omega_i^c$  conditional on the data and the most current values of all the other components of  $\Omega$ , i.e.

$$\omega_i^c \sim p\left(\omega_i \mid \tilde{\Omega}_{-i}^c, Y\right),$$

where

$$\tilde{\Omega}_i^c = (\omega_1^c, \omega_2^c, \dots, \omega_{i-1}^c, \omega_{i+1}^{c-1}, \dots, \omega_r^{c-1}).$$

Hence, in each step we are drawing on the distribution of  $\omega_i$  conditional on the data, and on the rest of parameters taking the values of the most recent draws we have made of those parameters. Note that here  $p\left(\omega_i \mid \tilde{\Omega}_{-i}^c, Y\right)$  is the true conditional distribution.

These steps are then repeated a large number of iterations until we believe that the chain has converged. The output of the algorithm can then be taken as a sample drawn from  $p(\Omega|Y)$ . A detailed description of these methods as well as proofs of the convergence of the algorithm can be found in Johanes and Polson (2009).

The importance of the priors that we chose for the parameters becomes clear once we note that for each of the  $\omega_i \in \Omega$  we will have<sup>1</sup>

$$\begin{aligned} \mu \mid \Omega_{-\mu} &\sim N(\mu_{\mu}, \sigma_{\mu}^2), & \rho_J \mid \Omega_{-\rho_J} &\sim N(\mu_{-\rho_J}, \sigma_{-\rho_J}^2), \\ (\kappa, \kappa\theta) \mid \Omega_{-(\kappa, \kappa\theta)} &\sim N(M, \Sigma), & \lambda_v \mid \Omega_{-\lambda_v} &\sim \beta(\alpha_{\lambda_v}, \beta_{\lambda_v}), \\ \sigma_v^2 \mid \Omega_{-\sigma_v^2} &\sim IG(\alpha_v, \beta_v), & \lambda_y \mid \Omega_{-\lambda_y} &\sim \beta(\alpha_{\lambda_y}, \beta_{\lambda_y}), \\ \mu_y \mid \Omega_{-\mu_y} &\sim N(\mu_{\mu_y}, \sigma_{\mu_y}^2), & \xi_t^v \mid \Omega_{-\xi_t^v} &\sim \mathbf{1}_{\xi_t^v > 0} N(\mu_{\xi_t^v}, \sigma_{\xi_t^v}^2), \\ \sigma_y^2 \mid \Omega_{-\sigma_y^2} &\sim IG(\alpha_y, \beta_y), & \xi_t^y \mid \Omega_{-\xi_t^y} &\sim N(\mu_{\xi_t^y}, \sigma_{\xi_t^y}^2), \\ \mu_v \mid \Omega_{-\mu_v} &\sim \exp(\lambda_{\mu_v}), & J_t \mid \Omega_{-J_t} &\sim \text{Bernoulli}(p_{J_t}). \end{aligned}$$

Thus, we can now see clearly the advantage of the prior distributions that were chosen: they have conjugate posteriors that are easy to draw from. Notice however,

---

<sup>1</sup>A detailed derivation of the posteriors and the procedure used to draw from  $\rho$  and  $V_t$  can be found in Appendix A.

that we are missing the posteriors for  $\rho$  and  $V_t$ .

These elements of  $\Omega$  unfortunately do not have posteriors that follow a distribution that is known in closed form. Thus, in order to draw from them, accept-reject methods must be used. In the case of  $\rho$  we use an independence Metropolis-Hastings algorithm to draw from  $\rho|\Omega_{-\rho}$ , and in the case of  $V_t$  a random walk Metropolis-Hastings algorithm is used. These Metropolis-Hastings algorithms belong to the family of accept-reject methods used to draw from non-standard distributions. A candidate  $\omega_i^c$  is drawn from a known distribution and it is accepted or rejected with a certain probability that is a function of the ratio of its likelihood and the likelihood evaluated at  $\omega_i^{c-1}$ . A disadvantage of these methods is that if each new draw takes steps that are too far from the previous iteration, the algorithm will tend to reject new draws very often and thus convergence will be very slow. On the other hand, if steps are too small, the algorithm will not reject often, but convergence will still be slow, since it will take the program a long time to explore all the potential values where the posterior distribution of  $\omega_i$  has positive mass. Therefore, when using these Metropolis-Hastings algorithms one must calibrate them, and the rule of thumb suggested in Johanes and Polson (2009) is that 30%-60% of the candidates should be rejected.

Putting all the steps together we have the following:

**Algorithm 1** *Estimation of stochastic volatility and jumps model.*

1. Set starting values for  $\Omega$
2. For  $c = 1 \dots C$ 
  - For  $i = 1 \dots \#\Theta$ 
    - Draw parameter  $i$  from:  $p(\Theta_i | \Theta_{-i}, J, \xi_t^y, \xi_t^v, V, Y)$
  - Draw jump sizes:
    - For  $t = 1 : T$  draw from:  $p(\xi_t^y | \Theta, J_t = 1, \xi_t^v, V, Y)$
    - For  $t = 1 : T$  draw from:  $p(\xi_t^v | \Theta, J_t = 1, V, Y)$
  - Draw jump times:

- For  $t = 1 : T$  draw from:  $p(J_t = 1 | \Theta, \xi^y, \xi^v, V, Y)$
- Draw volatility:
  - For  $t = 1 : T$  draw from:  $p(V_t | \Theta_i, J, \xi_t^y, \xi_t^v, V_t, V_{t-1}, Y)$

3. Discard the first  $\tilde{C}$  iterations.

By writing the algorithm in this expanded form, we can clearly appreciate that this method is very computationally intensive. Drawing the parameters is not the issue, since the number of parameters is less than 11 for all models. However, for each iteration of the algorithm at least  $T$  latent variables have to be drawn, and for the models with jumps in both equations, up to  $5T$  state variables have to be simulated for each iteration. Now,  $T$  is frequently in the thousands, and convergence criteria are usually not met unless  $C$  is in the tens of thousands. Hence, it becomes clear that computational cost will be one potential drawback of this method. Nonetheless, this procedure remains the best estimating option available, in spite of its computational cost, precisely because we need the draws of the state variables for our option pricing application and for the Bayesian model selection procedure we describe below.

Finally, Johannes and Polson (2009) show that this algorithm will in fact produce a Markov chain whose stationary distribution will be the posterior distribution  $p(\Omega | Y)$  that is of interest here. Precisely because we are interested in the stationary distribution, we discard the first  $\tilde{C}$  iterations and take the draws of  $\Omega^c$  from  $\tilde{C}$  to  $C$  as a sample from the stationary distribution.

Still, we need some criteria to know that the algorithm has achieved convergence. Although there are no set rules on how to guarantee convergence, there are still several things we can check. First, we can use very different starting points and check that the estimators that the algorithm produces are sufficiently close regardless of the starting point. Second, assuming we believe that  $C$  iterations are enough, we can run the procedure for a considerably longer time (up to an order of magnitude larger) and compare the distribution of the last  $C$  draws of the chain with the  $C$  draws where we thought that convergence had occurred.

## 1.4 Data

The IPC is the main index of the Mexican Stock Exchange. The exchange describes it as a "representative indicator of the Mexican stock market" (Bolsa Mexicana de Valores (2011)). It includes 35 stocks chosen for their market capitalization and liquidity. Nonetheless, a handful of stocks dominates the index; in 2010-2011, the two most heavily-weighted stocks comprised over 36% of the index. These were America Móvil (Telecommunications) having a weight of 25% and Walmart-Mexico (Retail) having a weight of 11.5%. However, having single stocks with very large weights is not unusual for stock indices; a common example being the Nasdaq 100, where Apple Inc. had a 20% weight on the index during 2010. The rest of the IPC is well diversified sectorwise, including firms as diverse as Peñoles (the world's largest silver producer) and FEMSA (a beverage company producer of beer and soft drinks).

Moreover, a crucial characteristic of the IPC is that it is used as a benchmark by banks and investment firms in Mexico, and it is one of the most closely followed economic indicators in the country. Thus, when the Mexican stock exchange introduced options in 2004, European-style calls and puts on the IPC were some of the first contracts to be issued.

The data used for this paper consists of 6189 daily prices of the index ranging from 1987 to 2011. Although the index existed before 1987, the methodology used for the construction of the index was very different, so only data from 1987 onward was used. Still, the period considered comprises a wide range of economic conditions. Near the beginning of the series, in October 1987, one of the most volatile periods in world financial history occurred. For the Mexican case this was compounded by internal problems that led to an annual inflation of over 100% for that year. Next, in 1994 Mexico experienced one of the most politically turbulent years in recent history, which was the prelude to the 1994-95 financial crisis. Later, we will pay special attention to this period and what some of the models can tell us about the behavior of the index under these extreme circumstances.

Other periods of worldwide financial instability are also reflected on the IPC.

These include the Asian and Russian crises of 1997-98, the burst of the .com bubble in the early 2000s and the global financial crisis beginning in 2008. Finally, between these highly volatile times, there are periods of low volatility, and thus the returns exhibit the usual volatility-clustering that is common for stock indices.

### 1.4.1 Descriptive Statistics

Let us begin the analysis of the data by presenting simple descriptive statistics. Table 1.1 shows that mean returns for the index were almost 27% per year with an average annual volatility of 29%, both numbers higher than US stock indices for this period. More importantly, the series exhibits negative skewness and kurtosis that considerably exceeds 3. This is a common finding for financial returns, indicating that log returns are not normally distributed and thus not consistent with the Black-Scholes model.

<b>Table 1.1</b> Summary Statistics for Daily Returns	
Mean (Annualized)	0.106% (26.82%)
Volatility (Annualized)	1.84% (29.32%)
Skewness	-0.508
Kurtosis	21.13
Min.	-20.24%
Max.	23.58%

Figure 1-1 shows the time series for the returns. We can see that the data exhibits periods of very high volatility, particularly in 1987, but also in 1994-95, 1997-98 and 2008-09. This presence of volatility-clustering is yet another suggestion that the Black-Scholes model will not be able to describe the dynamics of the index and that stochastic volatility models may be appropriate. However, one can also observe days of extreme returns with a magnitude that exceeds 10 standard deviations above or below the mean return. This is an indication that a model without jumps will not be able to fully capture the dynamics of the index. Further, and especially in 1987, we observe periods when there are returns with very large magnitudes during consecutive days, suggesting that jumps in returns might not be sufficient, and a correctly specified model for this data will also have to include jumps in the volatility series.

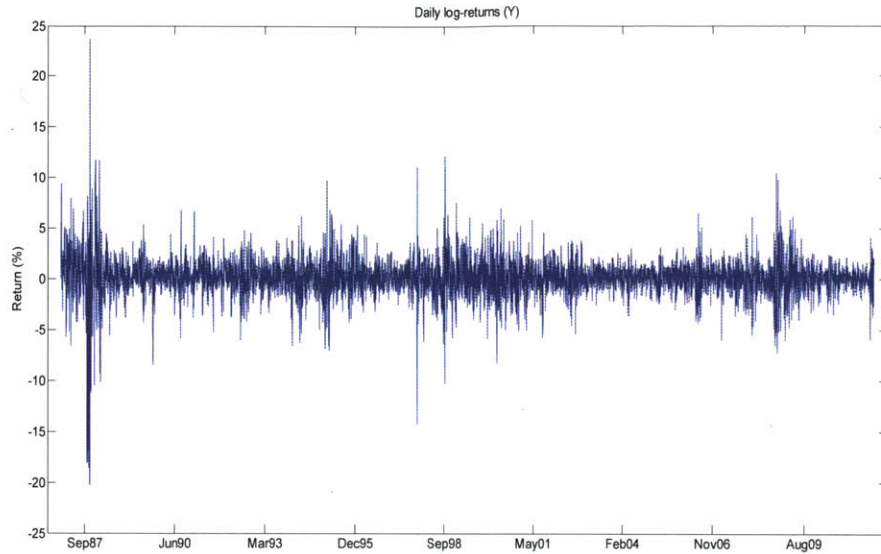


Figure 1-1: Daily log-returns of the IPC index 1987-2011.

## 1.5 Estimation Results and Model Selection

### 1.5.1 Estimation Results

The parameters are estimated by setting  $C = 150,000$  and discarding the first 100,000 iterations. Table 1.2 shows the median of the posterior distributions for each of the four estimated models.

The parameters that have an immediate interpretation are  $\mu$  and  $\theta$ . The estimates for  $\mu$  indicate a deterministic drift that remains close to 30% annually regardless of the model. Meanwhile the estimations for  $\theta$  imply a long term annual volatility of 27.72% for the SV model but only 26.48% in the SVJ model. These reductions in volatility are to be expected: under the SV specification all the variation in the return series is driven by the volatility of its diffusive element. However, in the SVJ model, part of this variation is now explained by jumps in the return series. This can also be seen in the estimates of  $V$ . Figure 1-2 shows the SV, SVJ and SVIJ estimates for  $V$  for two of the periods of particularly high volatility. Once more, we see that the volatility series is lower when part of the variation of the returns can be attributed to jumps. Yet, it is also important to note that volatility is higher when

**Table 1.2** Medians of the Posterior Distributions

Standard deviation of the posterior distributions in brackets

	SV	SVJ	SVCJ	SVIJ
$\mu$	0.1170(0.017)	0.1174(0.019)	0.1178(0.018)	0.1188(0.018)
$\sigma_v$	0.0430(0.001)	0.0417(0.001)	0.0423(0.001)	0.0425(0.001)
$\rho$	-0.0274(0.015)	-0.02510(0.015)	-0.02894(0.016)	-0.0283(0.018)
$\kappa$	0.0444(0.005)	0.0462(0.006)	0.04488(0.005)	0.0453(0.005)
$\theta$	3.051(0.230)	2.782(0.212)	2.958(0.222)	2.962(0.225)
$\sigma_y$		0.364 (0.12)	0.446 (0.24)	0.453 (0.27)
$\mu_y$		-0.124 (0.83)	-2.592 (1.00)	-3.420 (3.82)
$\lambda_y$		0.017 (0.006)	0.003 (0.001)	0.002 (0.001)
$\mu_v$			4.54 (16.61)	20.81 (9.85)
$\lambda_v$				0.007 (0.004)
$\rho_J$			-0.423 (1.14)	

jumps in the volatility series are considered. The explanation for this is that under the SVJ specification very large returns will be attributed to jumps in the return series, without volatility having to rise much. However, once we allow for jumps in the volatility series, a large return can be the consequence of a sudden increase in volatility and not necessarily a jump in the return series.

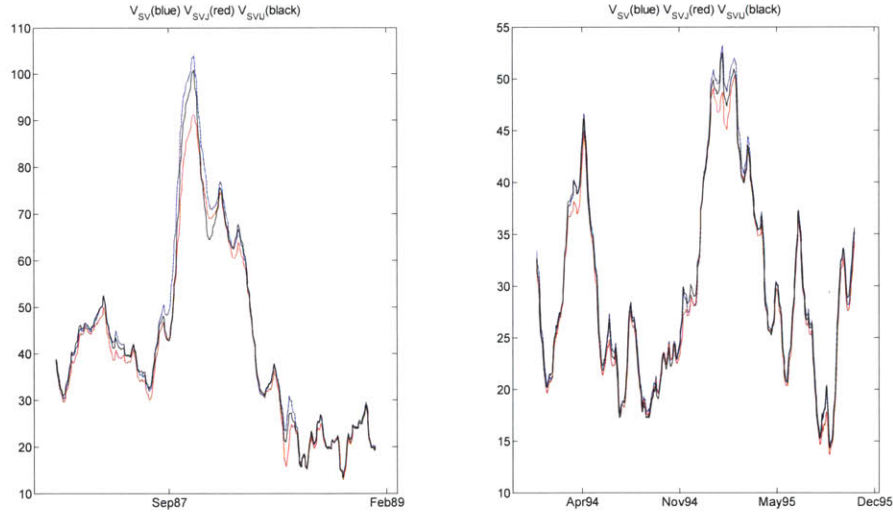


Figure 1-2: Estimated volatility for two sub-periods under the SV, SVJ and SVIJ specifications.

Finally, the estimates of  $\lambda = (\lambda_y, \lambda_v)$  also have an immediate interpretation. The frequency of the jumps, the estimations of  $\lambda_y$  imply 4.25 jumps in the return

series are expected every year under the SVJ specification, but only 2 jumps every 3 years are expected under the SVCJ and SVIJ specifications. Once again, this result is what we would expect: by introducing jumps in volatility, large returns can be attributed to a sudden increase in  $V$ , without the need for jumps in the return series.

## 1.5.2 Implied Jump Process

### Results for the SVJ Model

In Figure 1-3 we can see that the days with high probability of a jump are not necessarily the days with the largest returns, but days with large returns and relatively low levels of volatility. Thus, for high volatility periods, the model still assigns an important part of the variation in the returns to the diffusive volatility. In fact, in Figure 1-4 we can observe that on dates when a high probability of a jump is assigned, practically all the magnitude of the return can be attributed to the jump.

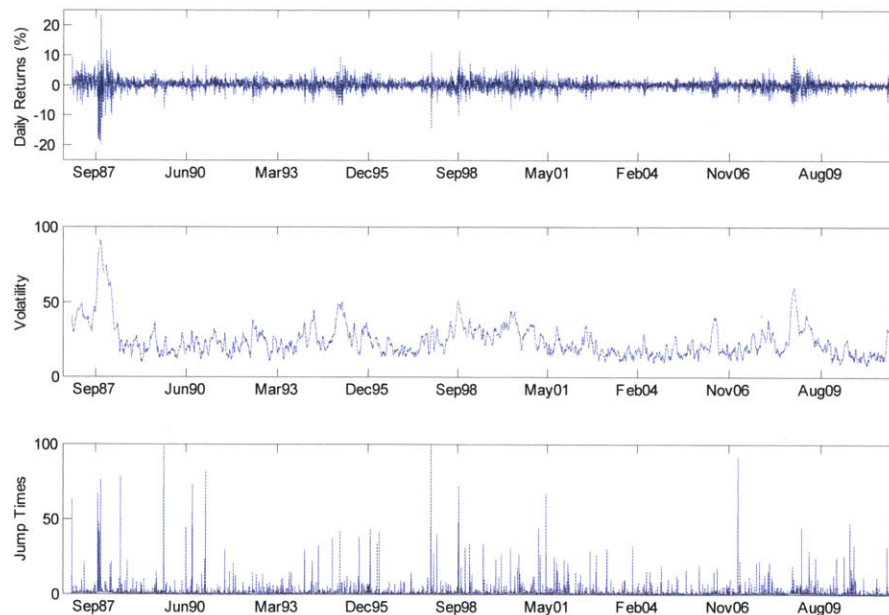


Figure 1-3: Volatility estimates and probability of jump times under the SVJ specification.



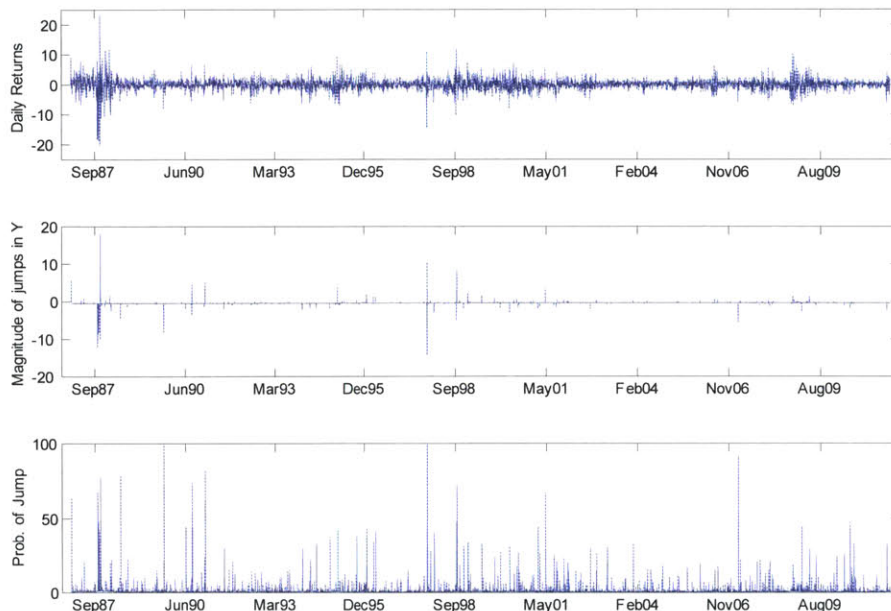


Figure 1-4: Magnitude and probability of jumps under the SVJ specification.

Additionally, the fact that the SVJ model very clearly identifies dates with high probabilities of jumps allows us to look at one of the most interesting periods in Mexican economic history and link historic events with predictions of jumps (or lack thereof). In Figure 1-5 we can see the results of the SVJ model for the 1994-1995 period. 1994 was one of the most politically unstable years in recent Mexican history. On New Year's Day 1994 Zapatista guerillas appeared in the southeastern state of Chiapas. This represented the first bout of civil insurgency in two decades in Mexico. Not surprisingly, our estimates assign a large probability of a jump having occurred on January 2nd (the first day of trading that year). The next date with a large probability of a jump is immediately after March 23rd. On this date the presidential candidate of the official party was assassinated. Given the political structure at the time, the candidate of the official party was practically assured to be the next president. Therefore, his assassination represented a particularly important shock to the confidence in Mexican financial markets. The next instances of high jump probabilities correspond to June 11th when the Zapatista guerrillas

rejected the terms of a proposed peace treaty and later on November 1994. The latter might be thought of as the beginning of the chain of events that triggered the 1995 financial crisis (see e.g. Aspe (1995)). On November 18th the central bank experienced one of the largest declines in foreign reserves on record. Confidence had been deteriorating throughout the year, but on November 15th a high ranking government official accused the official party of orchestrating a political assassination earlier that year. Given the political structure of Mexico during the 1990s, instability in the ruling party was perceived to be equivalent to instability in the government. Further, historically, the party had been characterized by ferocious loyalty of party members (see e.g. Preston and Dillon (2004)), so public accusations of serious misconduct coming from a high ranking party member were interpreted as a sign of deep instability within the party. Given the other problems discussed above, and that rumors of a potential sovereign default were beginning to circulate, it is no surprise that the model identifies a jump on November 18th.

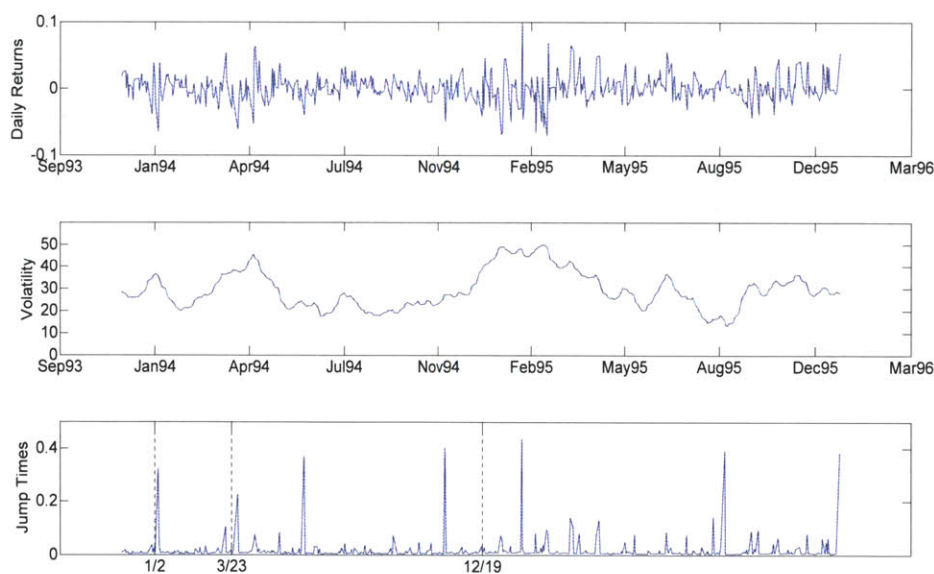


Figure 1-5: Volatility estimates and probability of jump times under the SVJ specification for the 1994-95 period.

Finally, the last date marked explicitly in Figure 1-5 is December 19, 1994. On this date the Mexican president announced rather abruptly that the peso would

be devalued shortly. This episode has become known colloquially as the "error de diciembre" or December's mistake. It is somewhat surprising that the model does not detect a jump on this date or the dates immediately afterward. Nonetheless, if we look at the return series we see that there were in fact very large returns around that date. However, after inspecting the volatility estimates, we see that volatility had been increasing from November onward. Thus, by the time the announcement was made, volatility was so high that jumps are not needed to explain the very large returns observed in late December and early 1995.

During early 1995, while the crisis was unfolding, we observe that volatility tended to be high and thus no jumps are needed to explain large swings in prices. Nonetheless, volatility decreased through the first half of the year and once it reached its lowest point of the year in August, a jump is once more detected. At this point volatility begins to increase once more, explaining the large returns during late 1995, and thus no further jumps are detected until 1996, which coincides with the end of the sub-period of interest.

### **Results for Models with Jumps in Both Returns and Volatility**

In Figure 1-6 we can observe graphically how under the SVCJ a much lower frequency of jumps is estimated (when compared to the result under the SVJ model). However, we still find that jumps are particularly important in explaining large returns during periods of low volatility. Also, it is important to note that, as we would expect, periods of high volatility tend to be preceded by large jumps in the volatility series.

