

Mean Reversion Models of Financial Markets

Inaugural–Dissertation

zur Erlangung der Würde eines *Doctor rerum politicarum*

an der

Universität Bremen
Fachbereich 7 — Wirtschaftswissenschaft

vorgelegt von: Eric Hillebrand
e–mail: erhil@aol.com
Gneisenaustraße 48
28201 Bremen
Telefon: 0421-550531.

Betreuer: Prof. Dr. Otto Steiger
Institut für Konjunktur- und Strukturforschung (IKSF)

Stanford, 12. Mai 2003

Acknowledgments

I am grateful to my adviser, Otto Steiger, for continuous support and advise during the last six years. His unconditional commitment and confidence in the success of this project was an invaluable source of motivation.

This thesis was written in Bremen and Stanford. From 1999 to 2001 I was at the Center for Complex Systems and Visualization (CeVis) in Bremen. It is a pleasure to thank Heinz-Otto Peitgen, Carl Everts and the team of CeVis and MeVis for creating a first rate working environment. Discussions with Carl Everts, Ralf Hendrych, Sven Kohle, Richard Rascher-Friesenhausen and Peter Singer were fruitful.

The support of Ludwig Arnold at the Department of Mathematics at the University of Bremen was crucial in many ways. He gave advise on early stages of the model for mean reversion in returns and his excellent contacts facilitated my way over the Atlantic.

I spent the academic years 2001/2002 and 2002/2003 at the Department of Mathematics at Stanford University. I am deeply indebted to George Papanicolaou, who invited me, gave intense and very motivating advise and, last but not least, supported me when I went on the U.S. job market for economists. This would have been impossible without him.

For many helpful discussions I thank Caio Almeida, Arnold Kim, Mordecai Kurz and the participants of his Ph.D. student's seminar at the Stanford Department of Economics, as well as Jonathan Mattingly and Knut Sølna.

Any remaining errors are mine, of course.

Kein Unternehmen wie dieses ohne Familie und Freunde: Vielen Dank an Marita, Edgar, Alexandra, Anni; Günter und Irmtraut (wo lässt man sein Zeugs wenn man für einige Jahre ins Ausland geht?); Mareike, Thomas, Ling, Christopher, Anita, Arnold, Caio, Ellen, Vero. Katrin, und ohne sie höchstwahrscheinlich alles ganz anders.

This analysis suggests that a more catholic approach should be taken to explaining the behavior of speculative prices.

Lawrence H. Summers

Contents

1	Introduction: Mean Reversion in Stock Market Prices	17
A	Mean Reversion in Prices and Returns	27
2	Measuring and Interpreting Mean Reversion in the Data	29
1	Definitions of Mean Reversion	29
2	Efficient Markets and Mean Reversion	32
3	Rationales of Mean Reversion	37
4	Measurement and Evidence of Mean Reversion	40
5	Conclusion	50
3	A Mean Reversion Theory of Stock Market Crashes	53
1	Mean Reversion and Stock Market Crashes	53
2	A Mean Reversion Theory of Stock Market Crashes	54
a)	Mean Reversion Expectations	54
b)	Mean Reversion Illusions and Disillusions	56
c)	Mean Reversion Disillusion and October 19, 1987	59
3	A Mean Reversion Model for Stock Returns	61
4	Mean Reversion and the Stock Market Crash of 1987 in Market Data	63
a)	The Mean Reversion Disillusion	65
b)	The Mean Reversion Illusion	68
5	A Note on the Stock Market Crash of 1929	73
6	Conclusions	74
B	Mean Reversion in Volatility	77
4	Volatility Persistence, Mean Reversion, and Long Memory	79
5	Mean Reversion and Persistence in GARCH(1,1)	85
1	Time Scales and Persistence in Financial Volatility Data	85
2	Persistence Estimation with GARCH Models	87

a)	The Model Formulation	87
b)	Measures of Persistence and Mean Reversion	88
c)	Maximum Likelihood Estimation	89
d)	GARCH(1,1) and Market Data: High Persistence in the Volatility of the Dow Jones and S&P500	90
e)	High Persistence as a Stylized Fact	91
3	Parameter Changes and Global GARCH(1,1) Estimations	93
a)	The Geometry of Almost-Integration	93
b)	The Analysis of Almost-Integration	94
c)	Simulations	101
4	Estimation of the Short Scale in Stock Volatility	103
a)	Synthetic Data	106
b)	Market Data	107
5	Summary and Conclusion	108
6	Generalization to GARCH(p,q)	111
1	Multiple Scales and Higher Order ARMA and GARCH	111
2	Measuring Mean Reversion in ARMA(1,1)	112
3	Aggregation of ARMA(1,1) Models	113
4	Aggregation of GARCH(1,1) Models	114
5	Unknown Parameter Regime Changes and Global GARCH(p,q) Modelling	118
6	Simulations	126
7	Conclusion	129
7	The Impact of Japanese Foreign Exchange Intervention on Level and Volatility of the Yen/Dollar Exchange Rate	131
1	Sterilized Intervention and Volatility	131
2	The Discussion of the Japanese Interventions in the Literature . .	132
3	Data	135
4	A GARCH(p,q) Model with Interventions as Exogenous Variables	137
5	Estimation Results	138
6	Changepoint Detection	139
7	Conclusion	141
8	Conclusions and Directions	143
1	Summaries	144
2	Future Research	148
	Bibliography	151
	Appendix I	158

Appendix II**160**

List of Figures

1.1	The concept of mean reversion.	18
1.2	Annualized returns and volatility of the DJIA 1985-2000.	22
3.1	The development of mean reversion expectations.	55
3.2	Mean reversion illusion and stock market crash.	59
3.3	Dow Jones 1901–2002 and mean reversion.	64
3.4	S&P500 between January 2, 1985 and July 23, 1990.	65
3.5	Likelihood-ratios over the 1987 crash.	68
3.6	The bull market 1982–1987.	70
3.7	Changepoint detector statistic 1982–1987.	71
3.8	DJIA between February 1, 1927 and December 31, 1954	74
4.1	Autocorrelation functions for Dow Jones volatility.	80
5.1	Volatility snippet.	86
5.2	The almost-integration effect.	94
5.3	The almost-integration effect II.	95
5.4	Histograms of GARCH(1,1) estimations of artificial series.	103
5.5	Power spectrum estimate of Scenario 1.	104
5.6	Power spectrum estimate of the Dow Jones series.	105
5.7	Power spectrum estimate of the S&P500 series.	106
5.8	Sample autocorrelation function of GARCH residual of the Dow Jones series.	107
6.1	Roots of the lag polynomial of the aggregation of two GARCH(1,1) processes.	116
6.2	Roots of the lag polynomial of the aggregation of two ARMA(1,1) processes.	117
6.3	The almost-integration effect in GARCH(2,2) I.	128
6.4	The almost-integration effect in GARCH(2,2) II.	129
6.5	The almost-integration effect in GARCH(2,2) III.	130
7.1	Yen/dollar rate and Japanese interventions	133
7.2	Yen/dollar volatility and Japanese interventions.	134
7.3	Japanese interventions and reserves 1991 to 2001.	135
7.4	Federal Reserve interventions and reserves 1991 to 2001.	136
7.5	Changepoint detection in yen/dollar volatility.	140

List of Tables

3.1	Estimation of model (3.3.2) on S&P500 sample periods before and after the 1987 stock market crash.	66
3.2	Estimation of model (3.3.2) on S&P500 sample periods before and after the 1987 stock market crash.	67
3.3	Sample distributions of the magnitude of the crash.	73
3.4	Estimation of model (3.3.2) on Dow Jones sample periods before and after the 1929 stock market crash.	75
5.1	GARCH(1,1) Parameters of artificial series.	101
5.2	Moments statistics of the four scenarios.	102
6.1	Setup of GARCH(2,2) simulation experiments	126
6.2	Setup of GARCH(2,2) simulation experiments II	127
7.1	Summary statistics for Japanese and US intervention series	137
7.2	Estimation of model (7.4.2)	139
7.3	Estimation of GARCH(1,1) on detected segments	142

List of Abbreviations

ARCH	Autoregressive Conditional Heteroskedasticity (Model)
ARMA	Autoregressive Moving Average (Time Series)
BBI	Barclay's Bank International
CPI	Consumer Price Index
CRSP	Center for Research on Securities Prices
DJIA	Dow Jones Industrial Average
FIGARCH	Fractionally Integrated \rightarrow GARCH
GARCH	Generalized \rightarrow ARCH
GLR	Generalized Likelihood Ratio
IGARCH	Integrated \rightarrow GARCH
S&P	Standard & Poor's

Chapter 1

Introduction: Mean Reversion in Stock Market Prices

“What goes up must come down.” This folksy wisdom turned out to be a highly non-trivial fact about stock markets.

In the late 1980’s DeBondt and Thaler documented the phenomenon that so-called contrarian strategies outperform the stock market. These are strategies where portfolios are selected according to past performance. “Contrarian” means that portfolios of former “losers” yielded substantially higher returns than portfolios of former “winners”. This was interpreted as evidence of mean reversion, that is, a force that drives prices back to a certain medium level after they went above or below it.

This spawned the discussion about whether there is mean reversion in returns. Intuitively speaking, any force that pushes the price process back to the mean would imply negative autocorrelation at some time scale and would thus induce the systematic success of contrarian strategies.

However, can there be mean reversion in stock market prices or returns and why? Or else, why should there be no mean reversion? It is intuitively clear that prices cannot take just any value:

“Surely, economic law tells us that the price of wheat—whether it be spot [...] or futures [...]—cannot drift sky-high or ground-low. It does have a rendezvous with its destiny of supply and demand, albeit our knowledge of future supply and demand trends becomes dimmer as the envisaged date recedes farther into the future.” (Samuelson 1965, p 44.)

The fact that prices are bounded is often taken to justify mean reverting behavior in prices. This is *not* the way I will consider mean reversion in this thesis. A series of standard normal variables will hardly ever leave the corridor $(-4,4)$, so

it is bounded (in some probabilistic sense) while it is statistically independent at any time horizon. There is no explicit or implicit “force” that drives the process back to its mean; its stationary, independent probability distribution keeps it within its boundaries. I will understand mean reversion in the sense that there is temporal dependency in the data, in some way or other, that is, either in the prices or in the first or in the second moment of the returns.

In the first moment of the returns, mean reversion means the change of the market return in the direction of a reversion level as a reaction to a prior change in the market return. After a positive change in the actual returns, mean reversion causes a negative change at some later point and vice versa. This reverting move can occur with different speeds, it can eliminate the prior change in, say, one day or in one year. Figure 1.1 illustrates the concept.

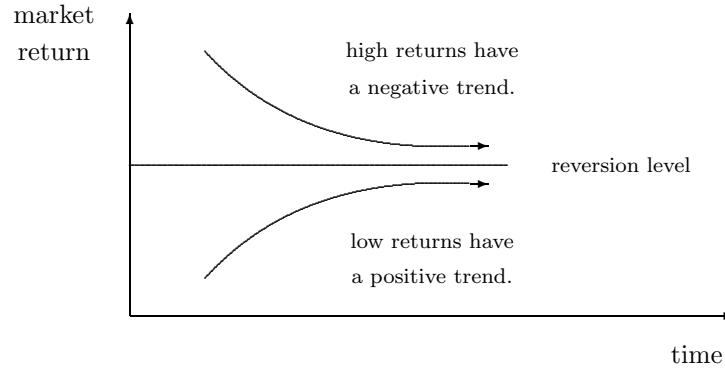


Figure 1.1: The concept of mean reversion.

Shortly after DeBondt and Thaler, Lawrence Summers (1986) showed that when there is a highly persistent mean reverting driver in stock prices, these are statistically indistinguishable from a random walk. Fama and French (1988) and Poterba and Summers (1988) presented corroborating evidence that k -period returns have a time dependent structure that is consistent with a mean reverting component in log prices. The evidence remained ambiguous, largely due to the fact pointed out by Summers (1986) that mean reversion in prices cannot be detected reliably.

Fama and French (1988) measured mean reversion indirectly, by regressing k -period returns on their own lags. Mean reversion would imply a negative relation at some lag k . Poterba and Summers (1988) used a different indirect method, variance ratios. If there is no mean reversion, the variance of k -period returns scales with k . The ratio of the k -period variance to the 1-period variance can thus be normalized to unity by dividing by k . If there is mean reversion, however, the covariance scales with less than k as the process is pushed back when it is deviating from the mean. Therefore, the variance ratio statistic falls below unity.

Some other possible measures of mean reversion have been discussed in the literature (Jegadeesh 1990, Duffee 1991, Kim, Nelson, and Startz 1998, Kim and Nelson 1998) but mean reversion has not become a stylized fact about asset prices and returns. Two contrasting statements of first rate theoreticians illustrate this

“I believe that there is normally considerable mean reversion in the market...” (Black 1988, p 271.)

“It is now well established that the stock market returns themselves contain little serial correlation [...] which is in agreement with the efficient market theory.” (Ding, Granger, and Engle 1993, p 85.)

Some theories were suggested to explain the economic cause of mean reversion. As mean reversion in returns contradicts the hypothesis of efficient markets, one self-suggesting explanation was investor irrationality (Griffin and Tversky 1992). Rational approaches like Black (1989) or Cechetti, Lam and Mark (1990) suggested that the reason for mean reversion is that investors maximize their consumption over their whole life span. When during a boom phase of the economy the expected returns increase (and the actual returns follow), the investor will expect higher income and increase her consumption by dissaving and selling stocks. This pushes the prices down right before they increase because of the boom.

In this thesis, I will not try to answer the question whether there is always mean reversion in prices or returns. I will take up an idea of Fischer Black how mean reversion may play a role in stock-market crashes for which it suffices that there is *sometimes* mean reversion in prices or returns. As this is at odds with the hypothesis of efficient markets, this is tantamount to the statement that the market is at times efficient and at times inefficient. Using daily prices of the Dow Jones Industrial Average for one century, I will show that this is indeed the case.

I think that the hypothesis of efficient markets is an indispensable building block for normative asset pricing theory. When fair prices are to be derived, the allowance of any arbitrage opportunities would render the result arbitrary. For descriptive and empirical finance, however, the assumption of efficient markets is too narrow, as the occasional mean reversion illustrates (or the existence of an arbitrage desk at most investment companies, for that matter). In this thesis, I will therefore walk in and out of efficient markets. The first part concerning mean reversion in prices and returns will deal with inefficient markets while the second part about mean reversion in volatility allows both, efficient and inefficient markets.

A Mean Reversion Theory of Stock Market Crashes

After the stock market crash of 1987, Fischer Black (1988) suggested that it may have been that the cause of the crash was an underestimation of the

expected mean reversion in the market. The idea is that during the boom phase 1982–1987 (as identified in the Brady-Report¹ 1988) investors inferred from the sustained upward moves that the expected mean reversion in the market is low. Had it been higher, the market would have gone down earlier. Even though most investors probably expected a faster reversion *a priori*, they acknowledged the low mean reversion they observed and lowered their own expectations.

The problem was, however, that they could not know that there was a huge hedge position in the market that protected a substantial volume of investments from a faster mean reversion, that is, from a falling market. The hedge position consisted of synthesized put options, positions in futures and bonds that did not show in an obvious way that they were set up to mirror put options.

When the market was disrupted by fundamental news during the week prior to the crash, the dynamic hedges had to be adapted, triggering very large transactions as the market lost up to ten percent in the course of trading. These transactions were reported in the press and their sheer volume surprised the market, as the Brady-Report remarks. This revealed the hedge position to the average investor who had to learn that a substantial group of market participants had had much higher expectations of mean reversion than she had perceived earlier from prices alone. Therefore, the investor had to re-adapt her mean reversion expectation upwards, more into the direction of the *a priori* estimate.

Not only that: the market had to be set into a position as if the mean reversion illusion had not happened. As this illusion had existed for quite a while (about nine months), the market had been allowed to trade below its proper mean reversion strength for a long time and the result was a large upward deviation away from the mean level. This upward deviation now had to be corrected from one day to the next. That was the crash.

In this thesis, I will specify a mean reverting process that is a discretized version of an Ornstein-Uhlenbeck process and examine daily prices of the S&P500 around the stock market crash of 1987. This process allows to estimate the mean reversion parameter directly. I will find Black's theory strongly supported. The estimates of the mean reversion parameter increase drastically from pre-crash periods to post-crash periods. Likelihood ratio tests of the null of a random walk against the mean reverting alternative show a highly significant rejection of the null after the crash.

A changepoint detection study locates the beginning of the illusion at about

¹After the stock market crash of 1987, then U.S. president Ronald Reagan set in the "Task Force on Market Mechanisms" under chairman Nicholas Brady. The Brady-Commission, as it was called, presented its final report on January 8, 1988. The publication of the report, oddly enough, triggered a correction of minus seven percent on the U.S. stock market on the release day. It was feared that restrictive measures like the introduction of circuit breakers would be taken to prevent further crashes.

the end of 1986. This corresponds to the phases of boom and exaggeration identified in the Brady-Report.

When I simulate the model with the parameter estimates obtained from these segments, I obtain a probability of about seven percent for a crash of 20 percent or more and a probability of over 40 percent for a correction of minus 10 percent or more.

In these arguments, I assume that when the market participants have certain preconceptions and expectations, they translate these into the markets by way of their sales and purchases. Economic time series are socially generated data. When investors change their minds, the data generating process changes. For this reason, rigorous equilibrium or no-arbitrage models are an excellent tool for normative theory, where irrationalities have no place, while descriptive theory has to acknowledge that apparent irrationalities exist, at least some of the time. Therefore, market data shows “anomalies” and presents puzzles. They may turn out to be completely rational and in line with equilibrium theory once we learn more about it, as Black (1989) and Cecchetti, Lam and Mark (1990) have suggested for mean reversion in prices and returns.

Mean Reversion in Asset Price Volatility

In the second part of the thesis, I turn to mean reversion in asset price volatility. In contrast to mean reversion in prices and returns, mean reversion in volatility is regarded a stylized fact. It is one possible explanation for the volatility clustering and leptokurtosis that was documented by Mandelbrot (1963), which spawned an intense discussion of the properties of stock returns (e.g. Fama 1965). When asset price variance mean reverts, it has time-dependent, autoregressive dynamics. That is, if a large fluctuation occurs at time t , the variance increases and via the autoregressive process, the variance in $t + 1$ also increases. Therefore, the probability of a large fluctuation in $t + 1$ has also increased, so that fluctuation clusters arise. There are other generators of leptokurtosis and volatility clustering, but this is a popular one. Mandelbrot himself suggested stable Paretian distributions that have infinite variance and that would render classical tools like regression analysis invalid.

For a long time it seemed, therefore, that a better description of volatility data than a constant in the random walk model would come at the cost of abandoning the hypothesis of efficient markets or the most frequently used econometric tools.

Only in 1982 Robert Engle succeeded in formulating a model that allowed for an explicit temporal dependence in volatility data while leaving the hypothesis of efficient markets untouched. This was the autoregressive conditional heteroskedasticity (ARCH) model that was generalized by Bollerslev (1986) to GARCH. Both ARCH and GARCH implicitly model mean reversion and their

good description of the data established mean reversion as a stylized fact of asset price volatility (Engle and Patton 2001).

Figure (1.2) shows the daily returns (upper plot) of the Dow Jones Industrial Average from 1985 to 2001. The lower plot shows the squares of the excess returns, that is the excess over the mean return. This is the volatility series.

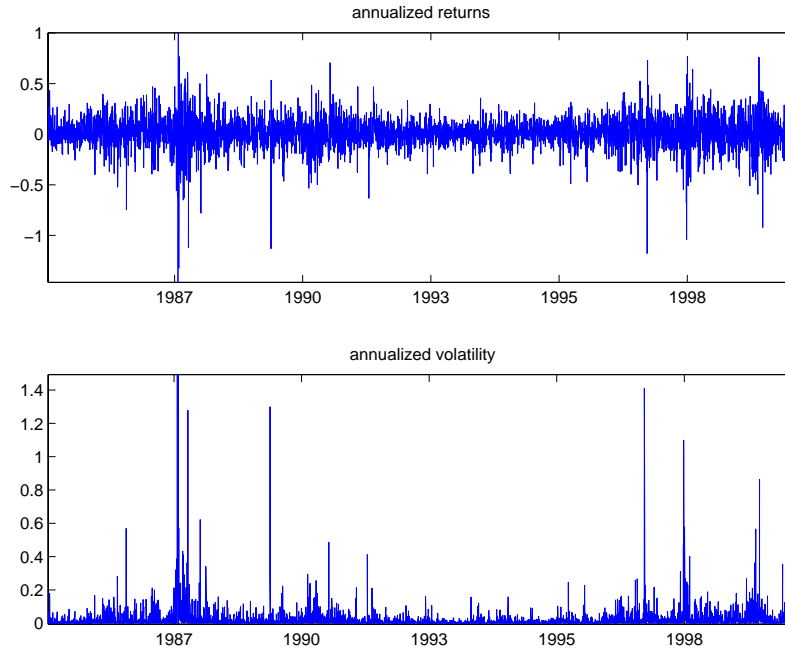


Figure 1.2: Annualized log-returns and volatility (squared distance of log-returns from their mean) of the Dow Jones Industrial Average’s daily closings from January 2nd, 1985 to January 2nd, 2001. The graph is chosen such that the events of October 1987 do not scale the rest of the series into invisibility. The 1987 crash marks in the return series at -4.05 and in the volatility series at 16.5. The upward jump of October 21st, 1987 at 1.53 in the return and at 2.31 in the volatility series.

The period about the stock-market crash of 1987, the time of the Gulf War in the early 1990s, the Asian Crisis of 1997 and the Russian insolvency crisis of 1998 stand out clearly as bulks of high volatility. Contrary to that, the mid-1990s were a period of prolonged calmness.

Unknown Parameter Changes and Almost-Integration in GARCH

ARCH and GARCH models conceptualize this apparent regularity by modelling a “memory” of the volatility process. In other words, the process has an inherent time scale, a certain period that it needs on average to return to its mean after a deviation from it. This length is an interesting piece of information for market participants as well as for researchers. In formulating models of volatility,

the question must be addressed what time scales are to be allowed at all. It is an empirical regularity that for long financial data sets that cover more than five or six years a long memory of the order of months is found.

Recently there has been a number of studies that also find a short correlation time scale of the order of days besides the well known long scale of the order of months (Fouque et al. 2002, Gallant and Tauchen 2001, for instance). Mostly, the short scale is found with methods that are not traditional econometrics tools, like spectral analysis or the Efficient Method of Moments. In this thesis, I will show the existence of the short scale with more traditional tools like GARCH, sample autocorrelation functions but also with spectral estimations.

In GARCH models the time scale is parameterized by the sum of all autoregressive parameters in the conditional variance equation. The longer the time scale, the closer this sum is to unity. For financial data that covers several years, usually a sum of the estimates of the autoregressive parameters very close to one is found, indicating the scale of the order of months. In this thesis, I will refer to this phenomenon as “almost-integration”.

Even if GARCH is a good model for asset price volatility, it would be the oddest thing if the data-generating parameters would remain constant over long time spans of several years. Not only because the series are socially generated, as emphasized above, but also because they are influenced by business cycles, different regimes of fiscal and monetary policy, and occasional financial or political crises, to name just a few.

The traditional econometric method of more or less arbitrarily choosing a sample period, motivated by the availability of the data, has substantial disadvantages in an environment in which regime changes can occur. Randomly picking the boundaries of the sample period is an arbitrary segmentation of the data.

I will show analytically and in simulations that unknown parameter changes in the data generating process cause the almost-integration effect in GARCH. That is, regardless of the data generating mean reversion, which may be quite fast, the GARCH estimation will indicate slow mean reversion when there are parameter changepoints in the data. The reason for this distortion lies in the geometry of the estimation problem, not in the statistical properties of the estimators about which little is known.

The hypothesis is, therefore, that the long scale in financial markets is caused by changes in the data generating parameters, which is quite likely for long range series, and that within periods of constance, mean reversion is fast. This is the short scale that is frequently found recently.

It is therefore very problematic to just specify a sample period. It will probably contain unknown parameter changes and if the instruments used are too

coarse, like a GARCH model, they will be blind to the short scale. A change-point detector study is one way to find a better segmentation. It is also possible to uncover the short scale even with an arbitrary sample period. I will do this in the GARCH(1,1) model by considering the difference of volatility measured from the data and volatility estimated from the GARCH(1,1) model. As the latter can pick up the long scale only, this difference contains the short scale. This is shown with spectral estimation and with the sample autocorrelation function.

