UNIVERSIDADE TÉCNICA DE LISBOA

INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

MESTRADO EM: ECONOMIA MONETÁRIA E FINANCEIRA

EXCESS RETURNS AND NORMALITY

JOÃO PEDRO DO CARMO BOAVIDA

Orientação: Doutor Paulo Brasil de Brito

Júri:

Presidente: Doutora Maria Rosa Borges

Vogais: Doutor José Dias Curto

Doutor Paulo Brasil de Brito

Junho/2011

Abstract

In this dissertation, I assess under which circumstances normality can be a good descriptive model for the U.S. excess returns. I explore two possible sources of deviations from normality: structural breaks and regime switching in long term aggregate time series. In addition, I study temporal aggregation (i.e., considering the frequency of data as a variable) for excess returns in short term time series. My main findings are summarized as follows. First, using long spanning monthly time series data from 1871 to 2010, I find that (1) there are structural breaks in monthly excess returns between pre-WWII and post-WWII data; and, (2) while pre-WWII data is consistent with normality, post-WWII data is not. Second, I provide evidence of two market regimes for excess returns in post-WWII data. These regimes may be seen as bull and bear market conditions. Third, using high frequency post-WWII data, I check for aggregational Gaussianity, from daily to annual data. I find that Gaussianity depends on the frequency of data: it may hold for highly aggregate data (starting from semi-annual to annual data) but it does not hold for high frequency data (less than semi-annual). My main contribution is to demonstrate the "normality survival" when frequency is taken as a variable. After a careful look at the available literature on aggregational Gaussianity, I found no previous applications and results for excess returns.

KEYWORDS: Excess returns, Normality, Structural breaks, regime switching, Time aggregation

JEL CODES: C1, G12

Resumo

Nesta dissertação, eu avalio sob que condições a normalidade pode ser um bom modelo descritivo para os excess returns nos E.U.A.. Para tal, exploro duas fontes potenciais de desvio da normalidade: quebras de estrutura e mistura de regimes para séries temporais longas agregadas. Adicionalmente, estudo a agregação temporal (i.e., tomando a frequência dos dados como variável) para os excess returns em séries temporais. Os principais resultados obtidos são os seguintes. Primeiro, utilizando dados mensais para um longo perodo temporal de 1871 até 2010, conclui-se que: (1) existem quebras de estrutura nos excess returns mensais entre o período antes e depois da Segunda Guerra Mundial (SGM) e, (2) enquanto os dados do período antes da SGM são consistentes com a normalidade, os dados do pós-guerra não são. Segundo, apresenta-se evidência de dois regimes de mercado para o período pós-SGM. Estes regimes podem ser vistos como descrevendo condições de mercado *bull* e *bear*. Terceiro, usando dados com frequências mais altas para o período pós-SGM, testa-se a *aggregational Gaussianity* para dados de diários até dados anuais. Conclui-se que a Gaussianity depende da frequência dos dados: pode ser válida para dados mais agregados (começando em dados semestrais até dados anuais) mas não é válida para dados com frequências mais altas (menores que semestrais). O principal contributo desta dissertação é demonstrar a sobrevivência da normalidade quando se toma a frequncia dos dados como variável. Após uma revisão aprofundada da literatura sobre *aggregational Gaussianity*, n ao encontrei resultados anteriores para os *excess returns*.

KEYWORDS: Excess returns, Normalidade, Quebras de Estrutura, Regime Switching, Agregação Temporal

JEL CODES: C1, G12

Agradecimentos

Ao Professor Doutor Paulo Brito pela sua preciosa colaboração e pela sua forma única de levar o conhecimento a um nível mais alto. Ao Professor Doutor Luís Costa pelos seus conselhos que me acompanharam ao longo da minha vida académica no ISEG. À Joana pela sua compreensão e o seu carinho demonstrado nos momentos mais difíceis. À minha irmã e aos meus pais pelas palavras de incentivo em superar este desafio. A Deus pela fé em tornar esta dissertação uma realidade.

Contents

List of Figures

List of Tables

1 Introduction

Normality of excess returns is maybe the central assumption in finance and macrofinance. Normality allows for the use of a powerful instrument such as the Itô calculus, which is the stepping stone of most of the results in theoretical asset pricing. However, empirical results have systematically pointed out to deviations from normality. In almost all the cases it has been established that empirical distributions of returns are skewed and leptokurtic (Mandelbrot, 1963; Fama, 1965; Campbell et al., 1997).

Should we consider most of finance theory as unrealistic, and therefore irrelevant? Should we just use empirical measurement without theory given the lack of appropriate theoretical background?

In this dissertation, I try to assess under which circumstances normality can be a good descriptive model for the U.S. excess returns. There are several reasons why normality may not hold. One branch of the literature studies deviations from normality assuming that parameters are time-varying, that is, assuming non-stationarity. In this context, deviations from normality in financial returns may arise if these are mean-reverting (Poterba and Summers, 1988) or characterized by GARCH features (Bollerslev et al., 1992). I explore other possible sources of deviations from normality: structural breaks and regime switching in long term aggregate time series. The existence of structural breaks or regime switching may provide a solution to obtain normality for excess returns conditional on subperiods or regimes, respectively. In addition, I study temporal aggregation (i.e., considering the frequency of data as a variable) for excess returns in short term time series.

The standard approach to test normality for excess returns uses the sample estimate of the mean and the sample estimate of the variance for a given sample period. This approach is acceptable if excess returns are assumed to be independent and identically distributed (i.i.d.) and, therefore, the estimated parameters - mean and variance - remain constant for the entire sample period. However, when long histories are being used, stationarity may not hold: for the sample period under investigation, the parameters of the probability distribution could be changing over time, experiencing shifts known as "structural breaks", or switching among a finite set of values, that is, they are "regime-switching".

Deviations from normality, such as structural breaks or regime switching, have been found for excess returns. Pástor and Stambaugh (2001) estimated the expected U.S. excess return in a Bayesian framework in which the prior distribution is subject to structural breaks. They reported a total of fifteen structural breaks for excess returns using a sample period covering 1834-1999. The most significant break, however, was found to be in the late 1990s.

Kim et al. (2005) adopted also a Bayesian likelihood analysis to compare four univariate models of excess returns using monthly data for a shorter sample period covering 1926-1999. They constructed Bayes factors based on the marginal likelihoods, which allowed them to make model comparisons and to test structural breaks. Their findings are similar to mine. First, they found statistical evidence of a structural break in excess returns around the WWII period. In particular, they found that there is a permanent reduction in the general level of stock market volatility in the 1940s. My methodology finds support of breaks in the variance but not in the mean of excess returns. Second, Kim et al. (2005) found that the empirical Bayes factors strongly favour the three models that incorporate Markov switching volatility over the simple iid model. My findings support evidence of two market regimes possibly connected with two different business cycle stages. In particular, I observe times where excess returns are expected to be low and highly volatile and times where, in contrast, excess returns behave in the opposite way.

Empirical research also indicates that the shape of the distribution of stock returns is not the same for different time scales (Cont, 2001 and Eberlein and Keller, 1995). For instance, it has been established that daily returns depart more from normality than monthly returns do (see, e.g., Blattberg and Gonedes, 1974; Eberlein and Keller, 1995; and Campbell et al., 1997). Campbell et al. (1997) compared the distributions of daily and monthly stock returns for two indices and ten stocks from the U.S. for the period 1962–94. They found that the non-Gaussian character (the skewness and kurtosis) displayed by the distributions of monthly data is significantly lower than that displayed by the distributions of daily data. This stylized fact, also known as "aggregational Gaussianity", indicates that as one increases the time scale over which returns are calculated their distribution looks more and more like a normal distribution (Cont, 2001). In particular, it has been observed that, as we move from higher to lower frequencies, the degree of leptokurtosis diminishes and the empirical distributions tend to approximate normality.

My main findings can be summarized as follows. First, using long spanning monthly time series data from 1871 to 2010, I find that (1) there are structural breaks in monthly excess returns between pre-WWII and post-WWII data; and, (2) while pre-WWII data is consistent with normality, post-WWII data is not. Second, I provide evidence of two different market regimes for excess returns in post-WWII data. These regimes may be seen as *bull* and *bear* market regimes. Third, using high frequency post-WWII data, I check for aggregational Gaussianity, from daily to annual data. I find that Gaussianity depends on the frequency of data: it may hold for highly aggregate data (starting from semi-annual to annual data) but it does not hold for high frequency data (less than semi annual).

In this dissertation, I incorporate two branches of literature that study deviations from normality in excess returns: (i) one that tests for the existence of structural breaks and regime switching and (ii) other that explores time aggregation. After a careful look at the available literature on aggregational Gaussianity, I found no previous applications and results on excess returns. My main contribution is to demonstrate the "normality survival" when frequency is taken as a variable.

This dissertation is organized as follows. In section 2, I define excess returns and I present the formal proof for the normality of log excess returns as well as some stylized facts for monthly excess returns. In Section 3, I test if non-normality is associated with the presence of structural breaks and regime switching for monthly data. In Section 4, I test for Gaussianity with disaggregated data. In Section 5, I conclude.

2 Excess returns and normality

2.1 Excess returns - a definition

Total returns for stocks are defined as the sum of returns from capital gains and reinvested dividends from a market stock index. Let P_t be the average price value of the stock index at period t and D_t the average dividend payment between periods $t-1$ and t. Thus, the continuously compounded return on the aggregate stock market at time t is

$$r_t = log(P_t + D_t) - log(P_{t-1}).$$

In addition, let r_t^f denote the average log-return on a riskless security at period t. Typically, yields from long-term Government Bonds have been used as a proxy for the relatively risk-free asset.

Excess return can be seen as the risk premium paid to investors for holding stocks - which are risky - instead of riskless securities. In this sense, the excess return may also be thought of as the payoff on an arbitrage market portfolio where investors go long on stocks and take short positions on riskless securities. In such market portfolio, the continuously compounded excess return at time t is defined by

$$z_t = r_t - r_t^f.$$

2.2 Normality and Itô calculus

Normality allows for the use of the Itô calculus, which is the stepping stone of almost all the results in theoretical asset pricing. In continuous-time asset pricing models, financial prices are assumed to be geometric Brownian motions, and, therefore, by the Itô's lemma, logarithmic returns follow a Normal distribution. This lognormality assumption is used, for instance, in the Black and Scholes model (1973).

In this part, using a continuous-time framework, I present the formal proof that the log excess returns are normally distributed.