In Figure 1-7 we can see that under the SVIJ specification the probability of jumps in the return series remains very high for very specific days. In contrast, the probability of jumps remains almost uniformly low for the volatility series. Nonetheless, this does not imply that there is little information in this series of probabilities. Since what is plotted in the last graph of Figure 1-7 is the probability of a jump occurring on each particular day, having a period of time with above average probabilities of jumps for each of those days tells us that there is a high probability of a jump having occurred in that period. So, instead of identifying precise days when

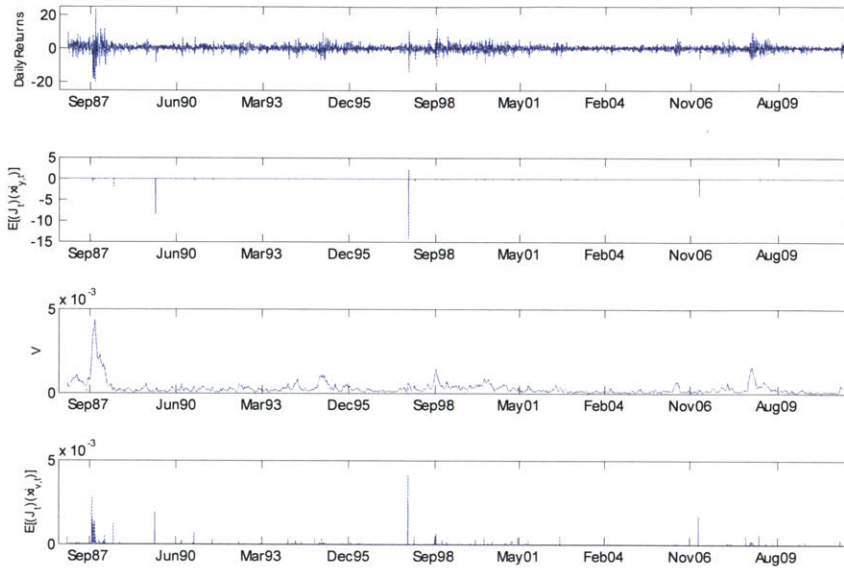


Figure 1-6: Returns, volatility and expected magnitude of jumps in both series under the SVCJ specification.

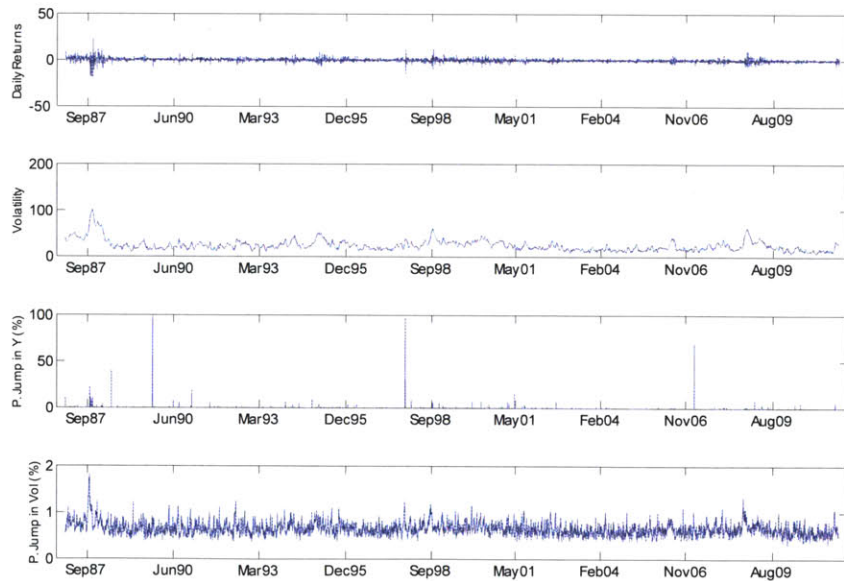


Figure 1-7: Returns, volatility, probability of jumps in returns and probability of jumps in volatility under the SVIJ specification.

a jump occurred (as we can with the return series) for the volatility series we can only identify an interval of days where it seems very likely that there was a jump. Further, Figure 1-8 shows the product of the probability of a jump with its expected magnitude. This gives us a sense of the expected jump on that day. Therefore, by looking at a given period we can get a measurement of the expected accumulated contributions of jumps to the volatility.

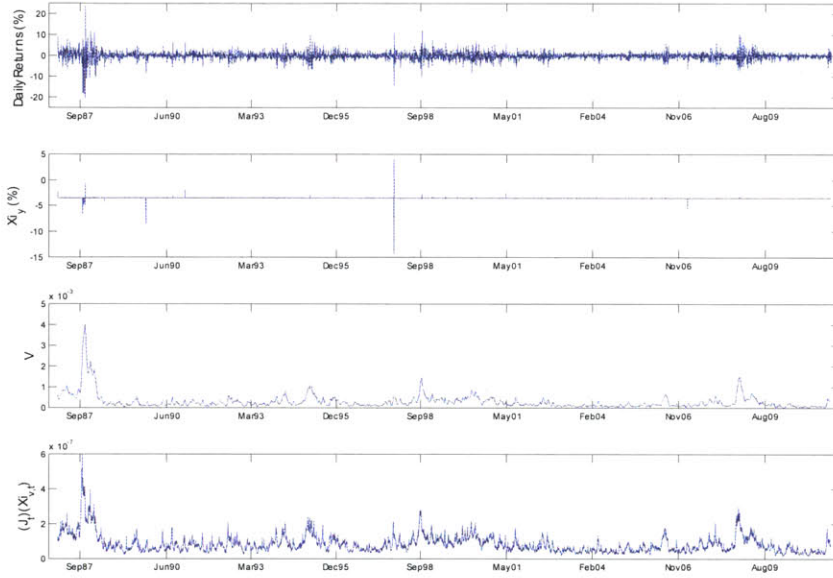


Figure 1-8: Returns, magnitude of jumps in returns, volatility and expected size of jumps in volatility under the SVIJ specification.

### 1.5.3 Model Selection: Test of Residuals

Through the MCMC procedure, we completely characterize the posterior distributions of the parameters. However, knowing the probability of the parameters lying in a certain interval will not be sufficient to test the suitability of the models or their relative performance. Thus, in order to test how well specified the models are, the following residuals are calculated:

$$\hat{\varepsilon}_t = \frac{Y_t - Y_{t-1} - \hat{\mu} - \hat{J}_t^y \hat{\zeta}_t^y}{\sqrt{\hat{V}_{t-1}}}$$

Note that, regardless of the model being considered, under the hypothesis that the model describes the dynamics of the returns, this residual should follow a standard normal distribution. Thus, a first approach to testing the fit of the different models is to construct a QQ plot for the residuals.

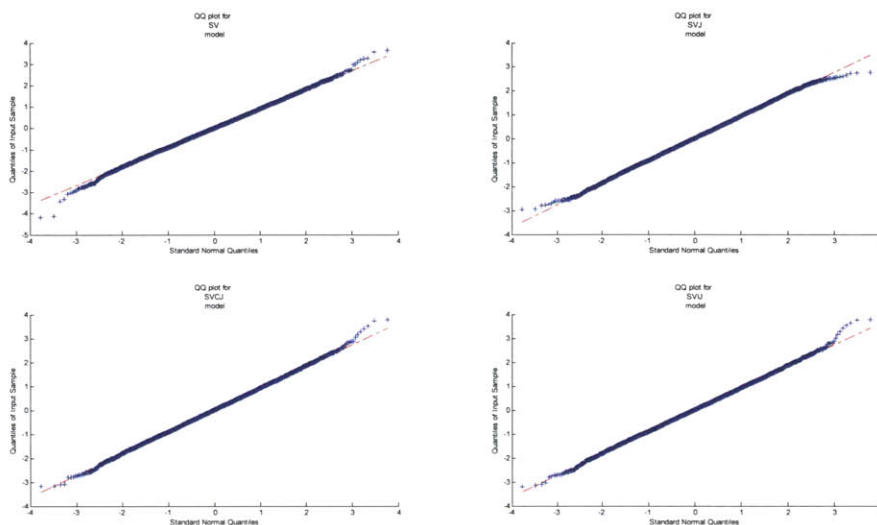


Figure 1-9: QQ plots for the residuals under the four models.

The QQ plots in Figure 1-9 show us how the distribution of the residuals is closer to normality for the SVJ than for the SV model and even closer for the models with jumps in both series. Moreover, beyond the graphical data represented by the QQ plots, we can perform a formal statistical test of the normality of  $\hat{\varepsilon}_t$  for each of the models. Table 1.3 shows the  $p$ -values for a Jarque-Bera normality test of the residuals.

<b>Table 1.3: Jarque-Bera Test</b>				
Model	SV	SVJ	SVCJ	SVIJ
$p$ -value	<0.001	0.0199	0.0763	0.0680

For the SV case the normality of the residuals can be rejected at the 99.9% level, confirming that in fact this model provides a very poor fit for the data. At the 95% level we can also reject the SVJ model, but not of the models with jumps in volatility. Thus, models with jumps in both series represent a considerable improvement relative to the SV model. This confirms the initial conjecture: that models with

jumps are better specified for the IPC than Heston's model.

#### 1.5.4 Model Selection: Bayes Factors

So far we have seen that the models with jumps offer a better fit than the standard Heston (SV) model. However, when assessing the merits of the different models, we still must take into account the fact that the SV model is the most tightly parametrized within the set. This should lead us to ask whether the improvements in fit outweigh the loss parsimony.

Fortunately, our estimation procedure produces estimates for all the latent variables. This, together with the fact that the SV model can be expressed as a restricted version of any of the other models, allows us to compare the performance of the models using Bayes factors.

In general, for two models  $M_1$  and  $M_2$  and data  $Y$ , the posterior odds

$$B(M_1, M_2) = \frac{p(M_1|Y)}{p(M_2|Y)}$$

represent the relative likelihood of the two models conditional on the data. Further, if we assume prior ignorance by setting  $p(M_1) = p(M_2)$ , a simple application of Bayes' theorem yields

$$B(M_1, M_2) = \frac{p(Y|M_1) p(M_1)}{p(Y|M_2) p(M_2)} = \frac{p(Y|M_1)}{p(Y|M_2)},$$

which is simply the ratio of the likelihoods of the data under the two model specifications. So, assuming we are comparing the SV model to any of the alternatives (SVX),  $B(SV, SVX)$  can be expressed as

$$\frac{p(Y|SV)}{p(Y|SVX)} = \frac{\int p(Y|\Omega, SV) p(\Omega|SV) d\Omega}{\int p(Y|\Omega, SVX) p(\Omega|SVX) d\Omega}.$$

Further, SV is equivalent to restricting the jump vector (or vectors) to zero in any of the other models, so that  $p(Y|\Omega, SV) = p(Y|\Omega, SVX, J = 0)$ . Thus, a

further application of Bayes' formula tells us that

$$\frac{p(J = 0|Y, SVX)}{p(J = 0|SVX)} = \frac{p(Y|J = 0, SVX)}{p(Y|SVX)} = \frac{p(Y|SV)}{p(Y|SVX)},$$

where the denominator can be computed directly from the prior distribution of  $\lambda$  (the rate of the jump process) and the numerator will be

$$P(J = 0|Y, SVJ) = \int_0^1 p(J_v = 0|\lambda, Y, SVX) p(\lambda|Y, SVX) d\lambda. \quad (1.5)$$

Now, since the Markov chain constructed for estimation completely characterizes  $p(\lambda|Y, SVX)$ , we can use the draws from this chain to perform a Monte Carlo integration of equation (1.5).

Following Smith and Spiegelhalter (1980), in Table 1.4 we report  $2 \ln(B)$  and the interpretation suggested in that same reference.

Models	$2 \ln(B)$	Interpretation
SV vs. SVJ	-7.5	Positive against SVJ
SV vs. SVCJ	9.1794	Positive for SVCJ
SV vs. SVIJ	15.164	Positive for SVIJ
SVCJ vs. SVIJ	5.9846	Positive for SVIJ

The negative value in SV vs. SVJ indicates that under this criterion the improvements in fit from SV to SVJ do not compensate for the less parsimonious model. However, the positive values for the rest of the factors indicate that there is otherwise positive evidence in favor of the more complex models.

## 1.6 Option Pricing Application

The starting point of this paper was the fact that MexDer uses Heston's model to price options on the days when there has been no trading. Now, under the naming convention used in the previous sections, Heston's model corresponds to the SV model, and so far I have shown that it is a poor choice for modeling the dynamics of the IPC and that models with jumps offer a much better description of the asset dynamics. However, we still have to check whether the option prices implicit in the



more elaborate models will significantly differ from those estimated using Heston's model.

Let us remember that one of the characteristics which make Heston's model is so attractive is that it has a closed form formula for the valuation of European-style options. In contrast, with more elaborate models we cannot always be sure that such valuation formulas exist. Fortunately, the four models estimated and tested above belong to the larger class of affine-jump diffusions. Duffie et al. (2000) show that the pricing formula for European options on assets that follow an affine-jump diffusion has a closed form (up to the solution of a system of complex-valued differential equations). Concretely, for an option expiration date  $T$ , at time  $t$  the conditional characteristic function is defined as:

$$CCF_t(z) = E \left[ e^{izY_T} | Y_t \right].$$

Using this definition, Duffie et al. (2000) show that if  $Y$  follows an affine process then  $CCF_t(z)$  will be of the form

$$CCF_t(z) = e^{\phi_{0t} + \phi_{Yt} Y_t}$$

where  $\phi_{0t}$  and  $\phi_{Yt}$  satisfy a pair of complex-valued ordinary differential equations. Since the  $CCF_t$  is fundamental for option pricing, in principle every time we wanted to price an option, one would have to solve  $(\phi_{0t}, \phi_{Yt})$  by some numerical method (e.g. as Runge-Kutta). However, for the models considered in this paper  $\phi_{0t}$  and  $\phi_{Yt}$  are known explicitly as a function of the parameters of the jump diffusions and the parameters of the options (see e.g. Sepp (2003)).

Further, Sepp (2003) also shows that the particular models estimated above yield a pricing formula that can be expressed as a weighted sum of the current price of the underlying asset and the present value of the strike of the option. Specifically, the price ( $F$ ) of a European option over an asset with current price  $S_t$ , and strike price  $K$  is given by

$$F = \varphi \left[ S_t P_1(\varphi) - e^{-r(T-t)} K P_2(\varphi) \right],$$

where  $\varphi = 1$  if the option is a call,  $\varphi = -1$  if the option is a put, and

$$P_j(\varphi) = \frac{1 - \varphi}{2} + \varphi \left( \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \mathcal{R} \left[ \frac{\phi_j(e^{iky})}{ik} \right] dk \right), \quad (1.6)$$

where  $\mathcal{R}$  denotes the real part of a complex number and  $(\phi_1, \phi_2)$  are functions of the  $CCF_t$  and parameters of the option being evaluated. When compared to equation (1.1) we see why Sepp (2003) calls this a Black-Scholes-style formula.

Using this pricing procedure and the parameters estimated above, prices are calculated for every call option issued between 2004 and 2009. Then the prices implied by Heston's model are compared to those implicit in the SVJ model.

Figures 1-10 and 1-11 show the percentage difference in price of call options under the SVJ specification relative to the prices under Heston's model. The graphs show the difference in price as it varies with the ratio of the strike price of the option to the current price (a measure of the "moneyness" of the options). In each case the set of options is separated into those that will expire in under 45 days, those that will expire in between 150 and 250 days, and those that will expire in more than 300 days. Figure 1-10 shows that for options that are out of the money, the difference in prices is moderate, although it increases as the expiration of the options becomes longer, and it also grows as the options come close to being in the money (as  $K/S_t$  approaches 1).

These differences in prices would be significant if one could actually trade in these options, but since the purpose of the prices is to mark to market them, the difference in prices is not that important.

However, in Figure 1-11 we see that for options that are in the money, the price differences are much larger and can substantially change the apparent value of the options that banks hold in their balance sheets. This clearly makes the choice of the model for the dynamics of the underlying assets a very important decision when imputing prices to options with no market price.

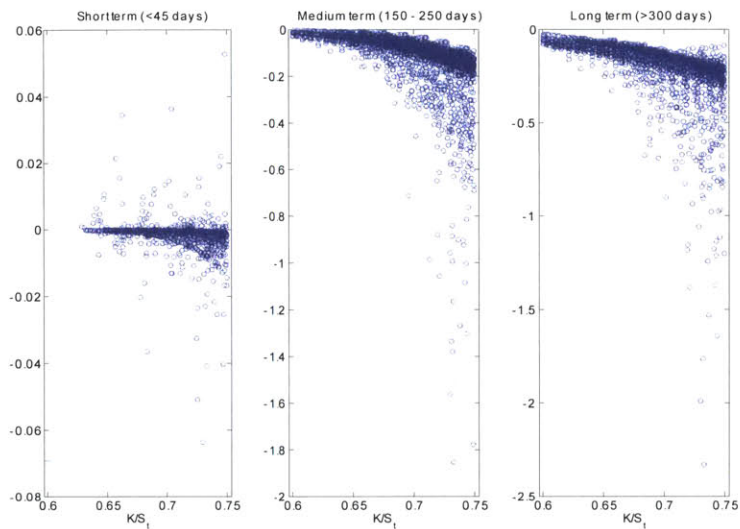


Figure 1-10: Percentage difference in price implied by the SVJ model relative to those implied by Heston's model for out of the money options.

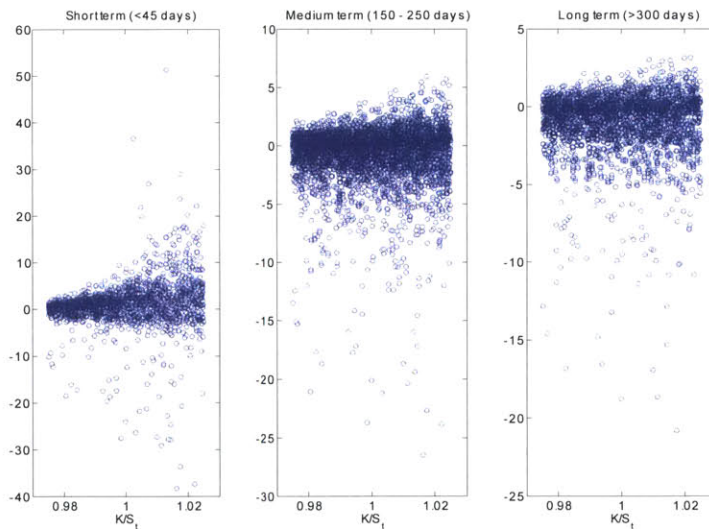


Figure 1-11: Percentage difference in price implied by the SVJ model relative to those implied by Heston's model for in the money options.

## 1.7 Conclusions and Direction of Future Research

I have shown that Heston's model is misspecified for the IPC and that models which incorporate jumps offer a better description of the dynamics of the index. Further, I showed that models with jumps in both returns and volatility improve the fit sufficiently to compensate for the loss in sparseness. Finally, transform pricing techniques allowed me to show that option prices will be significantly different under the assumption of jump-diffusions than under Heston's model.

The remaining issue is the suitability of the method in a real-life situation. The estimation procedure is very intensive computationally and may not be practical if the model needs to be estimated every day. Thus, implementing these results would require an estimation method that could incorporate new information, such as new prices at the end of each day, in a computationally-efficient manner. I conjecture that by combining the MCMC procedure presented here with Bayesian filtration methods such as those proposed in Johannes et al. (2009), a methodology for pricing the options can be constructed that is both based on a well-specified model for the asset dynamics and implementable in a real setting. This would then enable MexDer to provide more reliable prices on a daily basis.

## 1.8 Appendix A. Drawing from the Conditional Posteriors

Many of the results presented in this appendix can be found in standard textbooks, in Eraker et al. (2003) or in Johannes and Polson (2009). However, to ease replicability of my results, I present the derivation of all the posteriors.

We begin by re-stating the discretized model (1.4) as

$$\begin{aligned}\Delta Y_t &= \mu + \sqrt{V_{t-1}}\varepsilon_t^y + \xi_t^y J_t^y \\ \Delta V_t &= \alpha + \beta V_{t-1} + \sigma_v \sqrt{V_{t-1}}\varepsilon_t^v + \xi_t^y J_t^y\end{aligned}$$

where we have,  $\beta = -\kappa$  and  $\alpha = \kappa\theta$ ,  $\Delta Y_t = Y_t - Y_{t-1}$  and  $\Delta V_t = V_t - V_{t-1}$ . Then a distribution that will be useful throughout is

$$\begin{aligned}p(\Delta Y_t, \Delta V_t | V_{t-1}, \Theta) &= \\ &= \frac{1}{2\pi\sigma_v V_{t-1} \sqrt{1-\rho^2}} \exp \left( -\frac{1}{2(1-\rho^2)} \left[ \begin{array}{c} \frac{(\Delta Y_t - \mu - \xi_t^y J_t^y)^2}{V_{t-1}} \\ + \frac{(\Delta V_t - \alpha - \beta V_{t-1} - \xi_t^y J_t^y)^2}{\sigma_v^2 V_{t-1}} \\ - \frac{2\rho((\Delta Y_t - \mu - \xi_t^y J_t^y)(\Delta V_t - \alpha - \beta V_{t-1} - \xi_t^y J_t^y))}{\sigma_v V_{t-1}} \end{array} \right] \right)\end{aligned}$$

### Drawing from $\mu$

First, we have that the prior for  $\mu$  is

$$p(\mu) = \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp \left( -\frac{(\mu - \mu_\mu)^2}{2\sigma_\mu^2} \right)$$

and applying Bayes' rule,

$$p(\mu|Y, \Omega_{-\mu}) \propto$$

$$\begin{aligned}
& \propto p(Y|\mu, \Omega_{-\mu}) p(\mu) \\
& \propto \exp\left(-\frac{(\mu - \mu_\mu)^2}{\sigma_\mu^2}\right) \prod_{t=1}^T \exp\left(-\frac{(\Delta Y_t - \mu - \xi_t^y J_t^y)^2}{2V_{t-1}}\right) \\
& \propto \exp\left(-\frac{\mu_\mu^2 - 2\mu\mu_\mu}{2\sigma_\mu^2}\right) \exp\left(-\frac{1}{2}\left(\mu^2 \sum_{t=1}^T \frac{1}{V_{t-1}} - 2\mu \sum_{t=1}^T \frac{(\Delta Y_t - \xi_t^y J_t^y)}{V_{t-1}}\right)\right) \\
& \propto \exp\left(-\frac{\mu^2 - 2\mu\mu_\mu}{2\sigma_\mu^2}\right) \exp\left(-\frac{1}{2\left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1}}\left(\mu^2 - 2\mu \frac{\sum_{t=1}^T \frac{(\Delta Y_t - \xi_t^y J_t^y)}{V_{t-1}}}{\sum_{t=1}^T \frac{1}{V_{t-1}}}\right)\right) \\
& \propto \exp\left(-\frac{1}{2}\left(\mu^2 \left(\frac{1}{\sigma_\mu^2} + \frac{1}{\left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1}}\right) - 2\mu \left(\frac{\mu_\mu}{\sigma_\mu^2} + \sum_{t=1}^T \frac{(\Delta Y_t - \xi_t^y J_t^y)}{V_{t-1}}\right)\right)\right) \\
& \propto \exp\left(-\frac{1}{2\left(\frac{\sigma_\mu^2 \left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1}}{\left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1} + \sigma_\mu^2}\right)}\left(\mu^2 - 2\mu \frac{\left(\frac{\mu_\mu}{\sigma_\mu^2} + \sum_{t=1}^T \frac{(\Delta Y_t - \xi_t^y J_t^y)}{V_{t-1}}\right)}{\left(\frac{\left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1} + \sigma_\mu^2}{\sigma_\mu^2 \left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1}}\right)}\right)\right)
\end{aligned}$$

and thus,

$$\mu|Y, \Omega_{-\mu} \sim N\left(\frac{\left(\frac{\mu_\mu}{\sigma_\mu^2} + \sum_{t=1}^T \frac{(\Delta Y_t - \xi_t^y J_t^y)}{V_{t-1}}\right)}{\left(\frac{\left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1} + \sigma_\mu^2}{\sigma_\mu^2 \left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1}}\right)}, \frac{\sigma_\mu^2 \left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1}}{\left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1} + \sigma_\mu^2}\right)$$

**Drawing from  $\alpha, \beta$  and  $\sigma_v$**

Since

$$\begin{aligned}
\Delta V_t &= \alpha + \beta V_{t-1} + \sigma_v \sqrt{V_{t-1}} \varepsilon_t^v + \xi_t^y J_t^y, \\
\Delta V_t | V_{t-1}, \Omega &\sim N(\alpha + \beta V_{t-1} + \xi_t^y J_t^y, \sigma_v^2 V_{t-1}).
\end{aligned}$$

First, for the case of  $\sigma_v^2$  we have

$$\sigma_v^2 \sim IG(\check{\alpha}, \check{\beta}).$$

Which implies

$$p(\sigma_v^2) \propto (\sigma_v^2)^{-\check{\alpha}-1} \exp\left(-\frac{\check{\beta}}{\sigma_v^2}\right)$$

so,

$$\begin{aligned} p(\sigma_v^2|Y, \Omega_{-\sigma_v^2}) &\propto p(V|\sigma_v^2, \Omega_{-\sigma_v^2, V}) p(\sigma_v^2) \\ &\propto (\sigma_v^2)^{-\check{\alpha}-1} \exp\left(-\frac{\check{\beta}}{\sigma_v^2}\right) \prod_{t=1}^T \frac{1}{\sigma_v} \exp\left(-\frac{(\alpha + \beta V_{t-1} + \xi_t^y J_t^y - \Delta V_t)^2}{2\sigma_v^2 V_{t-1}}\right) \\ &\propto (\sigma_v^2)^{-\check{\alpha}-1-T/2} \exp\left(-\frac{\check{\beta} + \sum_{t=1}^T \frac{(\alpha + \beta V_{t-1} + \xi_t^y J_t^y - \Delta V_t)^2}{2V_{t-1}}}{\sigma_v^2}\right). \end{aligned}$$

Thus,

$$\sigma_v^2|Y, \Omega_{-\sigma_v^2} \sim IG\left(\check{\alpha} + T/2, \check{\beta} + \sum \frac{(\alpha + \beta V_{t-1} + \xi_t^y J_t^y - \Delta V_t)^2}{2V_{t-1}}\right).$$

For the case of  $\alpha$  and  $\beta$  the priors are

$$\begin{aligned} \alpha &\sim N(\mu_\alpha, \sigma_\alpha^2), \\ \beta &\sim N(\mu_\beta, \sigma_\beta^2). \end{aligned}$$

