



Università
Ca' Foscari
Venezia

Corso di Laurea Magistrale
(ordinamento ex D.M. 270/2004)
in Economia-Economics

Tesi di Laurea

Approaching Systemic Risk with Entropy

A new proposal for an Early Warning Measure

Relatore

Ch. Prof. Roberto Casarin

Laureando

Andrea Pasqualini

Matricola 828170

Anno Accademico

2013 – 2014

Master Thesis

Approaching Systemic Risk with Entropy

Andrea Pasqualini
828170@stud.unive.it

Ca' Foscari University of Venice
Department of Economics

Supervisor
Roberto Casarin

A THESIS SUBMITTED IN FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Laurea Magistrale
in
Models and Methods in Economics and Management

Abstract

This Thesis focuses on the Marginal Expected Shortfall (MES) and its application in Finance as early-warning reference. I will specify a logit model that allows to link MES to the conditional probability of financial crisis. The independent variable consists of three different measures of entropy, so that the degree of “disorder” in the financial markets is used to predict systemic events.

The analysis includes two bayesian density estimation approaches, which embed the frequentist one as a special case. Also follows a calibration of the model according to how much important the tails of the distribution of returns should be. Robustness of the results is also checked by the bayesian choice.

Acknowledgements

I would like to thank my supervisor Dr. Roberto Casarin for his valuable advice and guide, as well as for having inspired me for those choices related to my future. I am thankful to Dr. Michele Costola, who strongly supported me during my research and work. I also thank the faculty members at the Department of Economics and, in particular, Dr. Paola Ferretti for her tutorship and her advice, that considerably helped me take some key choices.

I would like to thank my friends, who experienced these last years with me. In particular, I thank Cristina, Giulia, Francesco, Marco, Kata, Tatiana, Alice, Matteo A. and all the people I met during my studies and during my AIESEC experience, who remarkably contributed to make my life at the University awesome, memorable and priceless.

Above all, I need to thank Matteo S., who accompanied me over the last five years. For the Thesis itself, he gave me important advice about 3D data visualization, without which I could have never considered some of the implications of this work. But for *Life, Universe and Everything*, it is just thanks to him that I can say I really am an Economist: without him, I could have never exploited those huge economies of scale that actually arose in the academic context, outside the University and everywhere in the last five years.

Thank you.

To Marco, Anna and Franco.

Contents

Abstract	iii
Acknowledgements	v
1 Introduction	1
1.1 Literature review	2
2 The theoretical framework and the data	5
2.1 Data	6
2.1.1 Crisis indicators	6
2.1.2 Assets and returns	7
2.2 Methodology	9
2.2.1 Density estimation	11
2.2.2 Entropy definition	12
3 The models and main findings	17
3.1 Model calibration	19
3.2 Robustness check	21
4 Conclusions	27
A The database in Reinhart and Rogoff (2008)	29
B The database in Laeven and Valencia (2012)	31
C Own defined functions in MatLab	35
C.1 Density estimation	35
C.2 Entropy formulations	36
C.2.1 Shannon (1948)	36
C.2.2 Rényi (1960)	37
C.2.3 Tsallis (1988)	38

List of Tables

2.1	Description of the ICB 8000 <i>financials</i> asset class.	7
2.2	List of financial markets and no. of assets collected.	9
2.3	The bayesian point estimates according to the uniform and the Jeffrey's priors.	14
3.1	Logit estimates on Shannon's Entropy.	17
3.2	Logit estimates on Renyi's Entropy ($\alpha = 1.7$).	21
3.3	Logit estimates on Tsallis' Entropy ($\alpha = 1.7$).	21
3.4	The p -values for the coefficients of the entropy indexes expressed in scientific notation $a \times 10^b$	26
3.5	Results from the logit fits for all the density estimation methods.	26
A.1	List of binary variables in Reinhart and Rogoff (2008).	30
A.2	List of countries analyzed in Reinhart and Rogoff (2008).	30
B.1	List of variables in Laeven and Valencia (2012).	32
B.2	List of countries analyzed in Laeven and Valencia (2012).	33

List of Figures

2.1	Distribution of returns in Europe over time.	8
2.2	Sample size and standard deviation of returns in Europe over time.	8
2.3	Distribution of MES in Europe over time.	10
2.4	Sample size and standard deviation of MES in Europe over time.	10
2.5	Estimated density of MES over time (MAP, with uniform prior).	13
2.6	Estimated densities of MES over time (posterior predictive with uniform and Jeffreys' priors).	13
2.7	Plot of Shannon's entropy for $\mathbf{p} = (p, 1 - p)$	15
2.8	Plot of Renyi's entropy for $\mathbf{p} = (p, 1 - p)$	15
2.9	Plot of Tsallis' entropy for $\mathbf{p} = (p, 1 - p)$	15
3.1	Scatter-plot of actual and estimated response variable against the Shannon's entropy.	18
3.2	Overview of actual and estimated response variable over time.	18
3.3	Sum of Squared Residuals as functions of α	20
3.4	Scatter-plot of actual and estimated response variable against the Renyi's entropy.	22
3.5	Overview of actual and estimated response variable over time.	22
3.6	Scatter-plot of actual and estimated response variable against the Tsallis' entropy.	23
3.7	Overview of actual and estimated response variable over time.	23
3.8	Sum of Squared Residuals as functions of Renyi's α	25
3.9	Sum of Squared Residuals as functions of Tsallis' α	25

Chapter 1

Introduction

As for the latest financial crisis, much attention has been paid to modeling systemic events. Many authors have come up with models to link crises across Countries and over time. A branch of the literature, reviewed in the next section, is focused on the joint probability of observing extreme, or tail, events. The theoretical foundation for this approach relies on representing a panel of financial returns, R_t^i , where t denotes time and i denotes an asset, as

$$\Pr(R_t^1, R_t^2, \dots, R_t^i, \dots, R_t^n), \quad t = 1, 2, \dots, T.$$

This formulation allows to study both marginal and conditional probabilities, so that the effects of one asset (or a set of them) on another asset (or another set) may be explored and discussed.

The motivation about this study relies on the willingness to find out whether, when and how we can predict — or, at least, explain — systemic events. On the macroeconomic perspective, this is relevant given the objectives of the most important institutions like OECD, ECB, Federal Reserve and so on. The ability of modeling systemic risk may enable financial and/or social regulators to pinpoint the cause of tail events. There are a number of reasons for which such regulators may need to predict, if not prevent, systemic crises. Acharya et al. (2010), for example, show that only a small tax may be necessary to reduce the probability of crises contagion among institutions, markets and, finally, countries.

On the statistical hand, the study of systemic events is the a challenge. It is well known that any (current) statistical application only relies on past information. This represents a strong limitation to the capability of predicting tail events. To overcome this weakness, the ongoing statistical research on rare outcomes tries to reach robustness, meant as the ability of the models to give the same — qualitative — results independently of the input samples.

1.1 Literature review

A variety of papers and publications try to model systemic crises on different bases, depending on the assumptions one may want to make. There are different views, in the literature, that take into account different aspects of systemic phenomena and, namely, different definitions of extreme systemic events. As for the crises, Reinhart and Rogoff (2008) created a dataset of binary variables that indicate when a certain country has been facing a currency, sovereign debt and/or banking crisis.

Broadly speaking, the idea of systemic crisis is associated to the joint variability into a whole economic system, where the linkages among institutions are the conditions for observing “domino” effects. That is, if an institution experiences major issues, another institution, linked to the former, will experience the same as a consequence of their economic relation.

Bisias et al. (2012) wrote a thorough survey about systemic crisis analytics, which gathers many of such approaches taken so far in the related research. Part of the literature sees an economic system as many interconnected subjects (consumers, firms, banks, etc.), where systemic risk is grounded on the basis of such connections. From this concept, a branch of the literature started investigating network-like linkages among financial institutions, where the main research question aims at discovering *how* a crisis unfolds and permeates an economic system. See, for example, Billio et al. (2011), who developed a Granger causality test to spot significant linkages among financial institutions, in order to tell apart which ones are systemically important and how those relate to the rest of their economic sectors.

Other authors, instead, see an economic system as a “portfolio” of institutions, where the performance of one of them impacts the others, and *how much* its components interact among themselves without actually questioning which are the means and the relations that make such interactions possible. See, for example, Adrian and Brunnermeier (2011), who developed the concepts of CoVaR and Δ -CoVaR.

For both visions, many measures and analytics have been conceived in order to first analyze and (hopefully) predict systemic issues. This Thesis embraces the view of Acharya et al. (2010), who provided a micro-founded model about what they called *Marginal Expected Shortfall* (MES). Their measure tell *how much* a specific institution is affected by another entity, be it another institution or the whole market. MES is closely related to the concept of Value-at-Risk (VaR), which measures the losses in case of extreme, or tail, events. This approach is grounded on two main facts. First, the sample variance of a distribution says how much the tails are fat, but it gives no clue about the direction of the variability. Second, these measures are basically relying on the joint probability, for institutions, of witness-

ing extreme events.

Apart from defining a way to measure systemic risk, Acharya et al. (2010) also introduce and solve a social welfare maximization problem in order to justify the research on such measures. They show that a wise study on systemic risk measures may allow the social planner to “force” financial institutions to also consider their impact on other ones, by introducing a simple tax. That is, by means of such tax, regulators may make banks and other financial institutions aware of how much their probability of bankruptcy may affect the economic environment in which they operate, whereas the same institutions would not care about the issue on their sole own. In this way, regulators may easily maximize the social welfare (i.e., minimize the *Systemic Expected Shortfall* the authors define).

In this Thesis, rather than considering the joint probability of tail events, I take a different approach to systemic risk. The approach follows the intuition that “disorder” in financial markets reinforces itself. During tail events the financial agents observe more chaos in the markets, and therefore they start panicking and become more sensitive about observed heterogeneity. So, spread behaviors and very different trends inside one market may signal the presence of important incoming (or ongoing) events.

In Chapter 2.1, I will present the data and the distribution of returns and MES at each time of the dataset. Specifically, in Section 2.2 I will show how to interpret a continuous random variable in terms of a multinomial one, and I will present four different bayesian ways to estimate the densities of MES, according to the which estimators and which priors are elicited. Basing on a crisis indicator, I will show in Chapter 3 how entropy measures are related to the conditional probability of a systemic event given the dispersion of the returns’ distribution at each state. Three different entropy measures are considered to assign different weights to the degree of randomness in the distributions. The logit specifications will include results of three of the four density estimation methods, and will show how the difference among the entropy formulations affect the final fit. In Chapter 4, I will draw some final considerations about the models and I will discuss possible future developments of this Thesis.

Chapter 2

The theoretical framework and the data

The reference measure in my Thesis is the *Marginal Expected Shortfall* (MES), as defined by Acharya et al. (2010). Starting from a series of asset returns R^i , where i denotes the asset, MES^i is defined as the expected value of R^i when a reference asset (or a reference market) is in its “worst state,” below a certain q_k quantile. That is, for $k = 0.05$,

$$MES^i = \mathbb{E} \left[R^i \mid R^m < q_{5\%} \right].$$

The authors, in their original formulation, put a minus in front of the expectation in order to meet consistency with the definition of “shortfall,” as the expected returns in case of a tail event are intuitively thought to be negative. Moreover, Acharya et al. (2010) considered MES as a measure of systemic risk, which assesses the expected losses in case the market faces a tail event. The intuition behind MES is that, if institution i is linked (no matter how) to a systemic event, the conditional returns should highlight it. Yet, the authors did not consider MES for hypothesis testing. They proposed and analyzed its properties at a firm-level risk management point of view. In particular, they analyzed its predictive power. However, as shown in Löffler and Raupach (2013), MES is successful in capturing systemic relations if calculated on the stock market returns, but it does not perform sufficiently well for other financial instruments, like bonds and derivatives. The authors do observe that MES, calculated on “mixed” portfolios, that is ones composed by not only assets, gives a biased picture of its risk. Even though no specific reason is offered in their paper, they suggest that derivatives, more than bonds, naturally induce non-linear behaviors in their returns that MES cannot capture, being it a linear estimator.