Let the time be continuous and consider two assets: $X_1(t)$ being the stock price and $X_2(t)$ the risk-free price[1]. For the sake of simplicity, I assume for these assets that there are no accruing dividends nor accruing interest. The stochastic processes for the risky asset $X_1(t)$ and for the riskless asset $X_2(t)$ follow two diffusion processes,

$$dX_1(t) = \mu_1 X_1(t)\,dt + \sigma_1 X_1(t)\,dW(t)$$

and,

$$dX_2(t) = \mu_2 X_2(t)\,dt + \sigma_2 X_2(t)\,dW(t)$$

where μ_1 and μ_2 represent the expected rate of return and σ_1 and σ_2 the expected volatility. I explicitly assume that these parameters are constant and, therefore, that the processes are stationary. In addition, I consider that these two asset prices are subject to the same source of uncertainty, dW, which follows a standard Wiener process[2].

Let the excess returns $Z(t)$ be defined by $Z = \frac{X_1}{X_2}$, or in logarithms by,

$$z = \ln\left(\frac{X_1}{X_2}\right) = \ln(X_1) - \ln(X_2)$$

Itô's lemma, which is a stochastic calculus theorem, yields an equivalent continuous-time equation for the $d\ln X_1(t)$ and $d\ln X_2(t)$. To illustrate this, let $F(X_1(t), t) = \ln(X_1(t))$ and $G(X_2(t), t) = \ln(X_2(t))$ be functions of the log prices of the underlying assets X_1 and X_2 and time t. Therefore, from the Itô's lemma, $d\ln X_1(t)$ and $d\ln X_2(t)$ are also stochastic

[1]Without loss of generality, assume a small value of volatility for the riskless asset.

[2]A continuous stochastic process $W(t)$ is a Wiener process if (i) each increment follow a normal distribution and (ii) increments over non-overlapping time intervals are independent and identically distributed random variables.

process, satisfying,

$$d\ln X_1(t) = \left(\frac{\partial F}{\partial t} + \frac{\partial F}{\partial X_1}\mu_1 X_1 + \frac{1}{2}\frac{\partial^2 F}{\partial X_1^2}\sigma_1^2 X_1^2\right) dt + \left(\frac{\partial F}{\partial X_1}X_1\sigma_1\right) dW(t) \tag{2.1}$$

and,

$$d\ln X_2(t) = \left(\frac{\partial G}{\partial t} + \frac{\partial G}{\partial X_2}\mu_2 X_2 + \frac{1}{2}\frac{\partial^2 G}{\partial X_2^2}\sigma_2^2 X_2^2\right) dt + \left(\frac{\partial G}{\partial X_2}X_2\sigma_2\right) dW(t) \tag{2.2}$$

From (2.1) and (2.2), and rearranging the terms, we may also find the stochastic process for the excess returns,

$$dz(t) = d\ln(Z(t)) = d\ln(X_1(t)) - d\ln(X_2(t)) = \left[\mu_1 - \mu_2 - \frac{1}{2}\left(\sigma_1^2 - \sigma_2^2\right)\right] dt + (\sigma_1 - \sigma_2) dW(t) \tag{2.3}$$

where, excess returns $dz(t)$ follow a generalized Wiener process with a constant drift rate $\mu_1 - \mu_2 - \frac{1}{2}\left(\sigma_1^2 - \sigma_2^2\right)$ and a constant variance rate $(\sigma_1 - \sigma_2)^2$[3].

Finally, integrating out equation (2.3) between two arbitrary moments t and $t+1$, it is obtained the approximation which allows to apply the theoretical results from stochastic calculus into discrete time,

$$\ln(Z(t+1)) - \ln(Z(t)) \sim \phi\left\{\left[\mu_1 - \mu_2 - \frac{1}{2}\left(\sigma_1^2 - \sigma_2^2\right)\right], (\sigma_1 - \sigma_2)^2\right\}, \tag{2.4}$$

[3]A similar result may be achieved using the following set up:
Let $z(t) = G(X_1(t), X_2(t)) = \ln(X_1) - \ln(X_2)$ with $G(\cdot)$ a continuous and differentiable function in relation to the variables X_1 and X_2. From a Taylor series expansion for $dz(t)$ it follows that,

$$dz(t) = \frac{\partial G}{\partial X_1}dX_1 + \frac{\partial G}{\partial X_2}dX_2 + \frac{1}{2}\left(\frac{\partial^2 G}{\partial X_1^2}dX_1^2 + 2\frac{\partial^2 G}{\partial X_1 \partial X_2}dX_1 dX_2 + \frac{\partial^2 G}{\partial X_2^2}dX_2^2\right)$$

and, since $dt^2 = 0$, $dtdw = 0$ and $dw^2 = dt$, then, rearranging this equation,

$$dZ(t) = \left[\mu_1 - \mu_2 - \frac{1}{2}\left(\sigma_1^2 - \sigma_2^2\right)\right] dt + (\sigma_1 - \sigma_2) dW(t)$$

where, excess returns $dz(t)$ follow a generalized Wiener process with a constant drift rate $\mu_1 - \mu_2 - \frac{1}{2}\left(\sigma_1^2 - \sigma_2^2\right)$ and a constant variance rate $(\sigma_1 - \sigma_2)^2$.

Equation (2.4), states that the change in log excess returns between time t and some future time $t+1$ is normally distributed. Therefore, from Itô's lemma, a variable $Z(t)$ has a lognormal distribution if the natural logarithm of the variable $\ln(Z(t))$ is normally distributed.

2.3 Monthly series and some *stylized facts*

In this subsection, I present some stylized facts for U.S. monthly excess returns using a long time series of data spanning from 1871:1 to 2010:12. First, I describe the data and provide some descriptive statistics. Then, I seek if a Normal distribution fits the empirical density distribution of excess returns and I quantify the difference between these distributions. Finally, I present some formal normality tests.

2.3.1 Data

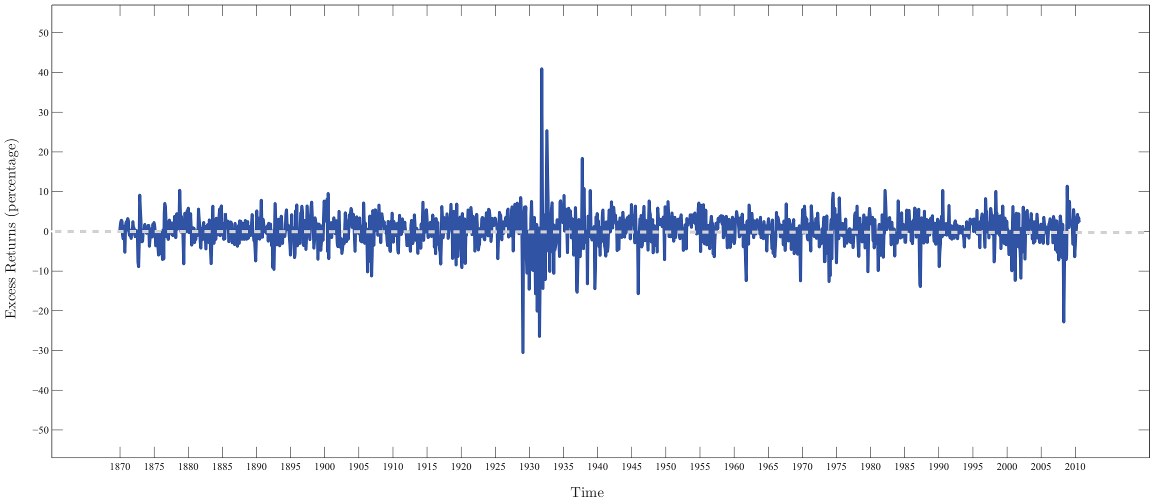
I use Robert Shiller's data (Shiller, 1989). These data are available for download in Shiller's website and consist of monthly averages of stock prices, dividends and interest rates, all starting in January 1871.

The stock price index is the Standard and Poor's Composite Index. The data source was Standard and Poor's Statistical Service Security Price Index Record, various issues, from Tables entitled "Monthly Stock Price Indexes - Long Term".

The dividend series is the total dividends per share for the period adjusted to the index. Monthly dividend data are computed from the S&P four-quarter tools since 1926, with linear interpolation to monthly figures. Dividend data before 1926 are from Cowles and associates (Common Stock Indexes, 2nd ed. [Bloomington, Ind.: Principia Press, 1939]), interpolated from annual data.

The interest rate series is the 10-year Treasury Bonds after 1953 published by the Federal Reserve. Before 1953, it is Government Bond yields from Sidney Homer "A History of Interest Rates", with linear interpolation to monthly figures.

Figure 1: Time series of U.S. monthly excess returns



2.3.2 Moments

Figure 1 plots the time series of U.S. monthly excess returns from 1871:01 to 2010:12.

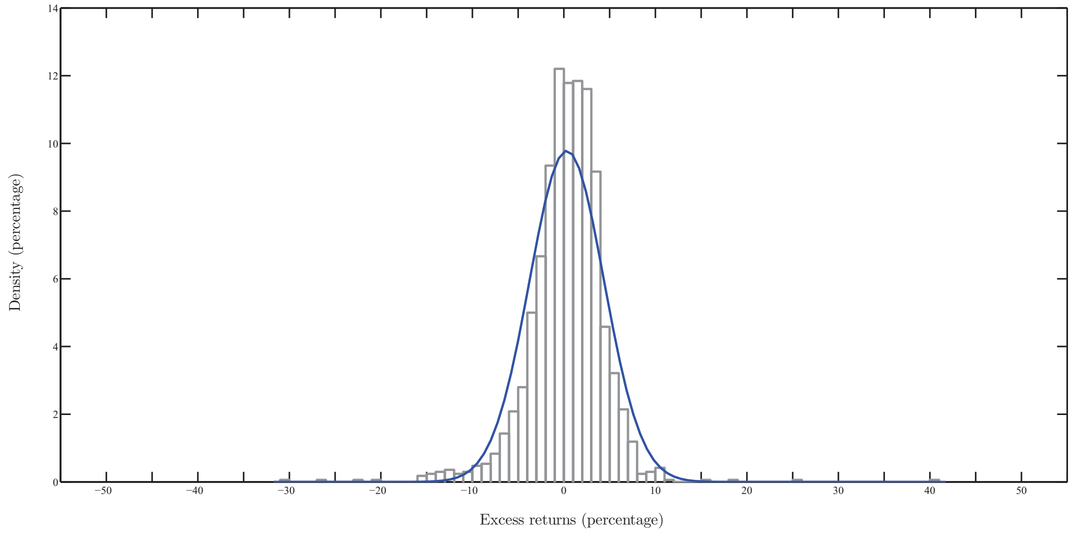
Excess returns have been highly volatile. Figure 1 displays the well-documented high volatility period around 1930, including the highest decline of 30% in October 1929, and the greatest increase of 41% in July 1932. Although to a lower extent, the post-war period has been also a period of high volatility. In particular, U.S. monthly excess returns decreased almost 23% in September 2008, which coincides with the start of the so-called Second Great Depression (Reinhart and Rogoff, 2008).