I generalize the proof that unknown parameter changepoints cause apparently slow mean reversion to models of order up to GARCH(2,2). Higher order GARCH models are better suited for multiple time scales in the data than GARCH(1,1) specifications but the considerations here show that this almost-integration effect prevents at least models up to the order of GARCH(2,2) from correctly picking up the short time scale. I explore this effect in simulations of the GARCH(2,2) model.

The Yen/Dollar Exchange Rate and Japanese Foreign Exchange Interventions

As another application, I consider the daily exchange rate of the yen against the dollar and newly released data on the foreign exchange interventions of the Japanese authorities. The hypothesis is that the interventions either directly cause shifts in the volatility parameter regimes or at least coincide with these shifts as both, shifts and interventions have a common cause (the Asian crisis of 1997, for instance). Thus, accounting for interventions should reveal the short scale.

I employ a GARCH model with the intervention series as exogenous variable in the conditional mean and the conditional variance equation as well as a changepoint detector for ARCH models (Kokoszka and Leipus 2000). Both methods indicate that the mean reversion within segments is substantially faster than the 62 days obtained by a GARCH(1,1) estimation with constant mean return. The short scale is of the order of 6–12 days.

There is mean reversion in the price data of financial markets, at least temporarily but recurrently in prices and returns and even on multiple, overlaying time scales in volatility. Investors seem to perceive it and seem to have expectations about its strength. Misestimations of these expectations may be a cause of stock market crashes, the crash of 1987 seems to be a case in point. In the case of mean reversion in volatility, GARCH models are not capable of resolving the short scale in global estimations. Together with spectral analysis of the difference of measured and modelled volatility, or when a changepoint detection study of volatility is employed, or when extensions with exogenous variables are considered, the short scale can be found in stock prices and exchange rates.

The structure of the thesis is as follows. Part A (Chapters 2 and 3) considers mean reversion in prices and returns, Part B (Chapters 4 to 7) considers mean reversion in volatility. Chapter 2 discusses various concepts of mean reversion, the relation to the efficient markets hypothesis, possible reasons for mean reversion, and the results of the seminal papers. Chapter 3 describes the mean reversion theory of stock market crashes motivated by Fischer Black's idea. Chapter 4 gives a short introduction to persistence, mean reversion, and long memory in volatility. Chapter 5 considers unknown parameter changes in GARCH(1,1). Chapter 6 extends the results to GARCH(p,q), and Chapter 7 investigates the yen/dollar exchange rate series under consideration of Japanese central bank interventions. Chapter 8 concludes with summaries and an outline of future research arising from the findings in this thesis.

Part A

Mean Reversion in Prices and Returns

Chapter 2

Measuring and Interpreting Mean Reversion in the Data

In this chapter, I will discuss some of the definitions of mean reversion suggested in the literature. The second section considers the relation of mean reversion to the efficient markets hypothesis. I will describe Samuelson's (1965) argument that the futures price of an asset must be serially uncorrelated even if the spot price is correlated. The third section presents the reasons for mean reversion that are discussed in the literature, and the fourth section describes the findings of the early and influential papers on mean reversion in prices and returns (DeBondt and Thaler 1985, Summers 1986, Fama and French 1988, Poterba and Summers 1988).

1 Definitions of Mean Reversion

There is a plethora of definitions of mean reversion. The concept of a process that returns to its mean is so general that many formal definitions can reproduce it. I will introduce three classes of linear time series models that I consider most important for mean reversion modelling. I will frequently refer later to these classes when I discuss the mean reversion models presented in the literature.

Class I: Consider a simple autoregressive process of order one with drift:

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \varepsilon_t, \quad (2.1.1)$$

where ε_t is a zero-mean variate and $\alpha \in (0, 1)$. The unconditional mean of the process is

$$\mathbb{E}x = \frac{\alpha_0}{1 - \alpha_1} \quad (2.1.2)$$

and the persistence parameter α_1 governs the reversion to this mean. Intuitively, the shock ε_{t-1} enters x_t with weight α_1 , it enters x_{t+1} with weight α_1^2 and so forth. That is, the fraction α_1 of the shock is carried forward *per unit of time*

and hence the fraction $(1 - \alpha_1) \in (0, 1)$ is washed out per unit of time. The inverse, $1/(1 - \alpha_1)$ is the average time for a shock to be washed out. It is the mean reversion time. (This argument is formalized in the context of GARCH models in Section 5.2.b).)

Notice that this is a concept of mean reversion that explicitly models *positive* autocorrelations.

Class II: Contrary to that, consider another autoregressive process of order one:

$$x_t = \mu + \alpha(\mu - x_{t-1}) + \varepsilon_t, \quad (2.1.3)$$

where $\mu > 0$ and, again, ε_t is a zero-mean variate and $\alpha \in (0, 1)$. The unconditional mean of this process is μ . The middle term on the right-hand side, the mean reversion term, measures the deviation of the process from the mean in the previous period and adds the correction with weight α . If the process was below the mean in the previous period, the process gets an α -kick upwards, if it was above the mean, downwards. If we substitute $\alpha_0 = \mu + \alpha\mu$ and $\alpha_1 = -\alpha$ in process (2.1.1), we see that the processes are equivalent up to the sign of the persistence parameter.

That is, process (2.1.3) captures a concept of mean reversion that explicitly models *negative* autocorrelations.

Also, the derivation of the mean reversion time is different: A shock to the process away from its mean is washed out with the fraction α per unit of time. The mean reversion time is therefore the inverse, $1/\alpha$.

Class III: Granger and Joyeux (1980) suggested a third important class of mean reversion models and with that they started the literature on long memory time series.

Consider again model (2.1.1). The solution conditional on the start value x_0 is given by

$$x_t = \frac{\alpha_0(1 - \alpha_1^t)}{1 - \alpha_1} + \alpha_1^t x_0 + \sum_{j=0}^{t-1} \alpha_1^j \varepsilon_{t-j}$$

and for large t and stationary AR(1) models where $0 < \alpha_1 < 1$ we have

$$x_t \approx \mathbb{E}x + \sum_{j=0}^{t-1} \alpha_1^j \varepsilon_{t-j}.$$

From this expression the autocorrelation coefficients $\rho(k) = \alpha_1^k$ can be read directly. They decay geometrically. In general, they decline geometrically for stationary ARMA(p,q) models (Brockwell and Davis 1991, p. 91ff).

Suppose now that x_t is an integrated AR(1) model of order $d \in \mathbb{N}$, that is, the d -th difference series

$$y_t := \nabla^d x_t = (1 - L)^d x_t,$$

where L is the lag operator, is AR(1) without drift:

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t.$$

Granger and Joyeux (1980) generalized the available theory for $d \in \mathbb{N}$ to $d \in (-0.5, 0.5)$, that is, to allow for fractional d . The shift operation is defined by the infinite binomial expansion

$$(1 - L)^d = 1 - dL + \frac{d(d-1)}{2!}L^2 - \frac{d(d-1)(d-2)}{3!}L^3 + \dots$$

The autocorrelations can be shown to have the order

$$\rho(k) \sim Ck^{2d-1} \text{ as } k \rightarrow \infty$$

where $C > 0$. The decay is thus slower than geometrical.

The process x_t nevertheless reverts to its mean zero. A non-zero mean μ can be introduced by considering

$$y_t = (1 - L)^d (x_t - \mu).$$

The case of $d \in (0.5, 1)$ is treated by differencing $x_t - \mu$ once:

$$y_t = (1 - L)^{d-1} (1 - L)(x_t - \mu),$$

so that for $y'_t := (1 - L)(x_t - \mu)$ the fractional parameter $d' := d - 1 \in (-0.5, 0)$ and the theory applies.

It must be emphasized that all three model classes are classes of so-called “reduced form” models, that is, they do not try to explicitly describe economic processes and therefore they do not have an immediate interpretation. Instead, they simply *model* the observed effect in the data.

A discussion of possible economic correlations or even causes has to consider other, so-called structural models which try to explicitly capture economic processes. Structural microeconomic models have to incorporate the agents’ expected utility maximization problems. Structural macroeconomic models are for the most part regressions with macroeconomic data (interest rates, GDP) as regressors. There are no “canonical” models of these types, the predominant models for stock price dynamics are all of reduced form (the standard Black-Scholes model for stock prices, stochastic volatility models, ARCH and GARCH,

etc.). Some interesting microeconomic models that describe mean reversion in prices and returns are discussed in Section 2.3.

The reason for the dominance of reduced form models is the complex form of the data. There is no hope to find even a reduced form model that describes all the characteristic features of stock price data, let alone a structural one. Therefore, model selection is highly subjective, it depends completely on the objectives of the respective investigation. For example, the standard Black-Scholes model was formulated with the interest to have a mathematically tractable model that allows option pricing while capturing some important data features, in particular its obvious stochasticity. It is well known that it does not capture asymmetry, leptokurtosis, and volatility clustering. Stochastic volatility models were formulated with the interest to have a generalization of the standard model that allows for leptokurtosis and volatility clustering while it is still tractable and still allows option pricing.

In the case of mean reversion, the primary interest is to find time scales in the data, and therefore, the simplest models that feature time scales are used. If we are to find an economic reason for the existence and determination of the time scale, we have to refer to other models.

With this categorization in hand, I will now discuss the concepts of mean reversion in prices and in the first moment of the return distribution, the expected return.

2 Efficient Markets and Mean Reversion

Why should there be no mean reversion in prices or returns in efficient markets? Consider an efficient market with a risk-free interest rate i . That is to say, the *discounted* stock price is subject to the martingale condition

$$e^{-iT} \mathbb{E}(S_{t+T} | S_t) = S_t. \quad (2.2.4)$$

In words, today at time t there is no information available whatsoever that allows one to predict tomorrow's stock price any better than by calculating the risk-free interest for one day and adding it to today's stock price. Then,

$$\tilde{r}_t = \mathbb{E}(r_t | \mathcal{F}_t) = \frac{\mathbb{E}(S_{t+1} | S_t) - S_t}{S_t} = e^i - 1 \approx i.$$

The covariance of the expected returns \tilde{r}_t is given by

$$\text{cov}(\tilde{r}_t, \tilde{r}_{t+T}) = \mathbb{E}[(\tilde{r}_t - \mathbb{E}\tilde{r}_t)(\tilde{r}_{t+T} - \mathbb{E}\tilde{r}_{t+T})] = \mathbb{E}(\tilde{r}_t \tilde{r}_{t+T}) - \mathbb{E}\tilde{r}_t \mathbb{E}\tilde{r}_{t+T}.$$

Since

$$\mathbb{E}(\tilde{r}_t \tilde{r}_{t+T}) = \mathbb{E} \left[\frac{\mathbb{E}(S_{t+1} | S_t) - S_t}{S_t} \frac{\mathbb{E}(S_{t+T+1} | S_{t+T}) - S_{t+T}}{S_{t+T}} \right] = (e^i - 1)^2,$$

and $\mathbb{E}\tilde{r}_t = \mathbb{E}\tilde{r}_{t+T} = e^i - 1$ we have that

$$\text{cov}(\tilde{r}_t, \tilde{r}_{t+T}) = 0 \text{ for any } T. \quad (2.2.5)$$

That is, in an efficient market there should be no temporal dependence between *expected* returns. It must be emphasized that this says nothing about actual returns. Only under rational expectations actual and expected returns will coincide.

It is common to attribute the proof that prices and returns must be serially uncorrelated to Samuelson (1965). Interestingly, Samuelson was not concerned about serial correlation in the spot price but in the futures price, a detail that often gets lost in the reference. That is, he proved that the sequence

$$\{e^{-iT}\mathbb{E}(S_{t+T}|\mathcal{F}_t); e^{-i(T-1)}\mathbb{E}(S_{t+T}|\mathcal{F}_{t+1}); \dots; \mathbb{E}(S_{t+T}|\mathcal{F}_{t+T})\},$$

with terminal condition $\mathbb{E}(S_{t+T}|\mathcal{F}_{t+T}) = S_{t+T}$ (up to commissions) is a martingale, as opposed to the series

$$\{S_t; e^{-i}S_{t+1}; \dots; e^{-iT}S_{t+T}\},$$

for which the martingale property proof took Fischer Black and Myron Scholes up to 1973 and a fully parameterized model for S .

I will give Samuelson's argument for two reasons: (1) it is a splendid example of the early applications (Bachelier aside) of formal probability to asset pricing, (2) Samuelson actually used a class I mean reverting model for the spot price and showed that even if the spot price is serially correlated, the future price is not.

Samuelson assumed two properties of the probability distribution of the price.

- First, that there is a probability law that governs the price movements, denoted by

$$\mathbb{P}(S_{t+T} \leq c | \{S_t, S_{t-1}, \dots, S_0\}). \quad (2.2.6)$$

- Second, that the probability for a certain price at a certain time is given by the sum of the probabilities of the different mutually exclusive ways by which the price process can arrive at that price:

$$\begin{aligned} & \mathbb{P}(S_{t+T} \leq c | \{S_t, S_{t-1}, \dots, S_0\}) \\ &= \int_{y=-\infty}^{y=\infty} \mathbb{P}(S_{t+T} \leq c | \{y, S_t, S_{t-1}, \dots, S_0\}) d\mathbb{P}(S_{t+1} \leq y | \{S_t, S_{t-1}, \dots, S_0\}) \\ & . \end{aligned} \quad (2.2.7)$$

Here, we sum over all the different possible prices that can occur at time $t+1$ (the integration indicator y stands for S_{t+1}) in order to assess the probability law for S_{t+T} .

Again, Samuelson derived his results for future prices only. He assumed that there is no risk-aversion and the future price $F(T, t)$, that is the price to be paid today (at time t) to receive one stock at the future time $t + T$ (when the price is S_{t+T}), is given by

$$\begin{aligned} F(T, t) &= e^{-iT} \mathbb{E}(S_{t+T} | \{S_t, S_{t-1}, \dots, S_0\}) \\ &= e^{-iT} \int_{y=-\infty}^{y=\infty} y d\mathbb{P}(S_{t+T} \leq y | \{S_t, S_{t-1}, \dots, S_0\}) \end{aligned} \quad (2.2.8)$$

with terminal condition

$$F(0, t + T) = S_{t+T},$$

that is, at maturity, the future price must equal the stock price, commissions aside.

These three basic assumptions: that there is a probability distribution, that it has a standard linear property, and that the future price is given by the discounted expected value of the spot price at maturity, allowed Samuelson to derive the “Theorem of Fair-Game Future Pricing” as he called it. That is, that the series of future prices is a martingale.

If the spot prices $\{S_t\}$ with probability law (2.2.6) are subject to (2.2.7), and the future price sequence

$$\{F(T, t); F(T - 1, t + 1); \dots; F(1, t + T - 1); F(0, t + T)\}$$

is given by the discounted expected values according to (2.2.8), then this sequence is a martingale, that is,

$$\mathbb{E}(F(T - 1, t + 1) | \{S_t, S_{t-1}, \dots, S_0\}) = e^i F(T, t). \quad (2.2.9)$$

To see why this must be, we simply apply the definitions:

$$\begin{aligned} F(T, t) &= e^{-iT} \int_{y=-\infty}^{y=\infty} y d\mathbb{P}(S_{t+T} \leq y | \{S_t, S_{t-1}, \dots, S_0\}) \\ F(T - 1, t + 1) &= e^{-i(T-1)} \int_{y=-\infty}^{y=\infty} y d\mathbb{P}(S_{t+T} \leq y | \{S_{t+1}, S_t, \dots, S_0\}) \\ &=: e^{-i(T-1)} f(S_{t+1}, S_t, \dots, S_0). \end{aligned}$$

Then, by the definition of the expected value,

$$\begin{aligned}
& e^{-i} \mathbb{E}(F(T-1, t+1) | \{S_t, S_{t-1}, \dots, S_0\}) \\
&= e^{-i} e^{-i(T-1)} \int_{-\infty}^{\infty} f(y, S_t, S_{t-1}, \dots, S_0) d\mathbb{P}(S_{t+1} \leq y | \{S_t, S_{t-1}, \dots, S_0\}), \\
& \quad \text{plugging in from the expression for } F(T-1, t+1) \text{ above,} \\
&= e^{-iT} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} z d\mathbb{P}(S_{t+T} \leq z | \{y, S_t, \dots, S_0\}) \right] d\mathbb{P}(S_{t+1} \leq y | \{S_t, S_{t-1}, \dots, S_0\}), \\
& \quad \text{and, provided that changing the order of integration is possible:} \\
&= e^{-iT} \int_{-\infty}^{\infty} z d \left[\int_{-\infty}^{\infty} \mathbb{P}(S_{t+T} \leq z | \{y, S_t, \dots, S_0\}) d\mathbb{P}(S_{t+1} \leq y | \{S_t, S_{t-1}, \dots, S_0\}) \right], \\
& \quad \text{finally, from (2.2.7),} \\
&= e^{-iT} \int_{-\infty}^{\infty} z d\mathbb{P}(S_{t+T} \leq z | \{S_t, S_{t-1}, \dots, S_0\}) \\
&= F(T, t),
\end{aligned}$$

which proves the proposition.

The intuition behind the theorem is that it is *not* the spot price over time that is considered, but the (discounted) expectation over time of the one price S_{t+T} . The theorem tells us that if the future price for delivery of the stock S at time $t+T$ is the discounted expected value of S_{t+T} , then this future price is a martingale.

The fascinating result is that a mean reverting spot price has a martingale future price series. Again, the reason is that for the future price it is not the dynamics of the spot that is of interest but the dynamics of the expectation of the spot price at a single point in time.

Let the stock price process be given by the autoregressive model

$$S_{t+1} = \alpha S_t + \varepsilon_t, \quad (2.2.10)$$

where ε_t is white noise with mean μ . Then

$$F(T, t) = e^{-iT} \mathbb{E}(S_{t+T} | \{S_t, S_{t-1}, \dots, S_0\}) = e^{-iT} (\alpha^T S_t + \mu T),$$

so that

$$\begin{aligned}
\mathbb{E}(F(T-1, t+1)) &= e^{-i(T-1)} \left[\int_{-\infty}^{\infty} \alpha^{T-1} S_{t+1} d\mathbb{P} + \mu(T-1) \right] \\
&= e^{-i(T-1)} \left[\int_{-\infty}^{\infty} \alpha^{T-1} (\alpha S_t + \varepsilon_t) d\mathbb{P}(\varepsilon_t) + \mu(T-1) \right] \\
&= e^{-i(T-1)} (\alpha^T S_t + \mu T) \\
&= e^i F(T, t).
\end{aligned}$$

For $\alpha \in (0, 1)$, the spot price model is class I and reverting to its mean $\mu/(1-\alpha)$.

Eugene Fama, in his first article on efficient capital markets in 1970, was just as fine with mean reversion in returns. His formalization of the idea that in an efficient market all information is “fully reflected” in the stock prices also allowed for temporal dependence. The efficiency condition read

$$\mathbb{E}(S_{t+1} | \mathcal{F}_t) = (1 + \mathbb{E}(r_t | \mathcal{F}_t)) S_t,$$

where $\mathbb{E}(r_t | \mathcal{F}_t)$ “projected on the basis of the information \mathcal{F}_t would be determined from the particular expected return theory at hand” (Fama 1970, p 384, notation adapted). This might very well include temporal dependence in returns. Fama considered the difference of the stock price from its conditional expectation:

$$x_{t+1} := S_{t+1} - \mathbb{E}(S_{t+1} | \mathcal{F}_t),$$

where by definition of the conditional expectation

$$\mathbb{E}(x_{t+1} | \mathcal{F}_t) = 0,$$

and also the difference in returns from their conditional expected value

$$y_{t+1} := r_{t+1} - \mathbb{E}(r_{t+1} | \mathcal{F}_t),$$

such that

$$\mathbb{E}(y_{t+1} | \mathcal{F}_t) = 0. \tag{2.2.11}$$

If the returns process is class I mean reverting

$$r_{t+1} = \mu + \alpha r_t + \varepsilon_t, \tag{2.2.12}$$

with ε_t some zero-mean white noise process and $\alpha \in (0, 1)$, still condition (2.2.11) will hold, again, by definition.

Formulated this way, efficient markets were a tautology mathematically. Economically, however, the notion of the conditional expectation in (2.2.11) meant

that all information about the future stock return that is given by today's return is *properly anticipated*. That is, model (2.2.12) is correctly specified and markets therefore correctly expect the return $\mathbb{E}(r_{t+1} | \mathcal{F}_t) = \mu + \alpha r_t$ so that $\mathbb{E}(y_{t+1} | \mathcal{F}_t) = 0$.

To summarize, mean reversion and temporal dependence of returns in general contradicts the hypothesis of efficient markets. The early works on the martingale property before Black and Scholes (1973), as exemplified by Samuelson's (1965) and Fama's (1970) argument here, were not able to pinpoint the martingale property for the discounted stock price process. To assume temporal dependence of the spot price process or even of the expected returns was a natural approach as prices often seem to follow seasonal or mean reverting patterns. Only with the development of the Black-Scholes-Merton analysis it became clear that this assumption violates efficiency.

3 Rationales of Mean Reversion

As we have seen, mean reversion in expected returns is at odds with the hypothesis of efficient markets. Therefore, it is not surprising that many explanations of mean reversion in returns draw on investor psychology. Keynes' (1936) famous "animal spirits" that are supposed to drive investor's decisions were put into a more scientific shape in the works of Tversky and Kahneman, for example in their (1981) article. Later, they developed a more economic explanation, the overreaction hypothesis.

Temporal dependence at some time scale that drives a return process back to the mean is often explained by the hypotheses of underreaction and overreaction. The hypothesis of underreaction states that investors are slow in appreciating good news about a stock. The news are incorporated slower into prices than appropriate and this leads to further positive returns. That is, underreaction explains positive autocorrelation in returns at some time scale.

Contrary to that, overreaction means that once a run of good news occurs, that is, a consecutive sequence of good news arrives, investors believe that this trend prevails and consequentially bid the price up higher than the appropriate level. The first arrival of negative news then induces a large negative jump. Overreaction thus corresponds to negative correlation of returns.

So which effect is prevalent, over- or underreaction? Griffin and Tversky (1992) argue that both can coexist and in fact explain the data. They distinguish between the "strength" and the "weight" of a signal. For example, in financial markets, a run of positive earning announcements has great "strength" but little "weight". That is, it is "strong" in the sense that it is much noticed and commented in the markets and in the financial press. It is "light" in the sense that even in an i.i.d. random Bernoulli chain, the probability of the occurrence of runs is surprisingly high. Thus, a run of positive earning announcements

provides little evidence that there is autocorrelation in the earnings process and that it is any more likely for the next announcement to be positive rather than negative. Contrary to that, a single positive earning announcement is “weak” in the sense that it is not much noticed and seen as a transitory bit of information but “heavy” in the sense that in fact this single piece of information has substantial weight for forecasting the level of earnings. The hypothesis is that investors tend to focus too much on “strength” and too little on “weight”. By that, they underreact to single positive earning announcements but overreact to runs. This creates positive autocorrelation on short scales between one month and one year and negative autocorrelation on long scales between three and four years, in line with the early findings of mean reversion in the 1980s that I will discuss in the following section.

Based on these concepts, Barberis, Shleifer and Vishny (1998) present their “model of investor sentiment” and calibrating it with real data, they are able to explain a substantial fraction of the later excess returns of “loser” portfolios.

Another explanation of mean reversion, still more in the tradition of economics, is given by Poterba and Summers (1988). It posits that equilibrium required returns may be time varying and thus cause mean reverting behavior of stock prices and returns.

The starting point is that in equilibrium, the stock price is given by the sum of the expected future dividends, discounted with the required future returns.

$$S(t) = \sum_{\tau=t}^{\infty} e^{-\tilde{r}_{\tau,t}(\tau-t)} \tilde{D}_{\tau,t}, \quad (2.3.13)$$

where $\tilde{r}_{\tau,t} := \mathbb{E}(r_{\tau} | \mathcal{F}_t)$ and $\tilde{D}_{\tau,t} := \mathbb{E}(D_{\tau} | \mathcal{F}_t)$.