As it turns out, MES basically filters data in order to “pick” specific returns, i.e., when the market is facing the so-called “black-swans.” To this matter, and

depending on the frequency of the data, MES does not take into account lagged influences. That is, MES only considers all those returns which are affected by the market performance only *during* the tail event. Consequences, which are lagged in time, are not captured by the measure. However, this “filtering” allows a specific analysis of tail events in which one can isolate data.

In this Thesis, I focus on MES as a filtering device, in order to obtain representations about the worst states of the financial markets. In the analysis that follows, the average in the definition of MES runs over time. To save data and avoid to come up with a unique number, it has been convenient to implement the *rolling window* technique as shown in Zivot and Wang (2007), that considers subsequent subsamples in order to preserve degrees of freedom. It takes as input a time series and returns, as output, a shorter and smoother times series, depending of the width of the window. Namely, if the time series is about returns R_t^i , with $t = 1, 2, \dots, T$, we can either average the data over the whole period and get a unique estimate, or consider smaller time intervals and get several numbers. For a window of width w such that $1 < w < T$, we have $T - w + 1$ averages over time, so that, for every subperiod of length w , we have a representation of the worst states.

2.1 Data

2.1.1 Crisis indicators

In the literature there are many databases that track the crises over years. Examples of these include Reinhart and Rogoff (2008), Reinhart and Rogoff (2010b), Reinhart and Rogoff (2010a), Laeven and Valencia (2008) and Laeven and Valencia (2012). For a description of the databases in Reinhart and Rogoff (2008) and Laeven and Valencia (2008) and their information, see Appendices A and B respectively. Such databases may be relevant whenever one wants to develop an Early Warning System, in order to have reference data for the crises. In particular, one may model the crisis in a given country as

$$C_t = \begin{cases} 1 & \text{if the country is in crisis at time } t \\ 0 & \text{otherwise.} \end{cases} \quad (2.1)$$

The main difficulty in this kind of modeling is giving a proper definition of “crisis” in a given country. As in Reinhart and Rogoff (2008), the idea of crisis is split into several perspectives. Specifically, they define sovereign debt crisis, banking crisis and currency crisis. As it will be clear in the next section, it is convenient to consider banking crises for this Thesis.

Supersector		Sector	
8330	Banks	8350	Banks
8500	Insurance	8530	Nonlife Insurance
		8570	Life Insurance
8600	Real Estate	8630	Real Estate Investment & Services
		8670	Real Estate Investment Trust
8700	Financial Services	8770	Financial Services
		8980	Equity Investment Instruments

Table 2.1: Description of the ICB 8000 *financials* asset class.

2.1.2 Assets and returns

This Thesis considers empirical data collected by DataStream.[®] As I wanted to consider systemic risk, the time series are about those firms which are classified under the ICB code class 8000. This is the class for financial firms. A description of this class of assets is in Table 2.1.

The data are about asset prices for many European countries, from 1st January 1985 to 12th May 2014 at a daily frequency, without holidays. Table 2.2 shows the list of considered financial markets (countries) and the amount of assets for each of them. For every country in the dataset, it is convenient to represent prices in the following way:

$$P = \begin{bmatrix} P_1^1 & P_1^2 & \dots & P_1^i & \dots & P_1^N \\ P_2^1 & P_2^2 & & & & P_2^N \\ \vdots & \vdots & \ddots & & & \vdots \\ P_t^1 & P_t^2 & & P_t^i & & P_t^N \\ \vdots & \vdots & & & \ddots & \vdots \\ P_T^1 & P_T^2 & \dots & P_T^i & \dots & P_T^N \end{bmatrix}. \quad (2.2)$$

The prices have then been converted to log-returns according to the formula in Equation 2.3. Returns are also arranged in the same matrix structure as for prices.¹

$$\log(R_t) = \log\left(\frac{P_{t+1}}{P_t}\right) = \log(P_{t+1}) - \log(P_t). \quad (2.3)$$

Considering the overall EU market, that is aggregating all the data in a unique array, the data are summarized in Figures 2.1 and 2.2. For that extent, each row has been considered as distinct vector where to compute averages, quantiles, standard deviation and so on.

For calculating MES, a reference asset is needed. As such, the MSCI Europe index provides a comprehensive image of 15 countries in Europe, for a total of 437

¹In the text, "returns" refers to log-returns, unless explicitly specified.

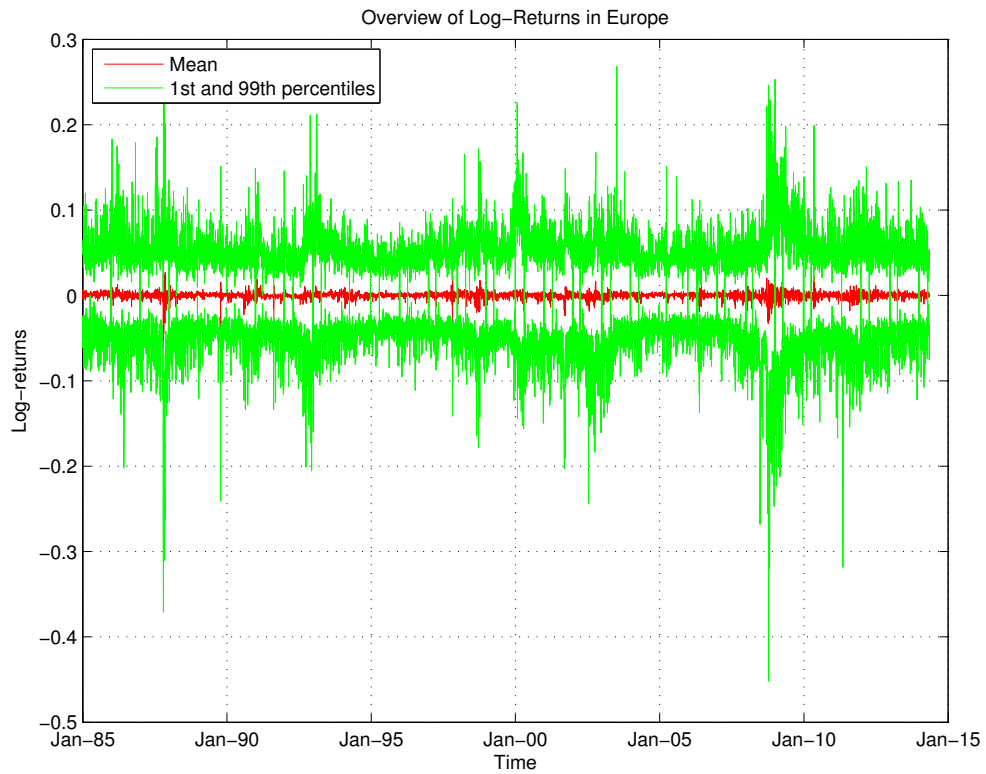


Figure 2.1: Distribution of returns in Europe over time.

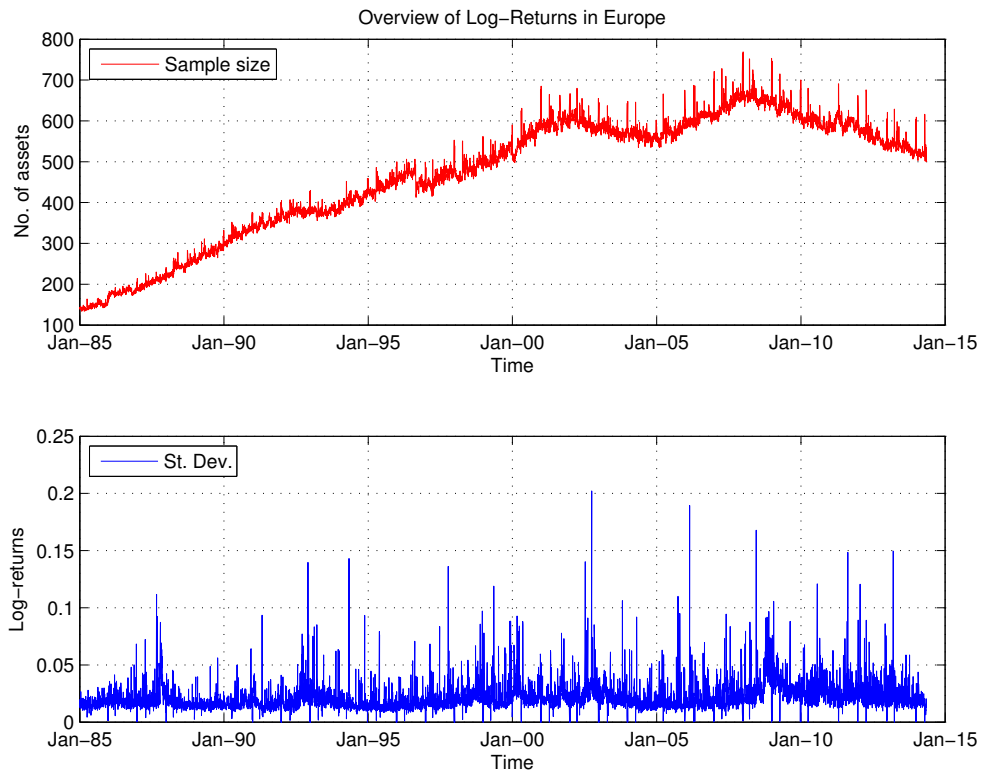


Figure 2.2: Sample size and standard deviation of returns in Europe over time.

constituents of any ICB class. Therefore, if R_t^m is the return computed on the MSCI Europe index, MES_t^i is given by:

$$MES_t^i = \mathbb{E} \left[R^i \mid R_s^m < q_{5\%} \right], \quad s = t - w + 1, t - w + 2, \dots, t - 1, t, \quad (2.4)$$

that is the average of returns in those days when MSCI is in its worst 5% during the time window of length w . The conditioning event basically tells over which $t = 1, 2, \dots, T$ we should average the returns of the i -th institution. The practical implementation of the rolling window technique has considered $w = 180$ -days long subsamples. That is, the first value of MES occurs at $t = 180$ and it is defined as

$$MES_{180}^i = \mathbb{E} \left[R^i \mid R_s^m < q_{5\%} \right], \quad s = 1, 2, \dots, 180,$$

the second value is

$$MES_{181}^i = \mathbb{E} \left[R^i \mid R_s^m < q_{5\%} \right], \quad s = 2, 3, \dots, 181,$$

the third is

$$MES_{182}^i = \mathbb{E} \left[R^i \mid R_s^m < q_{5\%} \right], \quad s = 3, 4, \dots, 182,$$

and so on. This allows us to save degrees of freedom and we obtain a vector of $T - 179$ values of MES for each institution. Figure 2.3 and 2.4 illustrate the main features of the distribution of MES over time.

2.2 Methodology

As we can see in Figure 2.1, assets in Europe featured a more spread distribution during the crisis years. This is highlighted in the representation for MES. As it turns out, most of the observations for each country are concentrated from 1st January 2000 to 12th May 2014. Before that period, small changes in the distributions are actually emphasized by the small amount of assets. For this reason, I will consider the aggregate European dataset, starting from the beginning of 2000.