Table 1: Descriptive statistics

	S&P 500	**Risk-free**	**Excess returns**
Mean	0.0071	0.0039	0.0032
Std. Dev.	0.0407	0.0019	0.0408
Skewness	-0.36	1.84	-0.34
Kurtosis	14.43	6.75	14.39

Table 1 presents some descriptive statistics for the log-returns of the S&P500 stock index, the log-returns of the long-term U.S. Government Bonds and the U.S. monthly excess returns.

Figure 2: Empirical density of excess returns and the Normal PDF



The figures reveal that stocks have consistently outperformed the risk-free security yielding a monthly expected excess return (i.e., a equity premium) of 0.32%. However, stocks have been riskier and this caused a high variability in U.S. excess returns over the same period, with a monthly standard deviation of 4.08%.

In addition, Table 1 displays the statistics for the skewness and kurtosis - the third and fourth standardized moments. Skewness and kurtosis are used to summarize the extent of asymmetry and the tail thickness in a given empirical density distribution. For a Normal distribution, these statistics correspond to 0 and 3, respectively. Therefore, it seems that the Normal distribution does not hold for U.S. excess returns. Instead, the empirical distribution of excess returns is leptokurtic, exhibiting a higher value of kurtosis (more than 14), and asymmetric, with a negative skewness of -0.34[4]. This conclusion holds true also for the S&P500 index and the long-term U.S. Government Bonds.

Figure 2 plots an empirical density histogram against a theoretical Normal distribution. The empirical density distribution of U.S. excess returns is more peaked in the center and has fatter tails when compared to a Normal distribution. Overall, the Normal distribution

[4]The value for skewness is statistically different from 0 at a standard significance level.

exhibits a very poor fit through all range of the data: at the center, frequencies are substantially underestimated; at a medium range from the center, the empirical frequencies approach zero faster than the Normal alternative, i.e., they are overestimated; and, finally, at the tails, empirical frequencies become again underestimated.

But how large is the error made if we assume a Normal distribution as a theoretical model for U.S. excess returns? To quantify deviations from normality I build an empirical frequency distribution of excess returns and compare it with what would be expected if the same frequencies were calculated with a Normal distribution. The frequencies correspond to the empirical proportions of excess returns within a given number of standard deviations from the mean. I split the frequencies into three groups: Center, Medium and Tails. These groups were chosen accordingly to the number of excess returns in the range of $(\mu \pm k\sigma)$, for $k \in \mathbb{R}$ and μ, σ are the sample mean and the sample standard deviation of excess returns.

Table 2: Excess returns: empirical frequencies *vis-à-vis* Normal distribution

Position	**Range**	**Observed**	**Normal**	**Difference**
Center	0 to 1	79.05	68.27	10.78
Medium	1 to 3	19.40	31.46	-12.06
Tails	3+	1.55	0.27	1.28
Left Tail		1.31	0.13	1.17
Right Tail		0.24	0.13	0.10

Empirical frequencies of excess returns and theoretical Normal frequencies. Range refers to the number of standard deviations beyond the mean. Column "Difference" refers to the difference, in percentage, of empirical relative frequencies given in column "Observed", and the expected normal frequencies represented in column "Normal".

The magnitude of the deviations can be seen in Table 2. The first row of this Table shows that the empirical distribution of excess returns contains 10.78% more observations in the center than it would be expected if the distribution was Normal. The fact that a large proportion of excess returns remains near its sample mean explains the higher peaked center in the empirical distribution. On the other hand, at a medium range, the empirical frequencies are overestimated which, in Table 2, corresponds to the negative difference of 12.06%. However, the most important characteristic of empirical frequencies is the evidence of two significant tails. A general observation about these extreme events is that 1.55% of total excess returns correspond to more than three standard deviations from the mean. This is about six times the value of the Normal distribution. Furthermore, the probability of occurring extreme downward and upward movements in excess returns is not the same. In Table 2, the left tail has a relatively higher density frequency which means that crashes are more likely than booms. The Normal distribution, being a symmetric distribution, cannot capture this feature.

2.3.3 Normality tests

In order to formally examining the normality assumption, I ran several classes of tests for monthly excess returns. These tests have been widely used to study normality hypothesis in empirical financial studies and they can be implemented in almost all standard statistical packages. I carried out these tests using *Eviews* and *Matlab*.

The first test used is the Jarque-Bera test (Bera and Jarque,1987). This test uses information on the third and fourth moments of a distribution and is based upon the fact that for a Gaussian distribution, whatever its parameters, the skewness is 0 and kurtosis is 3.

The second test applied is the Shapiro-Wilk test (Shapiro and Wilk, 1965). This test assumes that, under the null hypothesis, the data are coming from a Normal population with unspecified mean and variance. This test is generally considered relatively more powerful against a variety of alternatives. Shapiro-Wilk test is used, instead of Shapiro-Francia, because it has a relatively better performance for leptokurtic samples.

Finally, I conducted three additional tests to get a more precise picture of normality: the Anderson-Darling test (Stephens, 1974), the Lilliefors test (Lilliefors, 1967) and the Cramér-von-Mises test (Cramér, 1928 and Anderson, 1962). These goodness-of-fit tests, which are extensions of the Kolmogorov Smirnov test, find the discrepancy between the empirical distribution and the Normal distribution using the sample estimated mean and variance. Then, they assess whether that discrepancy is large enough to be statistically significant, thus, requiring rejection of the Normal distribution under the null hypothesis.

Table 3: Normality tests in excess returns

Normality test	**Statistic value**	**P-value**
Jarque-Bera	9051.1	0.0000
Shapiro-Wilk	0.9030	0.0000
Anderson-Darling	20.696	0.0000
Lilliefors	0.0707	0.0000
Cramér-von-Mises	3.1279	0.0000

Table 3 gathers the results of the normality tests performed in excess returns. All the results point into the outright rejection of the normality assumption for U.S. monthly excess returns.

According to the empirical literature, it is a "stylised fact" that distributions of stock returns are poorly described by the Normal distribution (see Mandelbrot, 1963 and Campbell et al., 1997). This empirical regularity was first reported by Mandelbrot (1963) when he discovered fundamental deviations from normality in empirical distributions of daily returns. More recently, Campbell et al. (1997) found contradictory evidence indicating that monthly U.S. stock returns are reasonably well described by a Normal distribution. My results seem to be closely related with Kim, Morley and Nelson (2005). They found that a Normal assumption fails at capturing the distribution of monthly excess returns on a value weighted NYSE portfolio during a smaller sample period (1926–1999). In particular, they found that historical excess returns are characterized by higher statistical moments with a negative sample skewness (-0.49) and an excess kurtosis (10.3).

3 Characterizing deviations from normality

So far I have found that normality does not hold for U.S. monthly excess returns. In particular, the empirical distribution of excess returns has long tails caused by a large kurtosis. Now, I attempt to assess under which circunstances normality may be statistically acceptable for U.S. excess returns.

When testing normality in the previous section I assumed - as many other financial researchers did - independent and identically distributed (i.i.d.) excess returns and, therefore, that the estimated parameters - mean and variance - remain constant for the entire sample period. However, especially when long histories are being used, the stationarity hypothesis may not hold: for the sample period under investigation, the parameters of the probability distribution can be changing over time, experiencing shifts known as ”structural breaks”, or switching among a finite set of values, that is, they are ”regime-switching”.

In this section, I explore those sources of deviations: structural breaks (subsection 3.1) and regime switching (subsection 3.2) in monthly excess returns data. The existence of structural breaks or regime switching may provide a solution to achieve normality for excess returns conditional on subperiods or regimes, respectively.

3.1 Structural breaks

Structural breaks are low frequency shocks with a permanent - long run - effect which tend to be related with significant economic and political events (Bai and Perron 1998, 2003).

In general, ignoring the presence of structural breaks may lead to misleading results. First, structural breaks might cause unit root tests to loose their power. In such case, one may, for instance, wrongly fail to accept stationarity for a given sample. In addition, structural breaks can be seen as extreme events which are shown to contribution to the leptokurtosis of financial returns. The degree of leptokurtosis, however, tends to disappear or to decrease substantially when structural breaks are taken into account by, for instance, simply estimating empirical models over subperiods. Finally, when structural breaks are ignored, making inference about ordinary descriptive statistics, such as the mean or the variance, may prove to be inaccurate. As referred by McConnel and Perez-Quiros (2000), taking theory to the data by confronting the moments generated from a calibrated Normal model with the moments of real data over a period with a structural break will lead to misleading conclusions. Therefore, even if the normality assumption for excess returns is not acceptable over the entire period, it could be so for subperiods between the breaks. That is, the normality theory may still be valid conditional to those subperiods.

This subsection starts with a description of the methodology undertaken to test and date structural breaks in monthly excess returns. Then, results are presented and normality is tested again for excess returns using smaller periods.

3.1.1 Methodology

I follow a flexible approach that is capable of simultaneously test and date structural breaks. It consists in a two-step procedure. In the first step, I use a generalized fluctuation framework to test for the presence of breaks in the conditional mean and in the conditional variance of excess returns. If there is evidence of breaks in the data, then, in the second step, I use an algorithm from Bai and Perron (1998, 2003) for dating the potential breakpoints and select the optimal number of breaks according to the Bayesian Information Criterion (BIC) (Schwarz, 1978). Such approach was suggested by Hornik et al. (2003) and may be easily implemented in their *strucchange* package for *R* software[5].