Investor tastes for current versus future consumption and the stochastic evolution of investment opportunities result in

$$\tilde{r}_{\tau,t} = \alpha \tilde{r}_{\tau,t-1} + \varepsilon_t. \quad (2.3.14)$$

Let $\tilde{D}_{\tau,t}$ be given by some function of today’s information and disturbance

$$\tilde{D}_{\tau,t} = f(\mathcal{F}_t) + \eta_t,$$

and assume

$$\mathbb{E}_{\mathcal{F}_t} \eta_t = 0.$$

Then, shocks to $\tilde{r}_{\tau,t}$ have no effect on expected dividends and as the $\tilde{r}_{\tau,t}$ ’s are mean reverting, there is no influence on the long run expectations. The cumulative effect of a shock must hence be exactly offset by an opposite adjustment in the *current* asset price. Consider for example a positive shock to the required return. The offsetting mechanism described above must result in a lower stock price and

hence negative returns today. Inasmuch as higher required future returns lead in fact to higher actual returns in the future, negative actual returns will be followed by positive actual returns.

Observe the difference to the statement (2.2.5). Here, we are considering the time series of the expectations of the return at a single point in time. This is to model *required* returns. Contrary to that, in (2.2.5) we considered the expectations of the returns time series itself, trying to model *expected* returns. Therefore, (2.3.14) does not violate the efficient market condition, and (2.3.13) is a rational equilibrium model.

We model both required and expected returns by conditional expectations even though they are economically distinct objects. There may be a better way to do it. Fama's statement is still valid:

But we should note right off that, simple as it is, the assumption that the conditions of market equilibrium can be stated in terms of expected returns elevates the purely mathematical concept of expected value to a status not necessarily implied by the general notion of market efficiency. (Fama 1970, p 384).

The autoregressive dynamics of required returns were not very well motivated in Poterba and Summers (1988) and their approach has therefore been refined to representative agent models that allow for mean reversion in fully rational equilibrium settings.

The basic idea in all of these models is that the agent maximizes his consumption over his *complete* life span whereas shocks to returns occur locally. The result is a smoothing effect. When the expected returns increase, for instance because the economy is in a sustained boom phase, the agent will expect higher income and increase his consumption by selling stocks, thereby pushing prices and returns.

The key parameter is the relative risk aversion which also controls the intertemporal substitution of consumption. The higher the risk aversion, the smoother the consumption dynamics and the stronger the mean reversion effect. Risk neutrality eliminates the consumption smoothing and hence the mean reversion effect. Fine examples of this class of models are Cechetti, Lam and Mark (1990) and Black (1990).

The state of the economy is assumed to be exogenous in these models and output fluctuations are captured in some reduced form manner, for example in a Markov chain indicating boom or recession. It is natural to extend the models to endogenously determined production and model changes in the state of the economy in the form of veritable output shocks. Examples of these type can be found in Basu (1993) and Basu and Vinod (1994). An excellent discussion of

representative agent models that accommodate mean reversion is presented in Bodmer (1996).

4 Measurement and Evidence of Mean Reversion

The discussion about whether stock returns actually exhibit mean reversion was started by the paper of DeBondt and Thaler (1985). They grouped stocks listed at the New York Stock Exchange between January 1926 and December 1982 into “winner” and “loser” portfolios based on the performance in the 16 independent three-year periods between January 1930 and January 1975. The basis was monthly returns. If a stock performed well in a three-year period it was grouped into the “winner” portfolio. Good performance meant that the stock belonged to the top 35, top 50, or top decile. Similarly, the “loser” portfolio consisted of the bottom 35, bottom 50, or bottom decile. For each portfolio, the cumulative excess returns for the next three years were calculated, that is the returns in excess of the average returns of all considered stocks in that period. The question of interest was if stocks that performed well in the first three years, the portfolio formation period, would also show significant excess returns in the next three years, the test period. Market efficiency implies that the performance in the past has no predictive power whatsoever for later periods.

The finding was spectacular as the winner portfolios significantly underperformed the market in the test periods. The loser portfolios yielded significantly higher returns in the test periods. The effect was asymmetric: The outperformance of the former losers was much larger than the underperformance of the former winners. The result was robust to the different top/bottom groupings as well as to different lengths of the formation and test periods.

One year later Summers (1986) gave a possible reason why the statistical techniques available had failed to reliably detect mean reversion. He suggested an alternative model that is able to capture rational or irrational deviations from fundamental asset values and showed that commonly employed statistical methods have virtually no power against this alternative. He points out that the fact that a vast literature has been unable to reject the null of efficiency does not mean that it can be accepted “as students of elementary statistics are constantly reminded.”

Summers starts with a model of efficient markets in which stock prices are the sums of the discounted expected dividends. This implies that the stock price at any given time is equal to its discounted future value plus expected dividend of the period considered.

$$S_t = Q_t := e^{-i} \mathbb{E} S_{t+1} + \mathbb{E} D_t, \quad (2.4.15)$$

where S_t denotes the stock price, i is the risk free interest rate, and D_t stands for the dividend in period t . The log-returns under efficiency are given by

$$r_t^* = q_t - q_{t-1},$$

where $q_t = \log Q_t$, the logarithm of the risk neutral fundamental value of the stock. Summers imposes the condition that the expected efficient return be the risk free interest rate

$$\mathbb{E}r_t^* = i$$

instead of deriving it from the martingale condition as in Section 2.2. The result is that the expected return is disturbed only by a single source of randomness that is coming from the dividend stream

$$r_t^* = i + \eta_t, \quad (2.4.16)$$

where η_t is white noise.

The alternative specification to account for “fads” or time varying required returns is a class I mean reverting model.

$$\log S_t = q_t + z_t, \quad (2.4.17)$$

where q_t is the logarithm of the fundamental value as above,

$$q_t = q_{t-1} + i + \eta_t, \quad (2.4.18)$$

disturbed only by random fluctuations in the dividends, and z_t is the class I mean reverting deviation from the fundamental value

$$z_t = \alpha z_{t-1} + \varepsilon_t. \quad (2.4.19)$$

Here, α is assumed to be less than but close to one, and ε_t is a white noise disturbance that reflects all the less rational factors or the influence of the time varying required returns. Rationality and irrationality are assumed to be orthogonal:

$$\mathbb{E}\varepsilon_t\eta_t = 0 \quad \forall t.$$

The logarithm of the stock price is thus given by two components, the fundamental value complying with market efficiency and the deviation that may or may not be irrational but in any case violates the efficiency hypothesis. The fundamental value is a random walk with drift given by the risk free interest rate; it is also called the permanent component as it has indefinite memory. The deviation is a stationary process, as the persistence parameter α is assumed to be less than one. It is therefore also called the stationary or transitory component. The deviation may cause large swings away from the fundamental value that die

out only very slowly because α is close to one. Nevertheless, it has no influence on the very long run as it is reverting to the mean zero.

To make his key argument, Summers is heading for the autocorrelations of the excess returns over the risk free interest rate. From (2.4.17) and (2.4.16) he obtains

$$\begin{aligned} r_t &= \log S_{t+1} - \log S_t \\ &= q_{t+1} - q_t + z_{t+1} - z_t \\ &= i + \eta_t + z_{t+1} - z_t, \end{aligned}$$

so that the excess returns are

$$r_t(i) := r_t - i = \eta_t + z_{t+1} - z_t. \quad (2.4.20)$$

Therefore, and as $\mathbb{E}\eta_t\varepsilon_t = 0$,

$$\begin{aligned} \mathbb{E}r_t(i)^2 &= \mathbb{E}(\eta_t^2 + z_{t+1}^2 + z_t^2 + 2\eta_t z_{t+1} - 2\eta_t z_t - 2z_{t+1} z_t) \\ &= \sigma_\eta^2 + 2(1 - \alpha)\sigma_z^2. \end{aligned} \quad (2.4.21)$$

Plugging this into the autocorrelation yields

$$\begin{aligned} \rho_k &= \frac{\text{cov}(r_t(i), r_{t+k}(i))}{\sigma_{r(i)}^2} = \frac{\mathbb{E}((\eta_t + z_{t+1} - z_t)(\eta_{t+k} + z_{t+k+1} - z_{t+k}))}{\sigma_\eta^2 + 2(1 - \alpha)\sigma_z^2} \\ &= \frac{-\alpha^{k-1}(1 - \alpha)^2\sigma_z^2}{\sigma_\eta^2 + 2(1 - \alpha)\sigma_z^2} \end{aligned} \quad (2.4.22)$$

Note that the autocorrelation coefficients implied by the model are all negative.

Now, Summers makes a striking point: Assume that you have roughly 16 years of daily data, 4000 observations. Let the null hypothesis be market efficiency, constant volatility, and normally distributed errors. Then, the standard error of the estimation of the first order autocorrelation coefficient ρ_1 is under the null hypothesis given by

$$\frac{1}{\sqrt{3997}} \approx 0.0158,$$

as three parameters, the mean, the standard deviation, and the first order autocorrelation coefficient are estimated.

Suppose now that the alternative given by the “fads”-model is true, that is, (2.4.17) through (2.4.19) is the data-generating process with mean reversion parameter $\alpha = 0.99$. That is to say, it will take on average 100 days for a deviation from the fundamental value to die out. Suppose further that $\sigma_z^2 = 0.09/20$, implying that the standard deviation of the stationary component for *monthly* stock prices is 30 percent. In other words, there is a $1 - 0.6827 = 0.3173$

chance that monthly stock price observations will be more than 30 percent off their fundamental value in either direction. This amounts to a quite substantial inefficiency. From the returns data, let $\sigma_{r(i)}^2 = 1.1\text{e-}4$. This is a realistic value for daily observations covering 16 years. Then, from (2.4.21) you can back out $\sigma_\eta^2 = 2\text{e-}5$. Plugging into (2.4.22) yields an expected value of

$$\rho_1 = -0.0041.$$

Compare this to the standard error of 0.0158 under the null. There is no hope to significantly tell the fads alternative apart from the efficiency null using only 4000 observations. Instead, from $0.0041\sqrt{n} = 2$ you need approximately $n \approx 238000$ observations or 950 years of data. If you assume a smaller variance of the mean reverting process and thereby a smaller average deviation from the fundamental value, matters get worse.

Summers adds a comment on the so called event studies literature. These investigate the speed with which specific news announcements that affect a stock are incorporated into the price. Most studies show that market prices absorb news instantaneously. Summers argues that if you assume that the stock price is composed of an efficient and an inefficient component, the question is beside the point, as the news that affect the rational valuation via dividend expectations have nothing to do with the “irrational” driver. In terms of the model, event studies look at how quickly η is absorbed into q . The deviations from fundamentals however arise from ε and z .

Along the same lines, the argument that the market is efficient because arbitrageurs fail to make sustained profits does not convince Summers. These arbitrageurs have the same problems as the researcher in finding reliable evidence of a significant mispricing. Large persistent valuation errors simply leave no statistically discernible trace and any arbitrageur who tries to exploit the small and noisy excess returns takes a high risk of being on the wrong side. Summers concludes that *“the hypothesis that market valuations include large persistent errors is as consistent with the available empirical evidence as the hypothesis of market efficiency.”*

Summers does not provide any reasons why the valuation error should take a mean reverting form. There is also no economic justification of the assumption that the process is highly persistent. The “almost” random walk quality of the data is certainly in support, but there is no economic theory behind Summers assumptions. Eventually, the finding is not that surprising as the constructed mean reverting driver deceptively resembles a random walk as the persistence parameter is so close to one.

Fama and French (1988) take up Summers’ model and use it to analyze market data. They observe that the first order autocorrelation of the k -period returns is negative and must have a U-shaped pattern with the k .

To see this, consider first the first order autocorrelation of k -period changes in z_t :

$$\rho_{z,k} = \frac{\text{cov}(z_{t+k} - z_t, z_t - z_{t-k})}{\text{var}(z_{t+k} - z_t)},$$

where

$$\begin{aligned} \text{cov}(z_{t+k} - z_t, z_t - z_{t-k}) &= \mathbb{E}(z_{t+k}z_t - z_{t+k}z_{t-k} - z_t^2 + z_tz_{t-k}) \\ &= 2 \text{cov}(z_t, z_{t+k}) - \text{cov}(z_t, z_{t+2k}) - \sigma_z^2. \end{aligned}$$

From (2.4.19) we have

$$\mathbb{E}(z_{t+k}z_t) = \mathbb{E}(\mathbb{E}(z_{t+k})z_t) = \alpha^k \sigma_z^2,$$

and for the same reason

$$\mathbb{E}(z_{t+2k}z_t) = \alpha^{2k} \sigma_z^2,$$

such that

$$\begin{aligned} \text{cov}(z_{t+k} - z_t, z_t - z_{t-k}) &= 2\alpha^k \sigma_z^2 - \alpha^{2k} \sigma_z^2 - \sigma_z^2 \\ &= -(\alpha^k - 1)^2 \sigma_z^2. \end{aligned}$$

The variance in the denominator of the autocorrelation is given by

$$\text{var}(z_{t+k} - z_t) = \mathbb{E}(z_{t+k}^2 - 2z_{t+k}z_t + z_t^2) = 2(1 - \alpha^k) \sigma_z^2. \quad (2.4.23)$$

The expected change in z from t to $t + k$ is

$$\mathbb{E}_t(z_{t+k} - z_t) = (\alpha^k - 1)z_t,$$

and its variance therefore

$$\text{var}(\mathbb{E}_t(z_{t+k} - z_t)) = (\alpha^k - 1)^2 \sigma_z^2.$$

Thus,

$$\rho_{z,k} = \frac{-\text{var}(\mathbb{E}_t(z_{t+k} - z_t))}{\text{var}(z_{t+k} - z_t)} = \frac{-(1 - \alpha^k)^2 \sigma_z^2}{2(1 - \alpha^k) \sigma_z^2} = \frac{\alpha^k - 1}{2}. \quad (2.4.24)$$

First of all, it is remarkable that the positive autocorrelation in the stationary component of the stock price implies negative autocorrelation in its first difference. This negative autocorrelation is (minus) the ratio of the variance of the expected change in z to the variance of the actual change in z .

From (2.4.24) it can be seen that when α is close to one, the autocorrelation for small lags is minute and converges slowly to -1/2 with growing k . This corresponds to Summers' observation that autocorrelations from small lags are

too small to be detected. Therefore, Fama and French concentrate on long-horizon returns.

A natural approach is to regress independent observations of k -period returns on their own lags. Denote the k -period return by $r_{t,t+k} := \log S_{t+k} - \log S_t$. Then, the regression slopes are the autocorrelations

$$\begin{aligned}\beta_k &= \frac{\text{cov}(r_{t,t+k}, r_{t-k,t})}{\text{var}(r_{t,t+k} - r_t)} \\ &= \frac{\text{cov}([q_{t+k} - q_t] + [z_{t+k} - z_t], [q_t - q_{t-k}] + [z_t - z_{t-k}])}{\text{var}([q_{t+k} - q_t] + [z_{t+k} - z_t])}.\end{aligned}\tag{2.4.25}$$

As $\mathbb{E}\eta_t \varepsilon_t = 0$, we have

$$\beta_k = \frac{\text{cov}(q_{t+k} - q_t, q_t - q_{t-k}) + \text{cov}(z_{t+k} - z_t, z_t - z_{t-k})}{\text{var}(q_{t+k} - q_t) + \text{var}(z_{t+k} - z_t)}.$$

As q is a random walk, $\text{cov}(q_{t+k} - q_t, q_t - q_{t-k}) = 0$, and thus

$$\begin{aligned}\beta_k &= \frac{\text{cov}(z_{t+k} - z_t, z_t - z_{t-k})}{\sigma_{\eta,k}^2 + \text{var}(z_{t+k} - z_t)} \\ &= \frac{\rho_{z,k} \text{var}(z_{t+k} - z_t)}{\sigma_{\eta,k}^2 + \text{var}(z_{t+k} - z_t)} \\ &= \frac{-\text{var}(\mathbb{E}_t(z_{t+k} - z_t))}{\sigma_{\eta,k}^2 + \text{var}(z_{t+k} - z_t)},\end{aligned}$$

where $\sigma_{\eta,k}^2$ denotes the variance of k -period differences in η .

Therefore, β_k measures the proportion of the variance of the k -period returns that is explained by the mean reversion of the stationary component z . If the price does not have a stationary component, the regression slope β_k is zero. If the price does not have a random walk component, $\beta_k = \rho_{z,k}$ and the regression slope converges to $-1/2$ for large lags k . If both components are present, the mean reverting term pushes the regression slope towards $-1/2$ and the random walk component pushes them towards zero. The variance $\text{var}(z_{t+k} - z_t) = 2(1 - \alpha^k)\sigma_z^2$ of the k -period differences in z in the denominator of the regression slope converges to $2\sigma_z^2$ for large k . The variance of the random walk component grows with k as $\sigma_{\eta,k}^2 = \sigma_\eta^2 \cdot k$. Therefore, the random walk component eventually dominates. At small lags, the regression slopes are small, as found by Summers, for medium lags, they grow more negative in the direction of $-1/2$, and move back towards zero as the random walk component begins to dominate for large k . Hence a U-shaped pattern of the β_k is expected. This is the pattern that Fama and French try to find in the data.

They use data from the Center for Research in Securities Prices (CRSP) consisting of monthly returns on the NYSE listed stocks in the period 1926–1985. They group the stocks into portfolios according to size and according to

industries. For each portfolio, the one-month returns are transformed into continuously compounded returns, adjusted for the inflation rate with the Consumer Price Index (CPI) and summed to obtain overlapping monthly observations of larger-horizon returns. The same procedure is applied to the two indices that the CRSP supplies, an equally weighted and a value-weighted index.

For both types of portfolios, industry and size, the negative autocorrelations of the k -period returns are clearly confirmed for k -horizons ranging from one to ten years. They are U-shaped as expected, with the minimum attained at about - 35 percent at the horizons three to five years. This means that about 35 percent of the variance of three to five year returns is caused by the mean reversion of the stationary component. From the size portfolios it can be seen that the finding is more pronounced for small firms.

The negative autocorrelations are significant for the indices, despite Summers' point, and the U-shape is confirmed. When the experiment is repeated on segmented data with breakpoint in 1941, the findings are not as pronounced and the negative autocorrelations are not significant for the second subperiod 1941–1985.

When the returns are regressed on the returns of the first decile portfolio of the size portfolios and the residuals analyzed, the negative autocorrelations vanish. This means that the effect is common to all portfolios and not an effect that is visible only in the cross section. This is supported by the fact that it is also found in the indices. Fama and French speculate that a common macroeconomic factor might generate the mean reversion.

Poterba and Summers (1988) use a different approach to obtain evidence from the model. Before going into market data, they compare the statistical power of two different statistics derived from the Summers (1986) model and another method, a parametric ARMA estimation as suggested by Campbell and Mankiw (1987) for output data.

The first statistic derived from the Summers (1986) model is a variance ratio test. Let r_t denote monthly returns. Then, the variance ratio is defined by

$$VR(k) = \frac{\text{var}(r_{t,t+k})}{\text{var}(r_{t,t+12})} \cdot \frac{12}{k}, \quad (2.4.26)$$

that is, the variance of the k -period returns is compared to the variance of annual returns. If the stock price follows a random walk, the returns are white noise and its variance grows with k . Then, $\text{var}(r_{t,t+12}) = \sigma_r^2 \cdot 12$ and $\text{var}(r_{t,t+k}) = \sigma_r^2 \cdot k$ and the statistic $VR(k)$ is one for all k . If the returns are mean reverting, the variance of returns with a long horizon will grow slower than for a white noise process. Therefore, the variance ratio will fall below one.

The second statistic is the regression slope (2.4.24) as used in Fama and French (1988). Poterba and Summers note that this statistic applies a negative weight

to autocorrelations up to order $2/3 k$, then increasing positive weight up to lag k , and then decaying positive weights. If the autocorrelations in the data are first positive and then negative, the regression test will reject the null hypothesis of serial independence more often than the variance ratio.

The third statistic is the likelihood ratio test of serial independence against a particular ARMA structure. This approach has the advantage that if the ARMA model is the data generating process, the likelihood ratio is the most powerful test according to the Neyman-Pearson lemma. It is exceedingly unlikely however that the stock prices are generated by any simple linear model like ARMA. Nevertheless, the excess returns (2.4.20) according to the fads model (2.4.17) through (2.4.19) are a linear combination of an ARMA(1,1) and a white noise (ARMA(0,0)) process and therefore have an ARMA(1,1) representation. So in conducting Monte Carlo studies with the fads model, likelihood ratios from the estimation of an ARMA(1,1) model for the excess returns give a lower bound on the estimation errors that can possibly be achieved.

The excess returns (2.4.20) are the starting point of the analysis of Poterba and Summers. As the solution for the stationary component z is given by

$$z_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j} = \frac{1}{1 - \alpha L} \varepsilon_t,$$

where L is the lag operator, the excess returns have the representation

$$r_t(i) = \eta_t + \frac{1 - L}{1 - \alpha L} \varepsilon_{t+1}. \quad (2.4.27)$$

Therefore, the variance of the excess returns is given by

$$\begin{aligned} \sigma_{r(i)}^2 &= \mathbb{E} \left[\eta_t + \frac{1 - L}{1 - \alpha L} \varepsilon_{t+1} \right]^2 \\ &= \sigma_{\eta}^2 + \mathbb{E} \left[(1 - L) \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j} \right]^2 \\ &= \sigma_{\eta}^2 + \frac{2\sigma_{\varepsilon}^2}{1 + \alpha}, \end{aligned} \quad (2.4.28)$$

where the last line is obtained after some algebraic manipulations. This representation of the variance of the excess returns decomposes it into the contributions of the random walk component and of the mean reverting component.

Poterba and Summers set up Monte Carlo simulations to compare the power of the three statistics to distinguish between serial independence and the fads model. They make three settings at the onset: σ_{η}^2 equal to one, $\alpha = 0.98$ and the

confidence level equal to 5 percent. Then, the ratio of the variance of the mean reverting component to the total variance of returns is given by

$$\delta := \frac{\frac{2\sigma_\varepsilon^2}{1+\alpha}}{\sigma_{r(i)}^2} = \frac{2\sigma_\varepsilon^2}{1 + \alpha + \sigma_\varepsilon^2}.$$

Poterba and Summers consider two scenarios: (1) $\delta = 0.25$, that is, the mean reverting component contributes a quarter to the total return variance, and (2) $\delta = 0.75$, the contribution of the mean reversion is three quarters. For both scenarios they generate 25,000 series of 720 monthly returns.

To evaluate the type II error, the failure to reject the null of serial independence when it is wrong, they use the empirical distribution of the test statistics when the data is generated with $\delta = 0$. In that case there is no mean reverting component. From these simulations they find the critical region for a one-sided 0.05 test of the random walk hypothesis against the fads alternative.

The type II error in the $\delta = 0.25$ -scenario using the parametric ARMA estimation is obtained as 0.924. In the $\delta = 0.75$ -scenario it is 0.76. As the excess returns have an ARMA(1,1) representation this is the best possible type II error according to the Neyman-Pearson lemma. It provides a lower bound for the type II errors for the two other statistics, the variance ratios and the regression slopes. The variance ratios are calculated for return horizons between 24 and 96 months and the type II error varies between 0.925 and 0.933 in the $\delta = 0.25$ -scenario, and between 0.813 and 0.863 in the $\delta = 0.75$ -scenario. The regression slopes are calculated for return horizons between 12 and 96 months and the type II error varies between 0.929 and 0.943 in the $\delta = 0.25$ -scenario, and between 0.841 and 0.914 in the $\delta = 0.75$ -scenario. Poterba and Summers also make a Monte Carlo study of the first order autocorrelation and find it to be least powerful, in line with Summers' (1986) finding: The type II error in the $\delta = 0.25$ -scenario is 0.941, and 0.924 in the $\delta = 0.75$ -scenario.

The upshot is that the type II error is huge, which supports Summers' (1986) point that a mean reverting component in the stock price leaves virtually no statistically discernable trace. It is questionable if the null hypothesis of a random walk is so valuable that the type I error should be kept at the relatively low level of 5 percent. Poterba and Summers show that a more equal weighting of the errors would place both at the order of 30 percent.

The authors exert variance ratio tests on four different data sets:

1. The equally weighted and the value-weighted index of the CRSP data. (Monthly returns on stocks listed at the NYSE since 1926.)
2. Annual returns on the S&P index since 1871 as reconstructed by the Cowles commission.