So we have

$$p(\alpha, \beta|\Omega_{-\alpha, \beta}) \propto p(V|\alpha, \beta, \Omega_{-\alpha, \beta}) p(\alpha, \beta),$$

which implies

$$\begin{aligned} p(\alpha, \beta|\Omega_{-\alpha, \beta}) &\propto \\ &\propto \exp\left(-\frac{1}{2} \left(\frac{(\alpha - \mu_\alpha)^2}{\sigma_\alpha^2} + \frac{(\beta - \mu_\beta)^2}{\sigma_\beta^2}\right)\right) \prod_{t=1}^T \exp\left(-\frac{1}{2} \frac{(\alpha + \beta V_{t-1} + \xi_t^y J_t^y - \Delta V_t)^2}{\sigma_v^2 V_{t-1}}\right) \end{aligned}$$

so,

$p(\alpha, \beta | \Omega_{-\alpha, \beta}) \propto$

$$\propto \exp \left( -\frac{1}{2} \begin{pmatrix} \alpha^2 \left( \frac{1}{\sigma_\alpha^2} + \sum_{t=1}^T \frac{1}{\sigma_v^2 V_{t-1}} \right) + \beta^2 \left( \frac{1}{\sigma_\beta^2} + \frac{1}{\sigma_v^2} \sum_{t=1}^T V_{t-1} \right) + 2 \frac{\alpha \beta T}{\sigma_v^2} \\ -2\alpha \left( \frac{\mu_\alpha}{\sigma_\alpha^2} - \frac{1}{\sigma_v^2} \sum_{t=1}^T \frac{(\xi_t^y J_t^y - \Delta V_t)}{V_{t-1}} \right) \\ -2\beta \left( \frac{\mu_\beta}{\sigma_\beta^2} - \frac{1}{\sigma_v^2} \sum_{t=1}^T (\xi_t^y J_t^y - \Delta V_t) \right) \end{pmatrix} \right). \quad (1.7)$$

Now,  $(\alpha, \beta | Y, \Omega_{-\alpha, \beta})$  are jointly normal if and only if

$$p(\alpha, \beta | Y, \Omega_{-\alpha, \beta}) \propto \exp(\mathcal{E}),$$

where,

$$\mathcal{E} = -\frac{1}{2(1 - \rho_{\alpha, \beta}^2)} \left[ \frac{(\alpha - \mu_\alpha)^2}{\sigma_\alpha^2} + \frac{(\beta - \mu_\beta)^2}{\sigma_\beta^2} - \frac{2\rho_{\alpha, \beta}(\alpha - \mu_\alpha)(\beta - \mu_\beta)}{\sigma_\alpha \sigma_\beta} \right].$$

Moreover, if

$$\tilde{\mathcal{E}} = -\frac{1}{2(1 - \rho_{\alpha, \beta}^2)} \left[ \begin{aligned} & \frac{\alpha^2}{\sigma_\alpha^2} + \frac{\beta^2}{\sigma_\beta^2} - 2(\alpha\beta) \frac{\rho_{\alpha, \beta}}{\sigma_\alpha \sigma_\beta} \\ & + 2\beta \left( \frac{\rho_{\alpha, \beta} \mu_\alpha}{\sigma_\alpha \sigma_\beta} - \frac{\mu_\beta}{\sigma_\beta^2} \right) + 2\alpha \left( \frac{\rho_{\alpha, \beta} \mu_\beta}{\sigma_\alpha \sigma_\beta} - \frac{\mu_\alpha}{\sigma_\alpha^2} \right) \end{aligned} \right],$$

we will still have that  $p(\alpha, \beta | Y, \Omega_{-\alpha, \beta}) \propto \exp(\tilde{\mathcal{E}})$  since  $\exp(\tilde{\mathcal{E}}) = K \exp(\mathcal{E})$  for some  $K$  that does not depend on  $\alpha$  or  $\beta$ . So if we have

$$p(\alpha, \beta | Y, \Omega_{-\alpha, \beta}) \propto \exp(\tilde{\mathcal{E}})$$

with

$$\tilde{\mathcal{E}} = \alpha^2 A + \beta^2 B - 2(\alpha\beta) C + 2\beta D + 2\alpha E. \quad (1.8)$$

Hence, to find the parameters of  $p(\alpha, \beta | Y, \Omega_{-\alpha, \beta})$ , all we need to do is solve the



following system of equations:

$$\begin{aligned}
A &= \frac{1}{\sigma_\alpha^2(1-\rho_{\alpha,\beta}^2)}, \\
B &= \frac{1}{\sigma_\beta^2(1-\rho_{\alpha,\beta}^2)}, \\
C &= \frac{\rho_{\alpha,\beta}}{\sigma_\alpha\sigma_\beta(1-\rho_{\alpha,\beta}^2)}, \\
D &= \left( \frac{\rho_{\alpha,\beta}\mu_\alpha}{\sigma_\alpha\sigma_\beta(1-\rho_{\alpha,\beta}^2)} - \frac{\mu_\beta}{\sigma_\beta^2(1-\rho_{\alpha,\beta}^2)} \right), \\
E &= \left( \frac{\rho_{\alpha,\beta}\mu_\alpha}{\sigma_\alpha\sigma_\beta(1-\rho_{\alpha,\beta}^2)} - \frac{\mu_\beta}{\sigma_\beta^2(1-\rho_{\alpha,\beta}^2)} \right).
\end{aligned} \tag{1.9}$$

Solving system (1.9) yields

$$\begin{aligned}
\mu_\alpha &= \frac{BE+CD}{C^2-AB}, \\
\sigma_\alpha^2 &= \frac{B}{AB-C^2}, \\
\mu_\beta &= \frac{AD+CE}{C^2-AB}, \\
\sigma_\beta^2 &= \frac{A}{AB-C^2}, \\
\rho_{\alpha,\beta} &= \frac{C}{\sqrt{AB}}.
\end{aligned} \tag{1.10}$$

Since expression (1.7) is of the form described in equation (1.8), we can conclude that  $(\alpha, \beta|Y, \Omega_{-\alpha,\beta})$  are distributed as a joint normal, and the parameters are given by matching the values of  $A - E$  in equation (1.7) and then using result (1.10).

#### **Drawing from $\mu_y$ , $\sigma_y$ , $\mu_v$ and $\rho_J$**

The priors of  $\mu_y$  and  $\sigma_y$  imply that

$$\begin{aligned}
p(\mu_y) &\propto \exp\left(-\frac{\mu_y^2}{2S}\right) \\
p(\sigma_y^2) &\propto \left(\frac{1}{\sigma_y^2}\right)^{\alpha_p+1} \exp\left(-\frac{\beta_p}{\sigma_y^2}\right).
\end{aligned}$$

Now, let  $\mathcal{J} = \{t : J_t = 1\}$ . Thus, the set  $\mathcal{J}$  will contain the values of  $t$  where a jump has occurred and we denote  $\#\mathcal{J}$  the number of elements in this set, i.e. the number of days when a jump occurred. Then, we restrict  $\xi^y = \{\xi_t^y : t \in \mathcal{J}\}$  i.e. the magnitude of jumps for days when in fact there was a jump. For ease of notation,

let us consider first the case where there are no jumps in the volatility series. Then

$$p(\xi^y | \mu_y, \sigma_y^2) \propto \exp\left(-\frac{\sum_{t \in \mathcal{J}} (\xi_t^y - \mu_y)^2}{2\sigma_y^2}\right).$$

So we have

$$p(\mu_y | \xi^y, \sigma_y^2) \propto p(\xi^y | \mu_y, \sigma_y^2) p(\mu_y),$$

which implies

$$\begin{aligned} p(\mu_y | \xi^y, \sigma_y^2) &\propto \exp\left(-\frac{\sum_{t \in \mathcal{J}} (\xi_t^y - \mu_y)^2}{2\sigma_y^2} - \frac{\mu_y^2}{2S}\right) \\ &\propto \exp\left(-\frac{\sum_{t \in \mathcal{J}} \mu_y^2 - \mu_y^2 \sum_{t \in \mathcal{J}} \xi_t^y}{2\sigma_y^2} - \frac{\mu_y^2}{2S}\right) \\ &\propto \exp\left(-\frac{1}{2} \left[ \frac{S(\#\mathcal{J})\mu_y^2 - 2\mu_y S \sum_{t \in \mathcal{J}} \xi_t^y + \mu_y^2 \sigma_y^2}{\sigma_y^2 S} \right]\right) \\ &\propto \exp\left(-\frac{1}{\frac{2\sigma_y^2 S}{(\#\mathcal{J} \cdot S + \sigma_y^2)}} \left[ \mu_y^2 - 2\mu_y \left( \frac{S \sum_{t \in \mathcal{J}} \xi_t^y}{(\#\mathcal{J} \cdot S + \sigma_y^2)} \right) \right]\right). \end{aligned}$$

For  $\sigma_y^2$  we have

$$\begin{aligned} p(\sigma_y^2 | \xi^y, \mu_y) &\propto p(\xi^y | \mu_y, \sigma_y^2) p(\sigma_y^2) \\ &\propto \left(\frac{1}{\sigma_y^2}\right)^{\alpha_p + 1} \exp\left(-\frac{\beta_p}{\sigma_y^2}\right) \left(\prod_{t \in \mathcal{J}} \frac{1}{\sqrt{\sigma_y^2}} \exp\left(-\frac{1}{2\sigma_y^2} (\xi_t^y - \mu_y)^2\right)\right) \\ &\propto \left(\frac{1}{\sigma_y^2}\right)^{(\alpha_p + \frac{\#\mathcal{J}}{2}) + 1} \exp\left(-\frac{1}{\sigma_y^2} \left(\frac{1}{2} \sum_{t \in \mathcal{J}} (\xi_t^y - \mu_y)^2 + \beta_p\right)\right). \end{aligned}$$

Now, for the models with jumps in volatility, we simply substitute  $\xi_t^y$  by  $\tilde{\xi}_t^y = \xi_t^y - \rho_J \xi_t^v$ .

For  $\mu_v$ , the prior implies

$$p(\mu_v) \propto \left(\frac{1}{\mu_v}\right)^{\alpha_{mv} - 1} \exp\left(-\frac{\left(\frac{1}{\mu_v}\right)}{\beta_{\mu v}}\right),$$

and we also have

$$\begin{aligned} p(\xi_t^v | \mu_v) &\propto \frac{1}{\mu_v} \exp\left(-\frac{1}{\mu_v} \xi_t^v\right) \\ p(\xi^v | \mu_v) &\propto \left(\frac{1}{\mu_v}\right)^{\#\mathcal{J}^v} \exp\left(-\frac{1}{\mu_v} \sum_{t \in \mathcal{J}^v} \xi_t^v\right), \end{aligned}$$

so,

$$\begin{aligned} p(\mu_v | \xi^v) &\propto \left(\frac{1}{\mu_v}\right)^{\#\mathcal{J}^v + \alpha_{mv} - 1} \exp\left(-\frac{1}{\mu_v} \left(\frac{1}{\beta_{\mu v}}\right) - \frac{1}{\mu_v} \sum_{t \in \mathcal{J}^v} \xi_t^v\right) \\ &\propto \left(\frac{1}{\mu_v}\right)^{\#\mathcal{J}^v + \alpha_{mv} - 1} \exp\left(-\frac{1}{\mu_v} \left(\frac{1}{\beta_{\mu v}} + \sum_{t \in \mathcal{J}^v} \xi_t^v\right)\right) \\ &\propto \left(\frac{1}{\mu_v}\right)^{\#\mathcal{J}^v + \alpha_{mv} - 1} \exp\left(-\frac{\left(\frac{1}{\mu_v}\right)}{\frac{\beta_{\mu v}}{1 + \beta_{\mu v} \sum_{t \in \mathcal{J}^v} \xi_t^v}}\right). \end{aligned}$$

Therefore,

$$\mu_v | \xi^v \sim \Gamma\left(\alpha_{mv} + \#\mathcal{J}^v, \frac{\beta_{\mu v}}{1 + \beta_{\mu v} \sum_{t \in \mathcal{J}^v} \xi_t^v}\right).$$

In the case of  $\rho_J$  we have

$$p(\xi^y | \Theta) \propto \exp\left(-\frac{\sum_{t=2}^T ((\mu_y + \rho_J \xi_t^v) - \xi_t^y)^2}{2\sigma_y^2}\right).$$

Further, the prior for  $\rho_J$  implies

$$p(\rho_J) \propto \exp\left(-\frac{1}{2\sigma_{\rho_J}^2} (\rho_J)^2\right).$$

Hence, we have

$$\begin{aligned}
p(\rho_J | \xi_t^y) &\propto p(\xi_t^y | \Theta) p(\rho_J) \\
&\propto \exp\left(-\frac{\sum_{t=2}^T (\rho_J \xi_t^v + (\mu_y - \xi_t^y))^2}{2\sigma_y^2} - \frac{1}{2\sigma_{\rho_J}^2} (\rho_J)^2\right) \\
&\propto \exp\left(-\frac{(\rho_J)^2 \sum_{t=2}^T (\xi_t^v)^2}{2\sigma_y^2} - \frac{\sum_{t=2}^T 2\rho_J \xi_t^v (\mu_y - \xi_t^y)}{2\sigma_y^2} - \frac{(\rho_J)^2}{2\sigma_{\rho_J}^2}\right) \\
&\propto \exp\left(-\frac{1}{\left(\frac{\sigma_{\rho_J}^2 \sum_{t=2}^T (\xi_t^v)^2 + \sigma_y^2}{\sigma_y^2 \sigma_{\rho_J}^2}\right)^{-1}} \left[ (\rho_J)^2 - 2\rho \frac{\sum_{t=2}^T \xi_t^v (\xi_t^y - \mu_y)}{(\sigma_{\rho_J}^2 \sum_{t=2}^T (\xi_t^v)^2 + \sigma_y^2) / \sigma_{\rho_J}^2} \right]\right).
\end{aligned}$$

Therefore,  $\rho_J | \Omega_{-\rho_J}, Y \sim N(m_{\rho_J}, s_{\rho_J}^2)$  where,

$$\begin{aligned}
m_{\rho_J} &= \frac{\sum_{t=2}^T \xi_t^v (\xi_t^y - \mu_y)}{(\sigma_{\rho_J}^2 \sum_{t=2}^T (\xi_t^v)^2 + \sigma_y^2) / \sigma_{\rho_J}^2} \\
s_{\rho_J}^2 &= \left(\frac{\sigma_{\rho_J}^2 \sum_{t=2}^T (\xi_t^v)^2 + \sigma_y^2}{\sigma_y^2 \sigma_{\rho_J}^2}\right)^{-1}.
\end{aligned}$$

### Drawing Jump Times and Jump Frequency

Drawing from  $\lambda_y$  and  $\lambda_v$  is an identical procedure except that in one case the posterior depends on  $J^y$  and in the other of  $J^v$  but both work as follows:

$$\begin{aligned}
\lambda &\sim B(\alpha_\lambda, \beta_\lambda) \\
p(\lambda) &\propto \lambda^{\alpha_\lambda - 1} (1 - \lambda)^{\beta_\lambda - 1},
\end{aligned}$$

and if

$$p(J|\lambda) \propto \lambda^{\#J} (1 - \lambda)^{T - \#J},$$

then

$$\begin{aligned}
p(\lambda|J) &\propto p(J|\lambda) p(\lambda) \\
&\propto \lambda^{\#J + \alpha_\lambda - 1} (1 - \lambda)^{T - \#J + \beta_\lambda - 1}.
\end{aligned}$$

Thus,

$$\lambda|J \sim B(a_\lambda + \#\mathcal{J}, \beta_\lambda + T - \#\mathcal{J}).$$

For the jump sizes, the prior distributions for  $\xi_t^v$  and  $\xi_t^y$  are given by

$$\xi_t^v \sim \exp(\mu_v)$$

and

$$\xi_t^y|\xi_t^v \sim N(\mu_y + \rho_J \xi_t^v, \sigma_y^2).$$

Therefore,

$$p(\xi_t^y|\xi_t^v) \propto \exp\left(-\frac{((\mu_y + \rho_J \xi_t^v) - \xi_t^y)^2}{2\sigma_y^2}\right).$$

Then

$$\begin{aligned} p(\xi_t^y|J_t = 1, \Theta_{-\xi_t^y}, V_{t-1}, Y_t) &\propto \\ &\propto p(Y_t|\xi_t^y, J_t = 1, V_{t-1}, \Theta) p(\xi_t^y|\Theta) \\ &\propto \exp\left(-\frac{((\Delta Y_t - \mu) - \xi_t^y)^2}{2V_{t-1}} - \frac{((\mu_y + \rho_J \xi_t^v) - \xi_t^y)^2}{2\sigma_y^2}\right) \\ &\propto \exp\left(-\left[-2\xi_t^y \left(\frac{(\Delta Y_t - \mu)\sigma_y^2 + (\mu_y + \rho_J \xi_t^v)V_{t-1}}{2\sigma_y^2 V_{t-1}}\right) + (\xi_t^y)^2 \frac{(\sigma_y^2 + V_{t-1})}{2\sigma_y^2 V_{t-1}}\right]\right) \\ &\propto \exp\left(-\frac{1}{2\frac{\sigma_y^2 V_{t-1}}{(\sigma_y^2 + V_{t-1})}} \left[-2\xi_t^y \left(\frac{(\Delta Y_t - \mu)\sigma_y^2 + (\mu_y + \rho_J \xi_t^v)V_{t-1}}{(\sigma_y^2 + V_{t-1})}\right) + (\xi_t^y)^2\right]\right). \end{aligned}$$

Thus,

$$\xi_t^y|\Omega_{-\xi_t^y} \sim N\left(\frac{(\Delta Y_t - \mu)\sigma_y^2 + (\mu_y + \rho_J \xi_t^v)V_{t-1}}{(\sigma_y^2 + V_{t-1})}, \frac{\sigma_y^2 V_{t-1}}{(\sigma_y^2 + V_{t-1})}\right).$$

Next, for the case of  $\xi_t^v$  we have that by two successive applications of Bayes' formula

$$p(\xi_t^v | J_t = 1, \Theta, V_{t-1}, Y_t, \xi_t^y) \propto$$

$$\propto p(Y_t, V_t | \xi_t^y, \xi_t^v, J_t = 1, V_{t-1}, \Theta) p(\xi_t^v | J_t = 1, \Theta, \xi_t^y)$$

$$\propto p(Y_t, V_t | \xi_t^y, \xi_t^v, J_t = 1, V_{t-1}, \Theta) p(\xi_t^y | J_t = 1, \Theta, \xi_t^v) p(\xi_t^v | J_t = 1, \Theta).$$

So,

$$p(\xi_t^v | J_t = 1, \Theta, V_{t-1}, Y_t, \xi_t^y) \propto$$

$$\begin{aligned} &\propto 1_{\xi_t^v > 0} \exp \left( -\frac{1}{2(1-\rho^2)} \left[ \frac{((\Delta Y_t - \mu) - \xi_t^y)^2}{V_{t-1}} + \frac{(\alpha_v + \beta_v V_{t-1} + \xi_t^v - V_t)^2}{\sigma_v^2 V_{t-1}} \right] \right. \\ &\quad \left. - \frac{2\rho((\Delta Y_t - \mu) - \xi_t^y)(\alpha_v + \beta_v V_{t-1} + \xi_t^v - V_t)}{\sigma_v V_{t-1}} \right. \\ &\quad \left. - \frac{((\mu_y + \rho_J \xi_t^v) - \xi_t^y)^2}{2\sigma_y^2} - \mu_v \xi_t^v \right) \\ &\propto 1_{\xi_t^v > 0} \exp \left( -\frac{((\Delta Y_t - \mu) - \xi_t^y)^2}{2V_{t-1}(1-\rho^2)} - \frac{(\alpha_v + \beta_v V_{t-1} + \xi_t^v - V_t)^2}{2\sigma_v^2 V_{t-1}(1-\rho^2)} \right. \\ &\quad \left. + \frac{2\rho((\Delta Y_t - \mu) - \xi_t^y)(\alpha_v + \beta_v V_{t-1} + \xi_t^v - V_t)}{(1-\rho^2)\sigma_v V_{t-1}} \right. \\ &\quad \left. - \frac{((\mu_y + \rho_J \xi_t^v) - \xi_t^y)^2}{2\sigma_y^2} - \mu_v \xi_t^v \right) \\ &\propto 1_{\xi_t^v > 0} \exp \left( -(\xi_t^v)^2 \left[ \frac{\sigma_y^2 + \rho_J^2 \sigma_v^2 V_{t-1}(1-\rho^2)}{2\sigma_v^2 \sigma_y^2 V_{t-1}(1-\rho^2)} \right] \right. \\ &\quad \left. - 2\xi_t^v \left[ \frac{\rho_J(\mu_y - \xi_t^y)\sigma_v^2 V_{t-1}(1-\rho^2)}{2\sigma_y^2 \sigma_v^2 V_{t-1}(1-\rho^2)} + \frac{\mu_v \sigma_v^2 \sigma_y^2 V_{t-1}(1-\rho^2)}{2\sigma_v^2 \sigma_y^2 2V_{t-1}(1-\rho^2)} \right] \right. \\ &\quad \left. + \frac{(\alpha_v + \beta_v V_{t-1} - V_t)\sigma_y^2}{2\sigma_v^2 \sigma_y^2 V_{t-1}(1-\rho^2)} - \frac{\rho\sigma_v \sigma_y^2 ((\Delta Y_t - \mu) - \xi_t^y)}{2(1-\rho^2)\sigma_v^2 \sigma_y^2 V_{t-1}} \right) \end{aligned}$$

and therefore  $\xi_t^v | J_t = 1, \Theta, V_{t-1}, Y_t, \xi_t^y$  is distributed as a truncated normal, with left truncation 0.

To draw from  $J$ , we use the fact that

$$p(J_t = 1 | V_t, V_{t-1}, Y_t, \xi_t^y, \Theta) \propto \lambda \cdot p(Y_t, V_t | J_t = 1, \Omega_{-J_t}) \quad (1.11)$$

$$p(J_t = 0 | V_t, V_{t-1}, Y_t, \xi_t^y, \Theta) \propto (1 - \lambda) \cdot p(Y_t, V_t | J_t = 0, \Omega_{-J_t}) \quad (1.12)$$

we define  $q = \Pr(J_t = 1 | \Omega_{-J_t})$ . Then we can define the odds ratio

$$O = \frac{q}{1 - q} = \frac{p(J_t = 1 | V_t, V_{t-1}, Y_t, \xi_t^y, \Theta)}{p(J_t = 0 | V_t, V_{t-1}, Y_t, \xi_t^y, \Theta)}.$$

Now, notice that the constants that would make (1.11) and (1.12) equalities instead

of proportions are indeed the same. Thus, we can calculate  $O$  as

$$O = \frac{\lambda \cdot p(Y_t, V_t | J_t = 1, \Omega_{-J_t})}{(1 - \lambda) \cdot p(Y_t, V_t | J_t = 0, \Omega_{-J_t})}$$

$$q = \frac{O}{(1 + O)}.$$

Therefore,

$$J_t | V_t, V_{t-1}, Y_t, \xi_t^y, \Theta \sim \text{Ber}(q).$$

Thus, covering all the state variables and parameters except  $V_t$  and  $\rho$ .

### Drawing from Non-Standard Distributions

First, for the case of  $V_t$  notice that by successive applications of Bayes' rule  $p(V_t | V_{t+1}, V_{t-1}, \Delta Y_{t+1}, \Theta, J, \xi^y, \xi^v) \propto$

$$\begin{aligned} &\propto p(V_t, V_{t+1}, V_{t-1}, \Delta Y_{t+1} | \Theta) \\ &\propto p(V_{t+1}, \Delta Y_{t+1} | V_{t+1}, V_{t-1}, \Theta) p(V_t, V_{t-1} | \Theta) \\ &\propto \left[ \begin{array}{c} p(V_{t+1} | V_t, V_{t-1}, \Delta Y_{t+1}, \Theta) p(\Delta Y_t | V_{t+1}, V_{t-1}, \Theta) \cdot \\ p(V_t | \Theta, V_{t-1}) p(V_{t-1} | \Theta) \end{array} \right] \\ &\propto p(\Delta Y_{t+1} | V_t, \Theta) p(V_{t+1} | V_t, \Theta) p(V_t | V_{t-1}, \Theta). \end{aligned}$$

We also have that

$$p(\Delta Y_{t+1} | V_t, \Theta) p(V_t | V_{t-1}, \Theta) p(V_{t+1} | V_t, \Theta) \propto \left( \begin{array}{c} \frac{1}{V_t^{1/2}} \exp \left\{ -\frac{(Y_{t+1} - (Y_t + \mu))^2}{2V_t} \right\} \\ \frac{1}{\sigma_v V_{t-1}^{1/2}} \exp \left\{ -\frac{(V_t - (\alpha_v + \beta_v V_{t-1}))^2}{2\sigma_v^2 V_{t-1}} \right\} \\ \frac{1}{\sigma_v V_t^{1/2}} \exp \left\{ -\frac{(V_{t+1} - (\alpha_v + \beta_v V_t))^2}{2\sigma_v^2 V_t} \right\} \end{array} \right) \quad (1.13)$$

$$\propto \left( \frac{1}{V_t} \exp \left\{ \begin{array}{c} -\frac{(Y_{t+1} - (Y_t + \mu))^2}{2V_t} \\ -\frac{(V_t - (\alpha_v + \beta_v V_{t-1}))^2}{2\sigma_v^2 V_{t-1}} \\ -\frac{(V_{t+1} - (\alpha_v + \beta_v V_t))^2}{2\sigma_v^2 V_t} \end{array} \right\} \right). \quad (1.14)$$

Thus,  $V_t$  does not seem to follow any standard distribution, so to draw from it we use a random-walk, Metropolis-Hastings algorithm, where a candidate  $V_t^c$  is drawn

as  $V_t^{c-1} + \zeta$  where  $\zeta$  is a mean 0 normal random variable. Expression (1.14) is then evaluated at the candidate  $V_t^c$  and at  $V_t^{c-1}$ . If the likelihood of  $V_t^c$  is greater than that of  $V_t^{c-1}$ ,  $V_t^c$  is immediately accepted, otherwise it is accepted with probability equal to the ratio of the two likelihoods. Finally, the variance of the step  $\zeta$  is calibrated so that candidate draws are rejected between 30% and 60% of the time.