Given that the banking crisis indicator in Reinhart and Rogoff (2008) has its last

Austria	43	Belgium	73	Denmark	179
Finland	30	France	285	Germany	344
Greece	82	Hungary	16	Ireland	30
Italy	139	Latvia	1	Lithuania	5
Luxembourg	40	The Netherlands	87	Norway	78
Portugal	29	Spain	84	Sweden	113
Switzerland	149	United Kingdom	1310		

Table 2.2: List of financial markets and no. of assets collected.

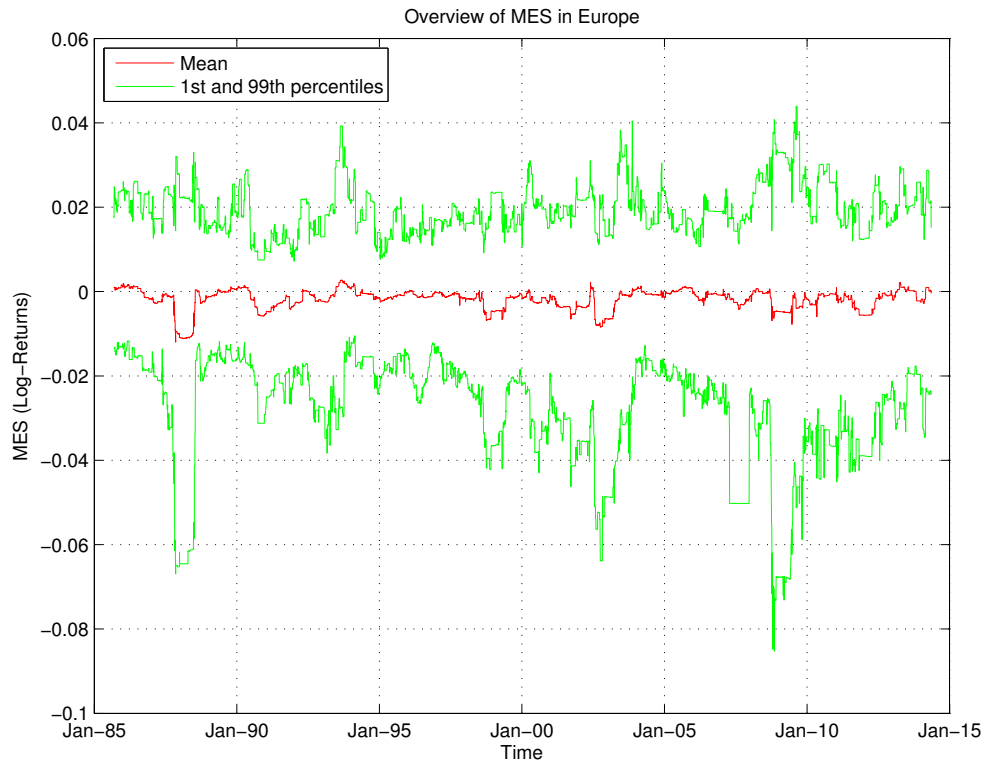


Figure 2.3: Distribution of MES in Europe over time.

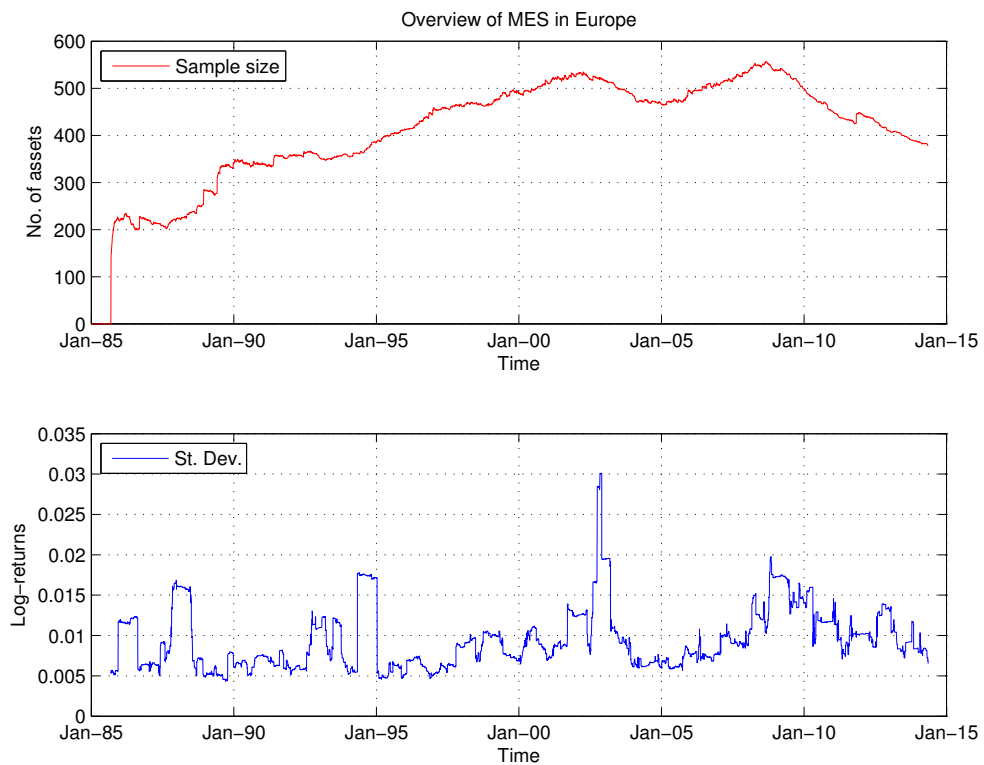


Figure 2.4: Sample size and standard deviation of MES in Europe over time.

record in 2010, I will also discard returns from 2011 to 2014. Also, the crisis indicator is given on a per-country basis. Given that most of the assets in the sample are registered under the British stock market, I have considered Europe to be facing a banking crisis whenever the United Kingdom is. Another assumption regards data frequency. Since the crisis indicator comes at a yearly frequency and the returns are observed daily, I will assume that the crisis indicator will equal 1 for all days in a given year, if the indicator equals 1 for that year. Although this is a very strong assumption, it will be discussed later on in Chapter 4, along with the conclusions.

The practical implementation in MatLab[®] is shown in Appendix C, where the core own-defined functions are presented.

2.2.1 Density estimation

One challenging task in doing statistics is estimating a density. This has a variety of applications and one of them is calculating an entropy index. The choice in this Thesis is to employ a *Bayesian non-parametric* approach, where different results can be obtained according to the prior knowledge one has when carrying out research. Müller and Quintana (2004) represented the state of Bayesian non-parametric methods known that far, with clear highlights about ongoing research in that field. As the bayesian approach implies the elicitation of prior distributions to represent prior knowledge about a parameter (or a whole parameter space), I will consider the uniform distribution and the Jeffreys' prior in order to state prior ignorance. In addition, the *frequentist* approach follows in this Thesis as a special case of the bayesian one.

Given the discretization of MES, which is a continuous random variable, it can be seen as version of a multinomial density, where, instead of a category, we place a specific value of MES into an interval. No specific meaning is actually assigned to any of the subintervals, whereas in the pure multinomial framework, each category may have some "label" attached. In this context, inference about the density of MES has been carried out using the bayesian multinomial density estimation approach as in Minka (2003). He started with the joint probability of a set of counts defined as:

$$\Pr(N_1, N_2, \dots, N_K | \mathbf{p}) = \binom{N}{N_1, N_2, \dots, N_K} \prod_{k=1}^K p_k^{N_k}, \quad (2.5)$$

with

$$N_k = \sum_{x=1}^K \delta(x = k), \quad \text{and} \quad \delta(x = k) = \begin{cases} 1 & \text{if } x = k \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathbf{p} = (p_1, p_2, \dots, p_K)$ it the vector of probabilities such that $p_k = \Pr(x = k)$.

As already introduced, one can compare the multinomial approach to the histogram one. Instead of a discrete-valued random variable, one can consider a real-valued variable that takes values on a set of intervals. In this context, N_k is exactly

the number of observations that fall within the k -th interval. Minka (2003) shows that a conjugate prior for \mathbf{p} is the Dirichlet distribution:

$$\Pr(\mathbf{p}|\alpha) \sim \mathcal{D}(\alpha_1, \dots, \alpha_K) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_{k=1}^K p_k^{\alpha_k-1}. \quad (2.6)$$

And the posterior distribution obtained by applying the Bayes' rule is:

$$\Pr(\mathbf{p}|\mathbf{X}, \alpha) \sim \mathcal{D}(N_k + \alpha_k) = \frac{\Gamma(\sum_k (N_k + \alpha_k))}{\prod_k \Gamma(N_k + \alpha_k)} \prod_{k=1}^K p_k^{N_k + \alpha_k - 1}. \quad (2.7)$$

The hyperparameter α_k can be thought of as a virtual count, on the same scale of N_k , for the k -th bin, before observing \mathbf{X} . As shown in Minka (2003), given a vector of data \mathbf{X} and a Dirichlet distribution $\mathcal{D}(\alpha_1, \dots, \alpha_K)$ as a conjugate prior, and rephrasing in terms of histograms, one can obtain the posterior predictive distribution

$$\Pr(x = k|\mathbf{X}, \alpha) = \mathbb{E}[p_k|\mathbf{X}] = \frac{N_k + \alpha_k}{N + \sum_{k=1}^K \alpha_k}, \quad (2.8)$$

or the Maximum *A Posteriori* (MAP)

$$\hat{p}_k = \frac{N_k + \alpha_k - 1}{N + K + \sum_{k=1}^K \alpha_k}, \quad (2.9)$$

where N is the number of IID samples $X = \{x_1, x_2, \dots, x_N\}$, N_k is the number of variables whose realizations are within the k -th bin and K is the total number of bins. Table 2.3 describes how Equations 2.8 and 2.9 are written when the Jeffreys' prior or the uniform one are employed. It can be noted that the Maximum *A Posteriori* (MAP) estimate with uniform prior corresponds to the standard frequentist histogram approach.

The point estimate of \mathbf{p} carried out using MAP and the Jeffreys' prior, however, may produce invalid values of p_k , as it may happen to be negative. Speaking in practical terms, if one arbitrarily decides to use specific intervals for discretizing, it may happen that, for some intervals, N_k is zero, so that the numerator becomes negative. Otherwise, by setting such intervals, one is also imposing a specific value for K , which may be such that $N < K/2$. In this case, the denominator becomes negative. Figures 2.5 and 2.6 show the estimated densities according to the employed method. Significant changes regard the probabilities of the modes, whereas the tails are somewhat equal. Qualitatively, however, the densities do not differ.