3. Post 1945 monthly returns for 17 stock markets outside the U.S. from the *International Financial Statistics* of the International Monetary Fund.
4. CRSP database: Monthly returns on individual firms in the U.S. since 1926.

The results are largely in support of Fama and French (1988).

1. The CRSP indices show negative serial correlation at long horizons (variance ratios below one). This result is significant at the 0.005 level for the equal-weighted index (33 percent of the variance of 8-year returns is due to mean reversion) and at the 0.08 level for the value-weighted index (58 percent of the variance of 8-year returns is due to mean reversion), despite the low power. For horizons below one year, the variance ratios are also less than one. As the standardizing unit is the variance of annual returns, this means that the returns are *positively* correlated for horizons below one year.
2. The annual returns on the S&P since 1871 show negative autocorrelation at long horizons. For 8-year returns about 17 percent of the variance is due to mean reversion. For the period before 1925 about 55 percent of the variance of 8-year returns is due to mean reversion.
3. The monthly returns data on the stock markets outside the U.S. show the same pattern: negative autocorrelation at long horizons and positive autocorrelations at horizons below one year. There are, however, a few exceptions: Finland, South Africa, and Spain have no negative autocorrelation at long horizons, Colombia has no positive autocorrelation in the short run. Unusual monetary events like exceeding inflation or political uncertainty may be one reason that keeps stock market prices more strongly fluctuating at longer horizons. For Spain, the hyperinflation followed by a strong deflation most probably caused the outlier pattern.
4. For 82 individual firms from the CRSP database, the negative autocorrelation pattern at long horizons is supported but not the positive autocorrelation at short lags. As the returns on individual stocks are not independent, excess returns over the value-weighted index are considered. Only about 12 percent of the 8-year excess returns are due to mean reversion. Unfortunately, Poterba and Summers give no explanation why the pattern is clearly weaker for individual securities or why there is no positive autocorrelation.

Poterba and Summers add some comments on the source of the mean reverting component. The possible reason they present are those explained in Section 2.3. They show within the fads model that if changes in required returns are the cause of the mean reverting component, these required returns must often exceed 20 percent in order to justify the observed magnitude of negative autocorrelation.

This is not very likely as risk factors like interest rates and market volatility can hardly account for such variation. Poterba and Summers conclude that the mean reverting component is probably due to fads and noise traders.

The positive autocorrelation at small lags may arise from positively autocorrelated expected dividends. An increase in expected dividends leads to higher stock prices and hence positive returns that may hold for a while until it leads to an increase in required returns that may push stock prices down again. There is however little theoretical justification or empirical evidence for these ruminations. It may also be possible that noise trader demand follows an autoregressive pattern, a kind of herd behavior that dominates on small horizons.

Cutler, Poterba, and Summers (1991) extended the analysis of Poterba and Summers (1988) to bonds, foreign exchange rates, real estate prices, collectibles, and precious metals. They found similar patterns across all markets, supporting the earlier findings.

Kim, Nelson, and Startz (1998) and Kim and Nelson (1998) however argued that both methods, autoregressions of k -period returns and variance ratios are severely biased in the presence of heteroskedasticity. They suggest a Gibbs randomization procedure to correct for the heteroskedasticity and repeat Fama's and French's as well as Poterba's and Summers' experiments with their procedure and find the evidence of mean reversion weakened.

5 Conclusion

Mean reversion in stock prices and returns is an elusive subject. The main point of the seminal papers on the topic is that many plausible mean reversion models cannot be told apart statistically from standard models.

There is a substantial body of theoretical literature on as to how mean reversion in prices and returns might come about, accommodating both, rational and irrational motives of investors. Conclusive evidence in the one or other direction cannot be found due to the same reasons that prohibit proper estimation of mean reversion models.

A shortcoming of the mean reversion in prices and returns literature is that it focuses very strongly on mean reversion models of class I as defined in Section 2.1. This is mainly because it does not connect very well to the much richer literature on mean reversion in volatility.

The data suggests that there is mean reversion in stock market prices and returns and many practitioners perceive it. In Fischer Black's words:

"I believe that there is normally considerable mean reversion in the market—but it's hard to estimate how much." (Black (1988) p 271.)

In this thesis, I will take up a model that is widely used in stochastic volatility analysis and apply it to the mean reversion in returns problem. This model is of class II according to Section 2.1 and allows much stronger inference statements as it models directly the observable stock prices and does not refer to an unobserved decomposition into permanent and transitory components. The application is motivated by the article of Fischer Black cited above. Shortly after the stock market crash of 1987, he explained the crash in terms of mean reversion expectations. I will use the model to test his assertions on daily stock market data and find his views strongly supported.

Chapter 3

A Mean Reversion Theory of Stock Market Crashes

Errors in the perception of mean reversion expectations can cause stock market crashes. This view was proposed by Fischer Black after the stock market crash of 1987. I discuss this concept and specify a stock-price model with mean reversion in returns. Using daily data of the Dow Jones Industrial Average and the S&P500 index I show that mean reversion in returns is a transient but recurring phenomenon. In the case of the crash of 1987 I show that during the period 1982–1986 mean reversion was higher than during the nine months prior to the crash. This indicates that mean reversion expectations were underestimated in 1987. This error was disclosed when in the week prior to the crash it became known that a surprisingly high volume of equities was under portfolio insurance and thus hedged against a faster reversion. Simulations of the model with parameter estimates obtained from the two periods show that a crash of 20 percent or more had a probability of about seven percent. Up to five years after the crash, mean reversion was higher than before. This supports Black's hypothesis. Contrary to that, the crash of 1929 cannot be explained by a mean reversion illusion.

1 Mean Reversion and Stock Market Crashes

In a stock market with mean reversion in returns, the participants will develop expectations about the speed of the reversion. When a market participant observes for instance a positive change in returns, her reaction to this change will depend on her expectation of the reversion speed. If she has a long position and expects the high returns to disappear quickly, she will probably sell in order to realize the high returns. If she has a short position, she will probably keep this position and cover later when prices are lower. She might even sell more short to gain the difference when prices come down again. If she on the other hand expects the reversion speed to be very slow, then in the case of a long position

she will probably hold the paper to get high returns. She might even want to buy when she thinks that more positive moves are possible. In the case of a short position she will probably cover earlier as there is the risk that prices will stay high or even rise. That is, after a positive change the expectation of a fast reversion leads to higher selling pressure than the expectation of a slow reversion.

The mean reversion expectations of the market participants are not directly observable, they can only be deducted from their sales. High sales after a positive change in returns indicate a fast expected reversion.

Black (1988) proposed that misperceptions in the development of mean reversion expectations can cause stock market crashes when the participants learn about their error. Black's work was based on a literature that emerged in the late 1980's and discussed the evidence of mean reversion in stock returns (DeBondt/Thaler (1985), Summers (1986), Fama/French (1988), Poterba/Summers (1988)).

In this thesis, I will propose a stock-price model with mean-reverting returns. Using daily data of the Dow Jones Industrial Average and the S&P500 I will show that there were recurring periods since 1901 where mean reversion was significant.

Examining the crash of 1987 in detail, I show that it was probably caused by a misperception of mean reversion expectations as for about five years after the crash, mean reversion was significantly higher than before the crash. During the period 1982–1986 which was identified as the bull market that led up to the crash by the report of the Brady-Commission, I measure a significantly higher mean reversion than during the year 1987. This supports the hypothesis that an illusion about the true mean reversion expectations in the market led to the high price level before the crash. The event that disclosed this illusion can be identified as the surprisingly high volumes of equities under portfolio insurance that became known in the week prior to the crash. Simulations of the proposed model using parameter estimates from the 1982–1986 and the January 1987–October 1987 periods as given by the Brady-Report result in a probability of more than seven percent for a crash of 20 percent or more. A correction of minus 10 percent or more had a probability of over 40 percent.

2 A Mean Reversion Theory of Stock Market Crashes

a) Mean Reversion Expectations

I consider the situation where at time t a positive change is observed (Fig. 3.1). An individual investor with conservative expectations might now think that returns will come down fast. If λ is some parameter in the return generating process that controls the reversion speed, her expectations can be represented by, say, the

parameter value λ_0 which stands for a fast reversion. As the individual investor is not alone on the market, her expectations are probably dependent on the behavior of other participants as well. Let us assume that between times t and $t + h$ she does not act in any way but observes the behavior of the other market participants to come up with an expectation which is some weighted average of her *a priori* expectation indicated by λ_0 and the observed market behavior. If the market's sales indicate a reversion expectation like the one represented by λ_2 , the investor recognizes that her *a priori* expectation was very conservative relative to the market and consequently adjusts it to λ_1 , for instance.

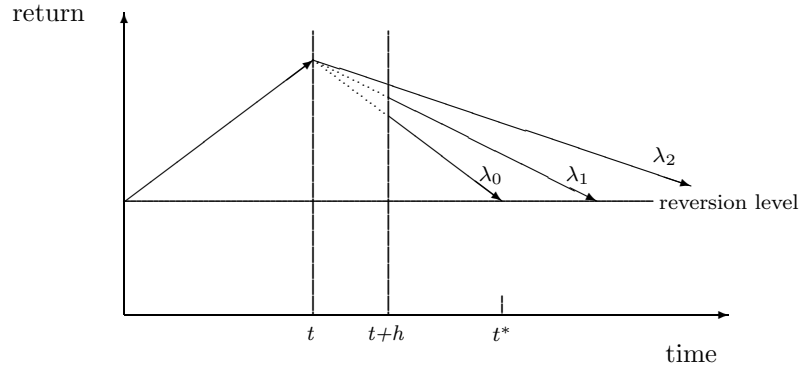


Figure 3.1: The development of mean reversion expectations.

The premise is that there are participants who act autonomously, i.e. who do not wait for others to act between times t and $t + h$. These might be institutional investors with predefined investment strategies which, explicitly or implicitly, induce certain reversion expectations. The mean of this implied expectations might be captured by the parameter value λ_2 . Another situation is conceivable. There may be investors who have the same expectation as represented by λ_0 but who are less risk averse than the individual investor considered. They may follow a strategy which hedges against the case that returns come down faster than according to λ_0 and at the same time act as if returns would follow λ_2 . This would allow them to participate in gains arising from a slow reversion behavior while at the same time the risk of a faster reversion than λ_0 would be hedged.

How to implement such a strategy? Each reversion speed $\lambda_{0,1,2}$ corresponds to a certain index (or stock) price at any time. For example, consider time t^* in Fig. 3.1 as the investment horizon. Let $S(t, \lambda_i)$ denote the index price at time t corresponding to reversion speed λ_i . Then in t^* I have the relation $S(t^*, \lambda_0) < S(t^*, \lambda_1) < S(t^*, \lambda_2)$ as at that time λ_2 implies a higher return than λ_1 than λ_0 , which in turn implies higher respective prices. One alternative would be to buy a put option at strike price $S(t^*, \lambda_0)$ with maturity t^* . The investor could control her positions as if she expects the price to behave according to λ_2 .

If the price drops below $S(t^*, \lambda_0)$ at time t^* , her exposure would be restricted to $S(t^*, \lambda_2) - S(t^*, \lambda_0)$.

The more risk averse individual investor could of course just as well hedge against the possibility that the price falls below her *a priori* level $S(t^*, \lambda_0)$. Her exposure would then be $S(t^*, \lambda_1) - S(t^*, \lambda_0)$ which is less than $S(t^*, \lambda_2) - S(t^*, \lambda_0)$ from which we see that her position is more risk averse.

In the case where a negative change was observed an investor who expects a fast improvement of returns would hold her long positions to avoid realizing temporary losses or buy more to exploit a cost-average effect. She would tend to close short positions to make use of the temporarily low prices. On the other hand, an investor who expects returns to come up slowly or to stay low for a while might want to sell her long positions in order to avoid possibly heavier losses in the future. She would keep or even enlarge short positions to participate from possible further downturns. In summary, after a negative change the expectation of a fast reversion implies higher buying pressure than the expectation of a low reversion.

After a negative change, an investor who wants to hedge against the possibility of a faster reversion while participating from stable low prices could assume short positions as if she expects returns to behave according to λ_2 (which now means that returns improve slowly). At the same time she could enter into a call option with strike price $S(t^*, \lambda_0)$ and maturity t^* . The price relation at time t^* would be $S(t^*, \lambda_0) > S(t^*, \lambda_1) > S(t^*, \lambda_2)$. The risk would be that the price rises quickly, so that the investor would have to cover at higher prices than she got when entering into the short position. If the price at time t^* rises above $S(t^*, \lambda_0)$, her exposure would be restricted to $S(t^*, \lambda_0) - S(t^*, \lambda_2)$.

Again, an individual investor who assumes *a priori* a reversion speed of λ_0 could wait till time $t + h$ to compare the market behavior. If she sees a reversion speed of λ_2 she would - just as in the case of a positive change - chose a weighted mean, for instance λ_1 . This would imply lower long and higher short positions than according to λ_0 . If she keeps her suspicion and hedges against prices higher than her *a priori* level $S(t^*, \lambda_0)$, her exposure would be $S(t^*, \lambda_0) - S(t^*, \lambda_1)$ which is lower than $S(t^*, \lambda_0) - S(t^*, \lambda_2)$, the exposure of the investor considered before. Thus, after negative changes as well as after positive changes it is risk-averse to assume a high reversion speed.

b) Mean Reversion Illusions and Disillusions

Assume that those investors who are less risk-averse and enter into an option contract while speculating on low reversion speeds give a public record of what they are doing. Then, when the individual investor considered develops her expectations between times t and $t + h$, she will not only look at the market to see

how the others play more risky. She will also look at these public records and will recognize that the mean reversion expectations of those investors who are already active on the market are not that different from her - *a priori* - own but that they have entered into appropriate hedges. Her perception of the market's expected speed of reversion would be higher and thus her own expectation, the weighted average of her *a priori* expectation and her market perception, as well.

The put-call ratio is a proxy for these imaginary public records. If after a positive change in returns investors look at a stable high market to develop mean reversion expectations, they can conclude from a high put-call ratio that the market's expected mean reversion speed is higher than indicated by stock sales alone. Conversely, if the market stays low after a negative change, a low put-call ratio indicates the same.

It gets problematic when the risk-tolerant investors choose to synthesize the options contracts. Then they hold hedge portfolios consisting of stocks (or futures) and bonds. It cannot be seen from the buy and sell orders that these transactions are designed to mirror an option and hence there is no record at all. In this case the individual risk-averse investor has no opportunity to infer her expectations from other sources than the stock sales itself. If the market stays high after a positive change or low after a negative change, she will systematically underestimate the market's expected mean reversion speed. In this case the information that the risk-tolerant investors are not confident of a low reversion speed but hedged against a high one is completely hidden.

How does the crash come about? Assume that the underestimation of the reversion speed is a mass phenomenon and not confined to a single investor because the expectations of the risk-tolerant investors are not or only rudimentary observable. For illustration, consider the extreme case where except for the small group of risk-tolerant investors, all others are more risk-averse. They wait between t and $t + h$ to observe the market without being able to infer the true expectations of the acting investors. The net effect of the market transactions of the risk-tolerant investors accounting for both, their purchases and their short sales from the portfolio that replicates the put option, will be positive after a positive change and negative after a negative change. This assertion is shown in the Appendix.

To stay within the picture of Figure 3.1 I assume that the λ_2 -position is now the net result of the risk-tolerant group's consolidated purchases and short sales. As the hedging cannot be perceived by the risk-averse investors, who *a priori* assume the speed λ_0 , they adjust their expectations to λ_1 . (Note that every single investor adjusts her expectations without knowing about the others. The move from the mean λ_0 to the mean λ_1 is the result of all these adjustments.) We are now able to formulate the mean reversion speed that will be effective on the market after $t + h$: Denote the proportion of the sales by the risk-averse

majority by $\alpha \in [0, 1]$ and the proportion of the sales by the autonomous, risk-tolerant group by $\beta \in [0, 1]$. Then the effective mean reversion speed will be $\lambda = \alpha\lambda_1 + \beta\lambda_2$. This holds for the illustrative, extreme case that the market consists of these two groups only, i.e. $\alpha + \beta = 1$. The theory set out here is valid as long as $0 < \lambda < \lambda_0$, i.e. the effective reversion speed is lower than the *a priori* reversion speed.

As mentioned in the previous section, it is not necessary for the argument that the risk-averse majority is not willing to enter into options contracts hedging the λ_0 -level. It is sufficient that the mean λ_0 of *a priori* expectations is larger than λ_2 , the expected reversion speed as implied by the sales of the autonomous group and that this λ_2 is mistaken to be the true expected reversion speed of the market.

This situation I will call *mean reversion illusion*. The mean reversion speed λ prevalent on the market is slower than it would have been if the group of risk-averse investors had seen the hedge activity of the autonomous group correctly. Of course this misconception is disclosed if the true expectations of the autonomous group and their hedge positions become known. Then every single investor readjusts her expectations. Yet another disillusion is conceivable: When the majority becomes aware of its majority, that is when it becomes known that a large number of investors had expected a faster reversion but adjusted it to a slower one after observing the activities of a small group. These two disillusiones are independent: the information about the true expectations of the autonomous group does not imply that many others followed them. If λ_0 were observable and the prevalent speed λ slower, then this would imply that most participants must have followed some group with seemingly slower expectations, but mostly λ_0 will be unobservable. Conversely, the information that a majority with high mean reversion expectations followed a minority with seemingly slow ones does not say anything about the true expectations of the minority.

In the light of this considerations a stock market crash will be defined as the *mean reversion disillusion*. If one of the two possible disillusiones happens at time t_c , it will become clear that the market assumed a false reversion speed since time t . That is, the price process followed a ‘wrong’ trajectory between t and t_c . ‘Wrong’ means that it did not properly reflect the average *a-priori* mean reversion expectations of the market. This wrong trajectory now has to be eliminated and the process has to be set into a position as if the illusion had not happened. The crash is thus not just a readjustment in one parameter. Instead, it is this readjustment *plus* a discontinuous correction for the difference in the trajectories induced by λ and λ_0 between t and t_c . The precipitousness of the crash depends therefore on the “depth” of the illusion ($\lambda - \lambda_0$) and its duration ($t_c - t$). Figure 3.2 illustrates the point.

The argument is symmetric: It might as well be that during the mean reversion illusion the price process follows a path below the one given by the higher

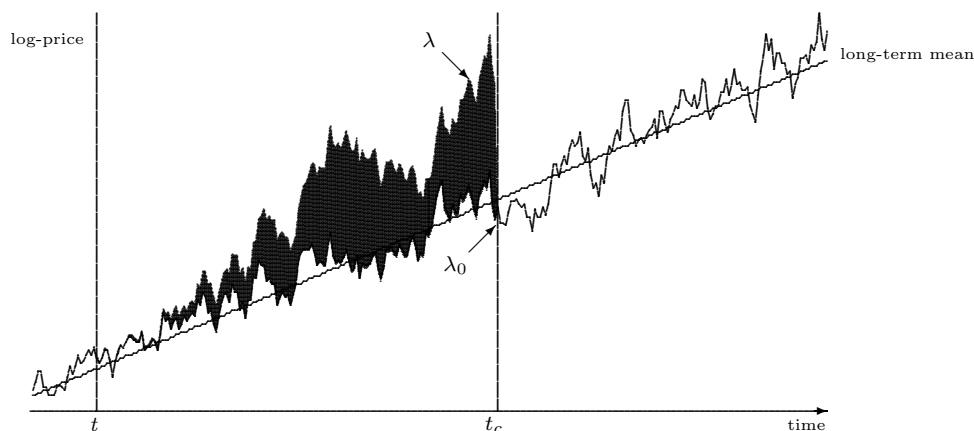


Figure 3.2: The mean reversion illusion between times t and t_c and the resulting difference in mean reversion velocities λ and λ_0 drive the log-price process above the λ_0 -level. The difference in the trajectories is shaded black and gives the potential crash at every point in time.

a-priori mean reversion speed. When the disillusion happens it will cause an upward jump. The magnitude of upward jumps is more restricted than that of downward jumps for a simple economical reason. Most investors have no large pile of money that can be unloaded onto the market in such an instance. They have to shift investments and restructure portfolios which leads to a delay between the decision to buy and the actual purchase. Contrary to that, in the case of a downward jump investors who have decided to sell will be willing to accept cash in any volume.

c) Mean Reversion Disillusion and October 19, 1987

Did errors in the perception of mean reversion expectations play any role in the stock market crash of 1987? This would mean that there was an illusion and later a disillusion about the market's average *a-priori* mean reversion expectation. In the notation of Figure 3.2, I am looking for the points t and t_c and the related events. I do not expect any particular event to cause the illusion at time t , so it will be difficult to identify t . The point t_c is the point immediately before the crash. The disillusion must be an event or a piece of information that is relevant for mean reversion expectations and that surprises the public. Following the argument set out in Section 3.2.b) that a mean reversion illusion is particularly likely to happen when hedges can be implemented that cannot be recognized by other market-participants, I look for the disclosure of a large hedge position. According to the hypothesis this would imply that a group of active market participants had been more risk-averse than the average investor perceived and that this group was hedged against a high mean reversion speed.

The three days prior to October 19, 1987, are of prime interest in this respect.

From Wednesday, October 14 to Friday, October 16, the U.S. stock market lost more than ten percent. The Dow Jones Industrial Average fell from 2.508 at closing on Tuesday to 2.246 at closing on Friday, the S&P500 from 314 to 282 over the same time. The loss on Wednesday was three percent, on Thursday two percent, and on Friday five percent.

These drops can be attributed to fundamental reasons, namely to the simultaneous budget and trade balance deficit and to the House Ways and Means Committee's plans to eliminate tax benefits for takeovers. On Wednesday, October 14, the U.S. government announced that the trade deficit was about ten percent higher than expected. The dollar fell sharply in reaction, this led to an expected decrease in foreign investment. Also on Wednesday it became known that the Committee actually filed legislation concerning the takeovers (Brady et al. (1988), p. III-2f). Mitchell/Netter (1989) observed that the losses were largely confined to the U.S. market, an indication of the fundamental cause.

Portfolio insurance companies reacted by increasing their cash positions through sales of index futures. They sold 530 million dollars on Wednesday, 965 million dollars on Thursday, and 2.1 billion dollars on Friday, the latter being eleven percent of the total daily sales on the futures market (Brady et al. (1988), p. III-16). At the same time, it became known that these sales were by far not sufficient to adjust the portfolio insurance positions adequately. The report of the Brady-commission that was set in after the crash to determine its causes mentions another eight billion dollars that were expected to be sold on the futures market. It is not clear from the report where these information came from. The implied volume of equities under portfolio insurance, 60 to 90 billion dollars, however seems to have surprised the market. This may have been the event that disclosed the average risk-aversion and the *a-priori* mean reversion expectations of the market participants (Brady et al. (1988), p. 29).

The Brady-Report and many other authors attributed the cause of the crash partly to the mere existence of portfolio insurance and associated program trading that cascaded in the crash. While this was probably important for the amplification of the downturn, the view proposed here is quite different. The unexpectedly high portfolio insurance volumes were *fundamental* information, not just a technical issue. They revealed that during the boom of 1987 a mean reversion illusion occurred. This view is closely related to the model of Jacklin/Kleidon/Pfleiderer (1992): They interpret the high volumes under dynamic hedging as a surprise to the market as well. In this fact they see the fundamental information that a large part of the stock purchases during the boom was not caused by fundamental information but noise. They construct a market model according to this hypothesis and show in simulations that underestimation of portfolio insurance results in a higher market level and that prices fall when the amount of portfolio insurance is revealed. Here, I will specify a stock return model with mean reversion and show that the movements in (actual) mean reversion can indeed be found in the

market data.

3 A Mean Reversion Model for Stock Returns

An intuitive way to think about mean reversion in stock prices is to assume that the return process reacts to any deviation from its long-term mean. If the return is above the mean in one period, there is a force that pushes it downwards in following periods, if the return is below the mean, it is pushed upwards.