A. Generalized fluctuation tests - first step

Structural breaks are permanent shifts that occur in the parameters of a data generating process. Although structural breaks are detected by finding changes in the parameters of a conditional model, they imply shifts in the unconditional moments. Following Hornik et

[5] I would like to thank Professor Achim Zeileis for his useful suggestions on this subject.

al. (2003), I consider the following standard linear regression model for testing parameter stability,

$$y_t = x_t^\intercal \beta_t + u_t \qquad (t = 1, ..., T) \tag{3.1}$$

where y_t is the observation of the dependent variable at time t, $x_t = (1, x_{t2}, ..., x_{tk})^\intercal$ is a $k \times 1$ vector of observations of the independent variables, u_t is $iid\,(0, \sigma^2)$ and β_t is a $k \times 1$ vector of regression coefficients which may vary over time.

Under this set up we are concerned with testing the null hypothesis of "no structural change"

$$H_0 : \beta_t = \beta_0 \qquad (t = 1, ..., T) \tag{3.2}$$

against the alternative that at least one coefficient vector varies over time.

To test this null hypothesis (are the parameters stable?), I use generalized fluctuation tests. Generalized fluctuation tests fit a parametric model (3.1) to the data via ordinary least squares (OLS) and derive an empirical fluctuation process which captures the fluctuation in the residuals or in the coefficient estimates. The null hypothesis of parameter stability is rejected if the fluctuation is considered improbably large.

I use two empirical fluctuation processes which are based in the estimates of regression coefficients β: the recursive estimate (RE) process in the spirit of Ploberger et al. (1989) and the moving estimate (ME) process introduced by Chu et al. (1995). From this point of view, the $k \times 1$-vector β is either estimated recursively with a growing number of observations, or using a moving data window of constant bandwidth h. For both RE and ME empirical fluctuation processes, fluctuations are determined in terms of their deviations from the full-sample estimate of the regression coefficient β.

These empirical fluctuation processes can capture departures from the null hypothesis of parameter stability (3.2). In particular, the null hypothesis of "no structural change" should be rejected when the fluctuation of the empirical process becomes improbably large compared to the fluctuation of its limiting process. Under the null hypothesis, the corresponding limiting processes are known (see Kuan and Hornik, 1995) and, from these limiting processes, boundaries can be found, whose crossing probability (under the null hypothesis) is a prescribed significance critical rejection level α.

It should be noted, however, that generalized fluctuation tests are not a dating procedure and a more general framework is needed to this purpose.

B. Dating multiple breakpoints - second step

Multiple breaks are likely to occur in long time series data. In such case, we may need to consider a more general framework than the one considered in (3.1). According to Bai and Perron (1998, 2003), it is assumed that there are m breakpoints and $m+1$ segments (in which the regression coefficients are constant) and that the following multiple linear regression model is given:

$$y_t = x_t^\intercal \beta_1 + u_t \qquad (t = 1, ..., T_1)\,,$$

$$y_t = x_t^\intercal \beta_2 + u_t \qquad (t = T_1 + 1, ..., T_2)\,,$$

$$...$$

$$y_t = x_t^\intercal \beta_{m+1} + u_t \qquad (t = T_m + 1, ..., T)\,,$$

where the breakdates indexed by $(T_1, ..., T_m)$ are explicitly treated as unknown.

The main goal here is to determine the number and location of the breakpoints. That is, we need to determine the estimates for the m breakpoints and the $m+1$ segment - specific regression coefficients β_{j+1}, when T observations for (y_t, x_t) are available.

I estimate the breakpoints by applying the algorithm of Bai and Perron (2003) which was recently implemented in R software by Hornik et al. (2003). The algorithm is based on the principle of dynamic programming and it is considered, among others, one efficient method to estimate breakpoints in an OLS regression context.

The algorithm proceeds as follows. Given a fixed number of breakpoints, the algorithm starts by calculating the sum of squared residuals for every possible segment in the regression model. Some restrictions may be imposed to the number of segments to be calculated, such as the minimum distance between each break. From all these possible segments, the algorithm evaluates which partition - combination of segments - achieves a global minimization of the overall sum of squared residuals. It is an iterative process solving a recursive problem and from which the breakpoints, or equivalently the optimal partition, correspond to the global minimisers.

For initialization, the algorithm requires a fixed number of breakpoints. In many applications, however, m is unknown and it must be chosen based on the observed data as well. One solution to this problem is to compute the optimal segmentation for a sequence of breakpoints $m = 1, 2, 3, ...m_{\max}$ and to choose m by optimizing some information criterion (Bai and Perron, 2003). Model selection for the number of breakpoints m is based on the minimization of BIC criteria. Further details on this algorithm can be found in Bai and Perron (2003) and Hornik et al. (2003).

3.1.2 Tests results

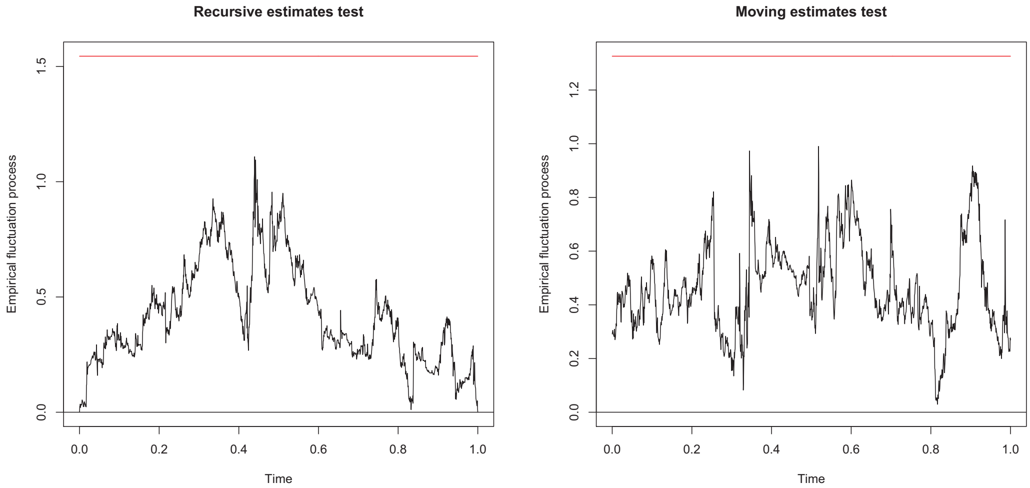
`A. Detecting evidence of structural breaks`

This part reports the results for the structural break tests undertaken to the conditional mean and the conditional variance of U.S. monthly excess returns. For each conditional model two fluctuation tests have been carried out: the recursive estimates (RE) and the moving estimates (ME). In the ME fluctuation tests, the moving date window of constant bandwidth h was set up to 5%[6].

Let the conditional mean of excess returns follow an autoregressive model of order 2 - AR(2). This model specification turns out to be necessary in order to remove autocorrelation

[6] Other values of h were used; however, the results appear to be qualitatively similar to the one obtained for the baseline value for h.

Figure 3: Stability tests in the conditional mean of excess returns



Monthly, from 1871:01 to 2010:12. The empirical fluctuation process with the coefficient estimates are plotted together with its boundaries at an (asymptotic) 5% significance level. Time refers to the fraction of the sample T.

from the residuals. Alternatively, we could assume a random walk model for the mean of excess returns. Or we could use more sophisticated models with a higher number of integration (e.g., AR(p) with $p > 2$). However, the empirical results remain qualitatively the same. Thus, the standard linear regression model upon which the structural break tests were based is,

$$y_t = \alpha + \beta^1 y_{t-1} + \beta^2 y_{t-2} + \epsilon_t, \qquad \epsilon_t \sim i.i.d\ N(0, \sigma^2) \tag{3.3}$$

with y_t the excess return in month t. Model (3.3) is a pure change model (Bai and Perron 1998, 2003) since all its regression coefficients - level α and persistence β - are allowed to change.

Figure 3 reports the structural break tests for the conditional mean of U.S. monthly excess returns. Clearly, the mean of excess returns did not vary much over the period. This is confirmed by both fluctuation tests; neither the RE nor the ME empirical processes cross the boundary line of critical rejection of 5% and, hence, one may conclude that there are no evidence of structural breaks in the mean of excess returns over the period from 1871:01 to 2010:12.

In order to run structural break tests in the variance, squared errors (ϵ_t^2) taken from equation (3.3) are considered as a proxy for the conditional variance of U.S. monthly excess returns. Other proxies of volatility may be considered instead. One alternative is to estimate a GARCH(1,1) model and work directly with the equation of conditional volatility. Or we may simply calculate the sample volatility using a moving data windows.

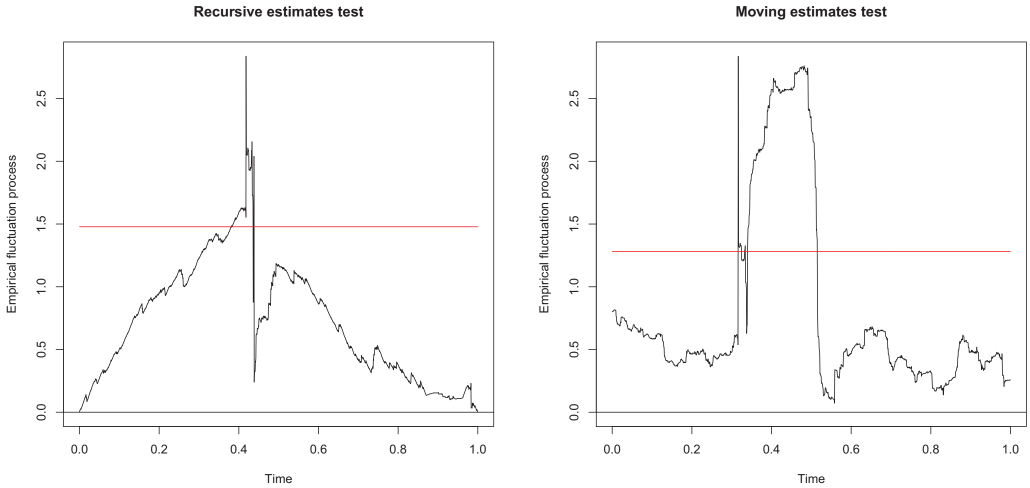
Let the conditional variance of excess returns follow an autoregressive model of order 1 - AR(1). Again, this model specification turns out to be necessary in order to remove autocorrelation from the residuals. Alternatively, we could assume a random walk model for the mean of excess returns. Or we could use more sophisticated models with a higher number of integration (e.g., AR(p) with $p \geq 2$). However, the empirical results remain qualitatively the same. Thus, the standard linear regression model upon which the structural break tests were based is,

$$\epsilon_t^2 = \delta + \eta\epsilon_{t-1}^2 + \zeta_t \quad with\ \zeta_t\ \tilde{}\ i.i.d\ N(0, \sigma^2). \tag{3.4}$$

The results, displayed in Figure 4, appear to be rather different for the conditional variance of U.S. monthly excess returns. At a 5% (asymptotic) significance level, both empirical fluctuation processes show significant departures from the null hypothesis of parameter stability, suggesting the existence of a structural break for the variance of excess returns.