2.2.2 Entropy definition

As noted by Shannon (1948) and many subsequent works by many authors, the entropy can be thought of a measure of disorder, or randomness. In fact, he noted that the entropy index applied to probability is maximum when the underlying

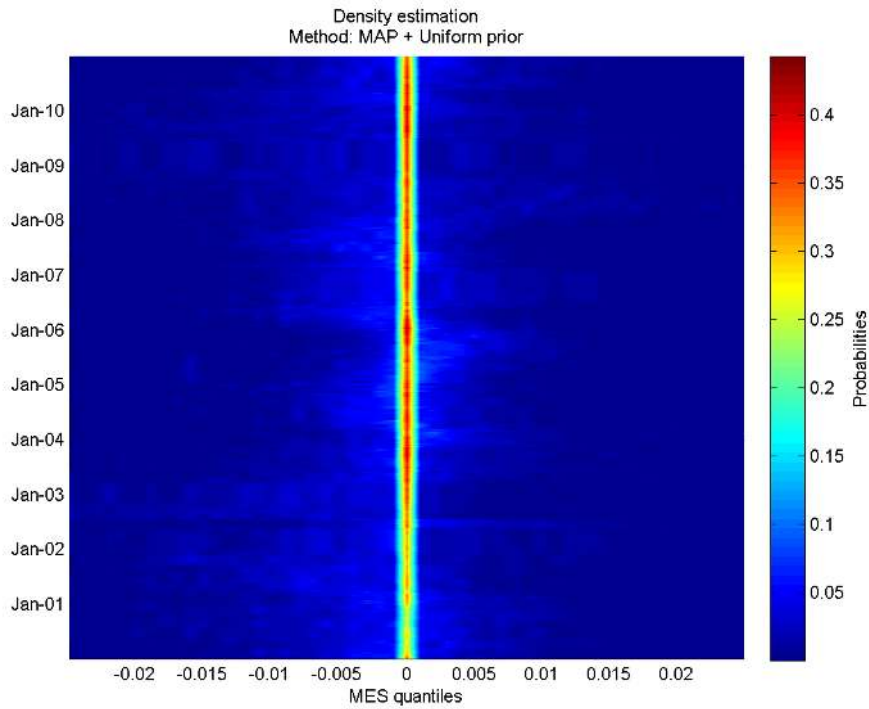


Figure 2.5: Estimated density of MES over time (MAP, with uniform prior).

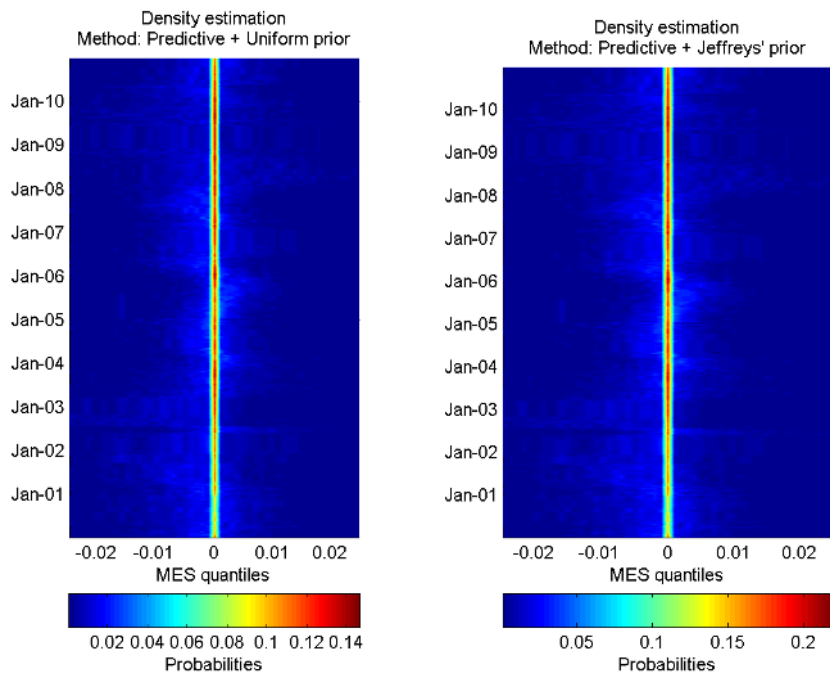


Figure 2.6: Estimated densities of MES over time (posterior predictive with uniform and Jeffreys' priors).

probability distribution is uniform. In other words, given the vector $\mathbf{p} = (p_1, p_2, \dots, p_K)$, if its elements are all equal, then the entropy associated to that vector is maximum. Otherwise, if only one element $p_k = 1$ and all the others are zero, the entropy is minimum.

Denoting the entropy measures in Shannon (1948), Rényi (1960) and Tsallis (1988), respectively, E_S , E_R and E_T , we have;

$$E_S = - \sum_{i=1}^n p_i \log(p_i), \quad (2.10)$$

$$E_R = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^n p_i^\alpha \right), \quad (2.11)$$

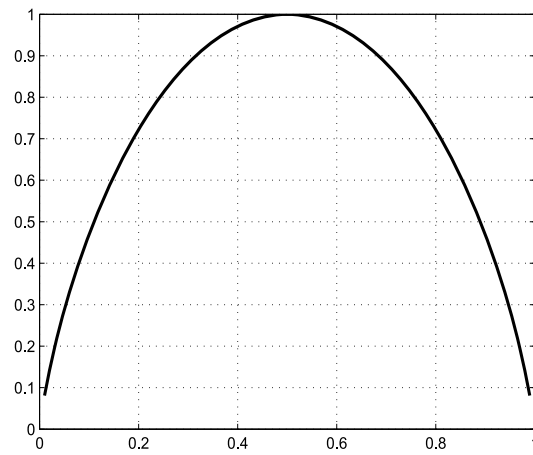
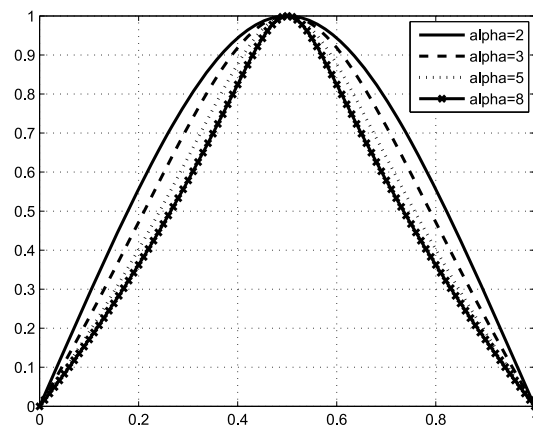
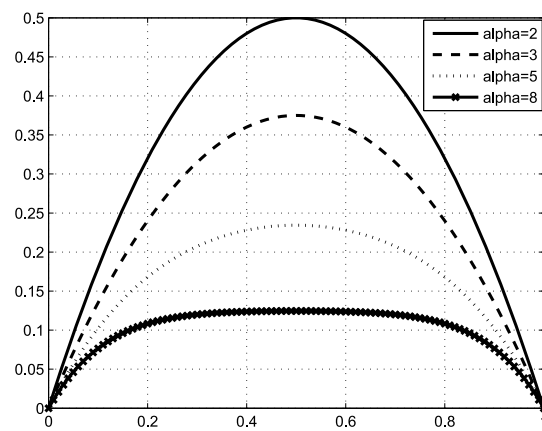
and

$$E_T = \frac{1}{\alpha-1} \left(1 - \sum_{i=1}^n p_i^\alpha \right). \quad (2.12)$$

As discussed in Maszczyk and Duch (2008), the entropy in Shannon (1948) is a special case of the other two formulations. In particular, according to the value of α , the measures in Equations 2.11 and 2.12 assign more or less weight to the tails of the distribution. To see this, assume that $\mathbf{p} = (p_1, p_2)$, with $p_2 = 1 - p_1$, and look at Figures 2.7, 2.8 and 2.9, which have been taken from Maszczyk and Duch (2008). Compared to the entropy index in Shannon (1948), and depending on the value of the parameter α , the entropy in Rényi (1960) penalizes the “mid-way” between the uniform and the impulse distributions, while the entropy in Tsallis (1988) assigns less importance to randomness, that is it penalizes uniformity in the distribution. Therefore, for the entropy in Rényi (1960), the higher the parameter α and the less the entropy for distributions “far” from the uniform, i.e., the tails of the distribution are penalized. In contrast, for the entropy in Tsallis (1988), the higher the parameter α , and the less the entropy for distributions “close” to the uniform, i.e., the tails of the distribution are emphasized. It is clear that the “farther” a distribution is from the uniform, the thinner its tails are. Entropies behave somewhat symmetrically.

Method	Uniform	Jeffreys
Posterior predictive, $\mathbb{E}[p_k \mathbf{X}] =$	$\frac{N_k + 1}{N + K}$	$\frac{N_k + 1/2}{N + K/2}$
Maximum <i>a posteriori</i> , $\hat{p}_k =$	$\frac{N_k}{N}$	$\frac{N_k - 1/2}{N - K/2}$

Table 2.3: The bayesian point estimates according to the uniform and the Jeffrey’s priors.

Figure 2.7: Plot of Shannon's entropy for $\mathbf{p} = (p, 1 - p)$.Figure 2.8: Plot of Renyi's entropy for $\mathbf{p} = (p, 1 - p)$.Figure 2.9: Plot of Tsallis' entropy for $\mathbf{p} = (p, 1 - p)$.

Chapter 3

The models and main findings

In this Thesis, I model systemic risk as a logistic function of entropy. For each definition of entropy defined in Section 2.2, many models have been estimated in order to take into account the effects of the parameter α in Rényi (1960) and Tsallis (1988).

If E_t is the entropy index (of any type) for the distribution of returns at time t and C_t is the crisis indicator at time t , then the specified models are of the form

$$\Pr(C_t = 1|E_t) = G(\beta_0 + \beta_1 E_t), \quad (3.1)$$

where $G(\cdot)$ is the logistic cumulative density function, namely

$$G(x) = \frac{e^x}{e^x + 1}. \quad (3.2)$$

In this Section, I analyze the results for the Shannon's Entropy estimated on the density obtained from the Maximum *A Posteriori* method with uniform prior described in Subsection 2.2.1, which corresponds to the frequentist approach. The results from the logit specification are presented in Table 3.1. Figures 3.1 and 3.2 give pictures of the models, the former in a scatter perspective (entropy *versus* crisis indicator, both actual and predicted) and the latter in a temporal perspective (time *versus* crisis indicator, both actual and predicted) and it emphasizes the use of such a model in terms of Early Warning System.

Variable	Coefficient	St. Error	<i>t</i> -ratio	<i>p</i> -value
<i>constant</i>	-32.3465	1.1631	-27.8106	0.0000
Entropy	10.9709	0.3963	27.6800	0.0000

No. of observations: 2870 (From 1-Jan-2000 to 31-Dec-2010)

Percentage of correctly predicted indicators: 86.48%

Table 3.1: Logit estimates on Shannon's Entropy.

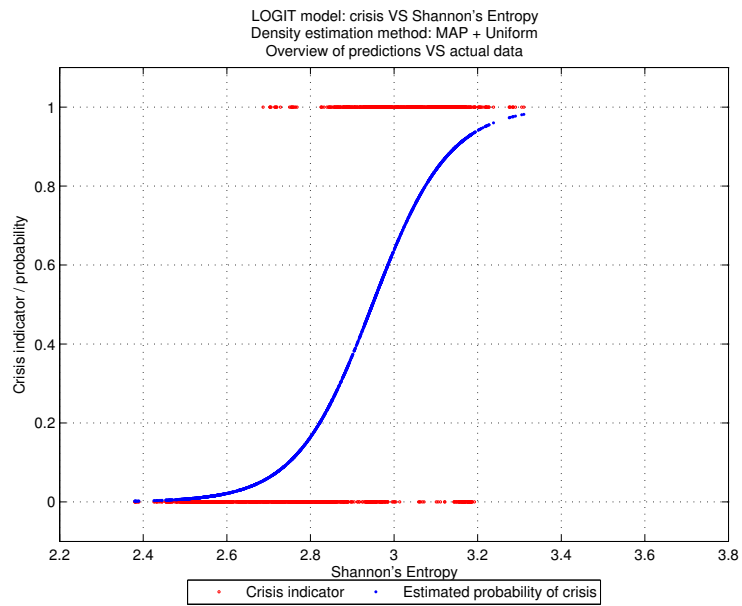


Figure 3.1: Scatter-plot of actual and estimated response variable against the Shannon's entropy.