The mean return induces a certain appropriate stock price, denoted by $\tilde{\vartheta}_t$, which can be interpreted as an estimator of the fundamental value of the underlying stock or stock index. I set

$$\tilde{\vartheta}_t = S_0 e^{\mu t}.$$

Consider the return process given by

$$\frac{dS_t}{S_t} = \mu dt + \lambda \frac{\tilde{\vartheta}_t - S_t}{S_t} dt + \sigma dW_t. \quad (3.3.1)$$

Here, the magnitude $(\tilde{\vartheta} - S)/S$ measures the deviation of the return process from the long-term mean μ . The parameter $\lambda \geq 0$ controls the speed with which the return is pushed back to the mean μ . The average mean reversion time is $1/\lambda$ units of time. W_t is standard Brownian Motion. It is shown in the Appendix that the expected value of the process satisfying (3.3.1) is

$$\mathbb{E}S_t = S_0 e^{\mu t} = \tilde{\vartheta}_t.$$

This is intuitively expected from a mean-reverting process. A similar model was proposed by Metcalf/Hasset (1991).

The process satisfying

$$d \log S_t = \tilde{\mu} dt + \lambda (\log \vartheta_t - \log S_t) dt + \sigma dW_t, \quad (3.3.2)$$

where $\tilde{\mu} = \mu - \sigma^2/2$ and $\vartheta_t = S_0 e^{\tilde{\mu} t}$ is a first-order approximation to (3.3.1). This is shown in the Appendix. The solution to model (3.3.2),

$$\log S_t = \log S_0 + \tilde{\mu} t + \sigma \int_0^t e^{-\lambda(t-u)} dW_u, \quad (3.3.3)$$

is an Ornstein-Uhlenbeck process. Hence, (3.3.2) is a Vasicek-type model for stock returns (Vasicek 1977). The unconditional distribution of the log-price process is given by

$$\log S_t \sim \mathcal{N} \left(\tilde{\mu} t + \log S_0, \frac{\sigma^2}{2\lambda} \right), \quad (3.3.4)$$

the process is non-stationary. The higher the speed of the mean reversion λ the smaller is the variance as the process will not leave a certain corridor around its mean. Interesting for purposes of time-series analysis is the conditional distribution of the log-returns $\log S_{t+1} - \log S_t$, given the knowledge of the time series through date t . It can be read directly from the model (3.3.2):

$$(\log S_{t+1} - \log S_t) \sim \mathcal{N}(\tilde{\mu} + \lambda(\log \vartheta_t - \log S_t), \sigma^2). \quad (3.3.5)$$

To estimate the model, I maximize the log-likelihood

$$L(\theta, \{S_t\}_t) = -\frac{T}{2} \log \sigma^2 - \frac{1}{2} \sum_{t=1}^T (r_t - \mu - \lambda(\log \vartheta_t - \log S_t))^2. \quad (3.3.6)$$

T denotes the number of observations, $\theta = (\mu, \lambda, \sigma)'$ is the parameter vector, $r_t = \log S_{t+1} - \log S_t$ denotes the logarithmic returns, $\vartheta_t = S_0 e^{\mu t}$ as above. I use the 'dfpmin' routine from Press et al. (2002) as well as 'fminunc' routine from the MATLAB optimization toolbox. Both implement a quasi-Newton method with line search using analytical gradients and numerical Hessians. The derivatives are readily calculated from (3.3.6).

The unconditional distribution of the log-returns is given by

$$(\log S_{t+1} - \log S_t) \sim \mathcal{N}\left(\tilde{\mu}, \frac{\sigma^2}{2\lambda}(e^\lambda - 1) + \frac{\sigma^2}{2\lambda}e^{-2\lambda t}(1 - e^{-\lambda})\right),$$

so that for $t \rightarrow \infty$ I obtain the stationary distribution

$$(\log S_{t+1} - \log S_t) \xrightarrow{t \rightarrow \infty} \mathcal{N}\left(\tilde{\mu}, \frac{\sigma^2}{2\lambda}(e^\lambda - 1)\right),$$

and thus the maximum likelihood estimates of $\theta = (\mu, \lambda, \sigma)'$ will be asymptotically normal and the usual statistical inference of the maximum likelihood estimation applies.

The model is of reduced form inasmuch as the parameter λ models mean reversion without having an interpretation as capturing an actual economic process. The mean reversion term is the only difference of the model to the standard Black-Scholes model that is already of reduced form. The parameters μ and σ model the mean return and the volatility. From equation (3.3.4) we can see, however, that the volatility as modelled by the diffusion (with parameter σ) now interferes with the mean reversion for the reason mentioned above.

To explain the measured mean reversion, a different model is needed, for example one of those described in Section 2.3. It is beyond the scope of this thesis to suggest a new model in this direction, but I will give a brief outline of a possible model in the spirit of this thesis in Chapter 8.

This model can be criticized for many reasons. It is not a model of efficient markets, the stock price process (3.3.3) is not a martingale. As I am interested in an explanation of stock market crashes, I accept a model that allows for non-efficiency locally.

It contains two magnitudes that are highly non-trivial to estimate, the mean return $\tilde{\mu}$ (see for example Merton 1980) and the mean reversion speed λ . Hence the considered samples must be chosen carefully to make sure that a mean return is estimated that is relevant to the analysis.

The analogon to the Ornstein-Uhlenbeck process in discrete time is the autoregressive process of order one. Our mean reversion model implies by $\lambda > 0$ that the autoregressive coefficient is negative. Campbell/Lo/MacKinlay (1997) 66f, and Lo and MacKinlay (1988) show that this coefficient is in fact often found to be positive. I will present evidence that the autoregressive parameter when estimated according to the model proposed here is negative. That is, the estimated mean reversion parameter $\hat{\lambda}$ is positive. Mean reversion in returns is rarely significant, though.

Also, one might argue that when $\vartheta_t = S_0 e^{\mu t}$ is an estimator of the fundamental value, why should the market trade an asset far above or below this value? In other words, why should a non-negligible distance $\log \vartheta_t - \log S_t$ occur at all in a market with mean reversion. White (1990) observed for the case of the 1929 stock market crash that during the boom that preceded the crash, fundamentals were very difficult to evaluate. This was mainly because many companies entered the stock market that had virtually no dividend history. A similar case can be made for the internet boom at the turn of the century. The quality of an estimator for the fundamental value that uses any type of historical long-term mean is questionable in situations like that. It is of course conceivable to extend the model to capture a higher mean reversion speed when the distance of the price process to its long-term mean is large. I will use the model (3.3.2) for the sake of simplicity.

4 Mean Reversion and the Stock Market Crash of 1987 in Market Data

The data are daily closings of the Dow Jones Industrial Average ranging from January 2, 1901, to October 2, 2002, covering 27,293 observations. The series was kindly provided by Dow Jones & Company. Also, I use daily closings of the S&P500 ranging from January 4, 1982 to December 30, 1991, covering 2563 observations. The series was obtained from Datastream. All holidays that repeat the price of the previous day were deleted.

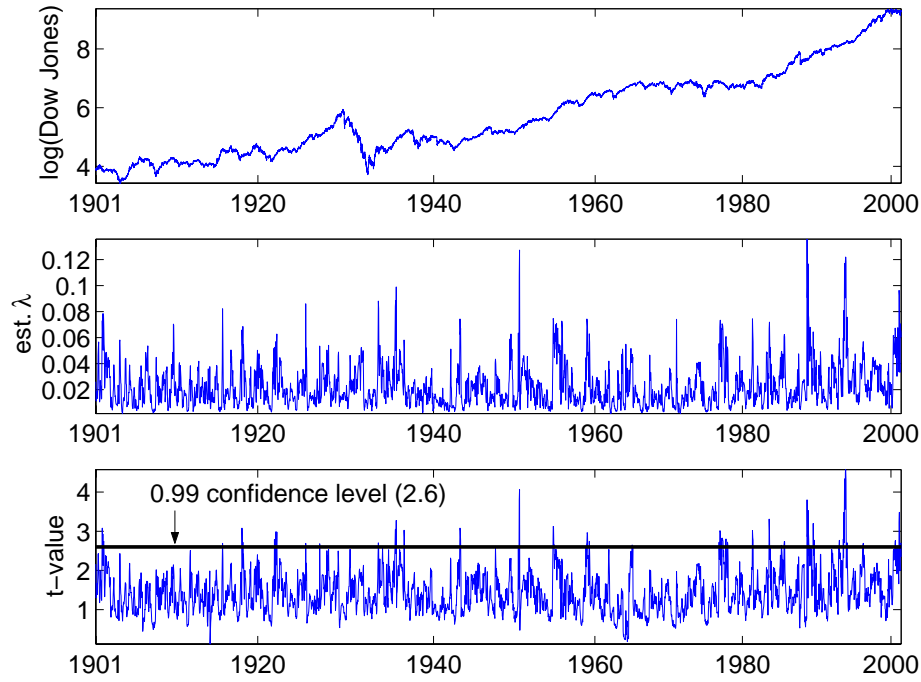


Figure 3.3: Log-price series of the Dow Jones Industrial Average (top plot), estimations of the mean reversion speed λ according to model (3.3.2) on a rolling 250-points window (middle plot), and t -statistics for the estimated mean reversion speed (bottom plot). The fact that all estimated mean reversion speeds are positive implies that there is no mean-aversion. Mean reversion is mostly insignificant but periods occur over the complete sample where it is highly significant.

Figure 3.3 shows the logarithmic price series of the Dow Jones over the 101 years that the series covers (upper plot). I estimated model (3.3.2) on a rolling window of 250 points length that is moved forward by 10 points every step. The middle plot shows the estimates of λ , the bottom plot the t -statistics of the estimations. The standard errors for this statistic were calculated according to the quasi-maximum likelihood method of White (1982).

The first observation is that all estimates of λ are positive. In the light of the findings reported by Lo and MacKinlay (1988) and Campbell/Lo/MacKinlay (1997) 66f, this is a surprising result. It implies that there is no mean-aversion in the daily log-returns of the Dow Jones.

Mean reversion is mostly insignificant but there are recurring periods over the whole century where mean reversion is highly significant. Among those are the 1920's and 1930's, the late 1950's, the late 1970's and early 1980's with a clear cluster around the crash, and this year. It must be emphasized that the method used here is based on a moving 250-points mean return. Other concepts of mean returns can lead to different results. Our results are qualitatively insensitive,

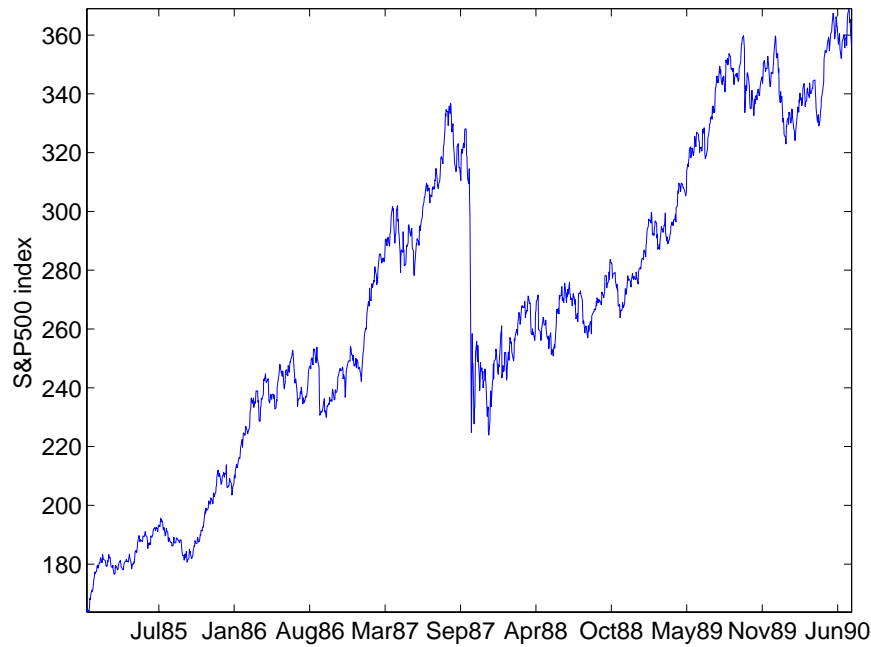


Figure 3.4: S&P500 between January 2, 1985 and July 23, 1990. The drop of about 20 percent on October 19, 1987 pushed the index from its closing point of 282.7 on Friday, October 16 to 224.8 at closing on Monday, October 19.

however, to varying window lengths.

From the considerations of Section 3.2.c) it is interesting whether there are movements in the mean reversion parameter occurring around the stock market crash of 1987 that could be attributed to the mean reversion illusion and disillusion. I will first look at the disillusion, that is, the crash itself. The hypothesis is that after the crash we should see a faster mean reversion, that is, a higher $\hat{\lambda}$, than before the crash.

a) The Mean Reversion Disillusion

The mean reversion disillusion is the crash itself. Figure 3.4 shows the S&P500 between January 2, 1985 and July 23, 1990. The crash on October 19, 1987 pushed the S&P500 down from 282.7 on October 16 to 224.8 at the end of October 19, a drop of 20.48 percent. It took about two years for the market to recover from the crash.

To estimate model (3.3.2), I deleted the observations October 16, 1987, to October 26, 1987, from the returns and the price series of the S&P500. By this, the crash itself did not affect the estimation of the mean reversion speed before and after. Then I estimated model (3.3.2) for the 100, 200, ..., 1000 observations

Table 3.1: Estimation of model (3.3.2) on sample periods before and after the 1987 stock market crash. The observations from October 16, 1987, through October 26, 1987, were deleted from the series. The numbers in parentheses are quasi-maximum-likelihood standard errors according to White (1982). The estimations of the mean returns and standard deviations are significant according to all common confidence levels with the single exception of the mean return of the 100 days before the crash. For the mean reversion parameter λ those estimates that are significant according to the two-sided 0.95 confidence level are marked with a single asterisk, the double asterisk denotes significance according to the two-sided 0.99 confidence level. Mean reversion speed clearly increased after the crash.

n	n days before Oct. 16, 1987			n days after Oct. 26, 1987		
	$\hat{\mu}$	$\hat{\lambda}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\lambda}$	$\hat{\sigma}$
100	0.000483 (0.000613)	0.010565 (0.011637)	0.009632 (0.000827)	0.001401 (0.000114)	0.16636** (0.061999)	0.016239 (0.001703)
200	0.001172 (0.000401)	0.005814 (0.005729)	0.010373 (0.000581)	0.000851 (7.6e-5)	0.077153* (0.03205)	0.013997 (0.001075)
300	0.001004 (0.000133)	0.024993 (0.015156)	0.010127 (0.000592)	0.000683 (4.9e-5)	0.052885* (0.021259)	0.012292 (0.000859)
400	0.000672 (0.000110)	0.016753 (0.008603)	0.009907 (0.000496)	0.000713 (3.1e-5)	0.052938** (0.019714)	0.011204 (0.000722)
500	0.001062 (0.000103)	0.013118 (0.008098)	0.009516 (0.000425)	0.000769 (2.9e-5)	0.044117** (0.014459)	0.010935 (0.000699)
600	0.000834 (6.4e-5)	0.015902* (0.006800)	0.009034 (0.000379)	0.000715 (2.8e-5)	0.033246** (0.011698)	0.01056 (0.000614)
700	0.000951 (4.7e-5)	0.017962* (0.007320)	0.008701 (0.000344)	0.000663 (3.2e-5)	0.023374* (0.009914)	0.010297 (0.000554)
800	0.000801 (4.3e-5)	0.015205* (0.0059002)	0.008523 (0.000314)	0.000541 (8.1e-5)	0.008240 (0.005264)	0.010536 (0.000496)
900	0.000728 (5.1e-5)	0.010165* (0.004539)	0.008546 (0.000289)	0.000563 (5.5e-5)	0.009867* (0.005016)	0.010475 (0.000451)
1000	0.000578 (9.6e-5)	0.003738 (0.002642)	0.008442 (0.000269)	0.000547 (4.5e-5)	0.009902* (0.004796)	0.010221 (0.000420)

before and after the crash. (More precise, before and after the gap.) Table 3.1 reports the estimations.

The findings clearly support the hypothesis. Up to 700 points before and after the crash, there is an increase in mean reversion speed. The estimations of the mean reversion speed λ in these samples are significant on the two-sided 0.95 confidence level, four out of seven on the two-sided 0.99 confidence level. As the sample size increase from sample to sample, different mean return concepts are applied here. Except for the 100 and 400 points samples I measure a slightly higher mean return before the crash than after.

As these findings are not independent, I estimated model (3.3.2) also on the corresponding opposite intervals of length 200, that is, for the observations *crash-1000* to *crash-800* and *crash+800* to *crash+1000*, then *crash-800* to *crash-600* and *crash+600* to *crash+800*, and so on. Table 3.2 reports the estimates. As the mean return concept applied here is a moving 200-days mean, the estimates of the first row are identical to those of the second row of Table 3.1. The other estimates are not comparable to that of Table 3.1. With the single exception of the sample

Table 3.2: Estimation of model (3.3.2) on sample periods before and after the 1987 stock market crash. The observations from October 16, 1987, through October 26, 1987, were deleted from the series. The numbers in parentheses are quasi-maximum-likelihood standard errors according to White (1982). The estimations of the mean returns and standard deviations are significant according to all common confidence levels. For the mean reversion parameter λ those estimates that are significant according to the two-sided 0.95 confidence level are marked with a single asterisk, the double asterisk denotes significance according to the two-sided 0.99 confidence level.

i	n_i ($n_0 = 1$)	day n_i through day n_{i-1} before Oct. 16, 1987			day n_{i-1} through day n_i after Oct. 26, 1987		
		$\hat{\mu}$	$\hat{\lambda}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\lambda}$	$\hat{\sigma}$
1	200	0.001172 (0.000401)	0.005814 (0.005729)	0.010373 (0.000581)	0.000851 (7.6e-5)	0.077153* (0.032050)	0.013997 (0.001075)
2	400	0.000672 (0.000110)	0.016753 (0.008603)	0.0099065 (0.000496)	0.000713 (3.1e-5)	0.052938** (0.019714)	0.011204 (0.000722)
3	600	0.000833 (6.4e-5)	0.015902* (0.006800)	0.009034 (0.000379)	0.000714 (2.8e-5)	0.033246** (0.011698)	0.010560 (0.000614)
4	800	0.000801 (4.3e-5)	0.015205* (0.005900)	0.008523 (0.000314)	0.000541 (8.1e-5)	0.008240 (0.005264)	0.010536 (0.000496)
5	1000	0.000578 (9.6e-5)	0.003738 (0.002642)	0.008442 (0.000268)	0.000546 (4.5e-5)	0.009902* (0.004796)	0.010221 (0.000420)

corresponding to $n_i = 800$, the estimates support the hypothesis, too.

As the estimates are sensitive to the mean return method, it is interesting to see how they behave when the samples are increased by a finer step-length than 100, as done in Table 3.1. Again, I delete the days around the crash from the S&P500 series as described above and estimate model (3.3.2) on the samples of day $crash - n_i$ through $crash - 50$. Then I increase the sample by one day until I estimate (3.3.2) on $crash - n_i$ through $crash + 50$. The result is an estimation series of length 100. I did these estimations for $n_i = 100, 200, \dots, 1000$. As a control, I estimated the standard model of geometric Brownian Motion with drift and using this as a null hypothesis, I calculated the likelihood-ratio test statistic.

Figure 3.5 shows the likelihood-ratio statistic for the estimation series corresponding to $n_i = 100$ (start date May 27, 1987), $n_i = 200$ (start date January 2, 1987), $n_i = 300$ (start date August 11, 1986), and $n_i = 700$ (start date January 8, 1985). For all runs except $n_i = 1000$, the likelihood ratio exceeded the 0.99 confidence level when the sample was increased over the time of the crash. This gives another piece of evidence that mean reversion significantly increased after the crash. There is no monotonous relationship between the time horizon of the mean return and the significance of the result: The two longest horizons in Figure 3.5 result in the highest peaks but the shortest, beginning in May 1987, scores higher than the $n_i = 200$ sample starting in January 1987.

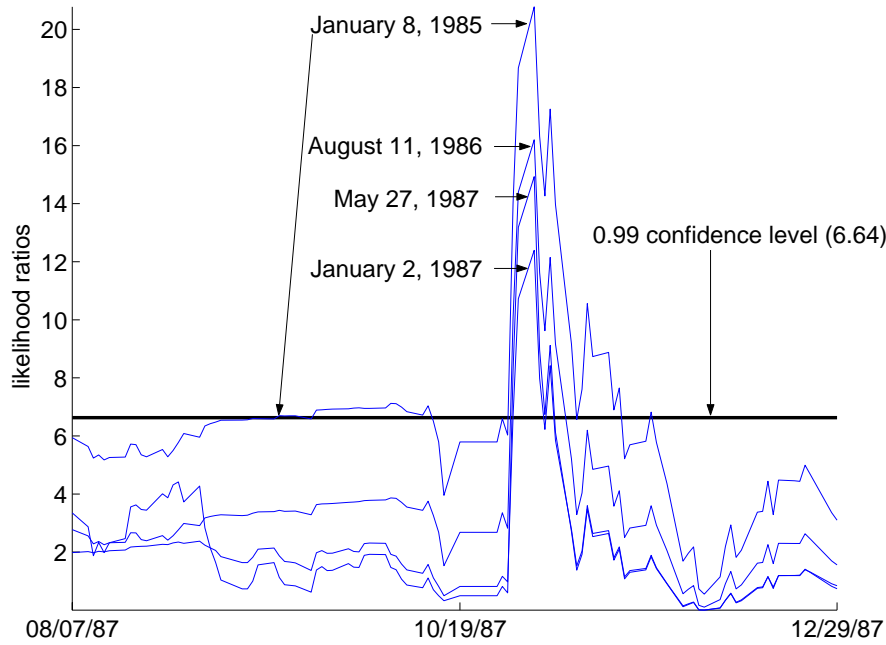


Figure 3.5: Likelihood-ratio statistics for the alternative hypothesis of (3.3.2) against the null hypothesis of geometric Brownian Motion with drift. Model (3.3.2) is estimated on the samples of day $crash - n_i$ through $crash - 50$. Then I increase the sample by one day until $crash - n_i$ through $crash + 50$. The result is a statistics series of length 100, corresponding to the days August 7, 1987, to December 29th, 1987. These estimations were calculated for $n_i = 100, 200, \dots, 1000$ and the figure shows $n_i = 100$ (start date May 27, 1987), $n_i = 200$ (start date January 2, 1987), $n_i = 300$ (start date August 11, 1986), and $n_i = 700$ (start date January 8, 1985). Except for $n_i = 1000$, all statistics series break through the line of the 0.99 significance level (Chi-square distribution with one degree of freedom: The alternative has one more parameter, λ , than the null). This shows the increase in the mean reversion speed after the crash.

b) The Mean Reversion Illusion

One of the defining characteristics of the situation of a mean reversion illusion is that mean reversion expectations can be implemented without being noticed by the other market participants, for example by synthesized options. Furthermore, the fundamental value of the assets in question is hard to evaluate in this situation. This means that finding the point of the start of the illusion is a much more subtle task than finding the disillusion.

In the notation of Figure 3.2 I look for the time t . That is, I search for a segment of a magnitude of years before the crash where mean reversion expectations were relatively high. As expectations cannot be measured, I use actual mean reversion as proxy. According to the hypothesis this segment should be followed by a segment with slower mean reversion that leads up to the crash.

The Brady-Report locates the beginning of the bull market that led up to the crash in 1982. The contributing factors are described as “*continuing deregulation of the financial markets; tax incentives for equity investing; stock retirements arising from mergers, leveraged buyouts and share repurchase programs; and an increasing tendency to include ‘takeover premiums’ in the valuation of a large number of stocks*”. The valuation of the U.S. stock market by the end of 1986 is described as high but not unprecedented in terms of price/earnings ratios. The appreciation from January 1987 through August 1987, however, “*challenged historical precedent and fundamental justification*” (Brady et al. (1988), p. 9, I-2).

Using this segmentation as a guideline, I estimate model (3.3.2) on the segments 01/02/82–12/30/86 and 01/02/87–10/15/87. That is, I set $t = \text{January } 2, 1987$. I assume that the mean return holds for the total period; the model (3.3.2) is estimated on the 1987-segment with the mean return set fix at the estimate from the period 1982–1986. Figure 3.6 illustrates the estimations. The estimate of the mean reversion speed on the 1982–1986 segment is significant at the one-sided 0.95 significance level. The estimates switch from a higher to a lower value, supporting the hypothesis.