Drawing from  $\rho$  implies a similar procedure called an independence Metropolis-Hastings. In this case, instead of drawing from a random walk, the algorithm draws candidate  $\rho^c$ s from a uniform distribution. The decision to accept or reject a given draw is then made following the same likelihood comparison step as in the case for  $V_t$ . The details of this drawing procedures can be found in Johanes and Polson (2009).

With this, we now have procedures for drawing from all the parameters and state variables.



# Bibliography

Aït-Sahalia Y. (2001): "Closed Form Likelihood Expansions for Multivariate Diffusions," Working paper, Princeton university.

Aït-Sahalia Y. (2002): "Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, Vol. 70, 223-262.

Aït-Sahalia Y. and R. Kimmel (2007): "Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Financial Economics*, Vol. 83, 413-452.

Aït-Sahalia Y. and J. Jacod (2009): "Estimating the Degree of Activity of Jumps in High Frequency Data," *Annals of Statistics*, Vol. 37, 2202-2244.

Aït-Sahalia Y. and A. Lo (1998): "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance*, Vol. 53, 499-547.

Andersen, T. G., L. Benzoni, and J. Lund (2002): "An Empirical Investigation of Continuous-Time Models for Equity Returns," *Journal of Finance*, Vol. 57, 1239-1284.

Andersen, T. G. and J. Lund (1997): "Estimating Continuous Time Stochastic Volatility Models of the Short Term Interest Rate," *Journal of Econometrics*, Vol. 77, 343-378.

Aspe, P. (1995): "Mexico en 1994: Las Razones de la Política Cambiaria," *Reforma*, Mexico July 15 1995.

Bakshi G., C. Cao, and Z. Chen (1997): "Empirical Performance of Alternative Option Pricing Models," *Journal of Finance*, Vol. 52, 2003-2049.

- Bakshi G., C. Cao, and Z. Chen (2000): "Do Call Prices and the Underlying Stock Always Move in the Same Direction?," *Review of Financial Studies*, Vol. 13, 549-584.
- Bates, D. (1996): "Jump and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options," *Review of Financial Studies*, Vol. 9, 69-107.
- Bates, D. (2000): "Post-'87 Crash Fears in S&P 500 Futures Options," *Journal of Econometrics*, Vol. 94, 181-238.
- Black, F. and M. Scholes (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, Vol. 81, 637-654.
- Bakshi G., C. Cao, and Z. Chen (1997): "Empirical Performance of Alternative Option Pricing Models," *Journal of Finance*, Vol. 52, 2003-2049.
- Bolsa Mexicana de Valores (2011): "Nota Metodológica del Índice de Precios y Cotizaciones," accessed October 30 2011, <http://www.bmv.com.mx/wb3/work/sites/BMV/resources/LocalContent/1476/7/NotaMetdIPCnov2011.pdf>.
- Duffie, D., J. Pan, and K. Singleton (2000): "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, Vol. 68, 1343-1376.
- Eraker, B. (2001): "MCMC Analysis of Diffusion Models with Application to Finance," *Journal of Business and Economic Statistics*, Vol. 19, 177-191.
- Eraker, B., M. Johannes and N. Polson (2003): "The Impact of Jumps in Volatility and Returns," *Journal of Finance*, Vol. 58, 1269-1300.
- Fisher M. and C. Gilles (1996): "Term Premia in Exponential-Affine Models of the Term Structure," Working paper, Board of Governors of the Federal Reserve System Washington DC.
- Gallant R. A. and G. Tauchen (1996): "Which Moments to Match?," *Econometric theory*, Vol. 12, 657-681.

- Heston, S. (1993): "A Closed-Form Solution of Options with Stochastic Volatility with Applications to Bond and Currency Options," *The Review of Financial Studies*, Vol. 6, 327-343.
- Honore, P. (1998): "Pitfalls in Estimating Jump-Diffusion Models," Working paper, University of Aarhus.
- Hull, J. and A. White (1987): "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance*, Vol. 42, 281-300.
- Jacquier E., N. G. Polson, and P. E. Rossi (1994): "Bayesian Analysis on Stochastic Volatility Models," *Journal of Business and Economic Statistics*, Vol. 12, 371-396.
- Johannes, M. and N. Polson (2009): "MCMC Methods for Continuous-Time Financial Econometrics," in Aït-Sahalia, Y. and L. Hansen, Ed. *Handbook of Financial Econometrics*, Vol. 2, North Holland, Oxford.
- Johannes, M., N. Polson and J. Stroud (2009): "Optimal Filtering of Jump-Diffusions: Extracting Latent States from Asset Prices," *Review of Financial Studies*, Vol. 22, 2759-2799.
- Kim, S., Shephard, N., and Chib, S. (1999): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, Vol. 65, 361-393.
- Ledoit, O., P. Santa-Clara, and S. Yan (2002): "Relative Pricing of Options with Stochastic Volatility," Working Paper, University of California at Los Angeles.
- McFaden D. A. (1989): "Method of Simulated Moments for Estimation of a Discrete Response Models Without Numerical Integration," *Econometrica*, Vol. 57, 995-1026.
- Melino A. (1994): "Estimation of Continuous-Time Models in Finance," in C. A. Sims, Ed.: *Advances in Econometrics: Sixth World Congress*, Vol. II Cambridge university press, Cambridge.
- Merton, R.C. (1980): "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, Vol. 8, 323-361.

- MexDer (2008): "Metodología para el Cálculo de Precios de Liquidación por Medio del Modelo de Heston," accessed October 30 2011, [http://www.mexder.com.mx/inter/info/mexder/avisos/Metodologia\\_de\\_Precios\\_de\\_Liquidacion\\_por\\_medio\\_de\\_Heston.pdf](http://www.mexder.com.mx/inter/info/mexder/avisos/Metodologia_de_Precios_de_Liquidacion_por_medio_de_Heston.pdf).
- Pakes, A. and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, Vol. 57, 1027-1058.
- Pan, J. (2002): "The Jump-Risk Premia Implicit in Options: Evidence From an Integrated Time-Series Study," *Journal of Financial Economics*, Vol. 63, 3-50.
- Preston, J. and S. Dillon (2004): "Opening Mexico: The Making of a Democracy," Farrar, Straus and Giroux, New York.
- Scott, L.O. (1987): "Option Pricing when the Variance Changes Randomly: Theory, Estimation, and an Application," *Journal of Financial and Quantitative Analysis*, Vol. 22, 419-438.
- Sepp A. (2003): "Fourier Transform for Option Pricing Under Affine Jump-Diffusions: An Overview," Working paper, University of Tartu.
- Smith, A. F. M. and D. J. Spiegelhalter (1980): "Bayes Factors and Choice Criteria for Linear Models," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 42, 213-220.
- Stein, J.C. (1989): "Overreactions in the Options Market," *Journal of Finance*, Vol. 44, 1011-1023.
- Wiggins, J. B. (1987): "Option Values Under Stochastic Volatilities," *Journal of Financial Economics*, Vol. 19, 351-372.

## Chapter 2

# Pessimistic Portfolio Selection with Many Assets

### 2.1 Introduction

The portfolio selection procedure of Markowitz (1952) was a milestone in the development of financial theory. Nonetheless, as a theory of individual choice it is inscribed in the rational choice paradigm and thus will be unable to reflect the preference of non-rational (or behavioral) individuals. Furthermore, from the perspective of institutional investors it rests crucially on using variance as a measure of risk. However, recent literature on risk measures (most notably Artzner et al. (1999)) has established a number of desirable properties for risk measures and shown that variance does not enter the set of so-called "coherent" measures.

In this context, Bassett et al. (2004) present a portfolio construction procedure that can be interpreted as the choice of a pessimistic individual or that of an institution concerned with a coherent risk measure. From the practical perspective, this procedure can be constructed as a quantile regression and can thus be implemented on standard statistical software.

However, if the number of available assets is very large and the optimal portfolio has non-zero positions in all the assets it could imply significant fees for a small investor. It is the purpose of this paper to extend Bassett's analysis by proposing

a procedure that will restrict the number of assets with non-zero positions. To do so I introduce an  $\ell_1$  constraint in the vector of portfolio weights. This idea has been implemented in the context of Markowitz portfolios by Fan et al. (2009) and by Brodie et. al. (2009). Moreover, in the context of regression analysis it has led to the LASSO regression of Tibshirani (1996) as well as  $\ell_1$ -Quantile regression as presented, for example, in Belloni and Chernozhukov (2011).

Now, in the context of Markowitz portfolios Fan et al. (2009) find that adding the  $\ell_1$  restriction is in fact equivalent to restricting the number of active assets (the  $\ell_0$  norm). I will show that this relationship is not so direct in the case of Choquet portfolios, but still useful in terms of restricting the number of active assets.

More importantly, using simulations and US market data I will investigate the effects of adding the  $\ell_1$  constraint to Choquet portfolios. My main focus will be to analyze how expected and realized performance is hindered by adding the  $\ell_1$  constraint. In order to do so, I construct portfolios without the  $\ell_1$  constraint and their performance is recorded as a benchmark. Then the expected and realized returns and risk are measured for portfolios with the  $\ell_1$  norm restricted.

I find that the loss in expected return or gain in expected risk tends to accentuate as  $\ell_1$  restriction approaches the no-short selling restriction ( $\ell_1$  norm equal to 1). This indicates that most of the expected loss of performance occurs only when we come close to this extreme case. Thus, as long as the restriction does not approach 1, portfolios can be constructed with the benefits of adding the  $\ell_1$  norm in terms of lower fees, and with relatively small losses in terms of expected performance.

Now, the available data for US markets allows me to perform ex-post performance analysis under very different environments. I present three scenarios, one baseline "non-stress" scenario with no significant market turmoil in the out-of sample period and two stress scenarios one with the 1987 crash and the second with the collapse of Lehman Brothers in the out-of-sample period. Under the stress scenarios I find that risk predictions tend to be unreliable regardless of the portfolio selection procedure being used. However, the difference between expected and realized risk tends to be much larger for unrestricted portfolios than for their restricted counterparts. This indicates that adding the  $\ell_1$  restriction might also have the benefit of

rendering more reliable estimates of risk.

The rest of the paper is organized as follows: section 2 introduces Choquet portfolio selection, both in the context of Choquet expected utility and coherent risk measures. Section 3 presents the main results illustrating the effect of adding an  $\ell_1$  constraint to the portfolio selection process and section 4 concludes.

## **2.2 Choquet Expectation and Choquet Expected Utility in Portfolio Selection**

### **2.2.1 Choquet Expected Utility as an Alternative to Expected Utility**

The study of choice under uncertainty dates back to at least Bernoulli's (1738) work. However, throughout the second half of the twentieth century expected utility theory as presented by von Neumann and Morgenstern (1944) and axiomatized by Savage (1954) became the standard tool used to model choice under risk (i.e. when probabilities are known). Later Anscombe and Aumann (1963) formalized the concept of subjective probability and thus extended the analysis to the uncertainty case.

Nonetheless, the limitations of expected utility have been widely documented in the experimental and behavioral literature (see for example Camerer (1995) for an extensive survey). Many of the violations seem to stem from the fact that the valuation of lotteries is not linear in probabilities as expected utility assumes. Hence, in order to model the situations where the standard expected utility theory is violated, a number of alternative choice theories have emerged.

In this context, one of the most popular alternatives to expected utility is prospect theory. It was introduced by Kahneman and Tversky (1979), generalized in Tversky and Kahneman (1992) and axiomatized in Wakker and Tversky (1993). One of the key tenets of prospect theory is that small probabilities tend to be overweighted and large probabilities tend to be underweighted. This has been modeled by introducing decision weights obtained from a nonadditive transformation of the

probability scale.

In Anscome and Aumann's (1963) exposition of subjective probability they show that if an individual has rational preferences over a set of lotteries with unknown probabilities of the outcomes, they will in fact be behaving as if they were maximizing expected utility even though they do not know the explicit probabilities of the outcomes they are facing.

In contrast, under cumulative prospect theory agents act as though they maximize expected utility with respect to a nonadditive measure (i.e. a functional where the measure of the union of disjoint sets is not necessarily the sum of the measure of the individual sets). In order to find expectations with respect to such a nonadditive measure (also called capacity) the Choquet functional (or Choquet integral) is used.

Formally, a Choquet Capacity (also called non-additive probability or simply capacity) and the Choquet integral are defined as follows:

**Definition 2** For a state space  $S$  and an associated  $\sigma$ -algebra  $\mathcal{E}$ , a **Choquet Capacity** will be a function  $W : \mathcal{E} \rightarrow \mathbb{R}$  with the following three properties:

1. (Monotonicity): If  $A, B \in \mathcal{E}$  and  $A \subseteq B$  then  $W(A) \leq W(B)$
2. (Normalization):  $W(\emptyset) = 0$  and  $W(S) = 1$

**Definition 3** Let  $W : \mathcal{E} \rightarrow \mathbb{R}$  be a Choquet Capacity. Then for any function  $\phi : S \rightarrow \mathbb{R}$  the **Choquet Integral** of  $\phi$  with respect to  $W$ , denoted as  $\int \phi dW$  is

$$\int_0^{\infty} W(\{i \in S : \phi(i) \geq t\}) dt + \int_{-\infty}^0 [W(\{i \in S : \phi(i) \geq t\}) - 1] dt$$

Note that a probability measure is a particular case of a capacity, and for any capacity that is also additive, the Choquet integral will be the usual Lebesgue integral (a proof is shown in Wakker (1989)).

Evaluating lotteries through Choquet expected utility encompasses a wide variety of models, such as those proposed by Quiggin (1982), Schmeidler (1989) and Yaari (1987). In all these models, cumulative, rather than individual probabilities are transformed.



Now, the relationship between expected utility and Choquet expected utility can be illustrated as follows. Let  $X$  be a lottery with monetary outcomes and an associated CDF,  $F_X$ . Under expected utility theory the value an agent assigns to  $X$  is given by the expectation of a monotonically increasing utility function  $u$  under  $F_X$ . A change of variables allows us to express expected utility in terms of the quantile functions as follows:

$$E_F u(x) = \int_{-\infty}^{\infty} u(x) F_X(x) = \int_0^1 u(F_X^{-1}(t)) dt.$$

This representation "adds" equally likely slices on the interval  $[0, 1]$  and since the integration is done with respect to the Lebesgue measure, we have that each of the "slices" will have the same weight. This representation is very useful to emphasize what changes once we move to Choquet expected utility. Bassett et al. (2004) show that the valuation of a lottery  $X$  can be expressed as

$$E_{v,F} u(X) = \int_0^1 u(F_X^{-1}(t)) dv(t),$$

where  $v : [0, 1] \rightarrow [0, 1]$  is monotonically increasing and  $v(0) = 0$  and  $v(1) = 1$ . The difference with expected utility is that the "slices" are no longer weighted equally. Thus, through the weighting function notions of pessimism and optimism can be introduced. That is, whenever  $v$  implies an overweighting of the left tail of the distribution, we will be in the domain of pessimism and when it overweights the right tail, we will be in that of optimism. Note that, since  $u$  is monotonically increasing,  $v$  will introduce pessimism if it is concave as it will overweight the probabilities of the worst outcomes.

Now, one of the simplest ways to introduce pessimism is by using the following weighting function:

$$v_\alpha(t) = \min \left\{ \frac{t}{\alpha}, 1 \right\}, \alpha \in [0, 1].$$

In this case we have that

$$E_{v_\alpha, F} u(X) = \frac{1}{\alpha} \int_0^\alpha u(F_X^{-1}(t)) dt.$$

Thus, the choice process is ignoring the set of the most favorable outcomes that has a probability of  $(1 - \alpha)$ .

### 2.2.2 Choquet Expectation and Coherent Risk Measures

Models of decision that include pessimism have been developed with individual decision makers in mind. However, one can interpret some of the most widely used tools of risk management as including elements of pessimism. In fact, we can think of risk measures such as Value at Risk (VaR) as being pessimistic in the sense that they are concerned only with the worst possible outcomes. Nonetheless, VaR might not be a very good risk measure. Since VaR is simply a statement of the  $\alpha$ -th quantile of the expected returns, it will not capture any information of the distribution of returns smaller than that level. Consider the case where two portfolios have identical distributions above the  $\alpha$ -th quantile, and both have a single possible outcome below the  $\alpha$ -th quantile. However, for one distribution, the single "catastrophic" outcome is very close to the VaR, while for the second distribution, the catastrophic outcome is extremely negative (e.g. it can represent a loss of all the investment). It stands to reason that one should consider the second alternative riskier than the first, nonetheless using the VaR criterion, both are determined to be equally risky.

Another commonly used measure of risk is the variance. Yet, this measure has the disadvantage that it is not monotone in returns. To see this, note that if a portfolio has a return that is always equal to that of a second portfolio plus a constant, using variance will assign both an equal risk. These examples have led to the idea that when using a risk measure to evaluate a portfolio, there are a number of desired properties this measure should hold. Following this train of reasoning, Artzner et al. (1999) provide an axiomatic foundation for "coherent" risk measures.

**Definition 4** (Artzner et al.) *For real valued random variables  $X$  a mapping  $\rho : \mathcal{X} \rightarrow \mathbb{R}$  is called a **coherent risk measure** if it is:*

1. *Monotone:  $X, Y \in \mathcal{X}$ , with  $X \leq Y \implies \rho(X) \geq \rho(Y)$*
2. *Subadditive:  $X, Y, X + Y \in \mathcal{X} \implies \rho(X + Y) \leq \rho(X) + \rho(Y)$*

3. *Linearly Homogeneous:*  $\forall \lambda \geq 0$  and  $X \in \mathcal{X}$ ,  $\varrho(\lambda X) = \lambda \varrho(X)$

4. *Translation Invariant:*  $\forall \lambda \in \mathbb{R}$  and  $X \in \mathcal{X}$   $\varrho(\lambda + X) = \varrho(X) - \lambda$ .

An immediate consequence is that neither VaR nor Variance will be coherent risk measures. VaR will violate subadditivity and variance will violate monotonicity. However, using Choquet expected value, we can construct an alternative risk measure that will be coherent.

**Definition 5** *The  $\alpha$ -risk of a random variable  $X$  is*

$$\varrho_\alpha(X) = - \int_0^1 F_X^{-1}(t) dv_\alpha(t) = -\frac{1}{\alpha} \int_0^\alpha F_X^{-1}(t) dt.$$

*Thus, the  $\alpha$ -risk is simply the negative of the Choquet expected value of  $X$ .*

Notice that if we denote

$$q_\alpha(u) = u(\alpha - \mathbf{I}(u < 0)),$$

(the typical quantile loss function) and recall that if  $\xi_\alpha$  is the  $\alpha$ -quantile of  $X$ ,  $\xi_\alpha$  will satisfy

$$\xi_\alpha = \arg \min_{\tilde{\xi}} E \left( q_\alpha \left( X - \tilde{\xi} \right) \right),$$

Then, following Bassett et al. (2004), note:

$$\begin{aligned} E \left( q_\alpha \left( X - \tilde{\xi} \right) \right) &= E \left[ \left( X - \tilde{\xi} \right) \left( \alpha - \mathbf{I} \left( X < \tilde{\xi} \right) \right) \right] \\ &= \alpha E[X] - \alpha \tilde{\xi} - \int_{-\infty}^{\tilde{\xi}} (x - \tilde{\xi}) F_X(x) dx. \end{aligned}$$

If we then evaluate at  $\tilde{\xi} = \xi_\alpha$ , we have

$$\begin{aligned} E \left( q_\alpha \left( X - \xi_\alpha \right) \right) &= \alpha E[X] - \alpha \xi_\alpha + \int_{-\infty}^{\xi_\alpha} \xi_\alpha F_X(x) dx - \int_{-\infty}^{\xi_\alpha} x F_X(x) dx \\ &= \alpha E[X] - \alpha \xi_\alpha + \alpha \xi_\alpha - \int_0^\alpha F_X^{-1}(t) dt \\ &= \alpha E[X] + \alpha \varrho_\alpha(X) \end{aligned}$$

then

$$E(q_\alpha(X - \xi_\alpha)) = \alpha E[X] + \alpha \varrho_\alpha(X)$$

Which can be rearranged as

$$\varrho_\alpha(X) = \frac{1}{\alpha} E(q_\alpha(X - \xi_\alpha)) - E[X] = \frac{1}{\alpha} \min_{\tilde{\xi}} E(q_\alpha(X - \tilde{\xi})) - E[X], \quad (2.1)$$

which will serve as the starting point to construct portfolios using  $\alpha$ -risk.

Moreover, we can go beyond this simple risk measure by defining a pessimistic risk measure as any one that can be represented as  $\int_0^1 \varrho_\alpha(X) d\varphi(\alpha)$  for some probability measure  $\varphi$  on  $[0, 1]$ . Then we can see that all pessimistic measures will be coherent.

### 2.2.3 Pessimistic Portfolios

We now consider the problem of an agent or a financial institution that faces the decision of investing in  $p$  available assets. Markowitz's procedure solves the problem of maximizing expected return subject to a constraint on the variance or minimizing the variance of the portfolio subject to an expected return constraint. In the context of expected utility this procedure will be consistent with the assumption of an individual either facing spherical returns or possessing quadratic preferences. Thus, Markowitz's procedure imposes very strong assumptions on the preferences of the agent. From the perspective of an institutional investor, it is implicitly assuming that the variance is used as a risk measure. This, as I discussed above, implies the use of a non-coherent risk measure.

In this context, we will first, present Basset et al.'s procedure for portfolio choice under pessimistic preferences (or a coherent risk measure) and then extend it to the many assets case.

Let us denote with  $R = \{R_i\}_{i=1}^p$  the set of returns of the  $p$  available assets and let us assume that we observe a random sample  $\{r_t = (r_{t1} \dots r_{tp})\}_{t=1}^T$  from the joint distribution of asset returns. We can use the empirical equivalent of  $\varrho_\alpha$  to denote

the  $\alpha$ -risk of a single asset as follows:

$$\hat{\varrho}_{v_a} = \min_{\xi \in \mathbb{R}} \frac{1}{T\alpha} \sum_{t=1}^T q_\alpha(r_t - \xi) - \hat{\mu}_n,$$

where  $\hat{\mu}_n$  is an estimator of  $E(R)$  and we can see that the  $\alpha$ -risk of a portfolio that assigns a weight  $w_j$  to asset  $j$  will be

$$\hat{\varrho}_{v_a}(w'R) = (T\alpha)^{-1} \min_{\xi \in \mathbb{R}} \sum_{t=1}^T q_\alpha \left( \sum_{j=1}^p w_j r_{jt} - \xi \right) - \hat{\mu}_n.$$

Thus, to find the portfolio that minimizes  $\alpha$ -risk we solve

$$\hat{\varrho}_{v_a}(w'R) = \min_{w \in \mathbb{R}^p, \xi \in \mathbb{R}} (n\alpha)^{-1} \sum_{i=1}^n q_\alpha \left( r_{ip} - \sum_{j=1}^{p-1} w_j (r_{i1} - r_{ij}) - \xi \right) - \hat{\mu}_n.$$

This would give us the minimum  $\alpha$ -risk portfolio. Nonetheless, in practice we are typically interested in achieving a minimum expected return  $\mu_0$ . Thus, the portfolio selection problem becomes

$$\min_w \hat{\varrho}_{v_a}(w'R) \tag{2.2}$$

subject to:

$$w' \bar{r} = \mu_0, \tag{2.3}$$

$$w' \mathbf{1} = 1, \tag{2.4}$$

which Basset et al. show can be solved by finding the parameters of a quantile regression (the procedure is detailed in Appendix B). This is in fact very convenient, since the portfolio selection problem can be solved in a straightforward manner, using standard statistical software.

However, a drawback to using the setup described in equations (2.2) - (2.4) is that in practice we might be more interested in maximizing expected return subject to a certain level of  $\alpha$ -risk. A classic example of this is an investment firm trying to maximize its expected return subject to a regulatory constraint on the amount of

risk it can be exposed to.

In this case, we now have

$$\max_w w' \bar{r} \tag{2.5}$$

subject to:

$$\hat{\rho}_{v_a}(w'R) \leq K, \tag{2.6}$$

$$w' \mathbf{1} = 1, \tag{2.7}$$

which can be reduced to a linear programming problem (shown in Appendix B).

#### 2.2.4 Portfolio Selection when $p$ is Large

Finding optimal portfolios under either Markowitz's procedure or under Choquet selection tends to yield solutions where all the available assets have non-zero weights in the solution. Brodie et. al. (2009) show that this can be a problem if the investor is facing a vast set of assets. First, consider the case where the agent is a small investor. In this case volume independent "overhead" costs can be important. Thus, restricting the amount of active assets (those with non-zero positions) in the optimal portfolio might be a good idea. This is equivalent to setting a restriction on the  $\ell_0$  norm of the portfolio. Nonetheless, adding a restriction on the  $\ell_0$  norm makes the portfolio selection problem an NP-complete combinatorial optimization problem. Therefore, it might become computationally very expensive, or even intractable if the number of assets is very large.

Further, for a large investor, the volume independent overhead costs will be less important. Instead, their principal cost will be a fixed bid-ask spread and therefore transaction costs will be proportional to the gross market value of the portfolio (which is proportional to the  $\ell_1$  norm of the portfolio).

For the case of Markowitz portfolios, it has been shown (Fan et al. (2009)) that adding a restriction on the  $\ell_1$  norm instead of  $\ell_0$  will result in an equivalent and feasible convex optimization problem. Further, in the context of quantile regression imposing  $\ell_1$  restrictions on the vector of estimated parameters has been used as a

form of obtaining models that use only a subset of the available variables (see for example (Belloni and Chernozhukov (2011)) and the references therein).