Yet the number of breaks seems to be different across the tests: whereas the RE test points to one peak, the ME test depicts two - or possibly three - peaks where the fluctuation of the empirical process is improbably large compared to the fluctuation of the limiting process under the null hypothesis. Therefore, it seems that there is more than one structural break in the conditional variance for the 1871-2010 period.

Figure 4: Stability tests in the conditional variance of excess returns



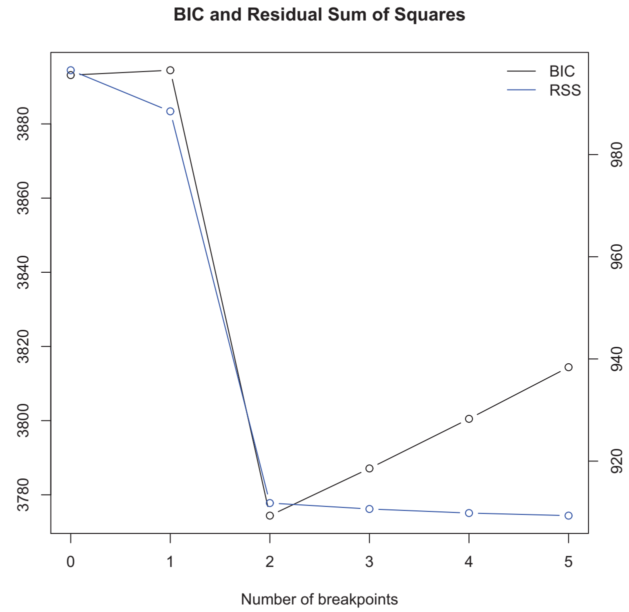
Monthly, from 1871:01 to 2010:12. The empirical fluctuation process with the coefficient estimates are plotted together with its boundaries at an (asymptotic) 5% significance level. Time refers to the fraction of the sample T.

B. Dating the breakpoints

After having found evidence of structural changes in the variance of excess returns, now I attempt to answer the question of when they have occurred. For dating the breakpoints an algorithm from Bai and Perron (1998, 2003) has been applied to the conditional variance model (3.4).

However, there are two important practical issues that we should take into account when applying the algorithm. One is related to the number of breaks to be considered in the dating procedure. I assume a maximum of $m = 5$ potential structural breaks and then select m according to a model selection which is based on the BIC criteria. Furthermore, I set up a minimum segment size of 96 months, corresponding to 5% of the 1871-2010 sample. According to Kim et al. (2005), this procedure turns out to be necessary to avoid any irregularities in the likelihood function resulting from a small subsample. Moreover, under this particular methodology, I am looking for permanent – rather than temporary – shifts in excess returns; that is, I have no particular interest in shocks that last for a short period of

Figure 5: BIC (and RSS) values for different number of breakpoints.



time[7].

Figure 5 displays the BIC criteria for models estimated up to $m = 5$ structural breaks. The BIC criterion selects two breakpoints for the variance of U.S. monthly excess returns. Looking at Figure 5, we may conclude that the value for which BIC is minimised is at $m = 2$. The Residual Sum of Squares (RSS) may be considered as an alternative information criterion for model selection. RSS, however, is monotonically decreasing suggesting a model with a higher number of breakpoints. Even though, I maintain my prior methodology to select structural breaks according to BIC criterion.

Table 4: Breakdates found in the conditional variance of excess returns

Breakdates estimates	
$m\left(\hat{T}_1\right)$	1929:06
$m\left(\hat{T}_2\right)$	1938:07

[7]Single or multiple outliers may be detected in algorithm Bai and Perron (2003) by using h=1 as the minimal length of a segment or by defining a large number for m, respectively.

Table 4 reports the corresponding breakdates for the model with $m = 2$ structural breaks. Breakdates for the variance of U.S. monthly excess returns were found at 1929:06 and 1938:07. These dates are related with the Great Depression in the late 1920s and the beginning of World War II.

These findings are somehow related with previous literature. Kim et al. (2005) found evidence of a structural break in U.S. monthly excess returns around the WWII period. In particular, this structural break is found to be associated with a permanent reduction in the general level of stock market volatility. My methodology finds support of breaks in the variance but not in the mean of excess returns. However, I found one additional break around the Great Depression period. Since Kim et al. (2005) studied a shorter sample period 1926-1999 and have set a wider length for each segment (10%), this may explain, in part, why they only found one – instead of two – structural break(s).

Pástor and Stambaugh (2001) reported a larger number of structural breaks for expected U.S. monthly excess returns (15 in total) over a longer sample period (1834-1999). The most significant break, however, was found to be in the late 1990s. In my procedure, I specify only 5 potential breaks. I also tried higher values for the maximum number of breaks to be considered but the findings were similar. Moreover, my model selection procedure is based on BIC criterion which rules out complex models (with a high number of segments), thus, favouring parsimony.

3.1.3 Normality in different subperiods

Breakdates for the variance of U.S. monthly excess returns were found at 1929:06 and 1938:07. This implies a model for the variance of excess returns with three subperiods. In this part, I assess normality for monthly excess returns over three subperiods: the first covering the time period until 1929:06, the second including the Great Depression period, starting in 1929:07 up to 1938:07, and finally, the remaining period 1938:08 - 2010:12.

Table 5: Descriptive statistics for excess returns over different periods

Time Period	Mean	Std. Dev	Skewness	Kurtosis
1871:01 – 2010:12	0.0032	0.0408	−0.34	14.39
1871:01 – 1929:06	0.0037	0.0324	−0.25	3.49
1929:07 – 1938:07	−0.0055	0.0916	0.34	7.52
1938:08 – 2010:12	0.0038	0.0363	−0.97	6.61

Table 5 reports the descriptive statistics for monthly excess returns for the whole sample period as well as for three different subperiods. There are several important findings in Table 5. First, excess returns during the Great Depression (second subperiod) were negative and extremely volatile. The average of excess returns during this subperiod was −0.55% with a standard deviation of more than 9.16%, which is almost three times higher than that for the other subperiods.

Second, normality in excess returns depends on the sample period over which data is observed. The subperiod before the Great Depression seems to be consistent with normality. For this subperiod, excess returns reveal a kurtosis of 3.49 and a slightly negative skewness of −0.25. On the other hand, normality is still not a reasonable assumption for the other two subperiods. The statistics for these two time spans reveal that the second subperiod exhibits the highest kurtosis (7.52) while the latter subperiod has the lowest value of skewness (−0.97).

Third, deviations from normality in excess returns for the three subperiods are smaller than those for the whole sample period. Looking at Table 5, the degree of leptokurtosis, in particular, is far less evident for the three subperiods when compared to the full-sample period. Therefore, it seems that the presence of structural breaks can, in part, explain the excess kurtosis for excess returns.

3.2 Regime Switching

The stationarity hypothesis may not hold due to the existence of structural breaks but also due to regime switching. Structural breaks are defined as irreversible (once and for all) changes as opposed to changes of a recurring nature which can be rather captured by regime switching models. In the latter, the parameters move and then revert from one regime to the other.

There are two different approaches for modelling regime switching. One approach uses Markov switching models, which define a probability transition matrix governing the shifts between regimes in a conditional regression model. A rather different approach uses a discrete mixture of distributions to fit the unconditional empirical distribution. Typically, two or three (or a finite number of) density distributions are defined, representing quiet and more turbulent regimes.

The purpose of this section is to find for market regimes in U.S. monthly excess returns. This is accomplished by estimating Normal mixture models up to $G = 5$ regimes for the excess returns data. Overall, it has been argued that the true distribution of stock returns may be Normal, although its parameters are "regime-switching" (Kon, 1984). In this sense, normality for excess returns may hold conditionally whether the data is arising from different market regimes.

I start this section by presenting the Gaussian mixture model and explaining the Expectation-Maximization (EM) algorithm (Dempster et al. 1977) used for the estimation of such model. Then, I report the estimation results for the excess returns in the sample period spanning from 1871:01 to 2010:12. Finally, using post WWII data, I assess if the Gaussian mixture model provides a reasonable model for the empirical distribution of monthly excess returns.

3.2.1 Gaussian mixture model

The Gaussian mixture distribution has been widely used as a suitable alternative for modelling distributions of financial returns. Gaussian mixture models as models for distributions of returns were first suggested by Kon (1984). This author proved that discrete mixture of Normal distributions can in fact explain both significant kurtosis (fat tails) and positive skewness (asymmetry) observed in the distributions of daily stock returns. In addition, mixtures of Normal distributions have the advantage of maintaining the tractability of a Normal distribution: the unconditional mean and variance are simply expressed as a linear combination of means and variances of different regimes.

Let $g = 1, ..., G$ be the components, or equivalently the regimes, of the mixture model and consider that each component has a Normal density $\phi\left(x; \mu_g, \sigma_g^2\right)$ with unknown means $\mu_1, ..., \mu_G$ and unknown variances $\sigma_1^2, ..., \sigma_G^2$. Under these circumstances, excess returns x_t are said to have density described by a finite mixture of Normal distributions,

$$f(x; \Psi) = \sum_{g=1}^{G} \pi_g \phi\left(x; \mu_g, \sigma_g^2\right) \tag{3.5}$$

where π_g is the (*a priori*) probability of an observation x_t belonging to the $g - th$ regime and Ψ includes $\theta = (\mu_1, ..., \mu_G, \sigma_1^2, ..., \sigma_G^2)$ and the probabilities $\pi_1, ..., \pi_G$ for G regimes.

The two main goals are: (i) to estimate the vector Ψ of unknown parameters and (ii) to select the number of regimes G. To do so, I apply a flexible two-step approach. In the first step, I estimate the parameters of Normal mixture models with a fixed number of regimes G using the Expectation-Maximization (EM) algorithm (Dempster et al. 1977). In the second step, I select the number of regimes G according to the Bayesian Information Criterion (Schwartz, 1977). Such approach has been shown to achieve a good performance to select the number of regimes in several applications (see Dasgupta and Raftery 1998) and it may be easily implemented in the MCLUST package for *R* software (Dasgupta and Raftery, 1998 and Fraley and Raftery, 1999 and 2002).