I use a Generalized Likelihood Ratio (GLR) scheme as a changepoint detector (Lai 1995). Let $S = \{S_t\}_{t \in \{1, \dots, N\}}$ be the considered time series of index prices. The GLR scheme sets a changepoint at

$$\inf_{n \in \{1, \dots, N\}} \left\{ \max_{1 \leq k \leq n} \sup_{\theta \in \Theta} \left[\sum_{i=k}^n \log \frac{f_{\theta}(S_i | S_1, \dots, S_{i-1})}{f_{\theta_0}(S_i | S_1, \dots, S_{i-1})} \right] > c \right\}, \quad (3.4.7)$$

where N is the number of observations and Θ is the open parameter set. f_{θ} is the probability density given the parameter vector θ . θ_0 is the parameter vector of the null hypothesis and c is an *a priori* constant. There is no analytical expression or distribution result for c so that it must be found by simulation methods.

I decomposed the problem (3.4.7) into the following steps. On a baseline segment of the first m points of the series I estimated model (3.3.2). Thereby I obtain the null hypothesis $\hat{\theta}_0 = (\hat{\mu}_0, \hat{\lambda}_0, \hat{\sigma}_0)'$. Then I estimated (3.3.2) on every single subseries $\{S_1, \dots, S_j\}$, $j = m + 1, \dots, N$. This gave us a series of $\hat{\theta}_j$ maximizing the likelihood functions (3.3.6) of the subseries. From this series I computed the probability densities $f_{\hat{\theta}_j}(S_j | S_1, \dots, S_{j-1})$ for every $j = m + 1, \dots, N$ and stored

$$Z_j := \log \frac{f_{\hat{\theta}_j}(S_j | S_1, \dots, S_{j-1})}{f_{\hat{\theta}_0}(S_j | S_1, \dots, S_{j-1})}.$$

From the resulting series $\{Z_j\}_{j \in \{m+1, \dots, N\}}$, the statistics series

$$\xi_n = \max_{m+1 \leq k \leq n} \sum_{j=k}^n Z_j, \quad n = m + 1, \dots, N \quad (3.4.8)$$

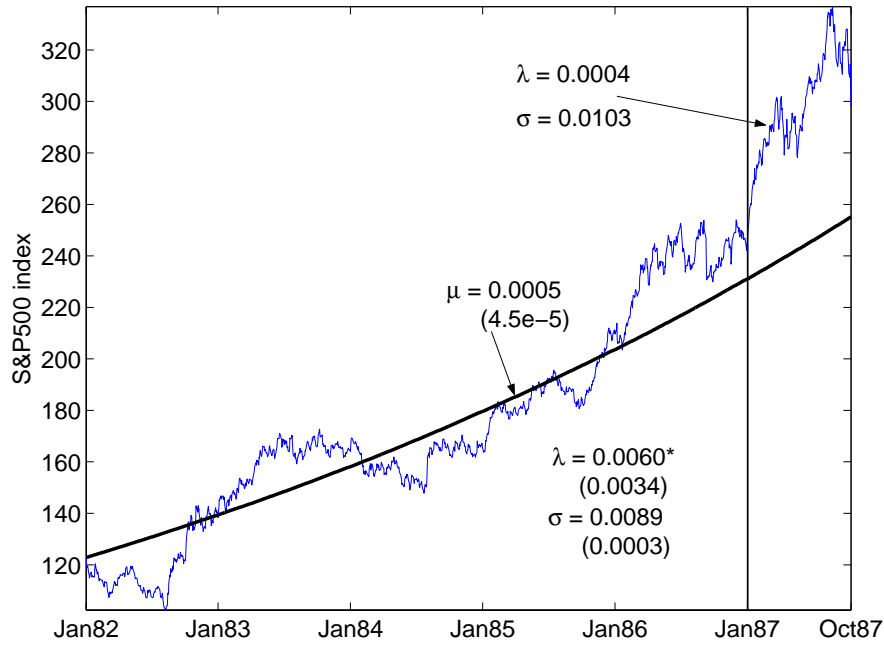


Figure 3.6: The bull market January 1982 to October 15, 1987 as seen in the S&P500. Using the segmentation of the Brady-Report, I estimate model (3.3.2) on the period January 1982 to December 1986 and January 1987 to October 15, 1987. I assume that the same mean return holds for the complete period, it is estimated at 0.0005 from the 1982–1986 segment. The figures in parentheses are standard errors according to White (1982). The estimate of the mean reversion parameter λ on the period 1982–1986 is significant at the one-sided 0.95 significance level. The estimate on the 1987 segment is much lower than the estimate before.

was calculated. As I search for a single changepoint only, it is interesting to plot the $\{\xi_n\}$ series. Figure 3.7 shows the series when the baseline distribution is estimated on the S&P500 observations January 2, 1982, through December 30, 1985. The series is then calculated for the observations January 2, 1986 through October 15, 1987. It can be seen that the estimated parameters move away from the estimated baseline parameters at two distinct speeds as the sample size increases. This is the interpretation of the two trends in the series that can be distinguished. The trend break is at the turn of the years 1986 to 1987. This supports the observation of the Brady-Report.

A simulation gives the significance levels: I generated 1,000 time series according to model (3.3.2) with the parameters obtained from the estimation of the sample period January 2, 1982 through December 30, 1985 ($\hat{\mu} = 0.0005$, $\hat{\lambda} = 0.006$, $\hat{\sigma} = 0.009$). This sample consists of 1,012 observations. The sample period January 2, 1986 through October 15, 1987 for which the detector series ξ_n in Figure 3.7 is depicted consists of 454 observations. Therefore, each of the 1000 simulated time series consisted of 1466 observations. On the first 1,012 observa-

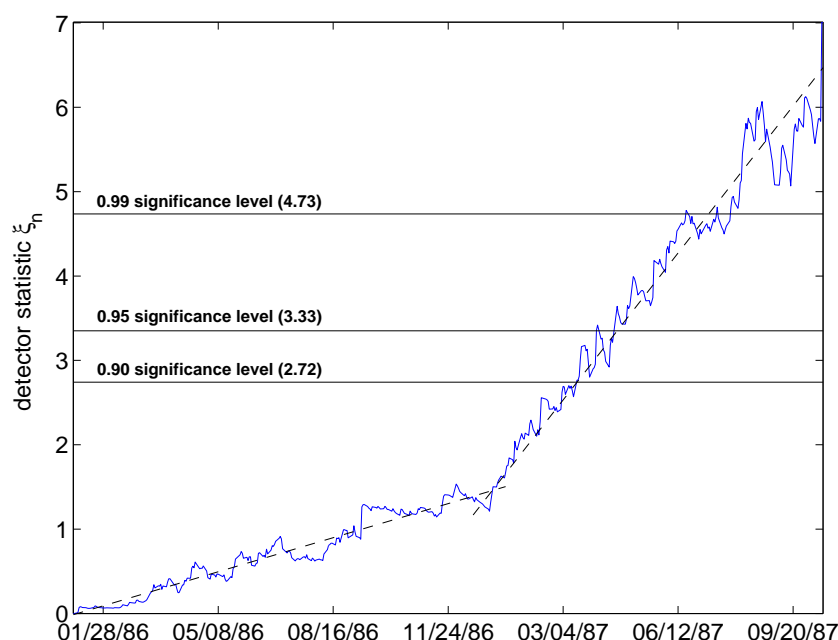


Figure 3.7: Change point detector statistic series $\{\xi_n\}$ as given by Equation (3.4.8). The baseline parameter vector θ_0 was estimated on the segment January 2, 1982 through December 30, 1985. The detector statistics series was then calculated for the observations January 2, 1986 through October 15, 1987. Two distinct trends can be observed in the statistic. This means that the estimated parameters move away from the estimated baseline parameters by two distinct speeds as the sample size increases. The trend break is almost exactly at the turn of the years 1986 to 1987, in line with the periods as given by the Brady-Report. The significance levels were obtained by simulation of the statistic.

tions of each series model (3.3.2) is estimated. Then for each series the detector statistic ξ_n is calculated for the remaining 454 observations, yielding 454,000 observations of the detector statistic. The significance levels reported in Figure 3.7 are the quantiles of these 454,000 observations.

With only this information in hand, what would have been the estimate on October 16, 1987, of the magnitude of a possible crash? More precise, with the information available on October 16, 1987, the question is: Given that the mean reversion illusion occurred at the beginning of the year 1987, about 200 days ago, and given that the mean reversion disillusion happens today, what will be the distance in the paths that must be corrected? In the notation of Figure 3.2 I now look for the distance in the trajectories that is shaded black, measured at the point immediately before the crash. Let me emphasize that I do not estimate the time of the crash, the disillusion is assumed to happen today for whatever reason.

I simulated model (3.3.2) with the estimated parameters as reported in Figure

3.6. I generated 10,000 paths of a random walk of length 200. Then I evaluated model (3.3.2) with the parameter vectors obtained from the 1982–1986 segment. The value 246.45 of the S&P500 on January 2, 1987, was set as the starting point. If a mean reversion illusion occurred in January 1987, it lasted for about 200 days up to October 16, 1987. That is, without the illusion the process would have continued for another 200 days under the old regime. The simulation thus gives an estimate of the distribution of the index value $S_{\text{no illusion}}(200)$ on October 16, 1987, without mean reversion illusion. The actual value of the S&P500 at the closing of October 15, 1987, was 298.08. I am hence interested in the sample distribution of the difference $\log(S_{\text{no illusion}}(200)) - \log(298.08)$. This is an estimate of the distribution of the magnitude of the crash.

Table 3.3 (left) shows the sample distribution of the difference $\log(S_{\text{no illusion}}(200)) - \log(298.08)$. There is still a substantial probability for an upward jump as even under the regime with stronger mean reversion there is a number of paths that end up above 298.08 after 200 days. The probability of a crash of minus 20 percent or more was more than seven percent. The probability of a correction of minus ten percent or more was more than 40 percent.

To put the somewhat random endpoint of 298.08 into perspective, I evaluated model (3.3.2) for 10,000 sample paths under both parameter regimes, that of the 1982–1986 period ($S_{\text{no illusion}}$) and that of the 1987 period (S_{illusion}). Table 3.3 (right) shows the sample distribution of the difference $\log(S_{\text{no illusion}}(200)) - \log(S_{\text{illusion}}(200))$. Even after only 200 days the difference in the mean reversion parameter λ results in substantial distances in the trajectories and thus substantial probabilities for large jumps when a mean reversion disillusion happens.

These sample distributions were calculated under the assumption that if the mean reversion illusion had not occurred, the Brownian sample path could have been different from the one that was realized between January 2, 1987, and October 15, 1987. One might argue that the stream of fundamental information that makes up the noise part would have been the same in either case. Under this assumption, I can reconstruct the Brownian sample path between January 2, 1987, and October 15, 1987, from model (3.3.2) by

$$\hat{\varepsilon}_t = \frac{1}{\hat{\sigma}} \left[\hat{\mu} + \hat{\lambda} \log \hat{\vartheta}_t + (1 - \hat{\lambda}) \log S_t - \log S_{t-1} \right]$$

using the parameter estimates from the 1987 segment.

Setting $\hat{\varepsilon}_t$ back in into the model with the parameters from the 1982–1986 segment, this gives a point estimate for the $S_{\text{no illusion}}(200)$ and thus a point estimate for the magnitude of the crash. In the case of 1987, I have $S_{\text{no illusion}}(200) = 273.78$ and thus

$$\log(S_{\text{no illusion}}(200)) - \log(298.08) = -0.085,$$

a correction of minus 8.5 percent.

Table 3.3: The left table shows the sample distribution of the difference $\log(S_{\text{no illusion}}(200)) - \log(298.08)$, the latter value is that of the S&P500 at the close of October 15, 1987. This gives an estimate of the distribution of the magnitude of the crash. The probability of a downward jump of 20 percent or more was more than seven percent. The right table shows the sample distribution of the difference $\log(S_{\text{no illusion}}(200)) - \log(S_{\text{illusion}}(200))$ when 10,000 Brownian sample paths of length 200 are evaluated under both regimes, that of the 1982–1986 period ($S_{\text{no illusion}}$) and that of the 1987 period (S_{illusion}). This shows that the difference in the mean reversion parameter leads to substantial probabilities for large moves when a mean reversion disillusion occurs.

r_i	$\mathbb{P}(r_i - 0.10 \leq r < r_i)$	r_i	$\mathbb{P}(r_i - 0.10 \leq r < r_i)$
		-0.5	0.0009
		-0.4	0.0053
-0.3	0.0029	-0.3	0.0221
-0.2	0.0753	-0.2	0.0751
-0.1	0.3652	-0.1	0.1572
0	0.4332	0	0.2333
0.1	0.1160	0.1	0.2297
0.2	0.0072	0.2	0.1687
0.3	0.0001	0.3	0.0775
		0.4	0.0244
		0.5	0.0052
		0.6	0.0006

It is conspicuous that the estimated magnitude of the mean reversion parameter λ is much higher after the crash than in the years 1982 to 1986. One reason for this may be that only a part of the mean reversion expectations after the crash depended on mean reversion expectations prior to the crash. A general increase in risk-aversion after the crash may have caused an autonomous increase in mean reversion expectations.

5 A Note on the Stock Market Crash of 1929

The stock market crash of 1929 occurred on two days, Monday, October 28 and Tuesday, October 29. During these two days, the market as measured by the Dow Jones Industrial Average lost about 23 percent from its closing level of 298.97 on Saturday, October 26 to its closing of 230.07 on Tuesday.¹ Figure 3.8 shows the DJIA between February 1, 1927 and December 31, 1954. The crash was followed by the Great Depression which pushed the DJIA to about 50 points in May 1932. It took until the 1950's to regain the losses of the Great Depression.

The stock market crash of 1929 can not be explained by a mean reversion illusion and disillusion. As the knowledge about the hedge-portfolio of the Black-Scholes analysis was not available and option trading was negligible, it was not possible to implement mean reversion expectations the same way like 1987. An estimation of model (3.3.2) in analogy to Table 3.2 supports this: the result

¹On Saturdays, the stock exchange was open in the morning by that time.

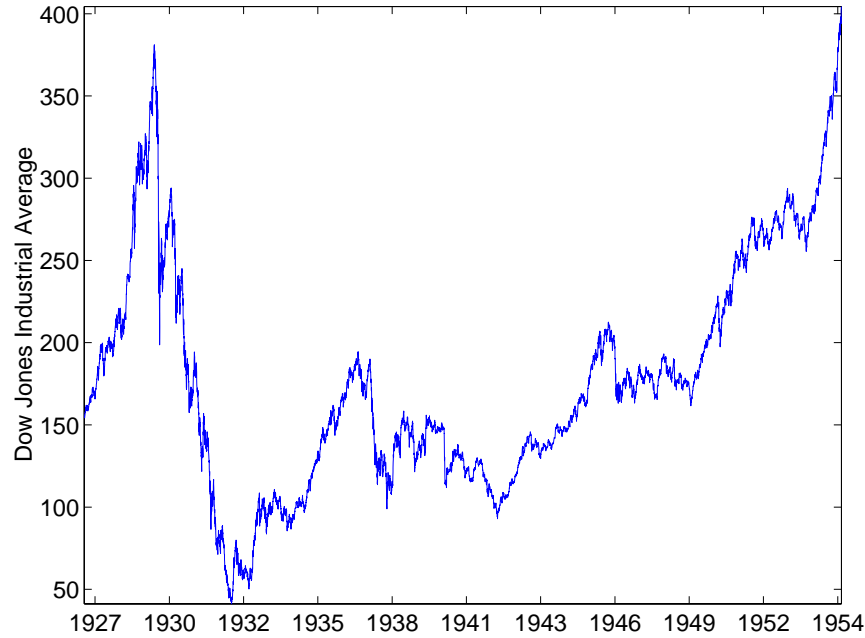


Figure 3.8: Dow Jones Industrial Average between February 1, 1927 and December 31, 1954. On Monday, October 28, 1929 the index dropped 12.8 percent from its level of 298.97 on Saturday, October 26, 1929 to 260.64 at closing. On Tuesday, October 29, 1929 it dropped another 11.7 percent to 230.07 at closing. It took more than 25 years for the market to recover.

is reported in Table 3.4. I deleted the observations October 26, 1929, through December 17, 1929, from the Dow Jones series as in this case it took almost two months before the Dow returned to normal daily changes. The results of Table 3.4 are qualitatively not sensitive to the choice of this gap, I obtained similar results for only ten deleted days. The change in the mean reversion parameter λ does not have the right sign to support the mean reversion theory except for the single instance corresponding to $n_i = 400$. The estimates are much less significant, at most at the one-sided 0.95 level and only in two instances *before* the crash. It is conceivable, however that a similar mechanism of error and correction worked for the expected return with coarser instruments like stop-loss orders. I will not pursue this question here, it may be the subject of a separate investigation.

6 Conclusions

Errors in the perception of the mean reversion expectations can cause stock market crashes. This view was proposed by Black (1988). When the *a-priori* expectation of the speed of the reversion is relatively high but market participants can hedge against a fast reversion and these hedge positions are not public infor-

Table 3.4: Estimation of model (3.3.2) on sample periods before and after the 1929 stock market crash. The observations from October 26, 1987, through December 17, 1929, were deleted from the series. The numbers in parentheses are quasi-maximum-likelihood standard errors according to White (1982). The estimations of the mean returns and standard deviations are significant according to all common confidence levels except for the mean return estimation of the 200 days before and after the crash. For the mean reversion parameter λ those estimates that are significant according to the one-sided 0.95 confidence level are marked with a single asterisk. The mean reversion theory cannot explain the crash of 1929.

i	n_i ($n_0 = 1$)	day n_i through day n_{i-1} before Oct. 26, 1929			day n_{i-1} through day n_i after Dec. 17, 1929		
		$\hat{\mu}$	$\hat{\lambda}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\lambda}$	$\hat{\sigma}$
1	200	9.6e-5 (0.000546)	0.014254 (0.014218)	0.013919 (0.001097)	-0.000230 (0.000550)	0.012919 (0.010533)	0.015425 (0.000980)
2	400	0.001116 (0.000364)	0.003449 (0.003398)	0.012798 (0.000739)	-0.001015 (0.000293)	0.008857 (0.005738)	0.016447 (0.000634)
3	600	0.000990 (0.000149)	0.009946 (0.007427)	0.011623 (0.000558)	-0.001654 (0.000496)	0.004993 (0.005851)	0.021101 (0.001057)
4	800	0.000950 (8.3e-5)	0.010573* (0.006111)	0.010688 (0.000471)	-0.001663 (0.000273)	0.008240 (0.003719)	0.023915 (0.000906)
5	1000	0.000751 (8.6e-5)	0.005945* (0.003410)	0.010145 (0.000401)	-0.001415 (0.000224)	0.005034 (0.003759)	0.025005 (0.000881)

mation, a situation may occur that I call *mean reversion illusion* in this thesis. That is, after a market movement, a group A of investors observes the market before they react. They adapt their high *a-priori* mean reversion expectations to the low expectations that they infer from the market behavior, which, at this point, reflects only the activities of another group B , that is acting autonomously. Investors A do not know that the mean reversion expectations of those investors B already active on the market are about as high but hedged, for example by synthesized put options that cannot be distinguished from stock sales and purchases due to fundamental information. Investors A adapt their expectations to those that they believe are B 's and the stock-price process behaves according to a lower mean reversion speed. When after a while the true *a-priori* expectations of group B become known, for instance because a surprisingly high hedge volume becomes known, the *mean reversion disillusion* of group A sets in. It is now clear that the stock price process followed a path that did not properly reflect the true *a-priori* mean reversion expectations. The process has to be set into a position as if the illusion had not happened. This is a correction in trajectories, not only in process parameters and hence the switch can be of substantial magnitude. This is the stock market crash.

There are two aspects of the illusion. Apart from the illusion about the risk aversion of group B , group A may also discover that it is the vast majority, and that group B , which was mistaken to be representative of the market, is comparatively small. Both types of disillusion can occur at the same time and contribute to the crash.

I specify a stock-price model with mean reversion in stock returns and estimate it on one hundred years of daily data of the Dow Jones Industrial Average. I show that there are recurring periods where mean reversion is highly significant. There is no mean-aversion, that is the mean reversion parameter is always positive.

Using daily data of the S&P500 index I examine the stock market crash of 1987 in detail. Using the periods of the bull market as proposed by the report of the Brady-Commission, I show that in 1987 mean reversion was much lower than during the period 1982–1986. This supports the hypothesis of a mean reversion illusion. Simulations of the model with the estimated parameters of the two segments show that a crash of 20 percent or more had a probability of more than seven percent. A correction of minus 10 percent or more had a probability of more than 40 percent.

There was a significantly higher mean reversion after the 1987-crash than before. This supports the hypothesis that a mean reversion disillusion occurred. The cause of the disillusion can be identified as the surprisingly high volumes of equities under portfolio insurance schemes that became known during the week prior to the crash. Not the mere existence of portfolio insurance and cascading program trading caused the crash but the fundamental information that the average *a-priori* mean reversion expectations in the market were much higher than commonly perceived.

The stock market crash of 1929 cannot be explained by errors in the perception of mean reversion expectations. Apart from the fact that synthesized put options were unknown by that time, no significant change in mean reversion before and after the crash can be measured. The question whether in this case a similar pattern of error and correction concerned the expected return is left for future research.

Part B

Mean Reversion in Volatility

Chapter 4

Volatility Persistence, Mean Reversion, and Long Memory

In contrast to mean reversion in prices and returns, for which explanatory models are predominantly microeconomic, mean reversion in volatility is mostly explained in a macroeconomic way. Probably the best example is the large outburst of volatility during the Great Depression in the 1930's. It is commonly believed that the deflation – loan crisis – contraction spiral described by Fisher (1933) caused the sustained increased volatility on the stock market.

Volatility is seen as the best proxy for the information stream that enters the market. In the efficient market model, the discounted expected value of the dividend stream is the stock price. New information changes the expectations of the dividends, thereby causing price fluctuation.

In this view, uncertainty about the dividend stream is the prime suspect to cause stock price volatility. The dividend stream is itself stochastic and the necessity to estimate it may even exaggerate the stochasticity (Barsky and DeLong 1993). Memory in volatility arises then from memory in the underlying fundamentals: bad news are followed by bad news, good news by good news.

Mean reversion in volatility is a much better established phenomenon than mean reversion in returns. One of the most frequently reported facts about volatility is that it exhibits long memory, or high persistence (Ding, Engle and Granger 1993, Engle and Patton 2001). That is, it possesses an autoregressive dynamic structure with slowly decaying autocorrelations (such that the sum of the autocorrelation coefficients is possibly not converging).

Figure 4.1 illustrates this for daily observations of the Dow Jones Industrial Average between January 2, 1902, and December 30, 2001. The upper panel shows the sample autocorrelation function of the squared daily returns, the lower panel the sample autocorrelation function for the absolute daily returns, both being a measure for volatility. Both sample autocorrelation functions decay slowly

exponentially, with significant coefficients up to lag 1,000 or 4 years for the squared returns and up to lag 2,000 or 8 years for the absolute returns. (For the difference in the measures see Ding, Engle and Granger 1993.)

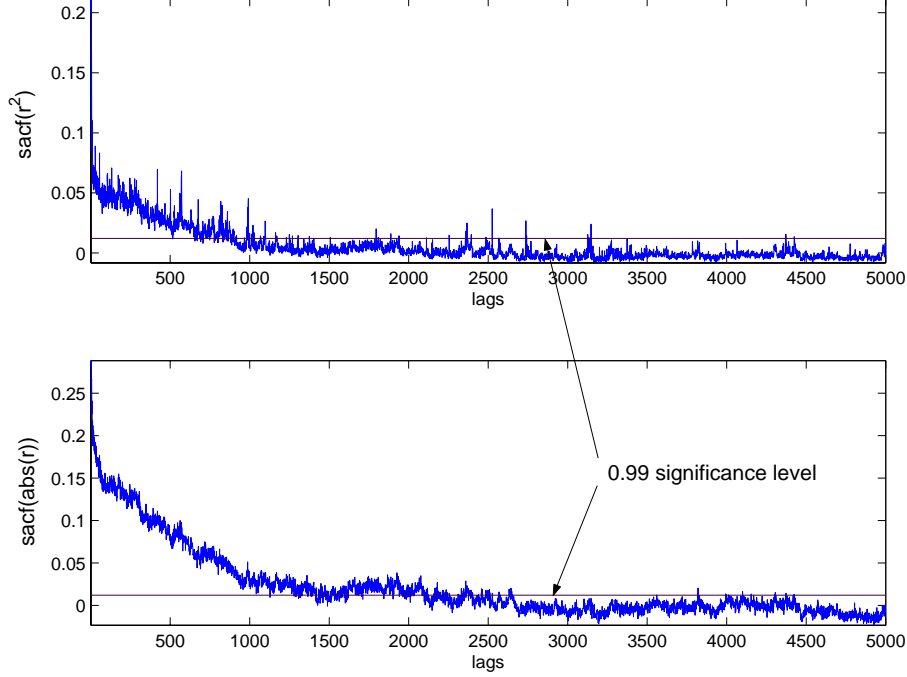


Figure 4.1: Sample autocorrelation functions for the square of the daily returns (upper panel) and for the absolute daily returns (lower panel) of the Dow Jones Industrial Average between January 2, 1902, and December 30, 2001.