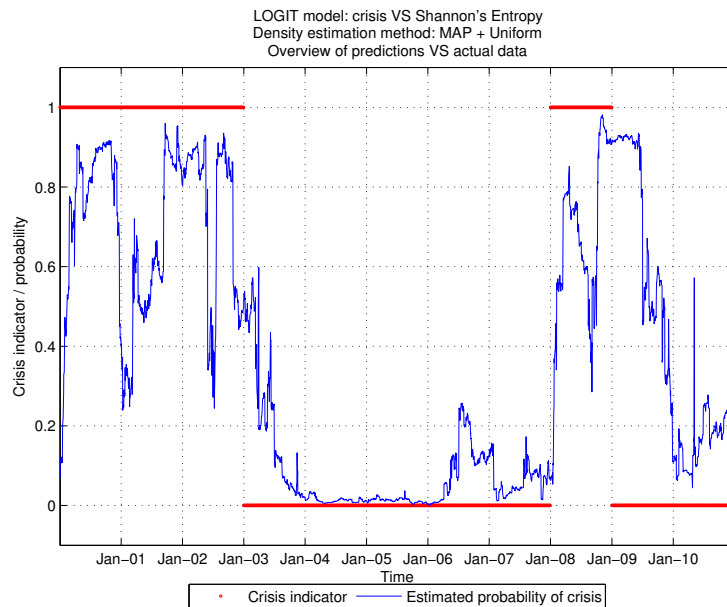


Figure 3.2: Overview of actual and estimated response variable over time.

The actual p -value for the coefficient associated to entropy is 1.2169×10^{-168} . That is, for any conventional confidence level, the entropy is statistically significant in explaining the crisis indicator. This result is intuitively depicted in Figure 3.2, where the overall impression is that the model does “work.”

For evaluating the goodness of the models, many approaches can be considered. See Wooldridge (2009) for a list of most common goodness-of-fit measures. For simplicity, the *percent of correctly predicted* indicators is here employed. It is built as follows. With reference to the dependent variable, i.e., the crisis indicator in Reinhart and Rogoff (2008), we can define a threshold as the percent of times where it is equal to one. Namely

$$\text{threshold} = \frac{\sum_{t=1}^T C_t}{T}. \quad (3.3)$$

This threshold is selected to take into account the intrinsic features of the crisis indicator: it is not completely separated by any entropy index, and the number of times in which Europe is deemed to be facing a banking crisis is less than half the considered time span. For the time between 1st January 2000 and 31st December 2010, such threshold is equal to 36.38%. Then, if \hat{C}_t is the predicted probability of crisis returned by the logit model, we can define a binary variable \tilde{C}_t such that:

$$\tilde{C}_t = \begin{cases} 1 & \text{if } \hat{C}_t \geq \text{threshold} \\ 0 & \text{otherwise.} \end{cases} \quad (3.4)$$

This way, we have T pairs of values (C_t, \tilde{C}_t) which, at any t , can form four possible permutations: either they are both equal to 1 or 0, or they are different. The percent of correctly predicted indicators is the number of times where $C_t = \tilde{C}_t$ relative to T .

3.1 Model calibration

More attention can be paid to the entropy indexes in Rényi (1960) and Tsallis (1988). As they embed the parameter α , they allow the researcher to calibrate (i.e., to fine-tune) the models.

Economically, it means identifying how much the tails of the distributions of MES are relevant to the prediction of the crises. Intuitively, the more spread the distribution, and the “fatter” its tails. As already discussed, the parameter α in the entropy definitions helps understanding how important are the tails. One may further develop the argument stating that such parameter assigns more or less weight to the degree of uncertainty — or “disorder” — of the scenarios.

Statistically, that means minimizing a loss function. Such function is identi-

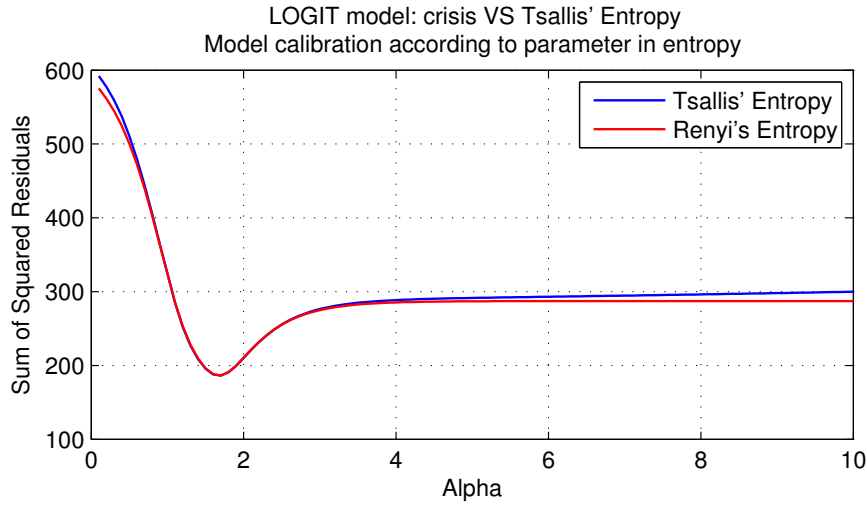


Figure 3.3: Sum of Squared Residuals as functions of α .

fied, in this Thesis, with the Sum of Squared Residuals (SSR), namely

$$\text{SSR} = \sum_{t=1}^T (C_t - \hat{C}_t)^2, \quad (3.5)$$

where \hat{C}_t is the estimated probability of crisis returned by the logit model in Equation 3.1. In other words, α is chosen in order to “stabilize” the estimates, so that they tend to be “stickier” to the true values of the dependent variable. It also means that changes in the regressors have greater impact in the estimates of the probability of crisis.

For the entropies in Rényi (1960) and Tsallis (1988), \hat{C}_t is a function of α , as it derives from the logistic regression run on those indexes. We can therefore want to minimize the SSR according to that parameter. That translates to

$$\min_{\alpha} \text{SSR}(\alpha) = \min_{\alpha} \sum_{t=1}^T (C_t - \hat{C}_t(\alpha))^2. \quad (3.6)$$

For many positive values of α , the logit models have been estimated and the SSR calculated. Figure 3.3 shows the behavior of SSR as a function of α for both Rényi’s and Tsallis’ entropy indexes. The implication of such behavior is twofold: asymptotically, the Rényi’s entropy has to be preferred to the Tsallis’ one. The second implication has to do with the common minimizer. For both entropies, $\min_{\alpha} \text{SSR}$ is achieved by setting $\alpha = 1.7$. This confirms the intuition that the two entropies behave somewhat symmetrically for values of α positive but not equal to the unit.¹ For the calibrated model, the logit estimates are described in Tables 3.2 and 3.3 and depicted in Figures 3.4, 3.5, 3.6 and 3.7.

¹If $\alpha \rightarrow 1$, the entropies converge to the Shannon’s one.

Variable	Coefficient	St. Error	<i>t</i> -ratio	<i>p</i> -value
<i>constant</i>	-60.2418	2.7521	-21.8896	0.0000
Entropy	27.8391	1.2809	21.7336	0.0000

No. of observations: 2870 (From 1-Jan-2000 to 31-Dec-2010)

Percentage of correctly predicted indicators: 88.89%

Table 3.2: Logit estimates on Rényi's Entropy ($\alpha = 1.7$).

Variable	Coefficient	St. Error	<i>t</i> -ratio	<i>p</i> -value
<i>constant</i>	-138.5621	6.3163	-21.9373	0.0000
Entropy	124.3537	5.6838	21.8785	0.0000

No. of observations: 2870 (From 1-Jan-2000 to 31-Dec-2010)

Percentage of correctly predicted indicators: 89.20%

Table 3.3: Logit estimates on Tsallis' Entropy ($\alpha = 1.7$).

An overall comment about the results is that the models perform fairly well. As expected by the minimization of SSR, the estimates tend to “stick” to the real values of the crisis indicator. The estimates correctly react to the crises at the beginning of the 21st century, where the *dot-com* bubble and the terrorist attack to the Twin Towers in New York have largely impacted the financial markets. Moreover, some sharp movements of the predicted probability of crisis can be noted to happen in 2006, where the first instabilities in the market for houses in the US appeared.² The model also highlights the instabilities that began in mid 2007. Recall that, on 14th September 2007, Northern Rock's shares had a crash in the stock market due to a bank run. See Wallop (2007) on this.

Comparing the results with the model on Shannon's entropy, one can see that different measures of entropy basically make the new estimates more responsive with respect to changes in the independent variable. That is, the intuition that the entropies in Rényi (1960) and Tsallis (1988) assign different weights to the tails of the distributions is confirmed. This can be easily seen in the second half of 2006.³ The entropy in Shannon (1948) reacts somewhat fairly, while the models with the Rényi's and Tsallis' entropy measures feature sharper changes in the estimated probability of crisis.

3.2 Robustness check

One may wonder whether or not the estimates presented so far are dependent on the available sample. Given that the exogenous variable is based on probabilities

²Recall that one of the ICB supersectors included in the dataset is made of firms active in the Real Estate financial markets (supersector 8600).

³Compare Figures 3.2 3.5 and 3.7.

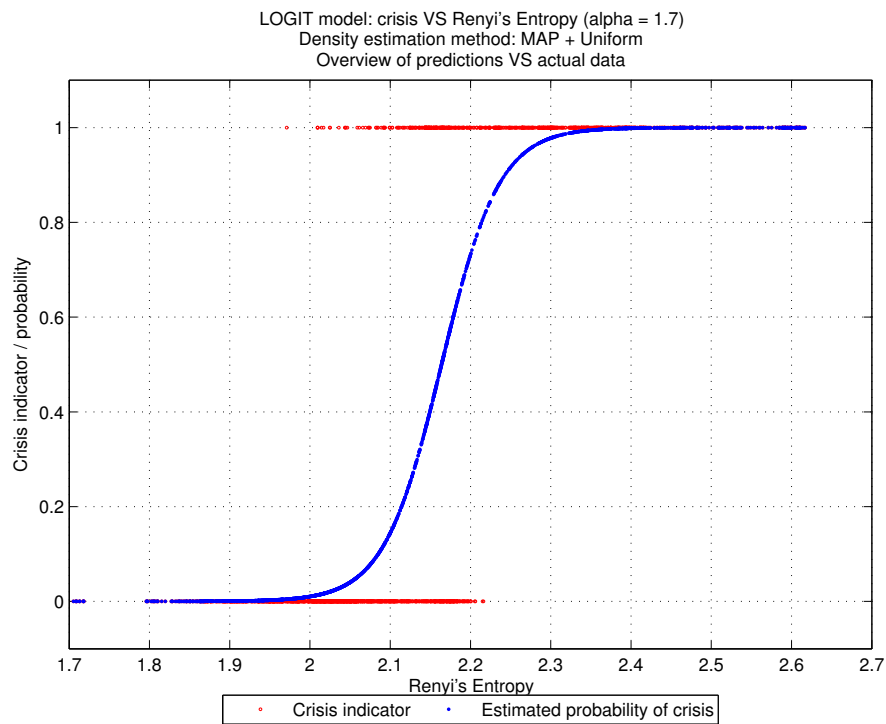


Figure 3.4: Scatter-plot of actual and estimated response variable against the Renyi's entropy.