Notice that an immediate effect of adding this restriction is that it prevents individual assets from having extremely large (either long or short) positions. For instance, by setting  $\|w\|_1 \leq 1.6$  restricts  $w_i \in [\underline{w}, \bar{w}]$  with  $\underline{w} \approx -0.3, \bar{w} \approx 1.3$  for every  $i$ . We can also note that the extreme case where we set  $\|w\|_1 \leq 1$  will correspond to the no-short selling restriction.

By restricting the  $\ell_1$  norm of the weights we are directly restricting the magnitude of fees that are proportional to the bid-ask spread. Further, following what is observed in the  $\ell_1$  constrained quantile regression setting, I would expect that by constraining the  $\ell_1$  norm of the weights, the number of active assets will also be restricted, contracting the overhead fees that would be significant for a small investor.

Nonetheless, by imposing this new constraint, we might be hindering the potential performance of the portfolios, since we are now choosing from a smaller set of possible investments. Because of this, in the next section, I will investigate the effect of adding an  $\ell_1$  restriction on the Choquet portfolio selection problem. I will be particularly interested in investigating the effect that restricting the number of active assets via an  $\ell_1$  restriction has on the performance of the portfolios relative to that of their unrestricted counterparts.

For the first comparison exercise, I will simulate data for 100 assets based on the three factor model of Fama and French (1993) calibrated to replicate the behavior of the US stock market from 2002-2005. This will allow me to analyze the performance of  $\ell_1$  constrained portfolios in a stylized model of the US stock market.

I will then use the daily returns of 100 industrial portfolios formed on size and book to market ratio from the website of Kenneth French. With this data I will study portfolio performance for three periods. First a non-stress period having March 1988-February 1992 as a sample period and February 1992 - February 1993 as an out-of-sample time frame. Then I will consider two stress scenarios having stock crashes in the middle of the out-of-sample period. One will contain the October 1987 crash and the second will contain the collapse of Lehman Brothers.

## 2.3 Performance Analysis

### 2.3.1 Simulation Exercise: Procedure

Following Fan et al. (2009), I simulate a three factor model based on the findings of Fama and French (1993):

$$R_i = b_{i1}f_1 + b_{i2}f_2 + b_{i3}f_3 + \varepsilon_i$$

where  $b_{ij}$  is the factor loading of the  $i$ 'th stock on the  $j$ 'th factor and  $\varepsilon_i$  is the idiosyncratic noise, independent of the three factors. The realizations of  $\varepsilon_i$  are assumed to be independent from each other and follow a Student's- $t$  distribution with 6 degrees of freedom and a standard deviation  $\sigma_i$ . Following the calibrations of Fan et al. (2009) the levels of idiosyncratic noise  $\{\sigma_i\}_{i=1}^p$  are assumed to be distributed as a Gamma with shape parameter 3.3586 and scale parameter 0.1876, conditioned on a noise level of at least 0.1950. For each of the  $p$  assets  $\sigma_i$  is drawn once and then kept constant throughout the simulations. Finally,  $(b_{i1}, b_{i2}, b_{i3})$  are assumed to be jointly normally distributed with mean  $\mu_f$  and variance  $\Sigma_f$  with the following parameters:

$$\mu_b = \begin{bmatrix} 0.7828 \\ 0.5180 \\ 0.4100 \end{bmatrix}, \quad \Sigma_b = \begin{bmatrix} 0.02914 & 0.02387 & 0.01018 \\ 0.02387 & 0.05395 & -0.00696 \\ 0.01018 & -0.00696 & 0.086856 \end{bmatrix},$$
$$\mu_f = \begin{bmatrix} 0.02355 \\ 0.01298 \\ 0.02071 \end{bmatrix}, \quad \Sigma_f = \begin{bmatrix} 1.2507 & -0.0350 & -0.2042 \\ -0.0350 & 0.3156 & 0.0023 \\ -0.2042 & 0.0023 & 0.1930 \end{bmatrix}.$$

These parameters were calibrated by Fan et al. (2008) to replicate the behavior of the US stock data for the May 2002 to August 2005 period.



### 2.3.2 Simulation Exercise: Results

First, a return  $\mu_0$  is set as  $\frac{1}{4}\bar{r}_{(\min)} + \frac{3}{4}\bar{r}_{(\max)}$ . Then I find the portfolio that minimizes 5%  $\alpha$ -risk subject to an expected return equal to  $\mu_0$ . Next, I construct a series of 100 portfolios adding a restriction of the form  $\|w\|_1 \leq c$  where  $c$  is allowed to vary from 1 (the no short-selling case) to the  $\ell_1$  norm of the unrestricted portfolio. For the purposes of comparison, this procedure is then repeated with the variance as objective function.

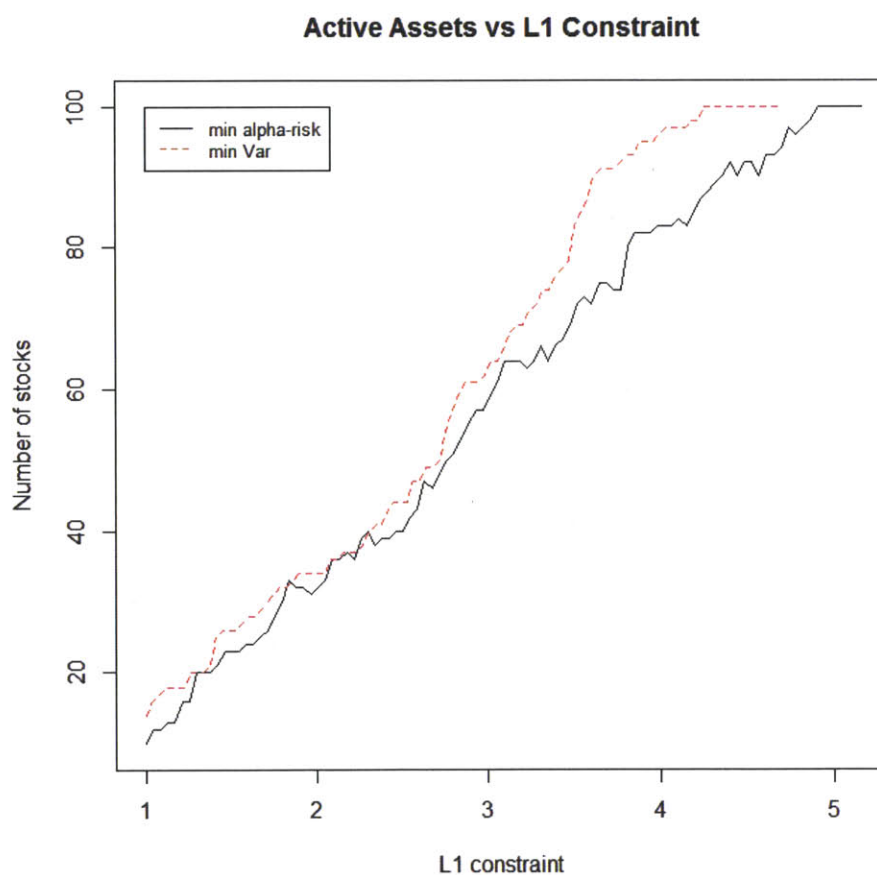


Figure 2-1: Active assets as a function of  $\ell_1$  constraint.

Figure 2-1 shows the number of active assets as a function of the restriction on  $\|w\|_1$ . We can see that when the objective function is variance, the number of active assets increases monotonically as the  $\ell_1$  constraint is relaxed. However, in the case of portfolios that minimize  $\alpha$ -risk, the relationship is not strictly monotonous.

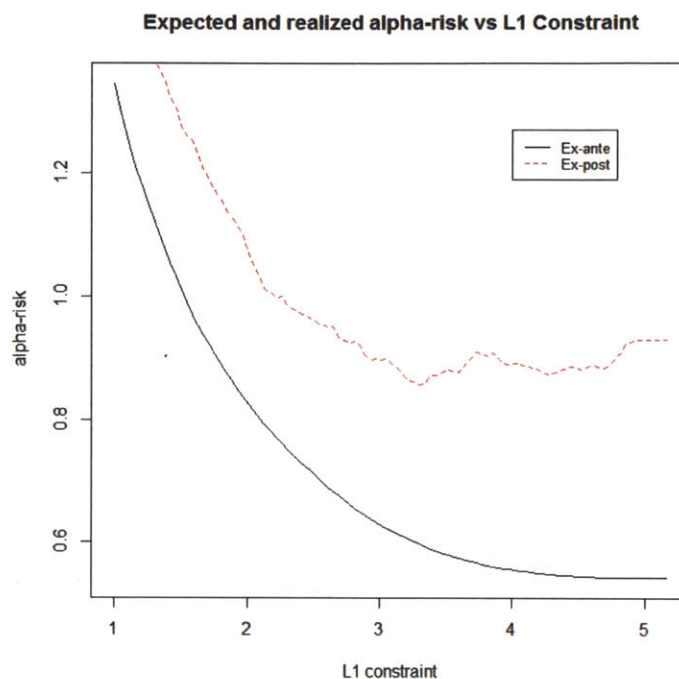


Figure 2-2: Expected and realized  $\alpha$ -risk as a function of the constraint on  $\|w\|_1$ .

Figure 2-2 shows how the expected  $\alpha$ -risk decreases monotonically as the constraint is relaxed. Further, the concavity of the expected  $\alpha$ -risk as a function of the constraint tells us that most of the increase in the expected risk occurs close to the maximum level of constraint, i.e. at  $c = 1$ . Moreover, the out-of-sample performance of the portfolios shows that under this stylized simulation, the empirical risk would follow a very similar pattern to its expected counterpart. Further, the out-of-sample alpha-risk is practically flat for  $c$  greater than 3, implying that there would be very small gains from relaxing the constraint beyond this point.

Meanwhile, Figure 2-3 shows the out-of-sample distribution of returns for 5 portfolios with  $\ell_1$  restrictions ranging from 1 (the no short-sell case) to 5 (equivalent to no restriction). The distributions become concentrated close to the mode as the restriction is relaxed. The red density corresponds to the  $c = 1$  case and we can clearly observe that there is much more mass in the tail of the distribution than in any of the other cases. This is once more an indication that the adverse effects of imposing the  $\ell_1$  restriction are intensified close to the no-short selling case.

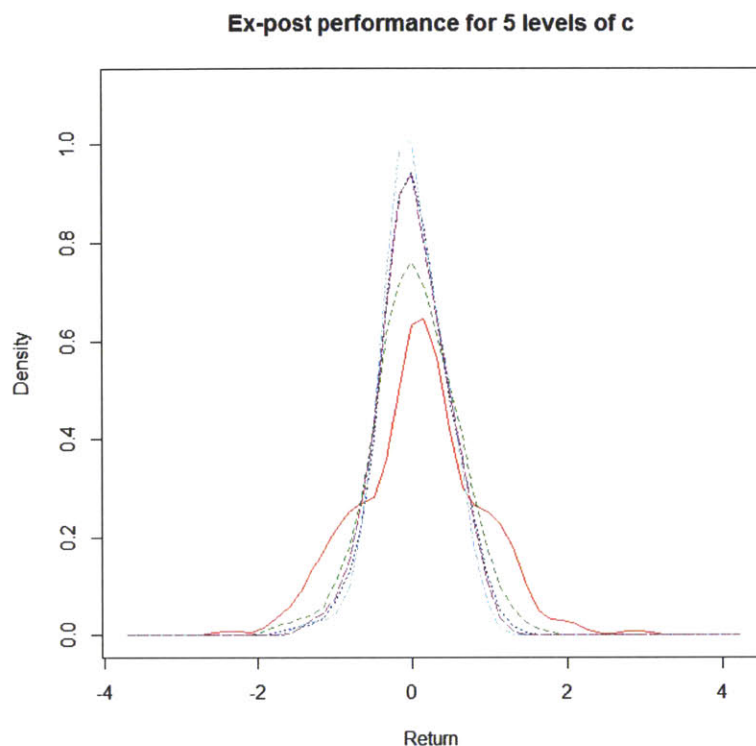


Figure 2-3: Out-of-sample distribution for 5 portfolios with different  $\ell_1$  restrictions.

Next, we move to the dual problem, where we select a portfolio that maximizes expected return subject to an  $\alpha$ -risk restriction. I consider two cases, in the first one I find the minimum  $\alpha$ -risk achievable while restricting the  $\ell_1$  norm to be 1 (we denote this value  $\alpha_0$ ). I then relax the restriction on the weights, while maintaining the restriction on the risk parameter. In the second scenario we find the expected  $\alpha$ -risk of the equally weighted portfolio (i.e. the portfolio that sets  $w_i = 1/p$  for all  $i$ ) and denote this value  $\alpha_E$ . Then I construct portfolios with this expected  $\alpha$ -risk with the  $\ell_1$  norm restriction ranging from 1 to the unrestricted case.

Figure 2-4 shows that, just like in the dual problem, the number of active assets does not increase monotonously as the restriction on the weights is relaxed. Moreover, as we can see in Figure 2-5, expected return is a concave function of the constraint on the weights. This shows, once more, that the effects of adding the constraint will be greater as we approach the no-short-selling level. Further, as we would expect, Case 2 exhibits larger expected returns and if the weights are left

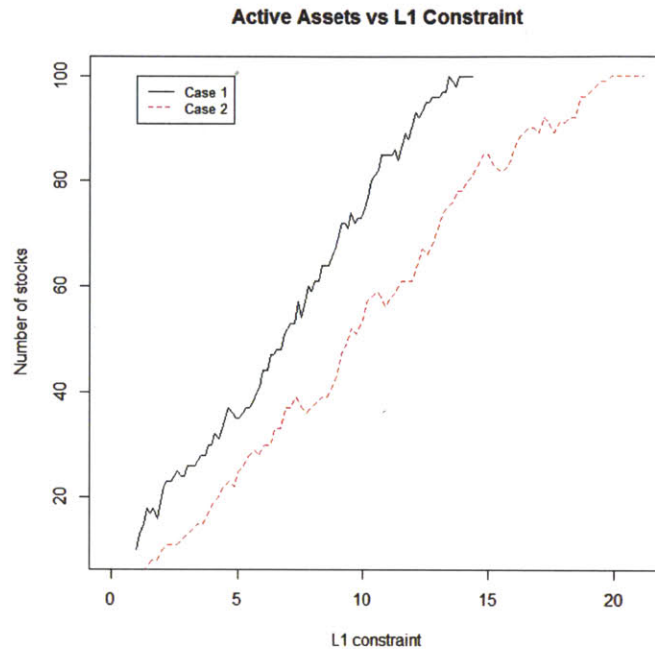


Figure 2-4: Active assets as a function of the  $\ell_1$  constraint when  $\alpha$ -risk is restricted to  $\alpha_0$  (Case 1) or  $\alpha_E$  (Case 2).

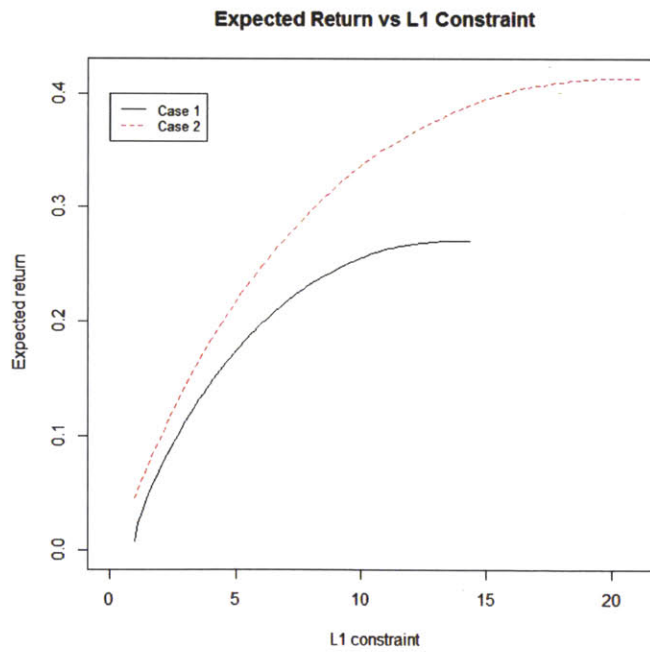


Figure 2-5: Expected return as a function of  $\|w\|_1$ .

unrestricted, it will lead to much more extreme positions.

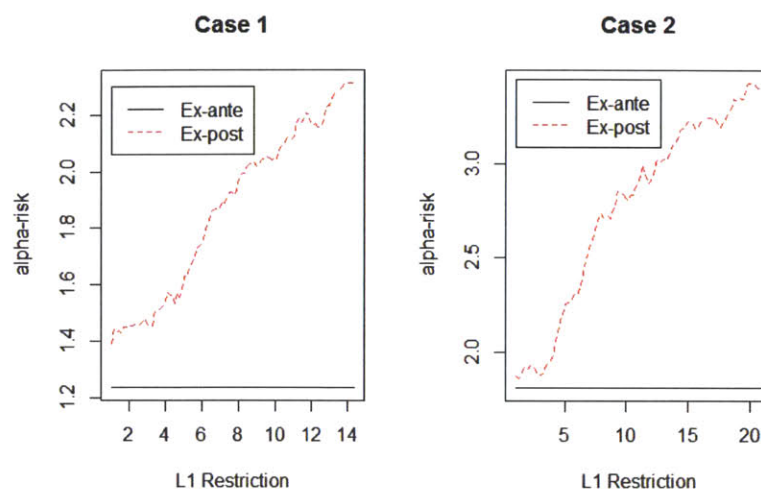


Figure 2-6: Ex-ante and ex-post  $\alpha$ -risk for both levels of risk restrictions.

Finally, Figure 2-6 shows that in both cases the realized empirical risk is much larger than its expected counterpart. Further, we can observe that this problem becomes more accentuated as the  $\ell_1$  restriction is relaxed.

### 2.3.3 Fama-French 100 Portfolios

In order to extend the analysis to real data I use the set 100 industrial portfolios from the website of Kenneth French. These portfolios are formed by sorting the stocks in the Center for Research in Security Prices (CRSP) database in two dimensions. Stocks are sorted according to market equity and according to the ratio of book equity to market equity. Ten categories are constructed for each variable, yielding the 100 portfolios.

#### Non Stress Scenario

For the baseline, non-stress case, I take a sample period of 1000 trading days, from March 2 1988 to February 12 1992, and an out-of sample period of 250 trading days, from February 13 1992 to February 11 1993.

I begin by constructing portfolios minimizing expected  $\alpha$ -risk subject to an ex-

pected return. In this exercise the number of active assets as a function of the  $\ell_1$  constraint behaves in a similar manner to the simulated data case.

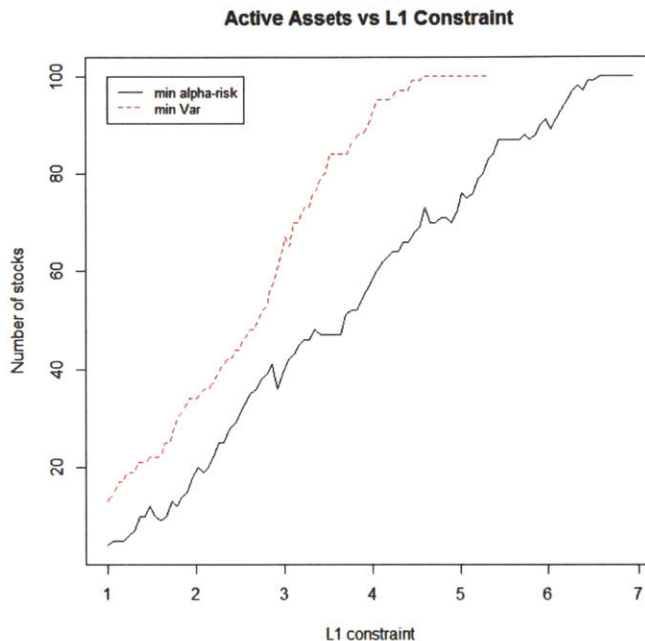


Figure 2-7: Active assets vs  $\ell_1$  constraint for the non-stress scenario.

Figure 2-7 shows that the number of active assets does not grow monotonously as the  $\ell_1$  constraint is relaxed. The same figure shows, in contrast, that when variance is minimized we get the familiar result of the number of assets growing monotonously as the  $\ell_1$  constraint is relaxed.

Meanwhile, Figure 2-8 shows that expected  $\alpha$ -risk is, once more, a convex function of the  $\ell_1$  constraint, implying that expected risk increases rapidly as we approach an  $\ell_1$  restriction of 1. Further, the fact that the realized  $\alpha$ -risk follows closely its expected counterpart, tells us that under non-stress situations expected  $\alpha$ -risk might be a good predictor of realized  $\alpha$ -risk. In particular when the weights are restricted so  $\|w\|_1$  is smaller than 4 (for this particular exercise). Moreover, the fact that ex-post risk diverges from its expected counterpart for larger values of  $\|w\|_1$  is not surprising, since taking extremely long or short positions makes portfolios much more sensitive to deviations from the distribution implied in the portfolio selection procedure.

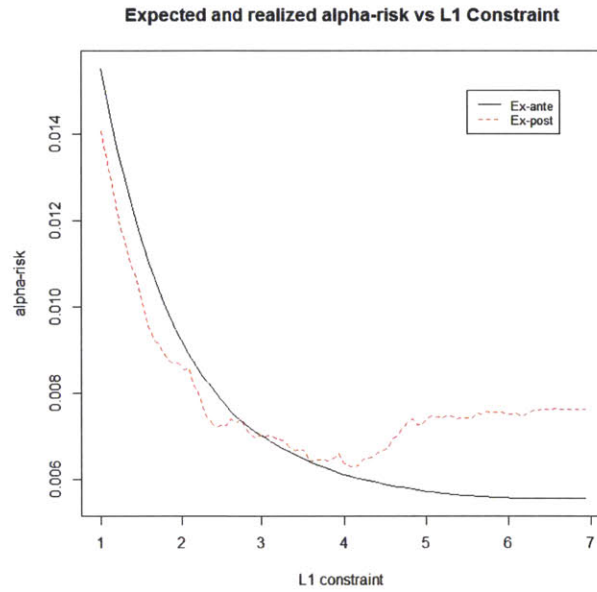


Figure 2-8: Ex ante and ex-post  $\alpha$ -risk for the non-stress scenario.

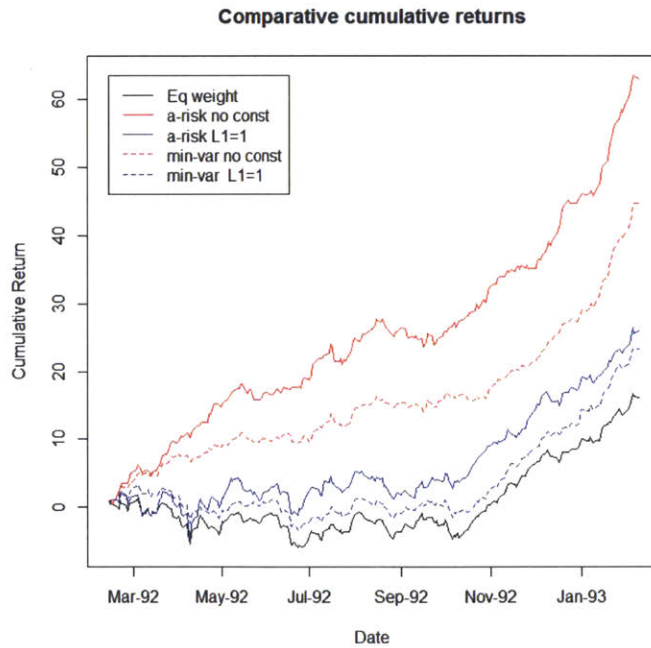


Figure 2-9: Ex-post cumulative returns for 5 portfolios: equally weighted, minimizing alpha risk (unrestricted and with no-short-sale restriction) and minimizing variance (unrestricted and with no-short-sale restriction).

Further, in Figure 2-9 we can see that the portfolio chosen by minimizing expected  $\alpha$ -risk with no restriction on  $\|w\|_1$  would have outperformed the equivalent portfolio with  $\|w\|_1$  restricted to 1, and the portfolio with the same expected return restriction but with variance as the objective function. In fact, all portfolios constructed by minimizing a risk measure, with or without restrictions on  $\|w\|_1$  would have outperformed the "naive" portfolio constructed by setting  $w_i = 1/p$  for all  $i$ .

Next, I consider the dual problem, that is, maximizing expected return subject to an expected  $\alpha$ -risk. As with the simulated data, I consider two cases, one the minimum  $\alpha$ -risk achievable subject to a no-short-sale restriction ( $\alpha_0$ ) and second the  $\alpha$ -risk of the equally weighted portfolio ( $\alpha_E$ ).

When we compare the expected and realized return under these scenarios we find that when  $\alpha$ -risk is most constrained (Case 1) the realized return is relatively flat as a function of  $\|w\|_1$ . Meanwhile, the second panel of Figure 2-10 shows that, for Case 2 (a less stringent restriction on expected  $\alpha$ -risk), realized mean return did rise as the  $\ell_1$  constraint was relaxed and was therefore much closer to its expected counterpart. Since portfolios are constructed by setting a restriction on the expected value of the  $\alpha$ -risk, I would expect that under this non-stress scenario the realized level of  $\alpha$ -risk would be relatively constant as the restriction on  $\ell_1$  is relaxed. Nonetheless, Figure 2-11 shows how, for both cases, the realized level of  $\alpha$ -risk increases as the restriction on the weights is relaxed. This indicates that expected  $\alpha$ -risk becomes a very weak predictor of the realized  $\alpha$ -risk when extremely long or short positions are allowed.