Therefore, volatility is a natural field for the application of mean reversion models. The most commonly used volatility models: GARCH, stochastic volatility models, and long-memory GARCH processes reflect this fact.

The volatility driver of the simplest ARCH specification (Engle 1982) has the structure

$$h_t = \omega + \alpha \varepsilon_{t-1}^2,$$

where ω is some positive real number, $\varepsilon_t^2 = (r_t - \mu)^2$ is the squared excess return, and α is a real number between zero and one. Intuitively, h may be called modelled volatility whereas ε^2 is the measured volatility. The model is set up such that the expected value of the measured volatility is equal to the expected value of the modelled volatility. Therefore,

$$\mathbb{E}_{t-k} h_t = \omega + \alpha \mathbb{E}_{t-k} h_{t-1}.$$

where $k \geq 2$. ARCH is thus a mean reverting model of class I according to Section 2.1 with α being the autoregressive parameter.

The volatility driver of the simplest GARCH specification (Bollerslev 1986) has the structure

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1},$$

where β is another parameter between zero and one and $\alpha + \beta < 1$. Exactly as in the ARCH specification, the expected value of the measured volatility is equal to the expected value of the modelled volatility. Therefore,

$$\mathbb{E}_{t-k} h_t = \omega + (\alpha + \beta) \mathbb{E}_{t-k} h_{t-1}.$$

GARCH, too, is thus a mean reverting model of class I according to Section 2.1 with the sum of the coefficients being the autoregressive parameter.

In stochastic volatility models, the volatility driver is often specified as an Ornstein-Uhlenbeck process (Fouque et al. 2001):

$$dV_t = \lambda(\mathbb{E}V - V_t) dt + \sigma dW_t,$$

where V_t is the volatility process at time t and λ is the coefficient that controls the “speed” of the mean reversion. The average time that the process V needs to revert to the mean $\mathbb{E}V$ is given by $1/\lambda$. A discretization of this process yields an autoregressive process of order one with *negative* autoregressive parameter $-\lambda$. Ornstein-Uhlenbeck processes of this type are mean reversion models of class II according to Section 2.1.

GARCH processes have been generalized to capture long memory behavior. The fractionally integrated GARCH (FIGARCH) approach of Baillie et al. (1996) as well as the multi-component model of Ding and Granger (1996) are essentially ARCH models with an infinite number of lags of the measured volatility ε^2 that has to be truncated at some arbitrary lag in estimations.

Baillie et al. arrive at the infinite ARCH order by fractional differencing as described in Section 2.1 in the context of mean reversion models of class III. Ding and Granger obtain it from adding up an infinite number of separate volatility drivers. The FIGARCH model is a mean reverting model of class III whereas the Ding and Granger approach is essentially class I. That the FIGARCH model has a class I representation shows the close relation of the model classes.

All of these models are, again, reduced form models and do not allow for an immediate economic interpretation. Three questions arise naturally:

1. What determines the level of volatility?
2. What determines the long memory of volatility?
3. Is there a connection between prices and volatility?

Except for the obvious fact that macroeconomic conditions influence the level of volatility, all of these questions remain largely elusive until today. Long memory is a “stylized fact” of the data, yet there are but a few explanations around (for example, Kirman and Teyssiere 2000 explain long memory by agent interaction and herd behavior).

The third question was addressed by Fischer Black (1976). In his view, a decrease in a company’s equity capital has two effects on the company’s stock price. On the one hand, the lower value induces a decrease in expected future dividends and thereby causes a decline of the stock price. On the other hand, everything else being the same, the debt-equity ratio decreases and this increases the uncertainty about whether the company will be able to meet its liabilities in the future. This increases the volatility of the stock.¹ This effect was later labeled the “leverage effect”.

This leverage effect would connect mean reversion in prices and returns with the volatility process. It is conceivable that this is another generator of mean reversion: transmission of mean reversion between volatility and returns.

A reduced form attempt to model this effect is the class of ARCH-in-mean models where the ARCH volatility driver described above also appears in the mean equation of the stock returns (Engle, Lilien, and Robins 1987). This may lead to a better fit to the data but it explains little. As to my knowledge, there is no explanatory model, for example a consumption smoothing approach, that connects mean reversion in returns to mean reversion in volatility.

In this thesis, I will concentrate on the first and second question posed above. I will employ GARCH models, which are the most commonly used models in practice. I will show that changes in the parameter regime of a GARCH process will cause the sum of the autoregressive parameters to be close to one, when the changepoint is not accounted for in global estimations. This is commonly viewed as evidence of high persistence, or long memory. When accounting for changepoints, however, I will show that the average data generating mean reversion of daily stock price volatility is quite fast, of the order of a few days.

This leads us to a partial answer of the second question: Long memory is caused by regime shifts, structural breaks in the time series. These change the level of volatility from regime to regime. We are left with two new questions: “What causes the changes in regime?” and “What determines the short memory of volatility within regimes?”

¹Fischer Black gives the following example: Suppose a company has 10 million worth of assets, 6 million capital and 4 million in bonds outstanding. Now let the value of the firm be cut in half but the own stock is shrinking to 2 million, while bonds drop only to 3 million. The debt-equity ratio has deteriorated significantly, which will cause an increase in volatility of the company’s stock.

One possible cause of regime changes is discussed in Chapter 7 in the context of exchange rates, where interventions by the monetary authorities are a good candidate. For stock prices, good candidates are harder to identify.

Substantially, short memory means that shocks to the stock price die out fast, the time scale involved is of the order of 5 to 10 days. It is conceivable that this is something like a weighted average of the investments horizons of the market participants. This would corroborate to the view expressed in Chapter 3 and in the introduction that investors implement their preferences, views and opinions in the data, and that the aggregate of this implementation over all investors constitutes the data generating process, in not too gross a deviation from certain rationality conditions. The investment horizon is definitely one of the most important parameters an investor has control over, besides his risk preference. At this point, however, there is no economic explanation available that would clearly link the time scale of volatility to the investment horizon of the agents. The issue was shrouded by the long memory ubiquitously measured in the data.

Finally, what are the effects of mean reversion in volatility? It could have been suspected that long memory in volatility carries forward uncertainty. Once a series of high fluctuations arrives, investors are shied away from the stock market, risk averse first. This causes a drop in prices, which, according to the leverage effect, causes another rise in volatility. In this view, long memory in volatility could turn the stock market into a catalyst of depressions. It must be emphasized, though, that this “causal” chain interprets *correlations* (as opposed to causal relations) found in the data quite liberally. The findings presented in this thesis contradict this picture anyway. When volatility has a short memory, it is much more likely that stock market volatility simply “measures” the state of the economy, without interfering much.

Chapter 5

Mean Reversion and Persistence in GARCH(1,1)

A common finding in the empirical literature is that the volatility of financial data exhibits high persistence, or slow mean reversion of the order of months. I present evidence that stock price volatility contains more than a single mean reversion time. After showing that the expectation of the sum of the estimates of the autoregressive coefficients of a GARCH(1,1) model is one when there are unknown parameter changes, I explore the phenomenon in simulations. For parameter changes within realistic ranges for stock price volatility I obtain global estimates indicating high persistence while the average data-generating mean reversion is of the order of a few days. Spectral analysis of the Dow Jones Industrial Average and the S&P500 index between 1985 and 2001 reveals a short time scale of the magnitude of 5-10 days present in the data. Thus, two different time scales exist in the data, one of the order of months corresponding to different volatility regimes, and one of the order of days corresponding to the mean reversion within regimes.

1 Time Scales and Persistence in Financial Volatility Data

There are at least two different ways to interpret volatility clustering. An investor with a long-term horizon will see relatively short periods of high or low volatility as jumps in the fluctuation level that have a persistent influence. He perceives a long term mean to which volatility reverts only slowly after a deviation. An investor with a short-term horizon will hold a different view. If the periods of high or low volatility last longer than his investment horizon, he is likely to see them as different states of the level of fluctuation. His idea of a mean level is short-lived and within each state volatility tends to revert fast to this level. The states are changing though, and they tend to be persistent.

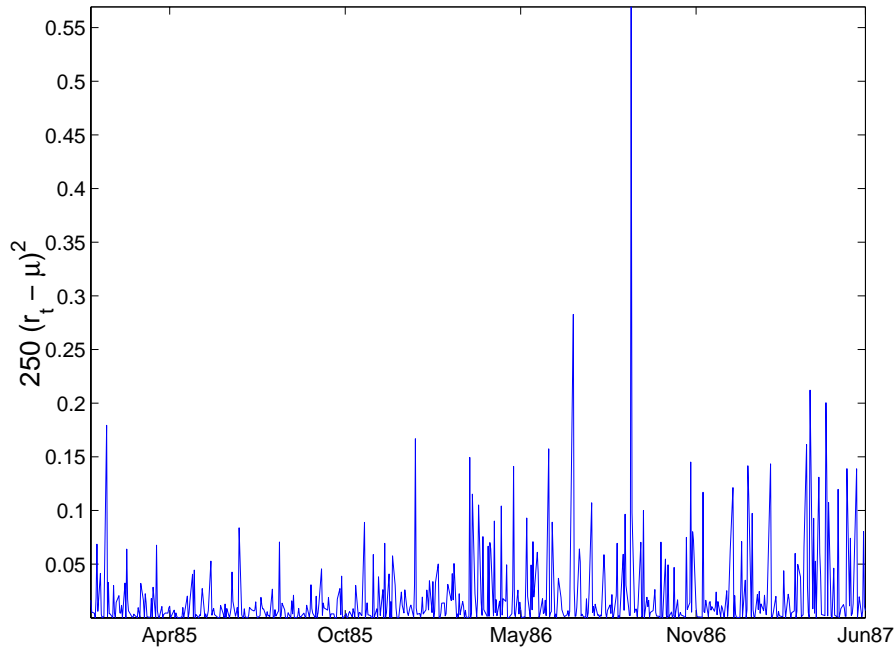


Figure 5.1: Annualized volatility series $250 \cdot (r_t - \mu)^2$ of the daily Dow Jones Industrial Average between January 3, 1985, and June 5, 1987. Clearly, there is a change in the volatility level in early 1986. An investor with a short investment horizon will be exposed to the small, fast mean reverting fluctuations within segments. He will perceive the change in early 1986 as a shift of the volatility mean. An investor with a long horizon of the order of months or years will not be exposed to the small fluctuations and perceive a global mean. In his view, the process moved below the mean in the first segment and above it in the second segment. The investor will interpret the change in the level as a jump that had persistent influence.

The changing states can be understood as the moves of a second process with a much longer time scale than the one governing the moves within the states. That is, contrary to only a single, long-range time scale we have to deal with two overlaying time scales. For modelling volatility it is thus desirable to have a method that can capture more than one time scale of the process under examination. When only one time scale can be modelled, one has to make a choice. The global estimation of such stationary processes will have to accommodate the changing states by assuming high persistence and this will mask the short time scale. The change in states has to be interpreted as a jump to a different level and this level persists. On the other hand, allowing for changes will capture the short-run dynamics of the volatility process better at the disadvantage that the long-term scale will be hidden and that the changepoints will have to be identified.

The most commonly employed time-series model for volatility estimation is the generalized model of autoregressive conditional heteroskedasticity, or GARCH

(Engle 1982, Bollerslev 1986). The sum of the estimates of its autoregressive parameters is often found to be almost unity. I will refer to this phenomenon as “almost-integration”. The implied high persistence of volatility is regarded as a stylized fact. This motivated the formulation of Integrated and Fractionally Integrated GARCH models (Engle and Bollerslev 1986, Baillie et al. 1996).

I will show that the ambiguity of persistent influences of jumps and changing states translates fully into the GARCH estimation. Analytically and numerically I will demonstrate that parameter changes that are not accounted for in global GARCH estimations lead to high estimated persistence close to integration. This is regardless of the data-generating persistence within segments and regardless of the estimation method. I find that a single changepoint between realistic values for stock market volatility can be sufficient for this effect to occur.

To find the short correlation structure, I use methods of spectral analysis that allow to detect time scales independently of the model formulation. I clear the volatility time series from the long time scale that was detected by GARCH and estimate the power spectrum of this properly defined residual. This method reveals a short time scale of the magnitude of 5 to 10 days present in the daily volatility of the Dow Jones and the S&P500.

2 Persistence Estimation with GARCH Models

a) The Model Formulation

Engle (1982) and Bollerslev (1986) suggested the following approach. The return r_t from a stock with price S_t at time t is modeled as

$$r_t := \log(S_{t+1}) - \log(S_t) = \mathbb{E}(r_t | \mathcal{F}_{t-1}) + \varepsilon_t = \mu(b) + \varepsilon_t. \quad (5.2.1)$$

Here, \mathcal{F}_t denotes the filtration modelling the information set and μ is the conditional mean function with argument b , for example a regression $\mu(b) = X_t^T b$, where X_t denotes a set of independent variables. The disturbance ε_t is assumed to be normally distributed, conditional on the information available at time $t - 1$:

$$\varepsilon_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, h_t), \quad (5.2.2)$$

i.e. $\varepsilon_t = \eta_t \sqrt{h_t}$, $\eta_t \sim \mathcal{N}(0, 1)$, where h_t denotes the conditional variance. The latter is determined by the difference equation

$$h_t = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i}, \quad (5.2.3)$$

with $\omega, \alpha_i, \beta_i \geq 0 \forall i$. This is the GARCH(p,q) model for the conditional variance. To obtain the unconditional expected variance, assume that the process $\{\varepsilon_t\}$ is

covariance-stationary. Then, $\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i < 1$ holds and

$$\mathbb{E}h = \frac{w}{1 - \sum \alpha_i - \sum \beta_i}. \quad (5.2.4)$$

We will restrict the arguments to the GARCH(1,1) specification

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} \quad (5.2.5)$$

with $\varepsilon_t = r_t - \mu$, $\mu \in \mathbb{R}$ fixed and $\varepsilon_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, h_t)$.¹

GARCH as a reduced form time series model has no economic interpretation of the α - and β -coefficients. The intuition of the model is that by allowing for time dependent volatility (heteroskedasticity), it can capture volatility clustering (high volatility is followed by high volatility, low volatility is followed by low volatility, via the autoregressive dynamics). By formulating the heteroskedasticity in a conditional way, that is, dependent on the data at the preceding observation point, the model becomes accessible with maximum likelihood estimation. This is because the transition probability density is explicitly modelled, from which the likelihood function can be obtained. An unconditional formulation of heteroskedasticity would require much more elaborate filtering techniques for estimation. However, the GARCH coefficients have a distinct interpretation as determining a time scale, this will be shown in the following Section.

b) Measures of Persistence and Mean Reversion

Consider the conditional variance at time $t + k$, $k \geq 2$, and take expectations conditional on \mathcal{F}_t :

$$\begin{aligned} \mathbb{E}_t h_{t+k} &= \omega + \alpha \mathbb{E}_t \varepsilon_{t+k-1}^2 + \beta \mathbb{E}_t h_{t+k-1} \\ &= \omega + (\alpha + \beta) \mathbb{E}_t h_{t+k-1}. \end{aligned} \quad (5.2.6)$$

Thus, the k -period forecast of the conditional variance according to the GARCH-(1,1)-model is a first-order linear difference equation with autoregressive parameter

$$\lambda := \alpha + \beta. \quad (5.2.7)$$

The closer λ is to unity, the more persistent the effect of a change in $\mathbb{E}_t(h_{t+k})$ will be. The parameter λ is the fraction of the forecast that is carried forward per unit of time, so $(1 - \lambda)$ is the fraction that is washed out per unit of time. Hence, $1/(1 - \lambda)$ is the average time needed to return to the mean when the time increment equals one. To formalize this, denote x_t as the distance of $\mathbb{E}_t h_{t+\Delta t}$ from its mean $\mathbb{E}h_t$, that is

$$\mathbb{E}_t h_{t+\Delta t} = \mathbb{E}h_t + x_t.$$

¹I will not report estimates of the constant mean return for the sake of brevity.

Then, from (5.2.6) we have for the case of $\Delta t = 1$ that

$$\mathbb{E}h_t + x_t = \omega + \lambda(\mathbb{E}h_t + x_{t-1}).$$

As $\mathbb{E}h_t = \omega + \lambda\mathbb{E}h_t$ for $\lambda < 1$, we obtain

$$x_t = \lambda x_{t-1}$$

or

$$x_{t+1} - x_t = (\lambda - 1)x_t \quad (5.2.8)$$

for the distance of the forecast from the unconditional mean. We may model x by a decreasing function y defined by the difference equation

$$y_{t+\Delta t} - y_t = -\kappa y_t \Delta t, \quad \kappa > 0. \quad (5.2.9)$$

In the continuous time limit, y_t has the form

$$\frac{y_t}{y_0} = e^{-\kappa t},$$

and the so-called e -folding time

$$\left(t_e \left| \frac{y_{t_e}}{y_0} = e^{-1} \right. \right)$$

is given by $t_e = 1/\kappa$. Comparing the coefficients in (5.2.8) and (5.2.9), we see that

$$\kappa = 1 - \lambda$$

and the e -folding time of the distance x of the forecast from the unconditional mean is

$$t_e = \frac{1}{\kappa} = \frac{1}{1 - \lambda}. \quad (5.2.10)$$

There are other ways to define and measure persistence in a discrete GARCH model (discussed e.g. in Engle and Patton 2001). Nelson (1990a) uses a variety of persistence definitions and shows that whether or not shocks are persistent depends crucially on the definition chosen.

c) Maximum Likelihood Estimation

The most common way to estimate a Gaussian GARCH(1,1) model with constant mean return given a sequence $\{S_t\}_{t \in \mathbb{N}}$ of prices is by maximum likelihood derived from equation (5.2.2). Let $\varepsilon_t(\mu) = r_t - \mu$, $\mu \in \mathbb{R}$ fix. Denote the parameter vector by $\theta = (\mu, \omega, \alpha, \beta)$. The log-likelihood function is given by

$$L_N(\theta, \{\varepsilon_t\}_t) := -\frac{1}{2N} \left[N \log(2\pi) + \sum_{t=1}^N \left(\log h_t(\theta, \varepsilon_{t-1}) + \frac{\varepsilon_t^2(\mu)}{h_t(\theta, \varepsilon_{t-1})} \right) \right]. \quad (5.2.11)$$

The GARCH model is not restricted to the conditionally normal case. Bollerslev (1987) suggests using the t -distribution and treating the number of degrees of freedom as additional parameter.

Consistency and asymptotic normality of the maximum likelihood estimator could only be proven in the conditionally Gaussian GARCH(1,1) case so far. The main results can be found in the papers by Weiss (1986), Bollerslev and Wooldridge (1992), and Lumsdaine (1996). There are no closed analytical expressions for the estimators.

In practice, the likelihood is maximized by numerical optimization methods. Most software packages implement a quasi-Newton method using linesearch and Hessian update algorithms. There are alternatives to this approach, like generalized least squares estimators (Gouriéroux 1997) or scoring methods (Harvey 1976, Greene 2000). I maximize (5.2.11) using code written in MATLAB and C++. The MATLAB code uses the ‘fmincon’ routine from the optimization toolbox which implements a quasi-Newton method. The C++ code uses the ‘dfpmin’ routine from the “Numerical Recipes” (Press et al. 2002) which also implements a quasi-Newton method. The gradients are computed using analytical expressions, the Hessians are approximated by finite differencing.

d) GARCH(1,1) and Market Data: High Persistence in the Volatility of the Dow Jones and S&P500

I use daily closings of the Dow Jones Industrial Average and the S&P500 ranging from January 2nd, 1985, to January 2nd, 2001.² The Dow Jones data was kindly provided by Dow Jones & Company, the S&P500 was downloaded from Datastream. When I globally estimate a Gaussian GARCH(1,1) model for the annualized daily returns of the 16 years that the series cover, I obtain

$$h_t = \underset{(0.00018)}{0.00049} + \underset{(0.0464)}{0.0872} \varepsilon_{t-1}^2 + \underset{(0.0393)}{0.8991} h_{t-1},$$

for the Dow Jones series. This implies a $\hat{\lambda}$ of 0.9863 ($1/(1 - \hat{\lambda}) = 73$ days). For the S&P500 series, I get

$$h_t = \underset{(0.00012)}{0.00037} + \underset{(0.0374)}{0.0888} \varepsilon_{t-1}^2 + \underset{(0.0316)}{0.9024} h_{t-1},$$

which implies a $\hat{\lambda}$ of 0.9912 ($1/(1 - \hat{\lambda}) = 114$ days). The numbers in parentheses are heteroskedasticity-robust standard errors (Bollerslev and Wooldridge 1992). The roots of the characteristic equations $1 - \alpha L - \beta L$, where L is the lag operator, are close to the unit circle for both indices. These estimations pick up a long time scale of the order of months.

²I deleted all holidays with zero returns.

I examined model selection criteria for GARCH(p, q) models of the class $p \in \{1, 2, 3\}$, $q \in \{1, 2, 3\}$. The Akaike and Schwarz information criteria favored GARCH(2,.) and GARCH(3,.) if any, but the margins were very small. For higher order models there was mostly only one β_i significant and it was β_1 in most of the cases. The exceptions were GARCH(3,3) for the Dow Jones (all three β_i 's significant) and GARCH(1,3) and GARCH(3,2) for the S&P500 (β_1 , β_3 and β_1 , β_2 significant). The characteristic equation of higher order GARCH models has more than a single root, so that these models might be able to capture multiple time scales. Possibly, this is the reason why I observe the slight advantage of higher order models according to the Akaike and Schwarz criteria.

e) High Persistence as a Stylized Fact

Global estimations of GARCH models usually indicate high persistence or slow mean reversion. For the GARCH(1,1) model, many studies report $\hat{\lambda} = \hat{\alpha} + \hat{\beta}$ close to unity, the majority of which base this observation on global estimations of long-range data sets.³ The conclusion that λ is indeed equal to one and that the constraint $\hat{\alpha} + \hat{\beta} < 1$ is active suggests itself. This gave rise to the formulations of Integrated GARCH (IGARCH, Engle and Bollerslev 1986) and Fractionally Integrated GARCH (FIGARCH, Baillie et al. 1996), which assume an indefinite memory.

The concern that the apparently high persistence in the observed data may be caused by structural changes was raised early. In a comment to the original IGARCH paper by Engle and Bollerslev (1986), Francis Diebold mentioned with regard to interest rate data that not accommodating shifts in monetary policy regimes, reflected in changes of the constant term ω in (5.2.5), might lead to an apparently integrated series of squared disturbances (Diebold 1986). Lamoureux and Lastrapes (1990) showed Diebold's conjecture to be right for stock data, obtaining their results by including dummy variables that indicate different states of the GARCH(1,1) constant ω , equidistant in time. Hamilton and Susmel (1994) used the regime switching model to improve volatility forecasts of ARCH models by incorporating state changes. Gray (1996) extended the regime switching approach to GARCH. These locally stationary approaches that segment the data obtained significantly lower estimates of the order of days.

Using high-frequency data of 5 minute returns estimated at different frequencies, Andersen and Bollerslev (1997) gave a concise overview of the irregular

³Engle/Bollerslev (1986): weekly returns on exchange rates over 12 years; Baillie/DeGennaro (1990): daily returns on stock index over 18 years; Bollerslev/Engle (1993): daily returns on exchange rates over 5 years; Baillie et al. (1996): daily returns on exchange rates over 13 years; Ding/Granger (1996): daily returns on stock index over 63 years; Andersen/Bollerslev (1997): 5 minute returns on exchange rates over *one* year and on stock index future prices over 4 years; Engle/Patton (2001): daily returns on stock index over 12 years.

picture. They obtained estimates ranging from two hours to 7 days for the half-life of the S&P500 1986–1989. Authors who were not using GARCH got still different results: Fouque et al. (2002) obtained about 1.5 days for the average mean reversion length from high-frequency S&P500 data 1994–1998 by spectral methods. Drăgulescu and Yakovenko (2002) obtained 22 days for daily returns of the Dow Jones 1982–2001 by estimating an explicit probability density.