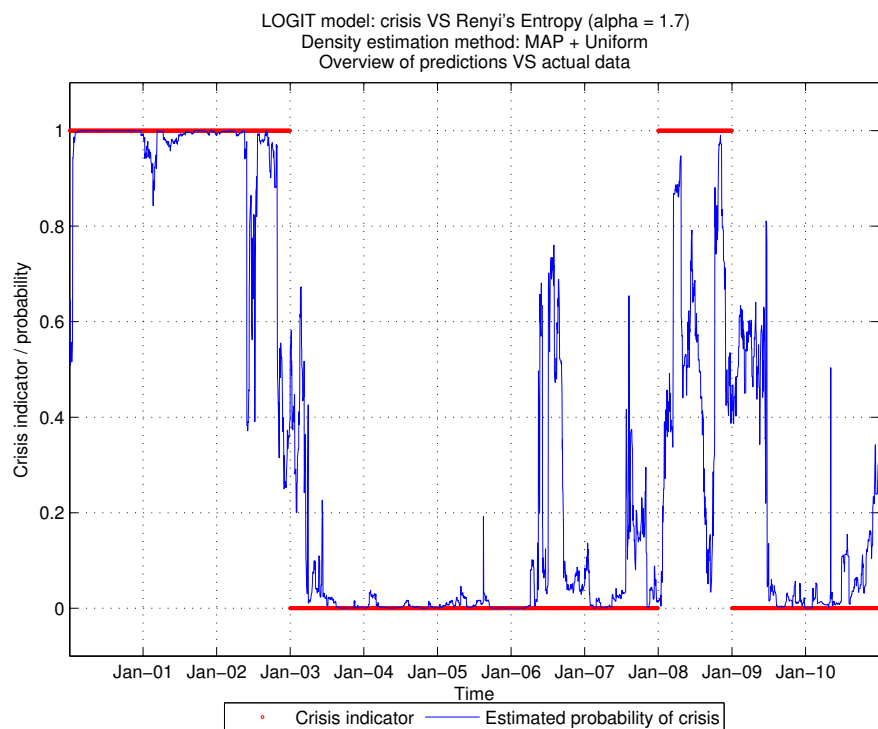


Figure 3.5: Overview of actual and estimated response variable over time.

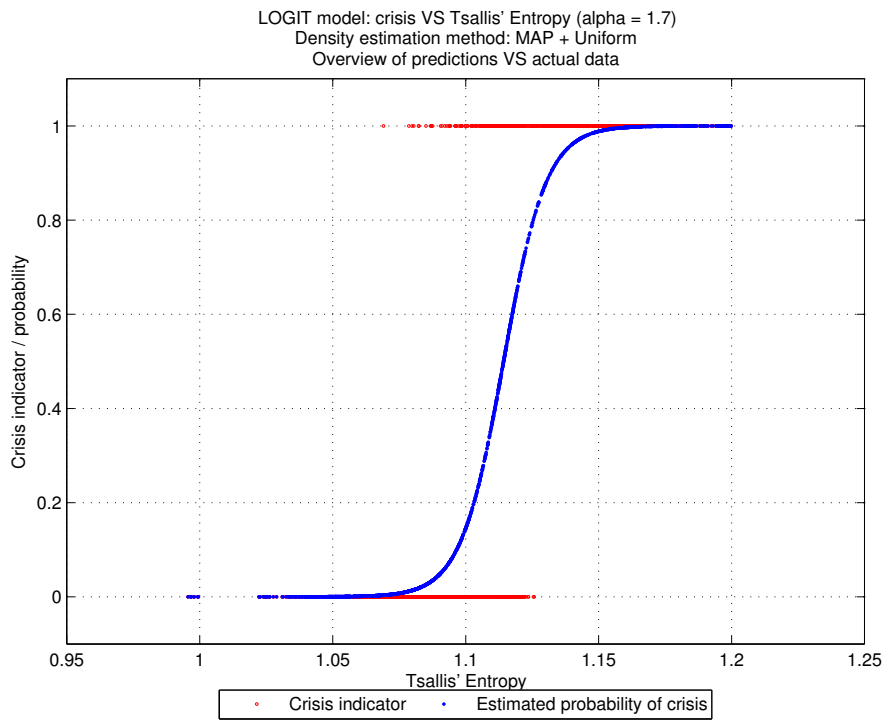


Figure 3.6: Scatter-plot of actual and estimated response variable against the Tsallis' entropy.

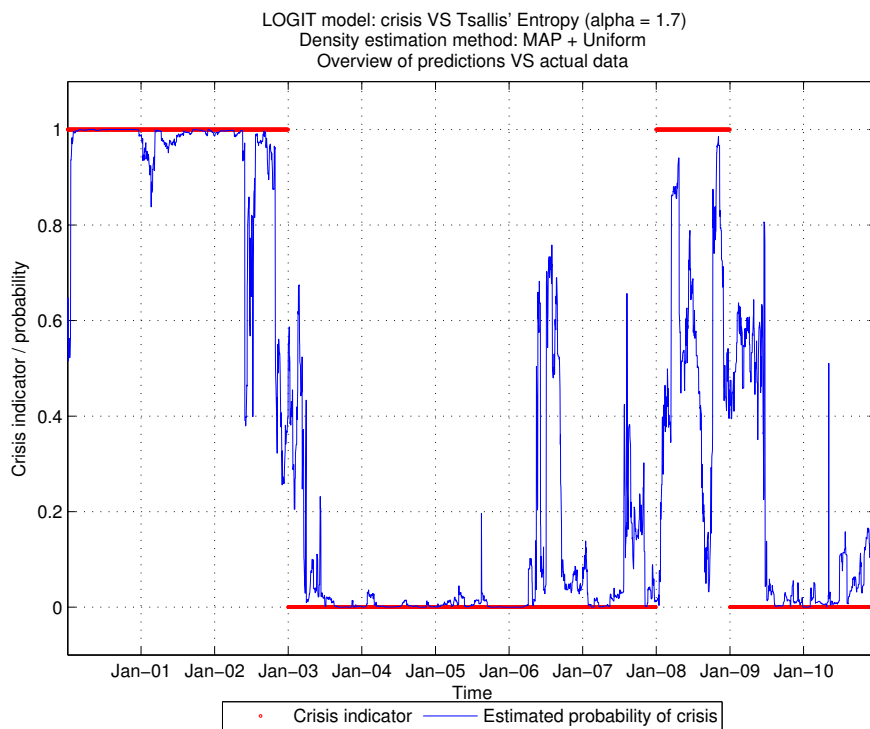


Figure 3.7: Overview of actual and estimated response variable over time.

(rather than directly on MES quantiles), a way of checking the robustness of the model is to change the way such probabilities are estimated.

In Section 2.2, different density estimation methods have been introduced, which have to do with the bayesian non-parametric multinomial approach, as thoroughly studied by Minka (2003). As already described in Subsection 2.2.1, MES has been discretized and it is assumed to come from a multinomial distribution.

Here I evaluate the same logit models as in Equation 3.1, taking into account the same entropy indexes. Different densities have been estimated using the posterior predictive function with a uniform prior and the Jeffreys' prior. To recall, such *posterior predictive function* is expressed as

$$\mathbb{E}[p_k|\mathbf{X}] = \frac{N_k + \alpha_k}{N + \sum_{k=1}^K \alpha_k} \quad (3.7)$$

$$= \frac{N_k + 1}{N + K} \quad \text{with a uniform prior,} \quad (3.8)$$

$$= \frac{N_k + 1/2}{N + K/2} \quad \text{with the Jeffreys' prior.} \quad (3.9)$$

As already noted in Figures 2.5 and 2.6, the effects of such methods is that the modes of the distributions are less likely, that is the probability of observing the mode is less than the same probability in the frequentist approach, while the probability associated to the tails is not zero, which is to say that the tails are “fatter” in these estimates.

Table 3.5 compares all the calibrated models, for the three entropy indexes and for the three different bayesian estimates. It can be noted that no much difference occurs between the models. The values of α that calibrates the model goes from 1.7 to 1.6. The minima of SSRs are at different heights, confirming that the frequentist approach is “stickier” to the dependent variable. Considering both the percent of correctly predicted probabilities and the SSR, all the models based on the Maximum *A Posteriori* estimate with uniform prior perform slightly better than the others, but the improvement is not of critical magnitude. In contrast, when evaluating which entropy index is to be preferred, the formulation in Rényi (1960) is marginally better than the others, apart for the frequentist density estimation case, where the entropy in Tsallis (1988) performs best.

In any case, however, the main insights of the model are proved to be robust, at least with respect to different estimation methods. All the t -ratios are at comparable levels and no p -value changes in a way that calls into question the significance of any regressor. Table 3.4 provides an insight about the magnitude of the p -values.

The main conclusion on this is that entropy, no matter how it is defined, is a reliable factor in explaining a crisis.

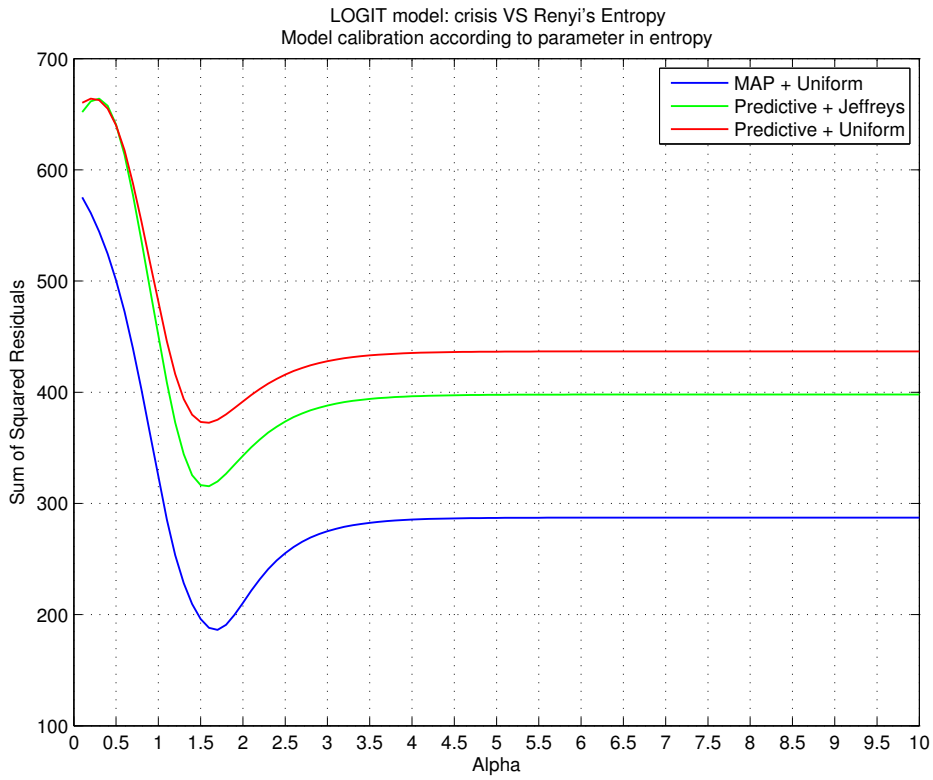


Figure 3.8: Sum of Squared Residuals as functions of Renyi's α .