### **Stress Scenarios**

Finally, I consider two stress scenarios. As in the baseline scenario, I take a 1000 day sample to construct the portfolios and an out-of-sample period of 250 days to test the ex-post performance of the portfolios. The first case has an in-sample data range from March 2004 to March 2008 and an out-of-sample range from March 2008 to March 2009. For the second stress scenario the sample period comprises May 1983 - April 1987 and the out-of sample period begins in April 1987 and ends in April 1988. These date ranges are chosen so that extreme stress dates occur in the middle



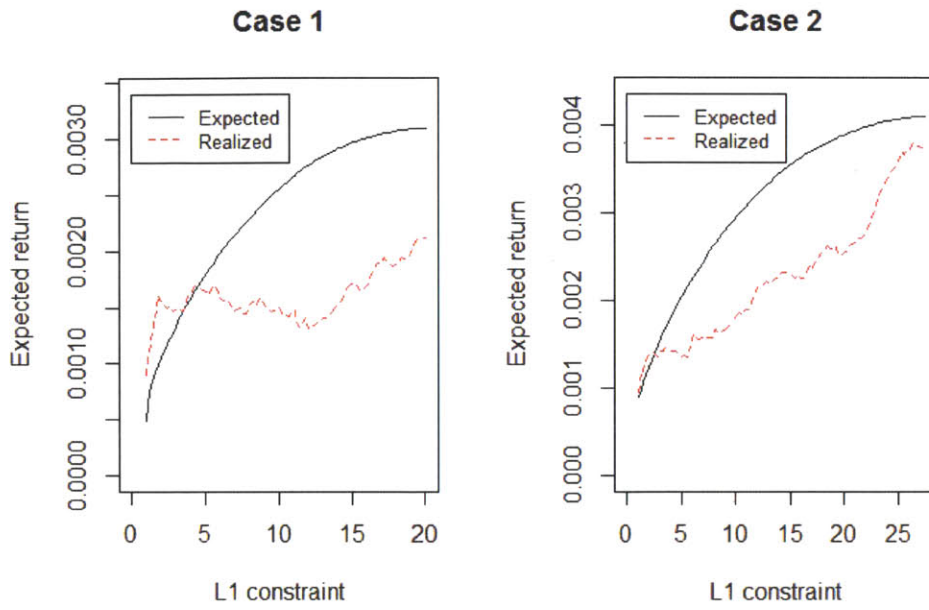


Figure 2-10: Expected and realized return under the two levels of risk restriction.

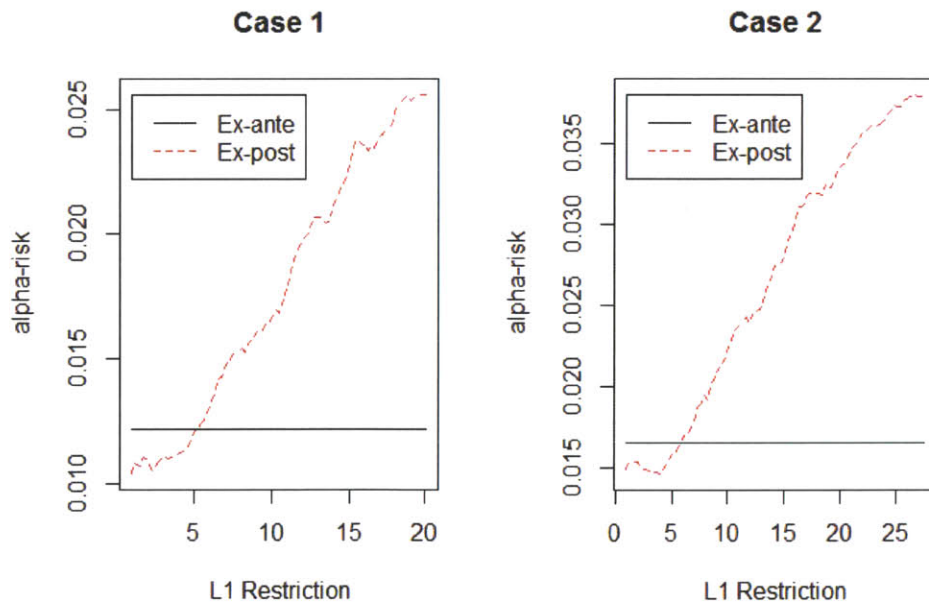


Figure 2-11: Expected and realized  $\alpha$ -risk under the two levels of risk restriction.

of the out-of sample period. For Case 1 the midpoint of the out-of-sample period is the date of the collapse of Lehman Brothers, and for Case 2 the corresponding date is the 1987 crash.

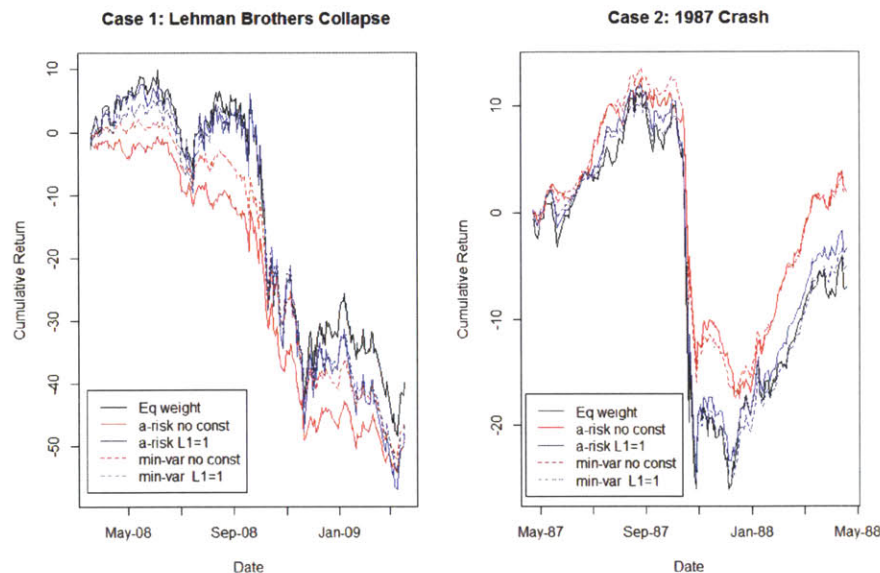


Figure 2-12: Out-of-sample cumulative returns for the stress scenarios.

Analyzing the performance of the different portfolios, I find that none of the portfolio selection procedures would have prevented significant losses from occurring during these periods. Nonetheless, in Figure 2-12 we can see that during the 1987 crash both the portfolios constructed by minimizing  $\alpha$ -risk and variance would have outperformed the naive portfolio. However, during the 2008 crisis, the equally weighted portfolio does better than any of the other portfolios being considered. This leads us to conjecture that even though both stress scenarios represent cases in which the underlying distribution of returns changed from the sample period to the out-of-sample period, the nature of the change was quite different for each of the cases. In particular, it leads us to suspect that in the 1987 case returns fell but the correlation structure of the returns remained relatively unchanged. Thus, the assets selected by the Markowitz or Choquet methods still outperform the naive portfolio. Meanwhile, in the 2008 case, it would seem that it was not only the level of the returns that changed from the sample to the out-of-sample period. Thus, the

information in the sample becomes practically useless to predict the performance in the ex-post period.

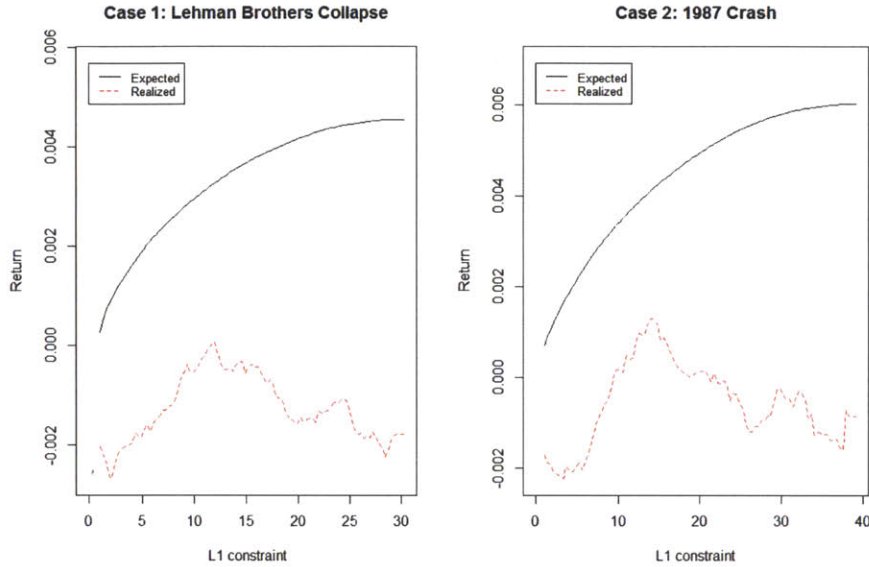


Figure 2-13: Expected and realized return in the stress scenarios.

Further, we can see in Figure 2-13 that regardless of the constraint on the  $\ell_1$  norm of the portfolio, the realized returns are considerably lower than the expected returns. Moreover, the realized returns are no longer monotonically increasing function of the  $\ell_1$  restriction. Nonetheless, we also see that, in both cases, the difference between the expected and realized returns accentuates as the restriction is relaxed beyond 12. This indicates that portfolios with extreme positions will tend to be more sensitive to extreme scenarios like the ones presented here.

Finally, we consider the dual problem, where the expected  $\alpha$ -risk is restricted and expected return is maximized. In this context, Figure 2-14 shows that, as expected, the observed  $\alpha$ -risk is larger than its expected counterpart regardless of the level of the  $\ell_1$  constraint. Further, the variation in realized  $\alpha$ -risk as the constraint is relaxed is smaller in magnitude than the difference with respect to the expected level. Thus, the effect of being in an extreme scenario is much more important than the effect that the  $\ell_1$  constraint might have.

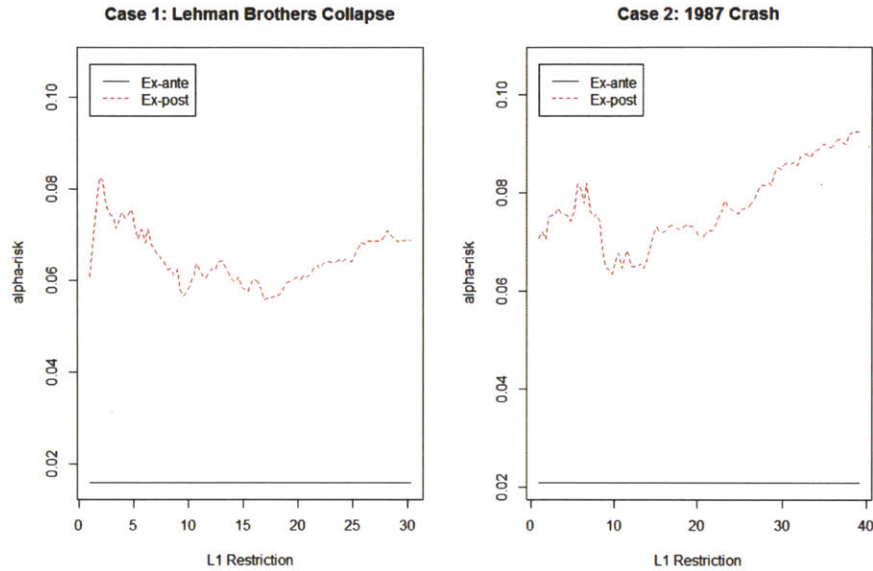


Figure 2-14: Expected and realized  $\alpha$ -risk in the stress scenarios.

## 2.4 Conclusions

I presented a generalization of the pessimistic portfolio allocation of Bassett et al. (2004) that introduces an  $\ell_1$  constraint on the vector portfolio weights. The initial objective was to minimize the exposure of the pessimistic investor to fees deriving from the number of active assets in her portfolio (for a small investor) or from the magnitude of the individual positions (for both small and large investors).

In contrast with the work on  $\ell_1$ -constrained Markowitz portfolios, I find that the amount of active assets in Choquet portfolios does not increase in a strictly monotonous fashion when the  $\ell_1$  constraint is relaxed. This will not be an issue in the context of a large investor whose fees depend essentially on the magnitude of the positions. However, if we consider a small investor who is looking to invest in at most  $\bar{p}$  assets, she will have to find all the levels of  $\ell_1$  restriction that imply an optimal portfolio with  $\bar{p}$  active assets.

My analysis also showed that loss in expected return or gain in expected risk tends to be accentuated when the  $\ell_1$  constraint approaches the no-short-selling constraint. Thus, indicating that we can get a significant portion of the advantage of having a constrained portfolio without losing much of the expected performance

as long as we do not get too close to the limit case where  $\|w\|_1 = 1$ .

Finally, the results also tend to indicate that when portfolios exhibit extreme positions ( $\|w\|_1$  large), expected  $\alpha$ -risk tends to be a poor predictor of future realizations of  $\alpha$ -risk. Therefore, beyond its application as a way to control the amount of fees an investor must pay, we now have indications that imposing the  $\ell_1$  restriction will lead to having a risk measure that is more reliable in the sense that the ex-post performance of the portfolio will be consistent with the expectations.

Therefore, I have generalized Bassett et al. (2004) portfolio selection procedure in a way that it can control the fees the investor might be exposed to. Beyond this, I have also found that restricted portfolios allow for a more reliable estimation of risk than their unrestricted counterparts.

## 2.5 Appendix B. Portfolio Selection as Regression

### 2.5.1 Markowitz Portfolios

In the classical portfolio selection theory (Markowitz - 1951), the agent wishes to allocate her wealth across the different assets in order to achieve an expected level of return  $\mu_0$ , while minimizing the variance of the portfolio

If we let  $\mathbf{R}$  denote the vector of returns on a set of  $p$  assets, let  $\bar{r}$  denote the vector of mean returns and  $\Sigma$  the variance-covariance matrix of the returns

Thus, the problem the agent must solve is

$$\min_{w \in R^n} w' \Sigma w$$

subject to:

$$w' \bar{r} = \mu_0,$$

$$w' \mathbf{1} = 1.$$

A way to find the optimal portfolio without the need of quadratic programming tools is by restating the optimization problem as a linear regression.

Note that the variance of the returns of a portfolio can be expressed as

$$\begin{aligned} Var(w' \mathbf{r}) &= \min_b E(w' \mathbf{r} - b)^2 \\ &= \min_b E(w_p r_p + w_1 r_1 + \dots + w_{p-1} r_{p-1} - b)^2 \end{aligned}$$

and since  $\sum_i w_i = 1$  we can express

$$w_p r_p = \left(1 - \sum_{i=1}^{p-1} w_i\right) r_p = r_p - \sum_{i=1}^{p-1} w_i r_p$$

so, we then have

$$\begin{aligned} Var(w' \mathbf{r}) &= \min_b E(r_p - w_1(r_p - r_1) - \dots - w_{p-1}(r_p - r_{p-1}) - b)^2 \\ &= \min_b E(Y - w_1 X_1 - \dots - w_{p-1} X_{p-1} - b)^2 \end{aligned}$$

where,

$$\begin{aligned} Y &= r_p, \\ X_j &= r_p - r_j, j \in \{1, \dots, p-1\}. \end{aligned}$$

Then finding the absolute minimum variance portfolio is simply a matter of finding the regression coefficient  $\beta = (w^*, b)$  where  $w^* = (w_1 \dots w_{p-1})$ . Now, to get back the original portfolio selection problem we need to add the restriction  $w' \bar{r} = \mu_0$ . To do so we add the pseudo-observation:

$$\begin{aligned} Y_{n+1} &= \kappa (\bar{r}_p - \mu_0) \\ X_{n+1} &= \kappa (0, \bar{r}_p - \bar{r}_1, \dots, \bar{r}_p - \bar{r}_{p-1}) \end{aligned}$$

so we have

$$Y = \begin{bmatrix} r_{p1} \\ \vdots \\ r_{pT} \\ \kappa (\bar{r}_p - \mu_0) \end{bmatrix}, \quad X = \begin{bmatrix} 1 & r_{p,1} - r_{1,1} & \cdots & r_{p,1} - r_{p-1,1} \\ \vdots & \vdots & & \vdots \\ 1 & r_{p,T} - r_{1,T} & \cdots & r_{p,T} - r_{p-1,T} \\ 0 & \kappa (\bar{r}_p - \bar{r}_1) & \cdots & \kappa (\bar{r}_p - \bar{r}_{p-1}) \end{bmatrix}$$

and this will guarantee that the objective return  $\mu_0$  is achieved (as long as  $\kappa$  is sufficiently large).

## 2.5.2 Portfolios that Minimize $\alpha$ -risk

For a random sample  $\{r_i\}_{i=1}^n$ , Bassett et. al. define the empirical analogue of  $\alpha$ -risk as

$$\hat{\rho}_{v_\alpha} = (n\alpha)^{-1} \min_{\xi \in \mathbb{R}} \sum_{i=1}^n q_\alpha(r_i - \xi) - \hat{\mu}_n$$

where  $\hat{\mu}_n$  is an estimator of  $Er$  and  $q_\alpha(\cdot)$  is the quantile loss function. Thus, for a  $p$  asset portfolio with weights  $w = \{w_i\}_{i=1}^p$  we have that the  $\alpha$ -risk will be

$$\hat{\rho}_{v_\alpha}(P_w) = (n\alpha)^{-1} \min_{\xi \in \mathbb{R}} \sum_{i=1}^n q_\alpha \left( \sum_{j=1}^p w_j r_{ij} - \xi \right) - \hat{\mu}_n \quad (2.8)$$

Now, if we note that  $\sum_{j=1}^p w_j = 1$  we have

$$w_1 = 1 - \sum_{j=2}^p w_j \quad (2.9)$$

so we can re-write (2.8) as

$$\begin{aligned} \hat{Q}_{v_a}(P_w) &= (n\alpha)^{-1} \min_{\xi \in \mathbb{R}} \sum_{i=1}^n q_\alpha \left( \left( 1 - \sum_{j=2}^p w_j \right) r_{i1} + \sum_{j=2}^p w_j r_{ij} - \xi \right) - \hat{\mu}_n \\ &= (n\alpha)^{-1} \min_{\xi \in \mathbb{R}} \sum_{i=1}^n q_\alpha \left( r_{i1} - \sum_{j=2}^p w_j (r_{i1} - r_{ij}) - \xi \right) - \hat{\mu}_n \end{aligned}$$

Thus, finding the portfolio that minimizes  $\alpha$ -risk subject to an expected return  $\mu_0$  can be expressed as

$$\min_{(\xi, w) \in \mathbb{R}^{p+1}} \sum_{i=1}^n q_\alpha \left( r_{i1} - \sum_{j=2}^p w_j (r_{i1} - r_{ij}) - \xi \right)$$

subject to:

$$\begin{aligned} r^T w &= \mu_0, \\ 1 - \sum_{j=2}^p w_j &= w_1. \end{aligned}$$

This problem has almost the form of a quantile regression optimization problem. In order to include the mean return constraint, we add a term  $\lambda(r_1 - \mu_0)$  to the response vector and a vector  $\lambda(0, r_1 - r_2, \dots, r_1 - r_p)$  to the  $r$  matrix (the 0 corresponds to the column of 1s associated with the constant term).

To set this up as a linear programming problem we define variables  $u_i$   $v_i$  as the positive and negative parts of the residual for the  $i$ 'th observation and let  $\gamma = (w^{+T}, w^{-T}, u^T, v^T)^T$  and  $c = (\mathbf{0}_{2p}, \alpha \mathbf{1}_{n+1}, (1 - \alpha) \mathbf{1}_{n+1})$  Finally, let  $A = [r; -r; I_{n+1}; -I_{n+1}]$ .



Thus, finding the portfolio becomes a problem of the form:

$$\begin{aligned} \min_{\gamma} c^T \gamma \\ A\gamma &= Y \\ \gamma &\geq 0 \end{aligned} \tag{2.10}$$

Finally, since many linear programming solvers only take inequality constraints, we divide (2.10) into two inequality constraints and so we get

$$\min_{\gamma} c^T \gamma$$

subject to:

$$\begin{aligned} A\gamma &\leq Y, \\ -A\gamma &\leq -Y, \\ \gamma &\geq 0. \end{aligned}$$

### 2.5.3 Portfolios that Maximize Expected Return with an $\alpha$ -risk Restriction

We wish to solve

$$\min_w -w' \bar{r} \text{ s.t. } \hat{\varrho}_{v_\alpha}(w'R) \leq K, w \in \mathbf{W} \tag{2.11}$$

where  $w \in \mathbf{W} \subset \mathbb{R}^p$  iff  $\sum_{i=1}^p w_i = 1$ . First, let

$$G_\alpha(X, \tilde{\xi}) = E\left(q_\alpha(X - \tilde{\xi})\right) - E[X].$$

Then, using result (2.1) from the main text, we know that

$$\varrho_\alpha(X) = G_\alpha(X, \xi_\alpha)$$

Based on the procedure in Krokhmal et. al. I will show that if  $(w^*, \xi^*)$  form a solution to

$$\min_{w, \xi} -w' \bar{r} \text{ s.t. } G_\alpha(w'R, \tilde{\xi}) \leq K, w \in \mathbf{W} \quad (2.12)$$

then  $w^*$  will also be a solution to (2.11).

First, note that the Karush–Kuhn–Tucker conditions for the optimization problem in (2.12) imply

$$-w^{*'} \bar{r} + \lambda G_\alpha(w^{*'}R, \xi^*) \leq -w' \bar{r} + \lambda G_\alpha(w'R, \tilde{\xi}) \quad (2.13)$$

and

$$\lambda (G_\alpha(w^{*'}R, \xi^*) - K) = 0, \lambda \geq 0, w \in \mathbf{W}.$$

Therefore, if  $w^*$  and  $\xi^*$  form a solution of (2.12). Then (2.13) implies that for any  $\xi \in \mathbb{R}$ ,

$$G_\alpha(w^{*'}R, \xi^*) \leq G_\alpha(w^{*'}R, \xi),$$

but we know from the main text, that  $\arg \min_{\xi} G_\alpha(w^{*'}R, \xi) = \xi_\alpha$  so,

$$G_\alpha(w^{*'}R, \xi^*) = \varrho_\alpha(w^{*'}R).$$

Then for every  $w \in \mathbf{W}$  and  $\xi \in \mathbb{R}$ ,

$$\begin{aligned} -w^{*'} \bar{r} + \lambda \varrho_\alpha(w^{*'}R) &= -w^{*'} \bar{r} + G_\alpha(w^{*'}R, \xi^*) \\ &\leq -w' \bar{r} + \lambda G_\alpha(w'R, \tilde{\xi}^*) \\ &= -w' \bar{r} + \lambda \varrho_\alpha(w'R) \end{aligned}$$

and since

$$\lambda (G_\alpha(w^{*'}R, \xi^*) - K) = \lambda (\varrho_\alpha(w^{*'}R) - K) = 0$$

and

$$\lambda \geq 0, w^* \in \mathbf{W}$$

we have that if  $w^*$  and  $\xi^*$  form a solution for (2.12) then  $w^*$  will also be a solution

to (2.11). So now all that rests is setting up the optimization problem (2.12) as a linear programming problem. To do so, we once more use the empirical equivalent of  $\varrho_\alpha(w^*R)$ . Hence, we have the following linear programming problem:

$$\min_{w, \xi} -w'\bar{r}$$

subject to:

$$(n\alpha)^{-1} \sum_{i=1}^n q_\alpha \left( r_{i1} - \sum_{j=2}^p w_j (r_{i1} - r_{ij}) - \xi \right) - \bar{r}'w \leq K.$$

## 2.6 Appendix C. Adding the $l_1$ Restriction

We wish to add the following restriction:

$$\|w\|_1 \leq k \tag{2.14}$$

for some  $k \geq 1$  (the case  $k = 1$  is equivalent to the no-short sale restriction). Now,

$$\begin{aligned} \|w\|_1 &= \sum_{i=1}^p |w_i| \\ &= |w_1| + \sum_{i=2}^p (w_i^+ + w_i^-) \end{aligned} \tag{2.15}$$

and given that portfolio weights must add to 1, we have

$$w_1 = 1 - \sum_{i=2}^p (w_i^+ - w_i^-). \tag{2.16}$$

Then if  $w_1 \geq 0$  (2.15) and (2.16) imply that

$$\begin{aligned} \|w\|_1 &= 1 - \sum_{i=2}^p (w_i^+ - w_i^-) + \sum_{i=2}^p (w_i^+ + w_i^-) \\ &= 1 + 2 \sum_{i=2}^p w_i^- \end{aligned}$$

and if  $w_1 < 0$  (2.15) and (2.16) imply that

$$\begin{aligned}\|w\|_1 &= -1 + \sum_{i=2}^p (w_i^+ - w_i^-) + \sum_{i=2}^p (w_i^+ + w_i^-) \\ &= 2 \sum_{i=2}^p w_i^+ - 1.\end{aligned}$$

Then in order to add condition (2.14), we add the following pair of restrictions to the linear programming problem:

$$1 + 2 \sum_{i=2}^p w_i^- \leq k \quad (2.17)$$

$$2 \sum_{i=2}^p w_i^+ - 1 \leq k \quad (2.18)$$

However, note that if for the optimal  $w$ ,  $w_1 \geq 0$  it is sufficient for (2.17) to hold and if  $w_1 < 0$  it is sufficient for (2.18) to hold.

What we need to check now, is that if  $w_1 \geq 0$  then (2.18) will hold (i.e. that the irrelevant constraint will not prevent the LP algorithm from finding the optimal  $w$ ) and that the converse is true if  $w_1 < 0$ .

Let us assume there is a case where the "true" optimal  $w$  has  $w_1 \geq 0$  but it violates (2.18). Then, if (2.17) is satisfied (since it is the solution to the optimization problem) and (2.18) is violated,

$$1 + 2 \sum_{i=2}^p w_i^- \leq k < 2 \sum_{i=2}^p w_i^+ - 1$$

$$\begin{aligned}\Rightarrow 2 &< 2 \sum_{i=2}^p (w_i^+ - w_i^-) \\ \Rightarrow 1 - \sum_{i=2}^p (w_i^+ - w_i^-) &< 0,\end{aligned}$$

but,

$$1 - \sum_{i=2}^p (w_i^+ - w_i^-) = 1 - \sum_{i=2}^p (w_i) = w_1$$

which we assumed as non negative. Thus, if the optimal has  $w_1 \geq 0$  (2.18) will be satisfied immediately. The proof of the converse is identical, only changing the direction of the inequalities.