3.2.2 Expectation-maximization

Maximum-likelihood (ML) estimates for the parameters of Normal mixtures cannot be written down in a closed form (Titterington, 1996); instead, these ML estimates have to be computed iteratively. I estimate the parameters of Normal mixture models using the Expectation-Maximization (EM) algorithm (Dempster et al. 1977).

The EM algorithm is an iterative procedure for maximizing likelihoods in statistical estimation problems involving incomplete data, or in problems which can be posed in a similar form, such as the estimation of finite mixtures distributions. In particular, for the purpose of the application of the EM algorithm, the observed-data vector of excess returns $x = (x_1, ..., x_n)$ is regarded as being incomplete; that is, we do not know which mixture component underlies each particular observation. Therefore, any information that permits classifying observations to components should be included in the mixture model. The component-label variables $z_t = (z_{t1}, ..., z_{tG})$ are consequently introduced, where z_{tG} is defined to be one or zero according to if x_t did or did not arise from the $G-th$ component of the mixture model, respectively. Moreover, if z_t is i.i.d. with a multinomial distribution of one draw from G categories with probabilities $\pi_1, ..., \pi_G$, we may represent the new complete-data loglikelihood for excess returns as

$$l\left(\theta_g, \pi_g, z_{tg}|x\right) = \sum_{t=1}^{n} \sum_{g=1}^{G} z_{tg} \log\left[\pi_g f\left(x_t|\theta_g\right)\right],$$

where the values of z_{tg} are unknown and treated by the EM as missing information on to be estimated along with the parameters θ and π of the mixture model.

The EM algorithm starts with some initial guess for the parameters in the mixture $\hat{\Psi}$ (component means $\hat{\mu}_g$, component variances $\hat{\sigma}_g^2$ and mixing proportions $\hat{\pi}_g$). Based on such guessed values, the algorithm finds the ML estimates by applying iteratively until convergence the expectation step (E-step) and the maximization step (M-step).

The E-step consists, at iteration h, in computing the expectation of the complete data loglikelihood over the unobservable component-label variables z, conditional on the observed data x_t and using the current estimates on the parameters $\hat{\Psi}^{(h)}$,

$$\hat{\tau}_{tg}^{(h)} = E\left[z_{tg}|x_t; \hat{\Psi}^{(h)}\right] = \frac{\hat{\pi}_g^{(h)} \phi\left(x_t; \hat{\theta}_g^{(h)}\right)}{\sum_{g=1}^{G} \hat{\pi}_g^{(h)} \phi\left(x_t; \hat{\theta}_g^{(h)}\right)},$$

for $g = 1, ..., G$ and $t = 1, ..., n$. Note that $\hat{\tau}_{tg}^{(h)}$ is the estimated posterior probability (at iteration h) that the $t-th$ observation belongs to the $g-th$ component.

In the M-step, new parameters $\hat{\Psi}^{(h+1)}$ are obtained by maximizing the expected complete data loglikelihood. These new parameters are estimated from the data given the conditional probabilities $\hat{\tau}_{tg}$ that were calculated from the E-step. Estimates of π_g and θ_g have simple closed-form expressions involving the data and $\hat{z}_{tg}$ from the E-step.

Once we have a new generation of parameter values $\hat{\Psi}^{(h+1)}$, we can repeat the E-step and another M-step. The E and M-steps are alternated repeatedly until the likelihood (or the parameter estimates) change by an arbitrarly small amount in the case of convergence. The EM algorithm increases the likelihood function of the data at each iteration, and under suitable regularity conditions converges to a stationarity parameter vector.

3.2.3 Number of regimes

When Normal mixture models and its corresponding parameters have been estimated, the number of regimes to consider in excess returns has to be decided. I assume a standard statistical problem in which the number of regimes G is known, but mixture models with different G should be compared based on some type of information criterion.

I select the number of regimes G according to the Bayesian Information Criterion (BIC) (Schwarz, 1978). The BIC chooses G to minimize the negative loglikelihood function augmented by some penalty function which rules out complex models with a large number of parameters. Overall, parsimony-based approaches have been proposed for choosing the optimal number of regimes in a given mixture model. In particular, Bayesian Information Criterion (BIC) has been shown to achieve a good performance to select the number of regimes in several applications (see Dasgupta and Raftery 1998). Medeiros and Veiga (2002) proposed a rather different approach to select the number of regimes in financial volatility. In such approach, the problem of selecting the number of regimes is solved by applying Lagrange multiplier type tests.

3.2.4 Parameter estimates

In summary, the following strategy for model selection has been performed: (i) a maximum number of regimes, $G_{\max}$ is considered; (ii) then, the parameters are estimated via EM for each model with a number of regimes up to $G_{\max}$; (iii) BIC is compared for each model; and (iv) finally, the number of regimes is selected such as to minimise BIC[8].

`A. Full sample period`

This part reports the results from the application of MCLUST package to monthly excess returns data from 1871:01 to 2010:12. Mixtures up to $G = 5$ Normal distributions were estimated via EM algorithm for monthly excess returns over the period 1871:01 to 2010:12.

Table 6: BIC values for different numbers of regimes in excess returns

Regimes	**BIC value**
G=1	$-5,969.69$
G=2	$\mathbf{-6,322.74}$
G=3	$-6,311.64$
G=4	$-6,285.11$
G=5	$-6,270.58$

Table 6 presents the corresponding BIC value calculated for each Normal mixture model. There are two regimes in U.S. monthly excess returns. Looking at Table 6, the mixture model for which the BIC value is minimised is at $G = 2$. Furthermore, the stationary Normal model with $G = 1$ has the worst performance among all the models. Hence, it seems that excess returns over the period from 1871:01 to 2010:12 may be described by a combination of $G = 2$ Normal distributions. The corresponding maximum likelihood estimators for the parameters are

[8]Concerning the problem of how to select suitable starting values for the EM algorithm, MCLUST provides a suitable solution. For initialization, the data is divided into quantiles equal to the number of regimes considered.

$$f_{GM}\left(x;\hat{\pi}_1,\hat{\pi}_2,\hat{\mu}_1,\hat{\mu}_2,\hat{\sigma}_1,\hat{\sigma}_2\right) = 0.9021\cdot\phi\left(x;0.0066,0.0009\right)+ \\ +0.0979\cdot\phi\left(x;-0.0286,0.0078\right) \quad (3.6)$$

The density model for excess returns in (3.6) reflects the existence of two market regimes: the first density, occurring with a probability of 90%, may be interpreted as a *bull* market regime - corresponding to months where excess returns have a relatively higher mean and a smaller variance - whereas the second density, occurring with a probability of 10%, may be interpreted as a *bear* market regime - corresponding to months where excess returns have a lower expected return and a greater variance.

One attractive property of mixture models compared to a stationary Gaussian model is that they can capture the leptokurtic and skewed characteristics of empirical data (Kon, 1984). To illustrate this flexibility, consider the probability density function of U.S. monthly excess returns in (3.6). Leptokurtosis arises from a relatively larger variance in regime 2 occurring with probability 10% which enables the mixture to put more mass at the tails. However, the majority of excess returns (90%) are coming from regime 1. Since this regime contains data which is expected to be closer to the (unconditional) sample mean this, therefore, explains the high peak observed in the empirical distribution of excess returns. On the other hand, skewness arises when means are different between regimes. In particular, there is a negative asymmetry in the density model (3.6) since the high–variance regime has both a smaller mean and a smaller mixing weight.

B. Post-WWII data

I have already found deviations from normality in U.S. monthly excess returns over the subperiod from 1938:08 to 2010:12. Now, I investigate if those deviations arise from regimes changes. That is, I test the viability of Normal mixture models as a descriptive model for U.S. monthly excess returns for this specific subperiod.

Figure 6: Empirical density of excess returns and the PDF mixture of two Normal distributions

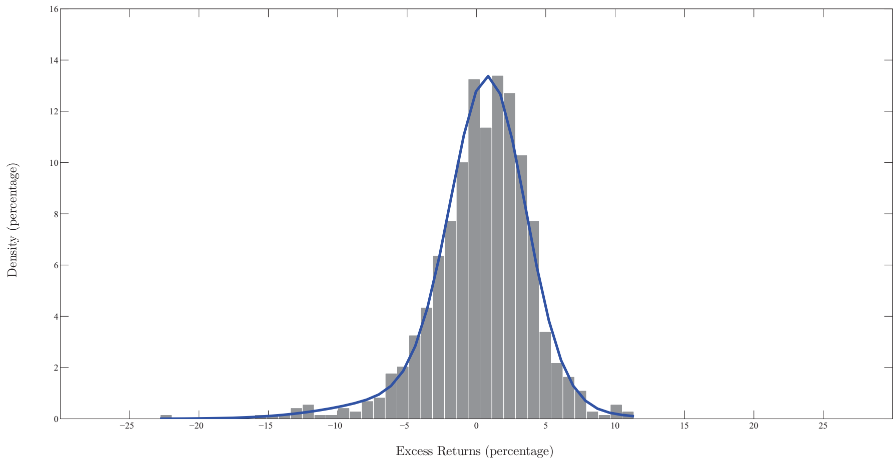


Table 7: BIC values for different number of regimes in excess returns over post-WWII data

Regimes	**BIC value**
G=1	$-3,286.63$
G=2	$\mathbf{-3,375.42}$
G=3	$-3,359.05$
G=4	$-3,344.23$
G=5	$-3,330.53$

Table 7 reports the BIC value for mixtures up to $G = 5$ Normal distributions estimated via EM algorithm for U.S. monthly excess returns over the period 1938:08 to 2010:12. Again, there are two regimes in U.S. monthly excess returns. The BIC criterion selects a model with $G = 2$ and, therefore, a mixture of two Normal distributions may be considered as a relatively good descriptive model to explain U.S. monthly excess returns over the subperiod from 1938:07 to 2010:12. In contrast, the stationary Normal model with $G = 1$ has the worst performance among all the models.

Figure 6 plots empirical density histograms for U.S. monthly excess returns together with a density of a mixture of two Normal distributions over the subperiod from 1938:07 to 2010:12. Overall, the Normal mixture model may be considered as a good descriptive model for U.S. monthly excess returns. Looking at Figure 6, it is particularly evident that the combination of two Normal distributions can capture the larger peakness and the heavier tails revealed in the empirical distribution of excess returns.