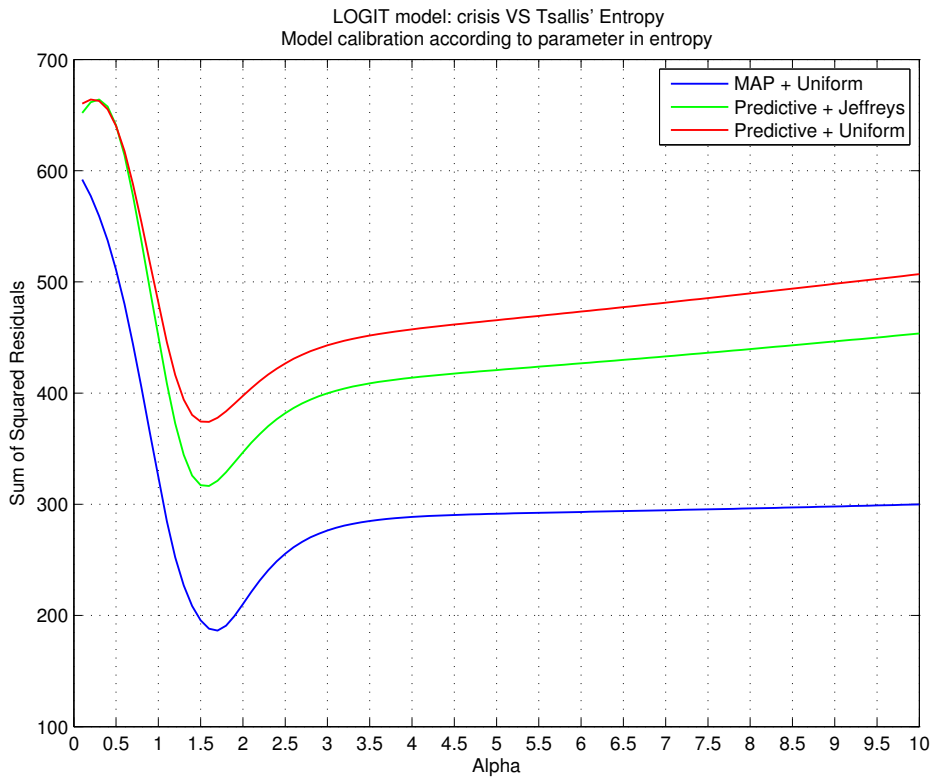


Figure 3.9: Sum of Squared Residuals as functions of Tsallis' α .

Entropy index	Estimator	<i>p</i> -values	
		<i>a</i>	<i>b</i>
Shannon	MAP + uniform	1.2168	-168
	Pred. + uniform	8.5118	-128
	Pred. + Jeffreys	1.4312	-148
Renyi	MAP + uniform	9.8718	-105
	Pred. + uniform	1.0449	-146
	Pred. + Jeffreys	8.2805	-150
Tsallis	MAP + uniform	4.1629	-106
	Pred. + uniform	2.5954	-149
	Pred. + Jeffreys	5.7417	-153

Table 3.4: The *p*-values for the coefficients of the entropy indexes expressed in scientific notation $a \times 10^b$.

	Variable	Coefficient	St. Error	<i>t</i> -ratio	<i>p</i> -value
Shannon	<i>constant</i>	-32.3465	1.1631	-27.8106	0.0000
	MAP + Uniform [86.48%]	Entropy 10.9709	0.3963	27.6800	0.0000
Shannon	<i>constant</i>	-185.0967	7.6778	-24.1080	0.0000
	Pred. + Uniform [73.94%]	Entropy 30.2427	1.2575	24.0492	0.0000
Shannon	<i>constant</i>	-118.9344	4.5665	-26.0449	0.0000
	Pred. + Jeffreys [75.40%]	Entropy 21.5589	0.8305	25.9592	0.0000
Renyi ($\alpha = 1.7$)	<i>constant</i>	-60.2418	2.7521	-21.8896	0.0000
	MAP + Uniform [88.89%]	Entropy 27.8391	1.2809	21.7336	0.0000
Renyi ($\alpha = 1.6$)	<i>constant</i>	-75.7430	2.9128	-26.0032	0.0000
	Pred. + Uniform [80.91%]	Entropy 15.6532	0.6069	25.7936	0.0000
Renyi ($\alpha = 1.6$)	<i>constant</i>	-69.7720	2.6512	-26.3167	0.0000
	Pred. + Jeffreys [83.17%]	Entropy 17.6278	0.6762	26.0686	0.0000
Tsallis ($\alpha = 1.7$)	<i>constant</i>	-138.5621	6.3163	-21.9373	0.0000
	MAP + Uniform [89.20%]	Entropy 124.3537	5.6838	21.8785	0.0000
Tsallis ($\alpha = 1.6$)	<i>constant</i>	-440.7371	16.9129	-26.0593	0.0000
	Pred. + Uniform [80.70%]	Entropy 279.8046	10.7515	26.0248	0.0000
Tsallis ($\alpha = 1.6$)	<i>constant</i>	-281.8222	10.6735	-26.4039	0.0000
	Pred. + Jeffreys [82.96%]	Entropy 186.4599	7.0774	26.3457	0.0000

Table 3.5: Results from the logit fits for all the density estimation methods. Numbers in brackets show the percent of correctly predicted indicators given the threshold in Equation 3.3. For all the models, the number of observations is 2870 (from 1st January 2000 to 31st December 2010).

Chapter 4

Conclusions

In Chapter 3, I have outlined the main numerical results for various logit specifications. For each density estimate, three measures of entropy have been used as independent variables. Considering the fact that each model has only one regressor, the models sport fairly well and each of them is statistically significant.

In the course of this Thesis, some assumptions have been made, sometimes also implicitly. Expanding the banking crisis indicator in Reinhart and Rogoff (2008) from its yearly frequency to a daily frequency required the strong assumption that, for a given year, if a country is facing a crisis, then it is facing it all the days in that year. This represents a choice that can be addressed in future works. A possible substitution of such dependent variable is the database about the *standing facilities* offered by the European Central Bank. In particular, given that banks are all connected to each other also through such facilities, one may pinpoint and further develop the idea of systemic risk, especially in the banking sector. The extent to which each bank resorts to the standing facilities may represent a good indicator of how much all the other banks are exposed to external risks.

Another assumption is the use of data about assets only. While such choice has been justified by the work of Löffler and Raupach (2013), who state that MES is not robust with respect to other financial instruments, working out the assumption may constitute a valid direction of advances. Exploring the role of bonds and derivatives in MES and in the entropy of returns may provide additional insights about the distributions of returns over time. An accurate study of the effects of non-asset instruments on returns may allow to capture relevant non-linearities in the behavior of the markets.

The models presented here rely on the logit specification, so that one may interpret how much “disorder” contributes to a banking crisis. Such reliance on a regression model may be relaxed. Along with the bayesian framework, Minka (2003) also provided a test for homogeneity in the distributions. He shows that the probability that two samples X and Y come from the same multinomial distribution can

be expressed in terms of the odds in favor of difference. The log-odds can be approximated by the Jensen-Shannon divergence. Therefore, one can use such test to understand when samples come from different distributions, and see whether or not there is correlation between such changes and the crisis indicator.

However, the main economic implications from the contents of this Thesis are two. On one hand, one understands that MES is a valuable starting point for analyzing systemic risk. It provides a simple, yet effective, “rule of thumb” for filtering data, so that one focuses on those states where an institution or a whole market are experiencing financial troubles. On the other hand, although the concept of entropy may be seen as a trivial expression of variability, it allows some considerations about the agents’ behavior. As in thermodynamics, entropy is interpreted as a measure of “disorder,” with the underlying idea that entropy acts as a count for the number of possible outcomes that can be witnessed. This directly follows from the conditions set in Shannon (1948) in order to define entropy in that specific way.

Thus, entropy can be employed in order to micro-found a model where agents in the financial market react to disorder. A behavioral approach in terms of Prospect Theory or Behavioral Finance may lead to the conclusion that agents not only obey the advice given by the pure mathematical understanding of the financial state, but also other biasing elements, linked to the situation of the financial market, are at play when making financial decisions.

Appendix A

The database in Reinhart and Rogoff (2008)

The database from Reinhart and Rogoff contains qualitative information built on quantitative sources, which are not available in the same database. The time frame ranges from 1800 to 2010 and features annually-paced binary variables.

Every variable in the Reinhart-Rogoff database is binary. Table A.1 lists all variables and their definitions. Table A.2 lists all analyzed countries.

Variable	Equals one when...
Independence Year	The first year of independence occurs and all subsequent years.
Inflation	The annual inflation is 20% or higher.
Currency crash	The annual depreciation rate against USD (or any more relevant anchor currency, like GBP, DEM, FRF or EUR) is 15% or higher.
Currency debasement (type 1)	The reduction in the metallic content of coins in circulation is 5% or higher.
Currency debasement (type 2)	A currency reform replaces a much-depreciated earlier currency (like in China, in 1948).
Banking crisis (systemic)	Bank runs occur and cause closure, merge or public takeovers of one or more financial institutions.
Banking crisis (distress)	No bank runs occur, but there anyway is a closure, merge or public takeovers of one financial institution that cause a series of similar events.
Debt crisis (external)	There is a failure to meet a principal or interest payment on the due date.
Debt crisis (domestic)	The definition for the external debt crisis applies <i>and</i> domestic bank deposits are frozen or forced to be converted from USD to the local currency.

Table A.1: List of binary variables in Reinhart and Rogoff (2008).

Algeria	Angola	Argentina
Australia	Austria	Belgium
Bolivia	Brazil	Canada
Central African Republic	Chile	China
Colombia	Costa Rica	Cote D'Ivoire
Denmark	Dominican Republic	Ecuador
Egypt	El Salvador	Finland
France	Germany	Ghana
Greece	Guatemala	Honduras
Hungary	Iceland	India
Indonesia	Ireland	Italy
Japan	Kenya	Korea
Malaysia	Mauritius	Mexico
Morocco	Myanmar	Netherlands
New Zealand	Nicaragua	Nigeria
Norway	Panama	Paraguay
Peru	Philippines	Poland
Portugal	Romania	Russia
Singapore	South Africa	Spain
Sri Lanka	Sweden	Switzerland
Taiwan	Thailand	Tunisia
Turkey	United Kingdom	United States
Uruguay	Venezuela	Zambia
Zimbabwe		

Table A.2: List of countries analyzed in Reinhart and Rogoff (2008).

Appendix B

The database in Laeven and Valencia (2012)

The database used by Laeven and Valencia (2012) features the variables of interest in their paper, which are mainly macroeconomic and related to GDP, liquidity in the banking system, fiscal and monetary policies.

The database is both included in the paper and as an Excel annex available at the IMF website.¹

The database does not provide “raw” data, but makes reference to aggregated and calculated summary data referred to as “Authors’ calculations” and “Staff reports.” Most of the figures are expressed in ratios to (or percentage of) GDP or other macroeconomic variables like foreign liabilities kept by the country’s domestic banking system.

No frequency is clearly stated in the paper, although (at least) monthly-paced data is suggested by a graph which indicates how many crises arose on each month of the year. The data relates to the period 1970-2011.

Table B.1 lists all macroeconomic variables used and discussed in the paper, along with a short description. Table B.2 lists all countries analyzed in the IMF paper. The database features 162 countries.