# Bibliography

Anscombe, F. J. and R. J. Aumann. (1963): "A Definition of Subjective Probability," *The Annals of Mathematical Statistics*, Vol. 34, 199-205.

Artzner, P. F. Delbaen, J. M. Eber and D. Heath (1999): "Coherent Measures of Risk," *Mathematical Finance*, Vol. 9, 203-228.

Bassett G. W., Koenker R. and Kordas G. (2004): "Pessimistic Portfolio Allocation and Choquet Expected Utility," *Journal of Financial Econometrics*, Vol. 2, 477-492.

Belloni, A. and V. Chernozhukov (2011): " $\ell_1$ -Penalized Quantile Regression in High-Dimensional Sparse Models," *The Annals of Statistics*, Vol. 39, 82-130.

Bernoulli, D. (1738): "Specimen Theoriae Novae de Mensura Sortis," *Commentarii Academiae Scientiarum Imperialis Petropolitanae* 5, 175–192 [translated by L. Sommer in *Econometrica* 22, 23–36].

Brodie J., I. Daubechies, C. De Mol, D. Giannone and I. Loris (2009): "Sparse and Stable Markowitz Portfolios," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 106, 12267-12272.

Camerer C. F. (1995): "Individual Decision Making," In *Handbook of Experimental Economics*, edited by Roth, A. E. and J Kagel. Princeton University Press.

Fama, E. and French, K. (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, Vol. 33, 3-56.

Fan, J., Y. Fan and J. Lv (2008): "Large Dimensional Covariance Matrix Estimation via a Factor Model," *Journal of Econometrics*, Vol. 147, 186-197.

- Fan, J., J. Zhang and K. Yu (2009): "Asset Allocation and Risk Assessment with Gross Exposure Constraints for Vast Portfolios," Working paper Princeton University.
- Kahneman D. and Tversky A. (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, Vol. 47, 261-292.
- Krokhmal. P., J. Palmquist, and S. Uryasev (2002): "Portfolio Optimization with Conditional Value-At-Risk Objective and Constraints," *The Journal of Risk*, Vol. 4, 11-27.
- Markowitz, H. M. (1952): "Portfolio Selection," *Journal of Finance*, Vol. 7, 77-91.
- Markowitz, H. M. (1959): *Portfolio Selection: Efficient Diversification of Investments*. John Wiley & Sons, New York.
- Quiggin J. (1982): "A Theory of Anticipated Utility," *Journal of Economic Behavior & Organization*, Vol. 3, 323-343.
- Savage, L. (1954): *Foundations of Statistics*. Wiley.
- Schmeidler, D. (1989): "Subjective Probability and Expected Utility without Additivity," *Econometrica*, Vol. 57, 571-87.
- Tibshirani, R. (1996): "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 58, 267-288.
- Tversky, A. and D. Kahneman (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, Vol. 5, 297-323.
- von Neumann, J. and O. Morgenstern (1944): *Theory of Games and Economic Behavior*. Princeton.
- Wakker, P. (1989): "Continuous Subjective Expected Utility with Non-additive Probabilities," *Journal of Mathematical Economics*, Vol. 18, 1-27.
- Wakker, P. and A. Tversky (1993): "An Axiomatization of Cumulative Prospect Theory," *Journal of Risk and Uncertainty*, Vol. 7, 147-75.

Yaari, M.E. (1987): "The Dual Theory of Choice Under Risk," *Econometrica*, Vol. 55, 95-115.



## Chapter 3

# Extended Pareto Law as a Parsimonious Alternative for Operational Risk Modeling

Joint with Victor Chernozhukov

### 3.1 Introduction

Banks and other major financial institutions are exposed to risks derived from failed internal processes, described as operational risk (Basel Committee on Banking Supervision (2008)). These risks include those derived from frauds and legal exposure and can therefore represent substantial losses for the institutions that suffer them. In fact, during the 1990s and early 2000s a number of very large operational losses plagued the financial system: de Fontnouvelle et al. (2006) describe more than 100 operational loss events exceeding \$100 million over the course of a decade.

These events led to an increasing regulatory interest in operational risk. Under the Basel II capital accord there are three possible approaches to estimating capital requirements for operational losses. These are the Basic Indicator, Standardized (TSA), and Advanced Measurement (AMA) approaches. The first sets capital requirements as 15% of enterprise-level gross income. The second sets a

capital requirement for separate business lines within the bank, with each capital requirement representing between 12% and 18% of the gross income of that particular line. Finally, the Advanced Measurement Approach allows the institution to develop empirical models of potential losses and use these models to calculate their capital requirements.

Clearly, the first two approaches are much simpler to implement. However, the capital requirements resulting from applying the AMA tend to be lower. In fact, the results from the Loss Data Collection Exercise for Operational Risk (Basel Committee on Banking Supervision (2008)) show that for AMA banks the typical ratio of operational risk capital to gross income is 10.8%, significantly below that of non-AMA banks. Therefore, even in situations where regulators give banks the choice of which approach to follow, there will be strong incentives to implement the AMA.

In spite of this, there are two important issues with the implementation of the Advanced Measurement Approach. First, institutions might not have sufficient data on their historical operational losses in order to propose a robust empirical model. Second, even when, sufficient data is available, the task of setting operational loss data to an empirical model is not trivial. We will focus on the latter problem.

It will be the main purpose of this paper to present the Extended Pareto Law (EPL) as an empirical model that is both statistically sound, and leads to reasonable capital requirements for operational risks. The primary motivation for the Extended Pareto Law comes from the empirical failure of Extreme Value Theory (EVT) to give adequate approximation to the operational losses in the tails. While EVT does well at describing moderately large losses, in many instances it does a rather poor job at describing extremely large losses.

In fact, Dutta and Perry (2007) examine a number of modeling techniques used to fit operational losses including simple parametric distributions, generalized parametric distributions, extreme value theory, and historical sampling. Their findings indicate that, in general, distributions with good statistical fit tend to produce irrationally large capital requirements; with one exception. The g-and-h distribution has good goodness of fit properties and also produces reasonable capital requirements.

However, the g-and-h distribution is highly parametrized: in the most general version used in Dutta and Perry (2007) it depends on 10 parameters. Thus, it will be the purpose of this paper to present the Extended Pareto Law (EPL) as a parsimonious alternative. We will show that even though the EPL depends on only two parameters it performs exceptionally well in terms of goodness of fit and produces reasonable capital requirement estimates.

Nonetheless, in practice, EVT is one of the most used tools for modeling operational losses. EVT implies the use of the Generalized Pareto Distribution (GPD) to model the tail behavior of losses. Since the GPD is a two parameter distribution, it also represents a parsimonious alternative. In this context, we will show that in spite of its widespread use, GPD tends to have poor statistical and economic properties when used as a model for operational losses.

In order to test the EPL we use a jittered version of the The Operational Riskdata eXchange Association (ORX) dataset, which compiles operational loss data for 61 financial institutions worldwide. Using this data, first, we check goodness of fit via Kolmogorov-Smirnov and Anderson–Darling tests. We then compare PP, QQ and Hill plots for the EPL and GPD distributions to show that the EPL captures the tail behavior of operational losses much better than the GPD.

Finally, capital requirements were estimated using the GPD and EPL models and we find that the capital requirements under the EPL model tend to be smaller. Therefore, we present the EPL model as a statistically sound alternative that will also produce reasonable capital requirement estimations.

The rest of the paper is organized as follows, section 2 introduces and characterizes the Extended Pareto Law, section 3 presents the empirical findings, and section 4 concludes.

## 3.2 The Extended Pareto Law

### 3.2.1 The EPL and its Essential Properties

A random variable  $X \geq t$  follows an Extended Pareto Law (EPL) with parameters  $(\alpha, \lambda)$  if

$$F(x) := P(X < x) = 1 - t^\alpha x^{-\alpha} e^{-\lambda(x-t)_+}.$$

The EPL was introduced by Pareto (1989) along with the canonical Pareto distribution. The EPL was previously used in geophysics to model the distribution of earthquake magnitude and in finance to model asset price dynamics (e.g. Heyde et al. (1996), Kagan and Schoenberg (2001), Wu (2006)). To the best of our knowledge, we are the first to use the EPL to model the severity of operational losses.

It is interesting to emphasize that, in the unbounded support case, EPL is at least as flexible as the standard Extreme Value Theory, which describes the asymptotic distribution for exceedances over high thresholds using the Generalized Pareto Distribution (GPD), i.e.

$$P(X < x) = \begin{cases} 1 - \left(1 + \frac{\xi(x-t)_+}{\sigma}\right)^{-\frac{1}{\xi}} & , \xi \neq 0 \\ 1 - \exp\left(-\frac{(x-t)_+}{\sigma}\right) & , \xi = 0 \end{cases}.$$

Note that GPD in the unbounded support case corresponds to either the Pareto Type II (when  $\xi \neq 0$ ) or exponential law (when  $\xi = 0$ ). Indeed, recall that EVT says that if the distributions have regularly varying (approximately power) tails, then the limit distributions of exceedances is a Pareto (power) law; if the distributions have gamma varying (approximately exponential) tails, then the limit distributions of exceedances is the exponential law. Now observe that, if  $\lambda = 0$ , the EPL variable  $X \geq t$  follows a Pareto Type I law, i.e.

$$P(X < x) = 1 - t^\alpha x^{-\alpha}.$$

On the other hand, if  $\alpha = 0$ , the EPL variable  $X \geq t$  follows exponential law:

$$P(X < x) = 1 - e^{-\lambda(x-t)_+}.$$

Moreover, in the unbounded support case, EPL is strictly more flexible than GPD in that it can capture the product behavior:

$$P(X > x) = t^\alpha x^{-\alpha} e^{-\lambda(x-t)_+},$$

which GPD cannot capture non-asymptotically. Remarkably, this flexibility comes at no cost in terms of additional parameters, since EPL relies on two parameters just like GPD does.

For analytical purposes as well as for simulating the loss distributions, the following stochastic representation for EPL variable is extremely useful. Let  $\mathcal{P}$  denote a Pareto variable with the shape parameter  $\alpha$  with a lower end-point  $t > 0$ , and  $\mathcal{E}$  denote an exponential variable with mean  $\lambda$ , which is independent of  $P$ . Then, by definition,

$$P\{X > x\} = P\{\mathcal{P} > x\} \cdot P\{\mathcal{E} + t > x\} = P\{\min(\mathcal{P}, \mathcal{E} + t) > x\}.$$

Hence,

$$X =_d \min(\mathcal{P}, \mathcal{E} + t) =_d \min((t/U_1)^{1/\alpha}, -\log(U_2)/\lambda + t),$$

where  $U_1$  and  $U_2$  are two independent  $U(0, 1)$  variables.

### 3.2.2 Motivation for EPL

Dutta and Perry (2007) offer a very extensive survey of the techniques that are available for the modeling of operational losses. Among the techniques that they consider, they find that the g-and-h distribution has the best statistical properties. Nonetheless, the g-and-h distribution is highly parametrized. Hence, the first important advantage of the EPL will be that it offers a parsimonious alternative to the g-and-h distribution.

Nonetheless, there are other parsimonious alternatives. In particular, Extreme Value Theory (EVT) has become one of the most widespread tools in modeling operational losses. However, it fails to give adequate approximation to the operational losses in the tails. Indeed, we can detect this failure by the means of the Hill plot, which is the standard device used in the EVT.

**Hill Plots:** The hill plot offers a very useful way to represent the tail behavior of a heavy tailed distribution. Let us begin by defining a heavy tailed distribution following Drees et al. (2000).

**Definition 6** *By a **heavy tailed distribution** we mean a distribution  $F$ , assumed for convenience to concentrate on  $[0, \infty)$  which satisfies*

$$1 - F(x) \sim x^{-\alpha} L(x)$$

for some  $\alpha > 0$  as  $x \rightarrow \infty$ , where  $L$  is a slowly varying function satisfying

$$\lim_{\tau \rightarrow \infty} \frac{L(\tau x)}{L(\tau)} = 1, \forall x > 0.$$

**Definition 7** *The **Hill estimator** of  $\gamma := \alpha^{-1}$  based on the  $k + 1$  upper order statistics is given by*

$$H_{k,n} := \frac{1}{k} \sum_{i=1}^k \log \frac{X_{(i)}}{X_{(k+1)}},$$

where  $X_{(i)}$  represents the  $i$ 'th upper-order statistic (i.e.  $X_{(1)} \geq \dots \geq X_{(n)}$ ).

If the  $X$ s are drawn from a Pareto Distribution the Hill estimator  $H_{n-1,n}$  will be the maximum likelihood estimator for  $\alpha^{-1}$ . Further, Mason (1982) shows that if the  $X$ s are i.i.d.,  $H_{k_n,n} \xrightarrow{p} \frac{1}{\alpha}$  when  $k_n \rightarrow \infty$  as long as  $k_n/n \rightarrow 0$ . Finally, the Hill plot is constructed by plotting  $H_{k,n}$  for different values of  $k$ .

Now, in the context of operational losses, instead of fixing  $k$ , a threshold  $t$  is used to determine the start of the tail. A very stable or flat Hill plot can suggest a canonical Pareto behavior. On the other hand, a "curved" or bending Hill plot can suggest Extended Pareto behavior, accentuating the importance of the exponential

deviation from the canonical Pareto shape. Thus, a useful diagnostic for the presence of EPL versus PL is the “curved” or bending Hill plot.

To prove the latter point we present the following simple computational experiment: First, we simulate a sample from the standard Pareto distribution with parameter values obtained using real data, and then construct a Hill plot using the simulated sample. We see from Figure 3-1 that the Hill plot is very stable or flat, suggesting that the stability of a Hill plot is a good indication for PL’s applicability. Second, we draw a sample of data points from the EPL with parameter values fitted on the real data (CPBP-2), and then construct a Hill plot using the generated data. We see from Figure 3-1 that the Hill plot has a curved or bending pattern, suggesting that the curved Hill plot is a good indication for EPL’s applicability over PL.

Diagnosing Extended Pareto Laws via Hill Plots (for parameter  $1/\alpha$ )

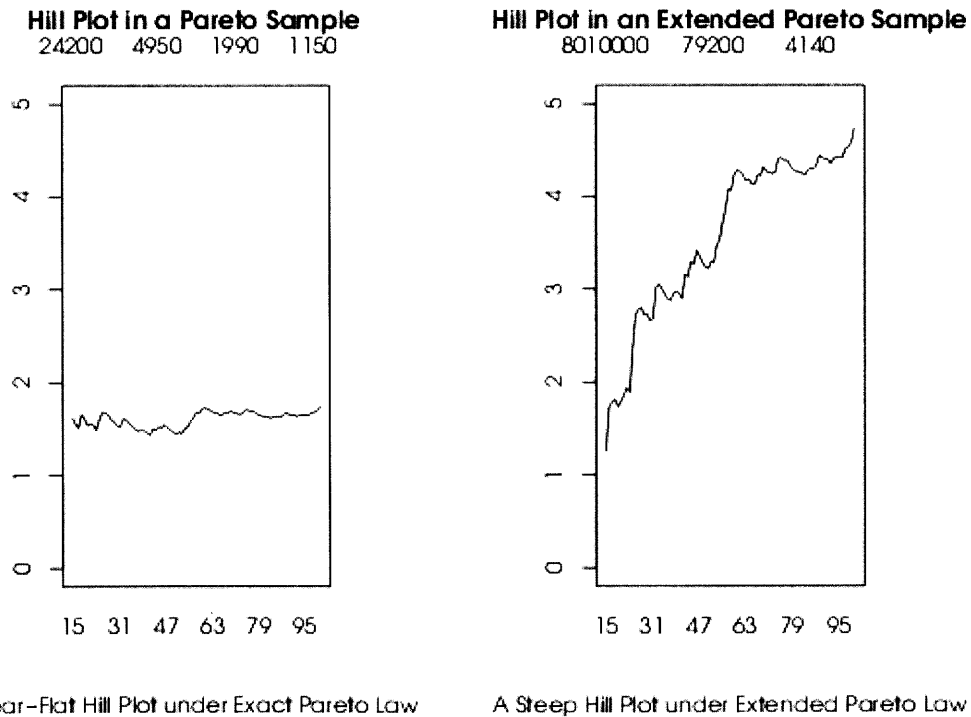


Figure 3-1: Hill plots for simulated Pareto and Extended Pareto samples. The bottom  $x$  axis denotes  $n$  (the number of upper order statistics used) and the top  $x$  axis represents  $X_{(n)}$  (the magnitude of the  $n$ 'th largest loss).

It is also useful to use qualitative judgement in assessing the probabilities of ultra-extreme losses. For instance, we may think that typical extremes may well follow a simple Pareto law, but that the ultra-extreme losses may obey a moderated version of the Pareto law. To see why this might be the case, consider a scenario where a bank is faced with a lawsuit in which vast amounts of money might be lost. In this case, one would expect the bank to defend itself against such a lawsuit with extreme rigor, making the probability of success for the plaintiffs smaller relative to less extreme lawsuits. Thus, moderation of success probabilities would lead to a deviation from the pure Pareto law, and this deviation can be modeled using the exponential function.

### 3.2.3 Formal Diagnostics and Estimation of the EPL

#### Estimation and Computation

Suppose we have an original sample of losses  $Y_1, \dots, Y_n$  and would like to fit the EPL to the losses located above the threshold  $t$ . For this purpose we create a truncated sample  $X_1, \dots, X_N$  by collecting any  $Y_i$  such that  $Y_i > t$ ; here  $N = \sum_{i=1}^n 1\{Y_i > t\}$ . Under the maintained modeling assumption, the density function of each  $X_i$  takes the form

$$f(x; \theta) = (x/t)^{-\alpha} e^{-\lambda(x-t)_+} (\alpha/x + \lambda) 1(x > t),$$

where we define

$$\theta = (\alpha, \lambda).$$

The log-density takes the form

$$\log f(x; \theta) = -\alpha \cdot \log(x/t) - \lambda(x-t)_+ + \log(\alpha/x + \lambda).$$

Under the i.i.d. sampling assumption on  $Y_1, \dots, Y_n$ , the sample  $X_1, \dots, X_N$  is also i.i.d. conditional on  $N$ , so that the log-likelihood function takes the form

$$L_N(\theta) = \sum_{i=1}^N \log f(X_i; \theta).$$



Next, we define the parameter space:

$$\Theta = \{\theta = (\alpha, \lambda) : \alpha \geq 0, \lambda \geq 0\}.$$

The maximum likelihood estimator is defined as

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L_N(\theta).$$

We also assume that the true value  $\theta_0 \neq (0, 0)$ , since the latter case would correspond to a degenerate case of  $X_i$  having a unit point mass at the threshold  $t$ . The log-likelihood function is also degenerate at  $\theta = (0, 0)$  taking value of  $-\infty$ . Therefore, the MLE  $\hat{\theta}$  can never take value  $(0, 0)$ . The MLE estimator can be computed by the standard constrained maximization algorithms. It turns out that the log-likelihood function is globally concave on  $\Theta$ , which means that the computation of MLE is an extremely tractable, efficient problem; practical experiences confirm this.

**Lemma 8** *The mapping  $\theta \mapsto \log f(x, \theta)$  is concave on  $\Theta$ , so that the mapping  $\theta \mapsto L_N(\theta)$  is also concave on  $\Theta$ .*

The first claim follows by observing that the mapping  $(\alpha, \lambda) \mapsto \log(\alpha/x + \lambda)$  is concave over  $\Theta$ , since it is a composition of a linear mapping  $(\alpha, \lambda) \mapsto \alpha/x + \lambda$ ,  $x \geq t > 0$ , whose values range over a subset of  $[0, \infty)$ , with a concave function  $u \mapsto \log(u)$  over  $[0, \infty)$ . The final mapping  $\theta \mapsto \log f(x, \theta)$  is then a sum of the mapping above and a linear mapping  $(\alpha, \lambda) \mapsto -\alpha \cdot \log(x/t) - \lambda(x-t)_+$ , and hence it is also concave. The second claim follows from the fact that the log-likelihood function is a sum of concave functions over  $\Theta$ , and hence it is also concave.

For computational purposes, it is useful to have expressions for the gradient and

the Hessian of the log-density function:

$$\frac{\partial}{\partial \theta} \log f(x, \theta) = \begin{pmatrix} -\log(x/t) + \frac{1/x}{\alpha/x+\lambda} \\ -(x-t)_+ + \frac{1}{\alpha/x+\lambda} \end{pmatrix}, \quad (3.1)$$

$$\frac{\partial^2}{\partial \theta \partial \theta'} \log f(x, \theta) = \begin{pmatrix} \frac{-1/x^2}{[\alpha/x+\lambda]^2} & \frac{-1/x}{[\alpha/x+\lambda]^2} \\ \frac{-1/x}{[\alpha/x+\lambda]^2} & \frac{-1}{[\alpha/x+\lambda]^2} \end{pmatrix}. \quad (3.2)$$

If the true parameter value  $\theta_0 = (\lambda_0, \alpha_0)$  for parameter  $\theta = (\lambda, \alpha)$  obeys  $\theta_0 > 0$ , that is, the parameter is away from the boundary of the parameter space, then the MLE is regular, namely it is root- $n$ -consistent, asymptotically normal and efficient.

**Theorem 9 (Regular Case)** *Under the maintained modeling assumptions, and if  $\theta_0 = (\lambda_0, \alpha_0)$  is in the interior of  $\Theta$ , then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(0, \mathcal{I}_{\theta_0}^{-1}),$$

where,

$$\mathcal{I}_{\theta_0} = -E_{\theta_0} \left[ \frac{\partial^2}{\partial \theta \partial \theta'} \log f(X_i, \theta_0) \right].$$

This result follows as an application of a general theoretical framework in Knight (1999) to our case.

If either  $\lambda_0 = 0$  or  $\alpha_0 = 0$ , that is, if the parameter is on the boundary, then the MLE estimates are still root- $n$ -consistent, although the limit distribution in this case is a censored normal.

**Theorem 10 (Non-regular Case)** *Under the maintained modeling assumptions, and if  $\alpha_0 = 0$  and  $\lambda_0 > 0$ , then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d \arg \min_{h \in \mathbb{R}^2: h_1 \geq 0} -Z'h + \frac{1}{2}h'\mathcal{I}_{\theta_0}h,$$

where  $Z =_d N(0, I_{\theta_0})$ . *Under the maintained modeling assumptions, and if  $\alpha_0 > 0$  and  $\lambda_0 = 0$ , then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d \arg \min_{h \in \mathbb{R}^2: h_2 \geq 0} -Z'h + \frac{1}{2}h'\mathcal{I}_{\theta_0}h.$$

The result follows as an application of a general theoretical framework in Andrews (1999) and Knight (1999) to our case.

### 3.3 Empirical Results

Our focus will be in examining the performance of the EPL model, and contrasting it with the main parsimonious alternative, namely extreme value theory (EVT). Since EVT leads to the modelling of tail behavior using the GPD model, we will present all the results both for the EPL and GPD models.

In order to do so we will use a jittered version of the ORX dataset. This is a large dataset that contains operational losses grouped in 11 units of measure (UOM). These units, and their descriptive statistics are shown in Table 3.1.

We can immediately note several things from the descriptive statistics. First, all the units of measure have minimum values that are very close to \$20,000. This simply reflects the fact that losses below that threshold are not included in the ORX dataset. This would be a problem if we were trying to characterize the entire distribution of the losses. However, since we are essentially interested in the tail behavior of the losses we do not need the information that is not reported. Next, note that the maximum loss tends to be several orders of magnitude larger than both the mean and standard deviation. This, together with the extremely large kurtosis for some of the series (in excess of 1000) confirms that the data, in fact, exhibits extremely heavy tails.

**Table 3.1:** Units of Measure and Descriptive Statistics of the Data

Unit of Measure	n	Mean \$1,000	Min. \$1,000	Max \$1,000,000	Std.Dev \$1,000	Skew	Kurtosis
Business Disruption & System Failure (BDSF-1)	1026	189	21.9	3.47	341	4.48	25.5
Business Disruption & System Failure (BDSF-2)	93	124	25.9	0.85	128	2.82	10.7
Clients, Products & Business Practice (CPBP-1)	1065	4400	21.7	557.00	28,600	13.30	207.0
Clients, Products & Business Practice (CPBP-2)	595	1880	23.9	297.00	14,000	16.80	333.0
Damage to Physical Asset (DPA-ALL)	42	267	25.8	3.95	638	4.70	23.8
Execution, Delivery & Process Mgmt. (EDPM-1)	17350	437	21.5	202.00	2,960	33.90	1710.0
Execution, Delivery & Process Mgmt. (EDPM-2)	2478	447	21.3	228.00	5,350	33.80	1350.0
External Fraud (EF-1)	1461	648	23.9	413.00	12,700	28.30	849.0
External Fraud (EF-2)	96	349	23.1	11.70	1,360	6.88	50.6
Employ. Practices & Workplace Safety (EPWS-ALL)	1020	480	24.7	38.50	1,780	13.60	240.0
Internal Fraud (IF-ALL)	216	11300	27.4	949.00	74,800	10.50	120.0

### 3.3.1 Model Validation

In order to assess the validity of both the GPD and EPL models we perform three sets of procedures.

1. Rigorously check goodness-of-fit using the formal Kolmogorov–Smirnov and Anderson–Darling tests.
2. Examine goodness-of-fit using PP and QQ plots.
3. Examine if the fitted model generates a Hill plot similar to that seen in the data.

The first set entails the use of the standard goodness-of-fit tests, which allows us to see if the fitted model provides an adequate fit from a statistical point of view. The second set entails the use of graphical diagnostics, which allow us to see if the fitted model provides an adequate fit from an economic point of view. This step is needed because, for example, if a very large sample is available, a highly parsimonious model such as the EPL can be rejected at the conventional significance level, but it can still provide an excellent fit from an economic point of view. The latter case occurs when the model's predictions deviate from the empirical quantiles by an amount that is not economically significant. The final procedure involves the application of the qualitative diagnostic based on the Hill plot, which we have already mentioned earlier. This procedure allows us to see if the fitted model can reproduce a pattern of the Hill plot seen in the data. The Hill plot is a summary of the tail behavior of the given data, and so it is important to have the model capturing this tail behavior.