In this section, I found evidence of two regimes with different means and variances for U.S. monthly excess returns and, at least for the last 70 years, these regimes can explain, in part, deviations from normality for this particular subperiod. These findings may be somehow related with the previous literature. Kim et al. (2005) compared four univariate models of monthly excess returns over the period 1926-1999 and they found that the three models that incorporate Markov switching volatility are preferred over the simple i.i.d. model.

There is a wide empirical evidence, in the literature, for the existence of regimes in stock returns. Kon (1984) examined daily returns of 30 stocks of the Dow Jones and estimated mixtures of Gaussian distributions with two up to four regimes which were found to fit appropriately. Aparicio and Estrada (1995) found partial support for a mixture of two Normal distributions to explain the distributions of stock returns of 13 European securities markets during 1990-1995 period. Using monthly S&P 500 stock index returns (1871–2005), Behr and Pötter (2009) investigate the viability of three alternative parametric families to represent both the stylised and empirical facts and they have concluded that the two component Gaussian mixture has the smallest (maximal absolute) difference between empirical and estimated distribution. They have suggested that, given its flexibility and the comfortable estimation using the EM algorithm, Gaussian mixture models should be considered more frequently in empirical financial analysis.

4 Time aggregation and Normality

I found consistent deviations from normality in monthly excess returns for the post-WWII data. These deviations were associated with regime switching. Since normality is a key assumption in finance and macrofinance and it allows for almost all the results in theoretical asset pricing, should we consider most of finance theory as unrealistic, and, therefore, irrelevant? That is, should we abandon normality for this particular subperiod? In this final section, I assess under which circumstances normality of excess returns may be reconsidered for the post-WWII data.

My main emphasis here is to see whether Gaussianity for excess returns depends or not on the frequency of the data. This is equivalent to test for aggregational Gaussianity. Aggregational Gaussianity means that as one increases the time interval over which returns are calculated, their distribution looks more and more like a Normal distribution. Therefore, normality in excess returns may not hold for high frequency data - up to monthly data - but it may hold for higher aggregate data.

In order to test aggregational Gaussianity for U.S. excess returns, I use post-WWII data measured at different time intervals. Ultimately, I try to determine the level of aggregation for which excess returns converge to a Normal distribution.

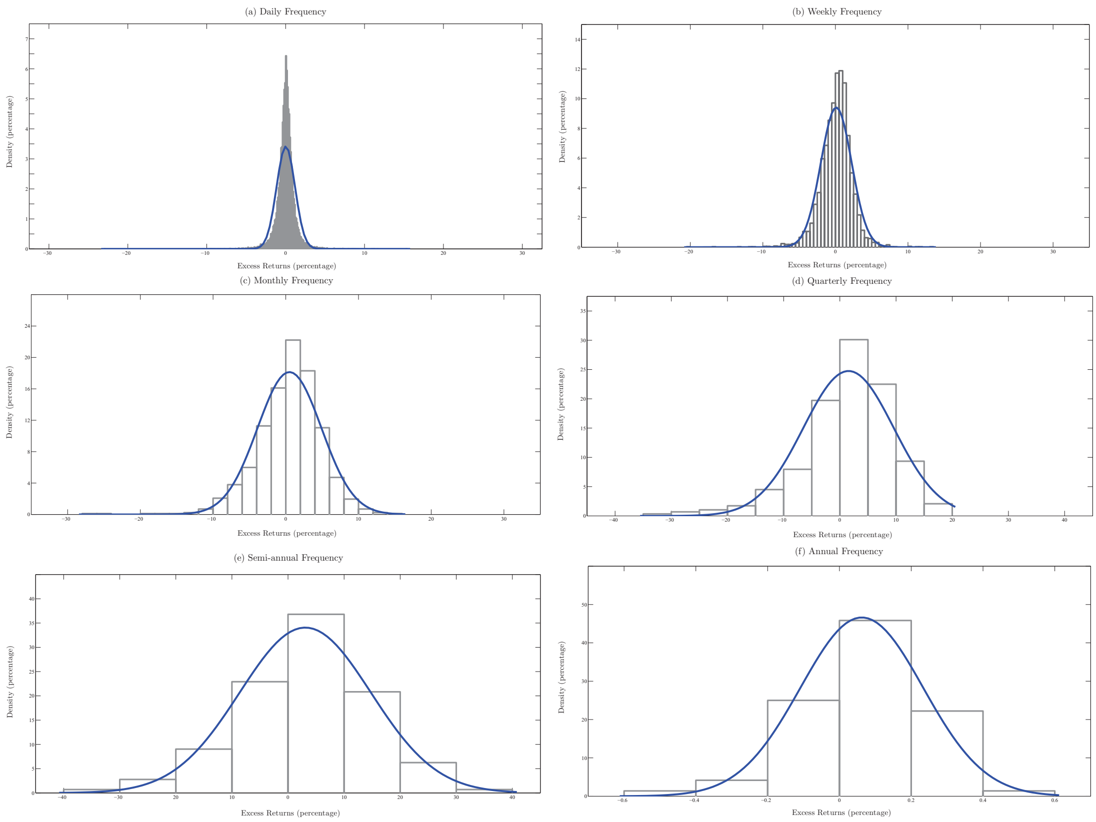
This section starts with a brief description of the data. Then, I test for the aggregational Gaussianity for U.S. excess returns. This section ends with some formal normality tests.

4.1 Data

I use excess returns data measured over daily, weekly, monthly, quarterly, semi-annual and annual frequencies for the sample period starting at 01/08/1938 until 31/12/2010. I use end-of-period data for all frequencies. From end-of-day figures, I calculate end-of-week excess returns when 7 calendar days are available. Similarly, using the last day of each month, I compute monthly excess returns. For quarterly excess returns, I assume, for each year, the end-of-month observation in March, June, September and December. Semi-annual observations are the end-of-month figure in June and December and, finally, annual observations correspond to the last day of each year.

The data sources are as follows. For stock prices, I use daily quotes from Bloomberg for the S&P 500 stock index. Monthly dividends are from Shiller's data. However, to obtain daily

Figure 7: Empirical density of excess returns and the Normal PDF for different levels of aggregation



and weekly observations, I apply linear interpolation to monthly data. For the risk-free asset, I use yields from the 3-month Treasury Bills published by Federal Reserve. In particular: (i) for daily and weekly frequencies and up to 04-01-1954, I use monthly observations which are linear interpolated to obtain daily and weekly observations (ii) for the remaining period, I use daily observations of the 3-month Treasury Bill secondary market rate (yield).

4.2 Aggregational Gaussianity

A simple way to check normality is to plot the empirical density histogram against the theoretical Normal distribution with the same empirical moments - mean and variance.

Figure 7 reports empirical distributions of excess returns recorded over time intervals of one day, one week, one month, three months, six months and one year. Also plotted in

Figure 7 is the fit of a Normal distribution with a mean and variance given by the respectively empirical moments.

There is evidence of aggregational Gaussianity in U.S. excess returns. Visual inspection of the empirical distributions suggests that these distributions are high-peaked and leptokurtic for lower levels of aggregation (up to quarterly data). However, the peak sharpness and the extent of leptokurtosis in the empirical distributions seem to decrease substantially and become more Gaussian-like as we move for higher levels of aggregation. Overall, normality seems to be a reasonable assumption for excess returns for frequencies starting at semi-annual data. At this level of aggregation, the shape of the empirical distribution of excess returns is closer to the bell-shape of the Gaussian distribution. In particular, the corresponding empirical distribution is characterized by thinner tails and a less peaked center compared to the empirical distributions for the lower levels of aggregation and, thus, it is more similar to the Gaussian benchmark.

Table 8: Descriptive statistics for excess returns over different time levels of aggregation

Time frequency	**Mean**	**Std. Dev.**	**Skewness**	**Kurtosis**
daily	0.0002	0.0117	-0.34	22.29
weekly	0.0012	0.0212	-0.66	9.03
monthly	0.0051	0.0440	-0.83	6.37
quarterly	0.0155	0.0806	-0.90	4.65
semi-annual	0.0308	0.1171	-0.46	**3.28**
annual	0.0624	0.1711	-0.77	**3.82**

The most evident fact about excess returns data is that as the level of aggregation increases so the kurtosis falls. Looking at the Table 1, the kurtosis coefficient is widely above 3 for daily frequencies, it slowly decreases up to the quarterly frequency and, finally, it reaches convergence to normality at semi-annual and annual frequencies. Skewness statistics do not yield a consistent pattern. These statistics suggest that excess returns are negatively skewed but this is not a major feature for these data. In addition, the higher the level of aggregation, the higher the value for the mean and for the standard deviation of excess returns.

Finally, it is worth mentioning that there are two particular forces working simultaneously. The first force stems from the fact that excess returns seem to belong to the domain of attraction of the Normal law, which is reflected in the tendency for the estimates of the kurtosis to approach the value of 3. The second one stems from the fact that, when lower frequencies are aggregated, the number of observations decreases and, so, the amount of precision[9]. Despite this lower precision, normality is obtained for semi-annual and for annual data.

4.3 Normality tests

To complement the visual evidence for aggregational Gaussianity presented in the previous subsection, I now take a more formal statistical analysis based on six normality tests.

Table 9: Normality tests in excess returns over different time levels of aggregation.

Frequency	JB	SW	AD	L	CvM
daily	0.0000	0.0000	0.0000	0.0000	0.0000
weekly	0.0000	0.0000	0.0000	0.0000	0.0000
monthly	0.0000	0.0000	0.0000	0.0000	0.0000
quarterly	0.0000	0.0000	0.0000	0.0017	0.0000
semi-annual	**0.0729**	**0.2037**	**0.0907**	0.0191	**0.0590**
annual	0.0170	0.0407	0.0339	**0.0804**	0.0348

Table 9 reports the normality tests performed to excess returns for different time levels of aggregation. A Normal distribution is statistically acceptable for semi-annual and, to a lower extent, for annual excess returns data. For lower levels of aggregation, from daily to quarterly frequencies, a Normal distribution is not acceptable according with all the tests. However, for semi-annual data and considering a p-value of 5%, the normality hypothesis is rejected only for the Lilliefors test.

[9]For example, for the time period 01/08/1938 to 31/12/2010, I have 20,099 daily observations but only 144 semi-annual and 72 annual observations.