¹<http://www.imf.org/external/pubs/ft/wp/2012/Data/wp12163.zip>

Variable	Definition
Output loss	In percent of GDP, it is the cumulative sum of the differences between actual and trend real GDP over the period $T, T + 1, T + 2, T + 3$, where T denotes the starting year of a crisis.
Fiscal costs	They include gross fiscal outlays related to the restructuring of the financial sector.
Gross outlays and recoveries	They measure fiscal outlays (and subsequent repayments by banks) in percent of GDP.
Peak liquidity	Liquidity is measured as the ratio [central bank claims on deposit money banks] + [liquidity from the Treasury] over [total deposits and liabilities to non-residents]. Total deposits are the sum of [demand deposits] + [other deposits] + [liabilities to non-residents]. Reference is made to IFS lines.
Liquidity support	See "Peak liquidity." Relates to support measures (other than fiscal outlays) which contributed to fuel the banking system with new liquidity (like large purchases of real and/or financial assets).
Peak NPLs	Non-Performing Loans in percent of total loans. Data come from IMF Staff reports and Financial Soundness Indicators.
Public debt	Measured in percent of GDP. Increases in public debt are measured over the interval $[T - 1, T + 3]$, where T denotes the starting year of a crisis.
Monetary expansion	Measured as the change in the monetary base between its peak during the crisis and its level one year prior to the crisis.
Credit boom	As defined in Dell'Ariccia et al. (2012)

Table B.1: List of variables in Laeven and Valencia (2012).

Albania	Algeria	Angola
Argentina	Armenia	Australia
Austria	Azerbaijan	Bangladesh
Barbados	Belarus	Belgium
Belize	Benin	Bhutan
Bolivia	Bosnia and Herzegovina	Botswana
Brazil	Brunei	Bulgaria
Burkina Faso	Burundi	Cambodia
Cameroon	Canada	Cape Verde
Central African Rep.	Chad	Chile
China, P.R.	Colombia	Comoros
Congo, Dem. Rep. of	Congo, Rep. of	Costa Rica
Côte d'Ivoire	Croatia	Czech Republic
Denmark	Djibouti	Dominica
Dominican Republic	Ecuador	Egypt
El Salvador	Equatorial Guinea	Eritrea
Estonia	Ethiopia	Fiji
Finland	France	Gabon
Gambia, The	Georgia	Germany
Ghana	Greece	Grenada
Guatemala	Guinea	Guinea-Bissau
Guyana	Haiti	Honduras
China, P.R.: Hong Kong	Hungary	Iceland
India	Indonesia	Iran, I.R. of
Ireland	Israel	Italy
Jamaica	Japan	Jordan
Kazakhstan	Kenya	Korea
Kuwait	Kyrgyz Republic	Lao People's Dem. Rep.
Latvia	Lebanon	Lesotho
Liberia	Libya	Lithuania
Luxembourg	Macedonia	Madagascar
Malawi	Malaysia	Maldives
Mali	Mauritania	Mauritius
Mexico	Moldova	Mongolia
Morocco	Mozambique	Myanmar
Namibia	Nepal	Netherlands
New Caledonia	New Zealand	Nicaragua
Niger	Nigeria	Norway
Pakistan	Panama	Papua New Guinea
Paraguay	Peru	Philippines
Poland	Portugal	Romania
Russia	Rwanda	São Tomé and Príncipe
Senegal	Serbia, Republic of	Seychelles
Sierra Leone	Singapore	Slovak Republic
Slovenia	South Africa	Spain
Sri Lanka	Sudan	Suriname
Swaziland	Sweden	Syrian Arab Republic
Switzerland	Tajikistan	Tanzania
Thailand	Togo	Trinidad and Tobago
Tunisia	Turkey	Turkmenistan
Uganda	Ukraine	United Kingdom
United States	Uruguay	Uzbekistan
Venezuela	Vietnam	Yemen
Yugoslavia, SFR	Zambia	Zimbabwe

Table B.2: List of countries analyzed in Laeven and Valencia (2012).

Appendix C

Own defined functions in MatLab

C.1 Density estimation

```
function p = minka(x,method,prior)
% MINKA      estimates data density as in Minka (2003)
%
%   P = minka(X,METHOD,PRIOR)
%
%   X is a vector of data, whose density has to be estimated.
%
%   The function provides two methods and two priors.
%   At this stage of the development of the script, no other
%   priors can be used without modifying the script.
%
%   METHOD is either 'predictive' or 'map'
%   PRIOR is either 'jeffreys' or 'uniform'
%
%   Note that P = minka(X,'map','uniform') corresponds to the
%   frequentist relative frequency approach.
%
%   Reference: Minka, Thomas P. (2003),
%               "Bayesian inference, entropy, and the multinomial
%               distribution"
%               Technical Report, Microsoft

k = -0.5:0.001:0.5;
K = length(k);
N = length(x(~isnan(x)));
Nk = hist(x(~isnan(x)),k);
```

```

switch [method prior]
    case 'predictivejeffreys'
        p = (Nk + 1/2) / (N + K/2);
    case 'predictiveuniform'
        p = (Nk + 1) / (N + K);
    case 'mapjeffreys'
        p = (Nk - 1/2) / (N - K/2);
        warning(['Make sure the output sums to 1 and every '...
                'element of the output is between 0 and 1!'])
    case 'mapuniform'
        p = (Nk / N);
    otherwise
        error('Method and/or prior have been misspecified')
end

end
end

```

C.2 Entropy formulations

C.2.1 Shannon (1948)

```

function H = shannon(p,base)
% SHANNON calculates the entropy index as in Shannon (1948)
%
% H = shannon(P,BASE)
%
% P is a vector of probabilities, or relative frequencies
% BASE can be either 'natural' or 'binary' and refers
% to the logarithm in the formula.
%
% Reference: Shannon, Claude E. (1948)
% "A Mathematical Theory of Communication"
% Bell System Technical Journal 27 (3): 379-423
%

switch base
    case 'binary'
        H = -sum(p(p>0).*log2(p(p>0)));
    case 'natural'

```

```

        H = -sum(p(p>0).*log(p(p>0)));
end

end

```

C.2.2 Rényi (1960)

```

function H = renyi(p,alpha,base)
% RENEYI calculates the entropy index as in Renyi (1961)
%
% H = renyi(P,ALPHA,BASE)
%
% P is a vector of probabilities, or relative frequencies.
% ALPHA is a scalar parameter.
% BASE can be either 'natural' or 'binary' and refers
% to the logarithm in the formula.
%
% Reference: Maszczyk, T. and Duch, W. (2008)
% "Comparison of Shannon, Renyi and Tsallis Entropy
% used in Decision Trees"
%

%% checking for errors
if nargin == 1
    error('A value of ALPHA must be specified')
end

%% checking for errors
if alpha == 1
    error('Infeasible inputs (ALPHA can''t be 1)')
end

%% checking for errors
if sum(p.^alpha) <= 0
    error('Infeasible inputs (domain of log)')
end

%% main part of the function
switch base
    case 'binary'

```

```

        H = (1/(1-alpha))*(log2(sum(p.^alpha)));
    case 'natural'
        H = (1/(1-alpha))*(log(sum(p.^alpha)));
end
end
end

```

C.2.3 Tsallis (1988)

```

function S = tsallis(p,alpha)
% TSALLIS calculates the entropy index as in Tsallis (1988)
%
% S = tsallis(P,ALPHA)
%
% P is a vector of probabilities, or relative frequencies.
% ALPHA is a scalar parameter.
%
% Reference: Maszczyk, T. and Duch, W. (2008)
%           "Comparison of Shannon, Renyi and Tsallis Entropy
%           used in Decision Trees"
%
%% checking for errors
if nargin == 1
    error('A value of ALPHA must be specified')
end

%% checking for errors
if alpha == 1
    error('Infeasible inputs (ALPHA can''t be 1)')
end

%% main part of the function
S = (1/(alpha-1))*(1-sum(p.^alpha));

end

```

Bibliography

- Acharya, V. V., L. H. Pedersen, T. Philippon, and M. P. Richardson (2010). Measuring systemic risk. *Working Paper, New York University*.
- Adrian, T. and M. K. Brunnermeier (2011). Covar. *Staff Report, Federal Reserve Bank, New York* (348).
- Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon (2011). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Working Papers, Department of Economics, Ca' Foscari University of Venice* (21).
- Bisias, D., M. Flood, A. W. Lo, and S. Valavanis (2012). A survey of systemic risk analytics. *Annual Review of Financial Economics* 4(1), 255–296.
- Boucher, C., P. Kouontchou, B. Maillet, and O. Scaillet (2013). The co-covar and some other fair systemic risk measures with model risk corrections, mimeo. *Working Paper, Variances and Univ. Lorraine*.
- Boudebbous, T. and J. E. Chichti (2013). Determinants of systemic banking crises in the countries of central and eastern europe. *Journal of Business Studies Quarterly* 5(1).
- Brownlees, C. T. and R. F. Engle (2012). Volatility, correlation and tails for systemic risk measurement. *Workin Paper, Department of Finance, Stern School of Business, New York University*.
- Burda, M., R. Liesenfeld, and J.-F. Richard (2008). Bayesian analysis of a probit panel data model with unobserved individual heterogeneity and autocorrelated errors. *Working Paper, Department of Economics, University of Toronto* (tecipa-321).
- Bussière, M. and M. Fratzscher (2006). Towards a new early warning system of financial crises. *Journal of International Money and Finance* 25(6), 953–973.
- Danielsson, J., K. R. James, M. Valenzuela, and I. Zer (2012). Dealing with systemic risk when we measure it badly. *Working Paper, London School of Economics*.

- Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47, 263–291.
- Laeven, L. and F. Valencia (2008). Systemic banking crises: a new database. *IMF Working Paper* (224).
- Laeven, L. and F. Valencia (2012). Systemic banking crises database: An update. *IMF Working Paper* (163).
- Löffler, G. and P. Raupach (2013). Robustness and informativeness of systemic risk measures. *Available at SSRN 2264179*.
- Maszczyk, T. and W. Duch (2008). Comparison of Shannon, Renyi and Tsallis entropy used in decision trees. In *Artificial Intelligence and Soft Computing—ICAISC 2008*, pp. 643–651. Springer.
- Minka, T. P. (2003). Bayesian inference, entropy, and the multinomial distribution. *Technical Report, Microsoft*.
- Minka, T. P. (2012). Estimating a Dirichlet distribution. *Technical Report, Microsoft*.
- Müller, P. and F. A. Quintana (2004). Nonparametric bayesian data analysis. *Statistical Science* 19(1), pp. 95–110.
- O’Brien, S. M. and D. B. Dunson (2004). Bayesian multivariate logistic regression. *Biometrics* 60(3), 739–746.
- Reinhart, C. M. and K. S. Rogoff (2008). This time is different: A panoramic view of eight centuries of financial crises. *Working Paper, National Bureau of Economic Research* (13882).
- Reinhart, C. M. and K. S. Rogoff (2010a). From financial crash to debt crisis. *Working Paper, National Bureau of Economic Research* (15795).
- Reinhart, C. M. and K. S. Rogoff (2010b). Growth in a time of debt. *Working Paper, National Bureau of Economic Research* (15639).
- Rényi, A. (1960). On measures of information and entropy. In *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, pp. 547–561.
- Shannon, C. E. (1948, July). A mathematical theory of communication. *The Bell System Technical Journal* 27(3), 379–423.
- Tabet, A. (2007). Bayesian inference in the multivariate probit model. *Master Thesis, University of British Columbia*.
- Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics* 52(1-2), 479–487.

- Wallop, H. (2007, September). Northern Rock shares crash as customers queue. <http://www.telegraph.co.uk/finance/newsbysector/banksandfinance/1563152/Northern-Rock-shares-crash-as-customers-queue.html>. Accessed: October 5, 2014.
- Wooldridge, J. (2009). *Introductory Econometrics: A modern approach* (Fourth ed.). South Western: Cengage Learning.
- Zivot, E. and J. Wang (2007). *Modeling Financial Time Series with S-PLUS*, Volume 191. Springer.