However, recall that we are interested in the distribution of extreme losses i.e. the tail of the distribution. Thus, a first step in the modeling procedure would be to decide where the tail actually starts. We will not focus on this point, and instead present results that are sufficient to show that the EPL model will perform better than the GPD regardless of where the threshold is set for the beginning of the tail.

## Goodness of Fit

We begin by running KS and AD tests on the tail data setting the threshold at different levels. First, we vary the number of elements that constitute the tail data from 30 to 150. However, the datasets for the different UOMs are of very different sizes so we repeated the experiment by setting the threshold as a quantile of the data instead of a fixed order statistic. Figures 3-2 and 3-3 show the results of the KS test for two of the datasets.

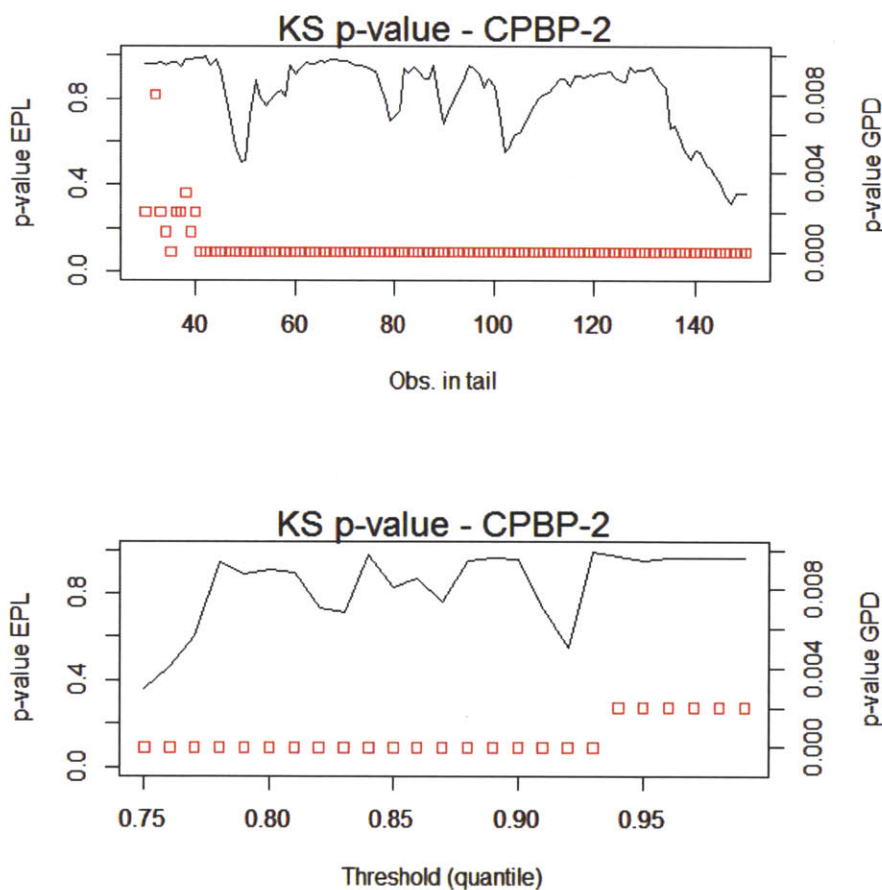


Figure 3-2:  $p$ -value of the KS test for the EPL (black line) and GPD (red dots) models as a function of the threshold level. Unit of measure: CPBP-2.

In Figure 3-2 we show the  $p$ -value for the KS test for the CPBP-2 data. This is one of the few series where the  $p$ -value of the GPD model is not zero at least for

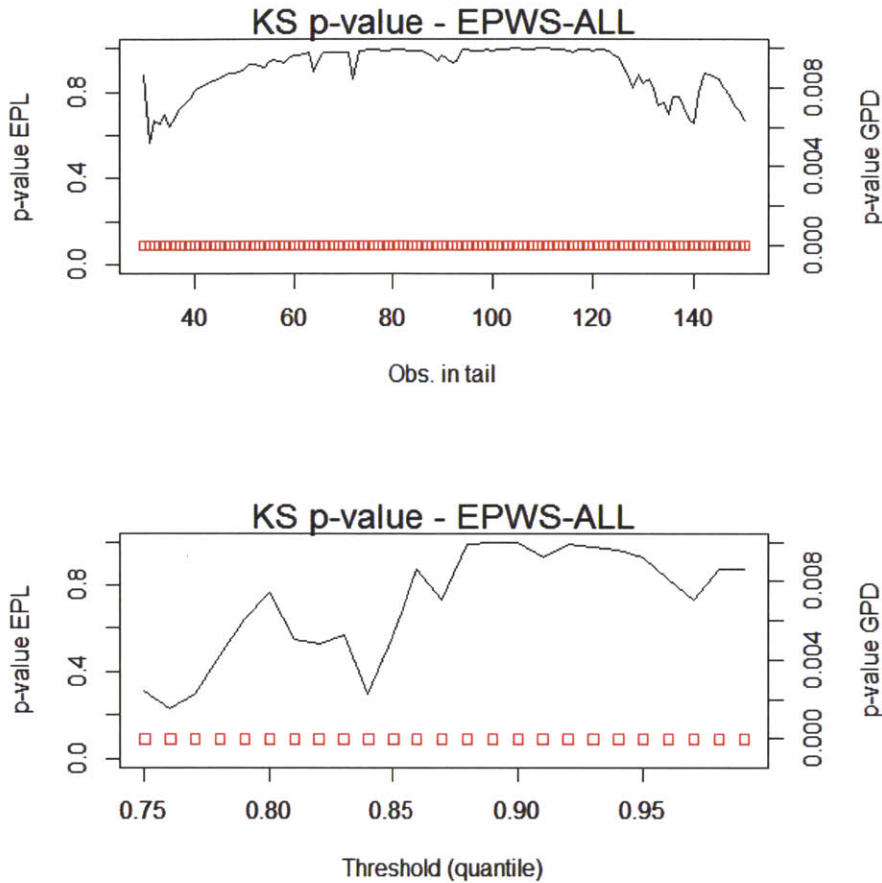


Figure 3-3:  $p$ -value of the KS test for the EPL (black line) and GPD (red dots) models as a function of the threshold level. Unit of measure: EPWS-ALL.

some threshold levels. In fact, Figure 3-3 is more representative of what we found for most series. We can see that regardless of the point at which the threshold is set, the  $p$ -value of for the KS test of the GPD model is zero. Meanwhile, for the EPL model there is a wide range of threshold levels for which we cannot reject the null hypothesis of the KS test, namely, that the data is drawn from an Extended Pareto Distribution.

Now, recall that we are not interested in describing a rule to set the threshold level (the beginning of the tail). However, we can still show that the EPL model will exhibit much better goodness of fit properties. Table 3.2 shows the maximum  $p$ -values in the AD and KS tests under the GPD and EPL models. We can see that

the GPD performs much worse than the EPL and that in fact the hypothesis that the tail data come from the GPD distribution can always be rejected regardless of the cutoff point of the tail. At the same time, we can see that there always exists a threshold level such that neither the KS nor the AD test reject the EPL model.

**Table 3.2:** Max.  $p$ -values in the KS and AD tests

Unit of Measure	GPD		EPL	
	KS	AD	KS	AD
BDSF-1	0	0	0.986	0.984
BDSF-2	0	0	0.950	0.991
CPBP-1	0	0.002	0.998	0.999
CPBP-2	0.013	0.040	0.993	0.992
DPA-ALL	0.002	0.015	0.955	0.968
EDPM-1	0	0	0.945	0.981
EDPM-2	0	0	0.976	0.958
EF-1	0.002	0.018	0.713	0.874
EF-2	0	0.001	0.999	1.000
EPWS-ALL	0	0	0.999	1.000
IF-ALL	0.013	0.062	0.994	0.998

We can make this result even stronger by fixing the threshold level at the point that maximizes the KS  $p$ -value for the GPD model and then keeping this cutoff point estimate the KS  $p$ -value for the EPL model. Since many  $p$ -values are 0 for the GPD estimations, many entries from Table 3.2 will be repeated.

**Table 3.3:**  $p$ -values with thresholds optimal for GPD

Unit of Measure	GPD		EPL	
	KS	AD	KS	AD
BDSF-1	0	0	0.986	0.984
BDSF-2	0	0	0.950	0.991
CPBP-1	0	0.002	0.998	<b>0.847</b>
CPBP-2	0.013	0.040	<b>0.963</b>	<b>0.970</b>
DPA-ALL	0.002	0.015	<b>0.955</b>	<b>0.963</b>
EDPM-1	0	0	0.945	0.981
EDPM-2	0	0	0.976	0.958
EF-1	0.002	0.018	<b>0.632</b>	<b>0.874</b>
EF-2	0	0.001	0.999	<b>0.991</b>
EPWS-ALL	0	0	0.999	1.000
IF-ALL	0.013	0.062	<b>0.883</b>	<b>0.933</b>

The bold numbers in Table 3.3 represent the values that are different from the corresponding values in Table 3.2. As expected, the  $p$ -values for the KS test of the



EPL model are smallest when the threshold is set at the point that maximizes the  $p$ -value for the GPD. Nonetheless, they remain higher than the corresponding values for the GPD model and they are never in the region that would lead to a rejection of the EPL model. Thus, even if the threshold is chosen to maximize the performance of the GPD model, the EPL model continues to perform better in terms of statistical fit.

Next, we present PP, QQ and Hill plots for the two models. Instead of presenting the plots for all the units of measure, we present only two chosen as follows. The first UOM is EPWS-ALL chosen since the KS  $p$ -value for the EPL model is one of the highest and it is achieved with more than 100 observations in the tail. The second UOM is CPBP-2 which was chosen since it was the UOM where the GPD achieved the highest KS  $p$ -value.

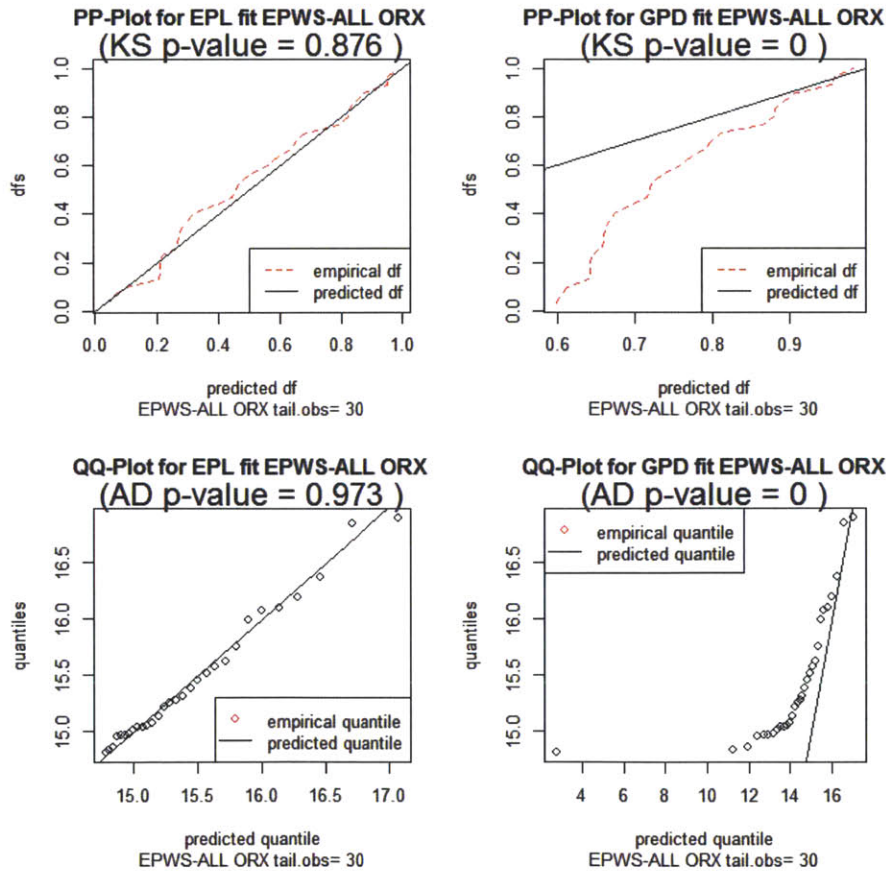


Figure 3-4: PP and QQ plots for EPL and GPD models. Unit of measure: EPWS-ALL.

Figure 3-4 confirms what the KS and AD tests had already shown, that the goodness of fit of the EPL model is far better for a unit of measurement like EPWS-ALL, where the GPD distribution performs particularly bad, while the EPL exhibits very high  $p$ -values on both tests. Further, Figure 3-5 shows the case when we choose a unit of risk where the GPD model does better and we set the threshold at the level that maximizes the KS  $p$ -value for the GPD model. Even under these conditions that are set to favor the GPD model, we find that EPL exhibits a much better fit. Finally, figures 3-6 and 3-7 show the Hill plots for these same units of measure. Each

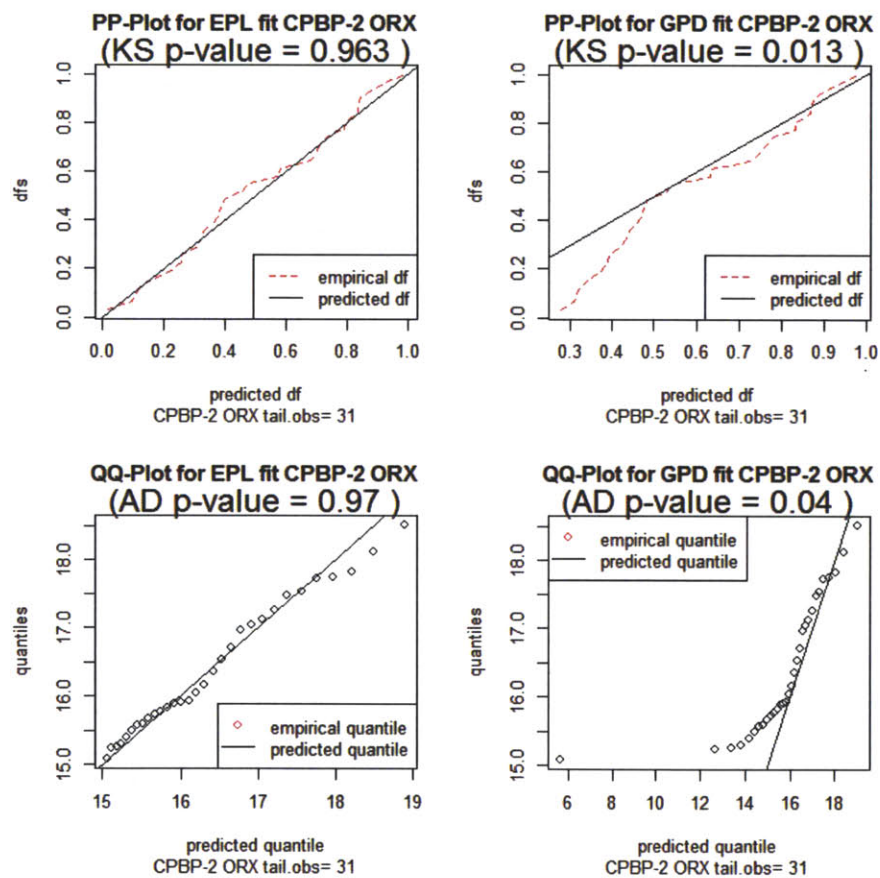


Figure 3-5: PP and QQ plots for EPL and GPD models. Unit of measure: CPBP-2.

figure has the Hill plot of the data, and of data simulated from the estimated GPD and EPL models. In both cases we can see that the qualitative behavior of the tail in the GPD model is far from the behavior of the data, and that the EPL model offers a much better approximation.

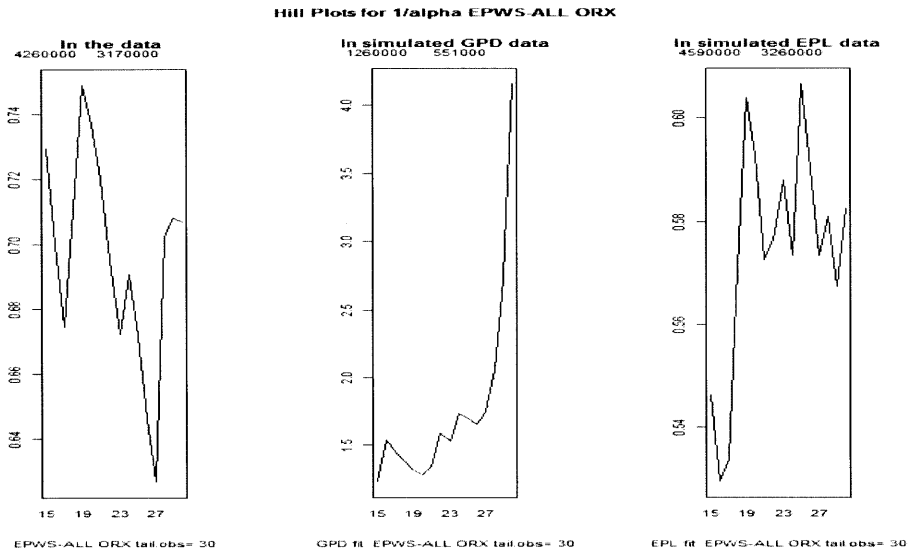


Figure 3-6: Hill plots for the EPWS-ALL unit of measure. In each case the bottom  $x$  axis denotes  $n$  (the number of upper order statistics used) and the top  $x$  axis represents  $X_{(n)}$  (the magnitude of the  $n$ 'th largest loss).

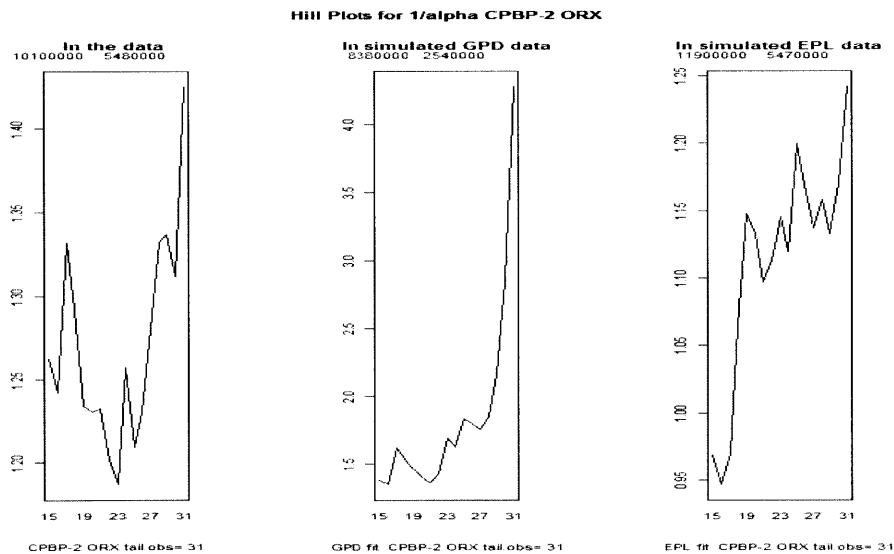


Figure 3-7: Hill plots for the CPBP-2 unit of measure. In each case the bottom  $x$  axis denotes  $n$  (the number of upper order statistics used) and the top  $x$  axis represents  $X_{(n)}$  (the magnitude of the  $n$ 'th largest loss).

### 3.3.2 Capital Reserve Estimation

We now compare the estimated capital reserve requirements under both the GPD and EPL models for each UOM. We estimate two canonical cases, in the first we assume the capital reserve requirements to be the 0.999 quantile of the loss distribution for a year and in the second we use the 0.9997 quantile.

In order to model the number of operational losses in a year, we assume a Poisson process with a frequency set by industry standards for each of the units of measure. We then simulate the magnitudes of the losses using the estimated values of the GPD and EPL models.

**Table 3.4:** Capital requirements at the 0.1% level (in \$1,000,000)

Unit of Measure	GPD	EPL	$CR_{GPD} - CR_{EPL}$
BDSF-1	4.08	6.41	-2.33
BDSF-2	1.00	0.97	0.03
CPBP-1	971.40	724.27	247.13
CPBP-2	1,288.73	385.36	903.36
DPA-ALL	112.30	9.46	102.84
EDPM-1	645.49	492.40	153.09
EDPM-2	7,161.37	455.18	6,706.18
EF-1	635.19	324.92	310.27
EF-2	6.26	5.36	0.90
EPWS-ALL	217.83	96.29	121.55
IF-ALL	818.54	697.36	121.18

In Table 3.4 we can see that with the exception of one unit of measure the EPL estimates of capital requirements are smaller than under the GPD case. We can even see that for some units of measure the change is so significant that there is a change in the order of magnitude of the capital requirements.

Furthermore, Table 3.5 shows that, as expected, at the 0.03% level the capital requirements are always higher than their 0.1% counterparts. Furthermore, we can see that the decrease in required capital occurs for the same units of measure. Note moreover that the units with irrationally high capital requirements under the GPD model (those with capital requirements in the billions of dollars) tend to have capital requirements of more reasonable orders of magnitude under the EPL model.

**Table 3.5:** Capital requirements at the 0.03% level (in \$1,000,000)

Unit of Measure	GPD	EPL	$CR_{GPD} - CR_{EPL}$
BDSF-1	4.62	7.48	-2.86
BDSF-2	1.61	1.23	0.39
CPBP-1	2,150.64	1,030.41	1,120.23
CPBP-2	3,209.69	510.19	2,699.50
DPA-ALL	370.55	11.84	358.72
EDPM-1	913.16	567.22	345.94
EDPM-2	33,893.99	630.46	33,263.52
EF-1	4,238.06	702.17	3535.89
EF-2	23.46	12.85	10.61
EPWS-ALL	445.46	124.96	320.50
IF-ALL	3,204.21	1,353.34	1850.87

### 3.4 Conclusions

We have presented the extended Pareto law as a parsimonious alternative that can be used to model operational losses. We also described how to estimate the model via maximum likelihood. Further, using the jittered version of the ORX dataset we have shown that the EPL model presents better goodness of fit properties than the GPD. In fact these results are extremely relevant since extreme value theory is one of the most widely-used tools in operational risk modeling and it implies the use of the GPD model for the tails of distributions. Hence, have shown that the EPL model performs much better than one of the most widely used tools.

Finally, comparing capital estimates implied by the GPD and EPL models we find that EPL produces smaller capital requirements. Thus, this indicates that estimating capital requirements via the EPL model will avoid one of the common problems of statistical modeling of operational losses: capital requirements that are excessively large.

One last step that would be very useful would be to compare the capital requirements to the assets or the gross income of the different institutions. This would allow us to get a better understanding of the magnitude of the capital requirements implied by the EPL model vis-a-vis other statistical models and also "rule of thumb" procedures such as the Basic Indicator and Standardized procedures described in the introduction. Unfortunately, this information is not easily available due to privacy concerns of the financial institutions.

Nonetheless, we have provided enough evidence to suggest that the EPL model is one of the most useful tools available for modeling operational risk: it is much more parsimonious than other models with good statistical fit such as the g-and-h distribution, and it has much better statistical fit than the main parsimonious alternative, extreme value theory.

# Bibliography

Andrews, D. W. K. (1999): "Estimation when a Parameter is on a Boundary," *Econometrica*, Vol. 67, 1341–1383.

Basel Committee on Banking Supervision (2008): *Results from the 2008 Loss Data Collection Exercise for Operational Risk*. Bank for International Settlements Press & Communications, Basel.

de Fontnouvelle P., V. DeJesus-Rueff, J. Jordan, and E. Rosengren (2003): "Capital and Risk: New Evidence on Implications of Large Operational Losses," FRB Boston Series, paper no. 03-5.

de Fontnouvelle P., J. S. Jordan, and E. S. Rosengren (2006): "Implications of Alternative Operational Risk Modeling Techniques," in M. Carey and R. Stulz, eds., *The Risks of Financial Institutions*, NBER/University of Chicago Press.

Drees H., S. Resnick, & L. de Haan (2000): "How to Make a Hill Plot," *Annals of Statistics*, Vol. 28, 254-274.

Dutta K. and J. Perry (2007): "A Tale of Tails: An Empirical Analysis of Loss Distribution," FRB of Boston Working Paper No. 06-13.

Heyde, C. C., Y. V. Prohorov, R. Pyke, and S. T. Rachev (eds.) (1996): Athens Conference on Applied Probability and Time Series Analysis., Vol. I, vol. 114 of Lecture Notes in Statistics. Springer-Verlag, New York, Applied probability, In honor of J. M. Gani, Papers from the conference held in Athens, March 22–26, 1995.

Kagan, Y. Y., and F. Schoenberg (2001): "Estimation of the Upper Cutoff Parameter for the Tapered Pareto Distribution," *J. Appl. Probab.*, Vol. 38A, 158–175.

Knight, K. (1999): "Epi-convergence and Stochastic Equisemicontinuity," Technical Report, University of Toronto, Department of Statistics (<http://www.utstat.toronto.edu/keith/papers/>).

Mason, D. (1982): "Laws of Large Numbers for Sums of Extreme Values," *Ann. Probab.*, Vol. 10, 754–764.

Pareto, V. (1989): *Statistique et Économie Mathématique, Œuvres Complètes de Vilfredo Pareto* [Complete Works of Vilfredo Pareto], VIII. Librairie Droz, Geneva, second edn., With a preface by René Roy, *Travaux de Droit, d'Économie, de Sciences Politiques, de Sociologie et d'Anthropologie* [Works in Law, Economics, Political Science, Sociology and Anthropology], 48.

Wu, L. (2006): "Dampened Power Law: Reconciling the Tail Behavior of Financial Security Returns," *The Journal of Business*, Vol. 79, 1445–1473.