Previous empirical research also indicates that the shape of the distribution of stock returns is not the same over different time intervals. (Cont, 2001 and Eberlein and Keller, 1995). For instance, it has been established that daily returns depart more from normality than monthly returns do (see, e.g., Blattberg and Gonedes (1974); Eberlein and Keller,1995; and Campbell et al.,1997). Campbell et al. (1997) compared the distributions of daily and monthly stock returns for two indices and ten stocks from the U.S. for the period 1962–94. They found that the non-Gaussian character (the skewness and kurtosis) displayed by the distributions of monthly data is significantly lower than that displayed by the distributions of daily data. Diebold (1986, 1988) referred that the convergence to normality in exchange rate and other financial returns is attained when the aggregation period tends to infinity; I find normality for excess returns data being the point of convergence around six months.

5 Conclusion

I presented evidence that the empirical density distribution of monthly U.S excess returns tends to be non-Gaussian, sharp peaked and heavy tailed for the period spanning from 1871:01 to 2010:12. This was my starting point. The purpose of this dissertation was to assess under which circumstances normality may still be acceptable as a descriptive model for the U.S. excess returns. In particular, I explored two sources of deviations from normality: (i) structural breaks and regime switching using a long time series and (ii) temporal aggregation for a shorter time series.

My main findings may be summarized as follows. First, I found two structural breaks in monthly excess returns. The breakdates are at 1929:6 and 1938:7. These breaks are mainly associated with two permanent shifts in the conditional variance of monthly excess returns. The volatility of excess returns is relatively low before the start of the Great Depression, increases sharply until the late 1930s and then returns to its prior levels. Second, the assessment of normality in excess returns seems to depend on the subperiods considered.

More specifically, normality may be acceptable for the subperiod before the start of the First Great Depression, i.e. 1871-1929, but not acceptable for the other two subperiods: the subperiod including the Great Depression period starting in 1929:7 up to 1938:7 and the remaining subperiod 1938:8-2010:12. Third, for the post-WWII data, I found that normality for excess returns may be recovered using mixtures of two Normal distributions. In particular, I provided evidence of switching regimes in monthly excess returns for this period. Those regimes may be seen as describing *bull* and *bear* market conditions. Finally, for the post-WWII data, I found that normality for excess returns depends on the frequency of the data. In particular, I tested for aggregational Gaussianity from daily to annual data, and concluded that normality for excess returns may also be recovered considering higher aggregated data, starting from semi-annual to annual data.

These findings may have possible implications for finance and macrofinance literature. First, if we accept normality only for frequencies starting at semi-annual up to annual, this means that calculations with VaR and other risk-based models, which rely on such assumption, could be misleading. In particular, for time horizons up to six months, investors, under a Gaussian alternative, would underestimate potential losses, since in reality empirical distributions are fat-tailed. In such case, investors must apply non-Gaussian probability tools to quantify their risks (relying, for instance, on extreme value theory and rare event analysis). On contrary, for time horizons starting at six months, investors face Gaussian risks and conventional risk management may hold. Second, they may also have possible implications for the equity premium puzzle. Using a general equilibrium asset pricing model, Mehra and Prescott (1985) proved that the magnitude of U.S. historical equity premium (the expected excess return) is greater than the one rationalized in the context of the standard neoclassical paradigm of financial economics. Their benchmark model assumes normality for asset returns. According to my findings, if the puzzle is tested for data frequencies starting at semi-annual up to annual, then the Normal distribution assumption cannot be pointed as a potential cause of this puzzle, since it tends to be a suitable assumption.

Yet these findings and its corresponding implications should be confirmed by further investigation. Possible directions for future research may include using other sources of data (end of month data, other financial assets) or testing the existence of structural breaks and regime switching for data frequencies other than monthly. In addition, the differentiation between structural breaks and regime switching is not always straightforward: whereas regime switching models may include a non-recurring state, which is well suited to model structural breaks, tests of structural breaks are capable of detecting regime shifts of a recurring nature. Future research may explore more powerful tests to identify the specific nature of the shocks underlying financial time series. Finally, it should be noted that structural breaks and more sophisticated models, such as GARCH and ARFIMA, tend to be interrelated. That is, they are very easy to confuse. When either GARCH or long memory are present, structural breaks may be spuriously detected; on the other hand, when structural breaks are ignored the degree of persistance and the long memory parameter may be overestimated. Hence, other possible direction for future research may be to explore this inter-relation for excess-returns.

References

Anderson, T.W. (1962). "On the Distribution of the Two-Sample Cramer-von Mises Criterion". *The Annals of Mathematical Statistics* (Institute of Mathematical Statistics) 33 (3): 1148–1159.

Aparacio, F. and Estrada, J. (2001), Empirical distributions of stock returns: European security markets, *European Journal of Finance*, 7, 1–21.

Bai, J. and Perron, P. (1998).“Estimating and Testing Linear Models With Multiple Structural Changes.*Econometrica*, 66, 47–78.

Bai, J. and Perron, P. (2003). “Computation and Analysis of Multiple Structural Change Models.*Journal of Applied Econometrics*, 18, 1–22.

Behr, A. and Pötter, U. (2009). Alternatives to the normal model of stock returns: Gaussian mixture, generalised logF and generalised hyperbolic models. *Annals of Finance*, 5, 49–68.

Bera, A. K. and Jarque, C. M. (1987). "A test for normality of observations and regression residuals." *International Statistical Review*. Vol. 55, No. 2, pp. 163–172.

Black, F. and M. Scholes (1973). “The Pricing of Options and Corporate Liabilities.*Journal of Political Economy* 81, 3, 637-654.

Blattberg, R. and Gonedes, N. (1974), A comparison of the stable and Student distributions as statistical models for stock prices, *Journal of Business*, 47, 244-280.

Bollerslev, Tim, Ray Y. Chou, and Kenneth F. Kroner (1992), "ARCH Modeling in Finance," *Journal of Econometrics*, vol. 52, 5-59.

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton, NJ: Princeton University Press.

Reinhart, C. and Rogoff, K. (2008), “This Time is Different: A Panoramic View of Eight Centuries of Financial Crises,” NBER Working Paper, No. 13882. (Cambridge, MA: National Bureau of Economic Research).

Cont, R. (2001), Empirical properties of asset returns: stylized facts and statistical issues, *Quantitative Finance*, I, 223–236.

Cramér, H. (1928), "On the composition of elementary errors. Second paper: Statistical applications. *Skand Aktuartdskr.*, 11, 141-180.

Dasgupta, A. and Raftery, A. (1998). Detecting features in spatial point processes with clutter via model-based clustering. *Journal of the American Statistical Association*, 93:294-302.

Dempster, A.P, Laird, N.M. and Rubin, D.B. (1977), Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society* B, 39, 1–38.

Diebold, Francis X.. (1986). Temporal aggregation of ARCH processes and the distribution of asset returns. Federal Reserve Board, Division of research and statistics, Special studies paper. August 1986, no. 200.

Diebold, Francis X.. (1988). *Empirical modeling of exchange rate dynamics, Lecture notes in Economics and mathematical systems*. Springer-Verlag, 1988, 303.

Eberlein, E. and Keller, U. (1995), Hyperbolic distributions in finance, *Bernoulli*, 1, 281–299.

Fama, E. F. (1965), 'The Behaviour of Stock-Market Prices', *Journal of Business*, 38, 34–105.

Fraley, C. and Raftery, A. E. (1998). How many clusters? Which clustering method? - Answers via model-based cluster analysis. *The Computer Journal*, 41:578–588.

Fraley, C. and Raftery, A. E. (1999). MCLUST: Software for model-based cluster analysis. *Journal of Classification*, 16:297–306.

Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis and density estimation. *Journal of the American Statistical Association*, 97:611–631.

Chu CSJ, Hornik K, Kuan CM (1995). "The Moving-Estimates Test for Parameter Stability.*Econometric Theory*, 11, 669–720.

Hornik, K., Kleiber, C., Krämer W. and Zeleis, A. (2003). Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis*, 44(1-2):109-123.

Hornik K., Kleiber C., Leisch F. and Zeileis, A. (2002). “strucchange: An R Package for Testing for Structural Change in Linear Regression Models.*Journal of Statistical Software*, 7(2), 1–38.

Kim C-J, Morley J, Nelson C (2005). ”The Structural Break in the Equity Premium”. *Journal of Business and Economics*, 23, 181-191.

Kon, S. J. (1984), Models of stock returns - a comparison, *Journal of Finance*, 39, 147-165.

Kuan CM, Hornik K (1995). “The Generalized Fluctuation Test: A Unifying View.*Econometric Reviews*, 14, 135–161.

Lilliefors, H. (1967), ”On the Kolmogorov–Smirnov test for normality with mean and variance unknown”, *Journal of the American Statistical Association*, Vol. 62. pp. 399–402.

Mandelbrot, B. B. (1963), ‘The Variation of Certain Speculative Prices’, *Journal of Business*, 36, 394–419.

Medeiros, M. C. and Veiga, A. (2009): ”Modeling multiple regimes in financial volatility with a flexible coefficient GARCH(1,1) model” *Econometric Theory*, 25, 2009, 117–161

McConnell, Margaret M, and Gabriel Perez-Quiros. “Output Fluctuations in the United States: What has Changed Since the Early 1980’s,*American Economic Review*, 90(5), 2000, pp. 1464 – 1476.

McLachlan, G. J. and Peel, D. *Finite Mixture Models*, Wiley, 2000.

Mehra, R., and E. C. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics*, 15 (1985): 145–161.

Pastor, Lubos, and Robert F. Stambaugh. “The Equity Premium and Structural Breaks”. *Journal of Finance*, 56 (2001):1207 – 1239

Ploberger, W., Krämer, W., Kontrus, K. (1989). “A New Test for Structural Stability in the Linear Regression Model.*Journal of Econometrics*, 40, 307–318.

Poterba, J. M. and L. H. Summers (1988). Mean reversion in stock prices: evidence and implications. *Journal of Financial Economics*, 22:27–59.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464, 1978.

Shapiro, S. S. and Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)", *Biometrika*, 52, 3 and 4, pages 591-611.

Shiller, R. (1989), *Market Volatility* (Cambridge, MA: MIT Press)

Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons, *Journal of the American Statistical Association*, Vol. 69, pp. 730-737.

Titterington, D. (1996). Mixture distributions(update). In Kotz, S., Johnson, N., and Banks, D., editors, *Encyclopedia of Statistical Sciences*, pages 399-407. New York:Wiley.

Zeileis, A. (2005). A unified approach to structural change tests based on ML scores, F statistics, and OLS residuals. *Econometric Reviews*, 24(4):445-466.