In the literature on long memory, the connection between structural breaks, aggregation, and long memory has been discussed for some time. Granger (1980) showed that aggregation over processes with different autoregressive parameters induced long memory properties. Lobato and Savin (1998) suggested that structural breaks may lead to overestimation of the parameter of fractional integration, labelling this effect “spurious long memory”. Granger and Hyung (1999) extended this approach to unknown breaks. Diebold and Inoue (2001) showed that stochastic permanent break models and Markov-switching models display behavior consistent with long memory. Granger and Teräsvirta (2001) showed that a simple nonlinear model that displays regime switching behavior also exhibits long memory properties. Sakoulis and Zivot (2000) and Choi and Zivot (2002) demonstrated that accounting for breaks and thereby reducing the estimated persistence can contribute to the solution of the so-called forward discount puzzle. These observations are closely related to the overestimation of the sum of the autoregressive GARCH parameters that I will consider here.

In the context of stochastic volatility models, the presence of multiple time scales is proposed recently. Fouque et al. (2002) suggest a multi-scale stochastic volatility model. They show that the estimation of such a model on the short time scale is not affected by the long-run dynamics. This corresponds to the observation that local GARCH estimations of properly segmented data do not reveal the long-term high persistence. On the other hand, LeBaron (2001) shows in a similar model with three short time scales that the three factors induce long memory properties. This corresponds to the occurrence of almost-integration in global GARCH estimations. Chernov et al. (2002) also discuss multi-driver stochastic volatility models. Gallant and Tauchen (2001) estimate a two-scale volatility model and find a long and a short correlation structure in daily returns on the Microsoft stock. The two concepts, overlaying long range processes and parameter switches, differ only in the continuity of their influence. A jump that occurs once every n units of time adds the time-scale of n to the process.

Francq et al. (2001) examine ARCH processes which are subject to Markov-switching parameters. They show in simulations that as a result of the stochastic nature of the Markov-switching process the ARCH parameters will be estimated in the neighborhood of integration. Mikosch and Starica (2000) show that the Whittle estimate of ARMA(1,1) parameters will imply almost-integration when there are changepoints.

3 Parameter Changes and Global GARCH(1,1) Estimations

I will show why global estimations of GARCH(1,1) models that do not account for a single changepoint in the constant ω will result in almost-integration. Therefore, it is not necessary that the long scale process has a specific stochastic structure, a single deterministic changepoint is sufficient for the effect to occur. This is regardless of the estimation method.

a) The Geometry of Almost-Integration

The reason why GARCH(1,1) exhibits almost-integration when there are unknown parameter changes is that a single estimation hyperplane is fitted through different means of volatility. These are the means within the segments of constant parameters.

Consider the case of a single changepoint, two segments of volatility data. In each segment, the data are centered approximately around the unconditional, stationary mean corresponding to the parameters of that segment. “Approximately” means up to terms that vanish with growing segment length.⁴ If a GARCH(1,1) model is estimated globally without accounting for the segmentation, the resulting estimation hyperplane (parameterized by $\hat{\omega}$, $\hat{\alpha}$, $\hat{\beta}$) must go through both means. If the means are sufficiently different, almost-integration must occur.

Figure 5.2 illustrates this for a synthetic GARCH(1,1) series with a single change in ω . The two different data-generating parameters induce two distinct expected values $\mathbb{E}h^{(1)}$ and $\mathbb{E}h^{(2)}$. The spheres in Figure 5.2 are centered at these expected values. The data points $\{h_t\}$, $\{\varepsilon_{t-1}^2\}$, and $\{h_{t-1}\}$ of the segments cluster around their respective means. The clusters exhibit slopes in both subspaces, reflecting the data-generating α in the $(h_t, \varepsilon_{t-1}^2)$ -subspace and the data-generating β in the (h_t, h_{t-1}) -subspace. These slopes cannot be captured by the single estimation hyperplane that has to go through both segments. The relative position of the two means dominates.

As the mean of the $\{h_t\}$ and the mean of the $\{h_{t-1}\}$ is equal for sufficiently long segments, a line connecting two different means in the (h_t, h_{t-1}) -subspace has slope equal to one. Therefore, β will be estimated close to one. The remaining parameter $\hat{\alpha}$ is chosen residually such that $\hat{\alpha} + \hat{\beta} < 1$ as the estimated process \hat{h}_t would blow up otherwise.

In real estimation problems, the ε_t and the h_t cannot be observed but have to be estimated along with the parameters: $\hat{\varepsilon}_t = \hat{\varepsilon}_t(\hat{\mu})$ and $\hat{h}_t = \hat{h}_t(\hat{\omega}, \hat{\alpha}, \hat{\beta})$. Figure

⁴Strictly, the stationary measure is not defined in the case of segmented data. For this reason, the largest part of the proof to follow is concerned with making this approximation precise.

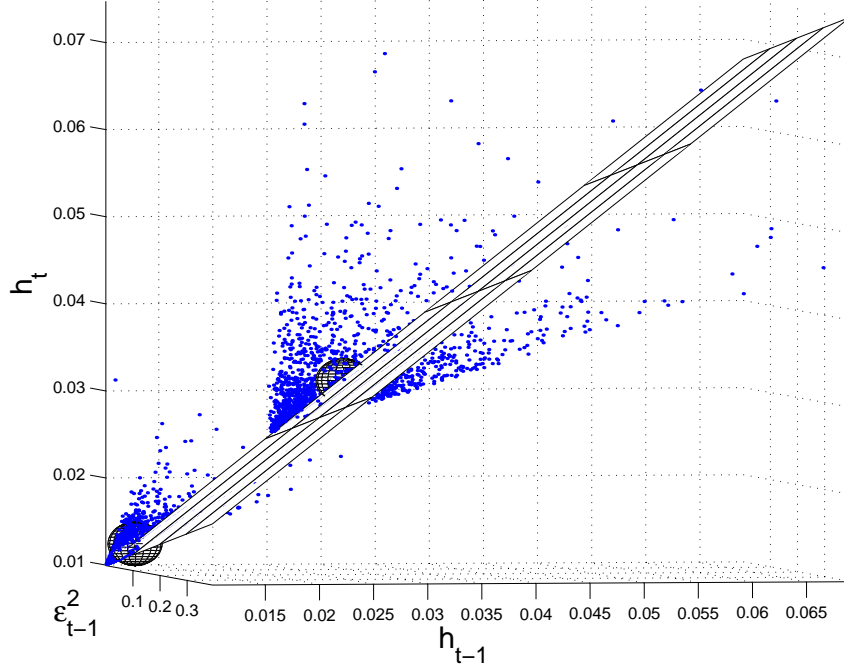


Figure 5.2: Plot of the data points $\{h_t\}$, $\{\varepsilon_{t-1}^2\}$, $\{h_{t-1}\}$ for a synthetic GARCH(1,1) series with a single changepoint in ω . The $\{\varepsilon_t\}$ and $\{h_t\}$ were generated by the parameters $\omega_1 = 2e-5$ and $\omega_2 = 5e-5$, $\alpha = 0.10$ and $\beta = 0.50$. The length of the entire series was $N = 4200$ and the changepoint N_1 was set at one half of N . The spheres are centered at the unconditional, stationary expected values $\mathbb{E}h^{(1)} = 250 * 2e - 5 / (1 - 0.1 - 0.5) = 0.0125$ and $\mathbb{E}h^{(2)} = 250 * 5e - 5 / (1 - 0.1 - 0.5) = 0.03125$. (The data were annualized, hence the multiplication by 250.) The fact that a single hyperplane is fitted through both segments, reflected in the two point clusters, leads to almost-integration. The slope of the clusters with respect to the (h_t, h_{t-1}) -subspace, which is $\beta = 0.5$, is largely overestimated. The slope of the clusters with respect to the $(h_t, \varepsilon_{t-1}^2)$ -subspace, which is $\alpha = 0.1$, is underestimated. The estimated parameters are $\hat{\omega} = 2.6e-5$, $\hat{\alpha} = 0.018$, and $\hat{\beta} = 0.981$.

5.3 shows the estimated data points $\{\hat{h}_t(\hat{\omega}, \hat{\alpha}, \hat{\beta})\}$, $\{\hat{\varepsilon}_{t-1}^2(\hat{\mu})\}$, and $\{\hat{h}_{t-1}(\hat{\omega}, \hat{\alpha}, \hat{\beta})\}$ for the same synthetic series shown in Figure 5.2. The viewpoint is chosen differently. By construction, all points lie on the estimation hyperplane. However, the two-cluster structure is still visible. The estimation hyperplane is, of course, the same as in Figure 5.2.

b) The Analysis of Almost-Integration

In this section, I will prove that if a GARCH(1,1) model is estimated on data that contain an unknown switch in the data-generating GARCH constant ω , almost-integration must occur.

First, I will summarize the idea of the proof that follows the geometric intu-

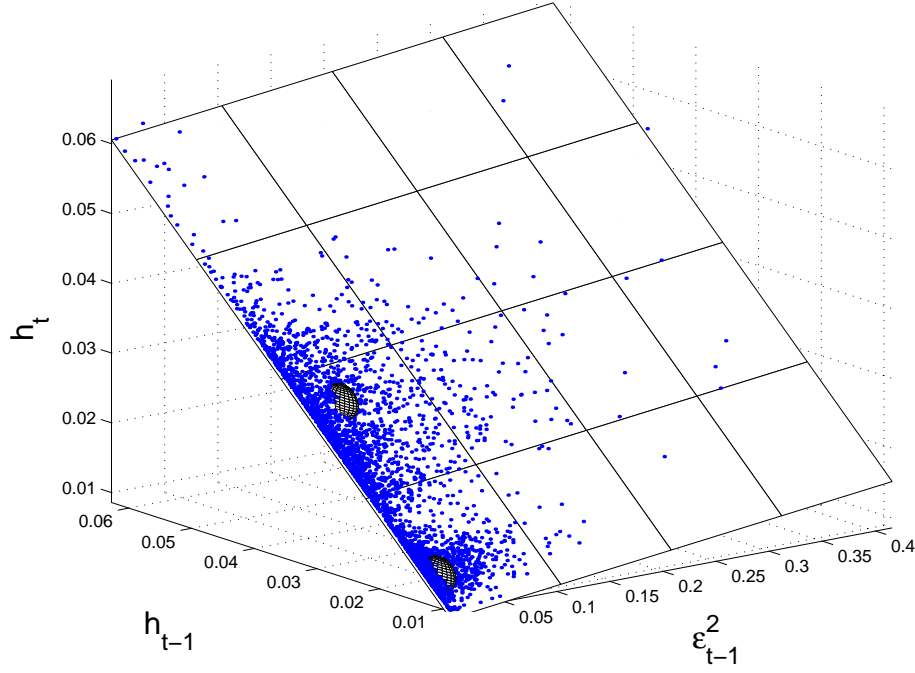


Figure 5.3: Plot of the estimated data points $\{\hat{h}_t(\hat{\omega}, \hat{\alpha}, \hat{\beta})\}$, $\{\hat{\epsilon}_{t-1}^2(\hat{\mu})\}$, and $\{\hat{h}_{t-1}(\hat{\omega}, \hat{\alpha}, \hat{\beta})\}$ for the same synthetic data series considered in Figure 5.2. By construction, all the points are lying on the hyperplane according to the estimates $\hat{\omega} = 2.6\text{e-}5$, $\hat{\alpha} = 0.018$, and $\hat{\beta} = 0.981$. However, the two-cluster structure is still visible. The viewpoint is chosen differently from Figure 5.2.

ition outlined in Section 5.3.a). I will assume that the processes $\{h_t\}$ and $\{\varepsilon_t\}$ can be observed without measurement error or with a measurement error that is independent of the parameter estimates and that vanishes with increasing sample size. This assumption is, of course, unrealistic. The process h_t is not observable and in real estimation problems h_t is estimated by $\hat{h}_t(\hat{\omega}, \hat{\alpha}, \hat{\beta})$ and ε_t by $\hat{\varepsilon}_t(\hat{\mu})$. The conjecture is, however, that if I can show that almost-integration would occur if h_t were observable, it will also occur when I have less information. Figure 5.3 supports this conjecture. For the sake of notational brevity, I will assume that the measurement is error-free. The case of an error that is independent of the parameter estimates would add a correction term that vanishes with growing sample size, without changing the argument.

Also, my only assumption about the estimators $\hat{\alpha}$ and $\hat{\beta}$ will be that their covariance with a single observation in the ε_t^2 series or in the h_t series vanishes with growing sample size. That is, the influence of a single realisation of the volatility process on the estimation vanishes with increasing sample size.

Write down the estimated GARCH equation with the correctly measured h_t

and ε_t :

$$h_t = \hat{\omega} + \hat{\alpha}\varepsilon_{t-1}^2 + \hat{\beta}h_{t-1}, \quad (5.3.12)$$

and subtract the sample mean

$$h_t - \bar{h} = \hat{\alpha}(\varepsilon_{t-1}^2 - \bar{\varepsilon}^2) + \hat{\beta}(h_{t-1} - \bar{h}).$$

One might argue that having exact measurement of h_t , it would be an obvious approach to plug in the values and just back out the parameters, thereby possibly finding parameter switches. However, as my ultimate interest is the case of h_t being unobservable, where this cannot be done, I will nevertheless proceed with the estimation of (5.3.12). The reader might think of an independent measurement error that vanishes with N , so that this trivial approach does not work.

There is a twofold dependency in the estimation of a GARCH model, one being the dependency of the estimated volatility process \hat{h}_t on the estimated parameters, which was assumed away above, the other being the standard dependency of the estimators on the data h_t and ε_t^2 . It is here where the assumption of the vanishing influence of a single realisation is used.

Applying these two assumptions and taking expectations, in the proof of Proposition 5 an expression similar to the following approximation is obtained

$$\mathbb{E}(h_t - \bar{h}) \approx \mathbb{E}\hat{\alpha}\mathbb{E}(\varepsilon_{t-1}^2 - \bar{\varepsilon}^2) + \mathbb{E}\hat{\beta}\mathbb{E}(h_{t-1} - \bar{h}).$$

According to the distribution assumption (5.2.2),

$$\mathbb{E}\varepsilon_{t-1}^2 = \mathbb{E}\mathbb{E}_{t-2}\varepsilon_{t-1}^2 = \mathbb{E}h_{t-1}$$

for $t \geq 2$. For essentially the same reason, $\bar{\varepsilon}^2 \approx \bar{h}$, as will be shown (Lemma 3). Also, $\mathbb{E}h_t \approx \mathbb{E}h_{t-1}$ and thus

$$\mathbb{E}(h_t - \bar{h}) \approx \mathbb{E}(\hat{\alpha} + \hat{\beta})\mathbb{E}(h_t - \bar{h}). \quad (5.3.13)$$

Now, if there are no parameter switches in ω and (5.3.12) is the correct specification, then $\mathbb{E}h_t = \mathbb{E}\bar{h}$. The condition (5.3.13) is trivial. However, in the case where there is a switch in ω , equation (5.3.12) is misspecified, and the expected value of \bar{h} is a weighted average of the means of h_t in the two segments (shown in Lemma 3). To be precise, we have to take care of the start values within each segment, as the stationary measure is not defined on segmented data.

That is, within each segment, the difference between the expected value of h_t and the expected value of the sample mean of h_t is non-zero. Hence, condition (5.3.13) is not trivial and thus

$$\mathbb{E}(\hat{\alpha} + \hat{\beta} \mid \text{segment's start values}) \approx 1,$$

regardless of the data-generating parameters and also regardless of the estimation method. This is a paraphrase of Figure 5.2, where condition (5.3.13) is reflected in the fact that the estimation hyperplane has to go approximately through $\mathbb{E}h^{(1)} = 0.0125$ and $\mathbb{E}h^{(2)} = 0.03125$ and therefore also through $\bar{h} = N_1/N \mathbb{E}h^{(1)} + (N - N_1)/N \mathbb{E}h^{(2)} = 0.5 \cdot 0.0125 + 0.5 \cdot 0.03125 = 0.021875$, where N_1 is the changepoint.

Along these lines, I will now present the proof.

Lemma 1. *Denote by $\mathbb{E}_0 h_t$ the expected value of a stationary Gaussian GARCH(1,1) model conditional on the start value $h_0 \in \mathbb{R}$. Then, the relation*

$$\mathbb{E}_0 h_t = \mathbb{E}h + O(\lambda^t), \quad (5.3.14)$$

holds for $t \in \{1, \dots, N\}$, where $\mathbb{E}h = \omega/(1 - \lambda)$.

Proof. The expected value conditional on the start value is given by

$$\mathbb{E}_0 h_t = \omega + \mathbb{E}_0(\alpha \eta_{t-1}^2 + \beta) \mathbb{E}_0 h_{t-1} = \omega + \lambda \mathbb{E}_0 h_{t-1} = \omega \frac{1 - \lambda^t}{1 - \lambda} + \lambda^t h_0,$$

as $\mathbb{E}_0 \eta_t^2 = 1$ for all t and the η_t are independent. Thus, substituting from equation (5.2.4)

$$|\mathbb{E}_0 h_t - \mathbb{E}h| = \left| \omega \frac{1 - \lambda^t}{1 - \lambda} + \lambda^t h_0 - \frac{\omega}{1 - \lambda} \right| = \lambda^t \left| h_0 - \frac{\omega}{1 - \lambda} \right| = O(\lambda^t). \quad \square$$

Assumption 2. *The processes $\{h_t\}$ and $\{\varepsilon_t\}$ are observable without measurement error, or at least with a measurement error that is independent of the parameter estimates $(\hat{\mu}, \hat{\omega}, \hat{\alpha}, \hat{\beta})$.*

Now, let $\{h_t\}$ be generated by

$$h_t = \begin{cases} \omega_1 + (\alpha \eta_{t-1}^2 + \beta) h_{t-1}, & t \in \{1, \dots, N_1\}, \\ \omega_2 + (\alpha \eta_{t-1}^2 + \beta) h_{t-1}, & t \in \{N_1 + 1, \dots, N\}, \end{cases} \quad (5.3.15)$$

where $\eta_t \sim \mathcal{N}(0, 1)$. This fact is unknown to the econometrician. The estimated model equation is

$$h_t = \hat{\omega} + \hat{\alpha} \varepsilon_{t-1}^2 + \hat{\beta} h_{t-1}. \quad (5.3.16)$$

Subtract the mean from (5.3.16):

$$h_t - \bar{h} = \hat{\alpha}(\varepsilon_{t-1}^2 - \overline{\varepsilon^2}) + \hat{\beta}(h_{t-1} - \bar{h}) \quad (5.3.17)$$

If the segmentation were known, the econometrician would insert a term for the difference in ω .

Let $\mathbb{E}_{(i)} h_t$ denote the expected values with respect to the start value in segment i , where i is 1 for $t \in \{1, \dots, N_1\}$ and i is 2 for $t \in \{N_1 + 1, \dots, N\}$. In other words,

$$\begin{aligned} \mathbb{E}_{(1)} h_t &= \mathbb{E}(h_t | \mathcal{F}_0), \\ \mathbb{E}_{(2)} h_t &= \mathbb{E}(h_t | \mathcal{F}_{N_1}). \end{aligned}$$

Lemma 3. Let $\mathbb{E}h^{(i)}$ denote the expected value of an indefinite process h'_t generated by $\theta = (\mu, \omega_i, \alpha, \beta)$ with respect to the stationary measure. Let h_t be generated according to (5.3.15). Then,

$$\begin{aligned}\bar{h} &= \frac{N_1}{N}\mathbb{E}h^{(1)} + \frac{N - N_1}{N}\mathbb{E}h^{(2)} + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1}, \\ \overline{\varepsilon^2} &= \frac{N_1}{N}\mathbb{E}h^{(1)} + \frac{N - N_1}{N}\mathbb{E}h^{(2)} + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1},\end{aligned}\quad (5.3.18)$$

where $o(1)_{N_1} \rightarrow 0$ as $N_1 \rightarrow \infty$ and $o(1)_{N-N_1} \rightarrow 0$ as $N - N_1 \rightarrow \infty$.

Proof. Write $h_t = \mathbb{E}_{(i)}h_t + x_t$, x_t being the deviation from the expectation conditional on the start values within segments such that

$$\begin{aligned}\frac{1}{N_1} \sum_{t=1}^{N_1} x_t &= o(1)_{N_1} \\ \frac{1}{N - N_1} \sum_{t=N_1+1}^N x_t &= o(1)_{N-N_1}.\end{aligned}$$

From this and Lemma 1 I obtain

$$\begin{aligned}\bar{h} &= \frac{1}{N} \sum_{t=1}^N h_t, \\ &= \frac{1}{N} \sum_{t=1}^{N_1} \mathbb{E}_{(1)}h_t + \frac{1}{N} \sum_{t=N_1+1}^N \mathbb{E}_{(2)}h_t + \frac{1}{N} \sum_{t=1}^{N_1} x_t + \frac{1}{N} \sum_{t=N_1+1}^N x_t, \\ &= \frac{1}{N} \sum_{t=1}^{N_1} \mathbb{E}h^{(1)} + \frac{1}{N} \sum_{t=N_1+1}^N \mathbb{E}h^{(2)} + o(1)_{N_1} + o(1)_{N-N_1} \\ &\quad + \frac{1}{N} \sum_{t=1}^{N_1} O(\lambda^t) + \frac{1}{N} \sum_{t=N_1+1}^N O(\lambda^{t-N_1}) \\ &= \frac{N_1}{N}\mathbb{E}h^{(1)} + \frac{N - N_1}{N}\mathbb{E}h^{(2)} + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1}.\end{aligned}$$

In the same manner, write $\varepsilon_t^2 = \mathbb{E}_{(i)}\varepsilon_t^2 + y_t = \mathbb{E}_{(i)}h_t + y_t$ by the distribution assumption (5.2.2). Then,

$$\begin{aligned}\overline{\varepsilon^2} &= \frac{1}{N} \sum_{t=1}^N \varepsilon_t^2, \\ &= \frac{N_1}{N}\mathbb{E}h^{(1)} + \frac{N - N_1}{N}\mathbb{E}h^{(2)} + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1}.\end{aligned}\quad \square$$

Assumption 4. *The influence of a single realisation of the processes ε_t^2 and h_t on the estimators $\hat{\alpha}$ and $\hat{\beta}$ vanishes with growing sample size:*

$$\begin{aligned} \text{cov}(\hat{\alpha}, \varepsilon_t^2) &= o(1)_N \quad \forall t \\ \text{cov}(\hat{\beta}, h_t) &= o(1)_N \quad \forall t. \end{aligned}$$

This assumption is not very restrictive, it holds for a very general class of estimators. If we apply the Cauchy-Schwarz inequality to the covariance

$$\text{cov}(\hat{\alpha}, \varepsilon_t^2) = \mathbb{E} [(\hat{\alpha} - \mathbb{E}\hat{\alpha})(\varepsilon_t^2 - \mathbb{E}\varepsilon_t^2)] \leq \sqrt{\text{var}(\hat{\alpha}) \text{var}(\varepsilon_t^2)},$$

for example, we see that the assumption is tantamount to a vanishing variance of the estimator as the sample size increases, given that the fourth moment of the ε_t series is finite (Bollerslev 1986). The variance of estimators when there are dependent errors is usually of the order $O(1/\sqrt{N})$ (for example, White 2001, Sections 5.3 and 5.4). For the specific case of conditional heteroskedasticity and GARCH, the same holds true (Weiss 1986, Bollerslev and Wooldridge 1992, Lumsdaine 1996). For instance, the asymptotic distribution of the maximum likelihood estimator $\hat{\theta}$ of the GARCH(1,1) parameters $\theta = (\mu, \omega, \alpha, \beta)^T$ is given by

$$\sqrt{N}(\hat{\theta} - \theta) \sim_{N \rightarrow \infty} \mathcal{N}(0, \Delta)$$

where Δ is a bounded symmetric positive definite matrix that arises from the outer product of the likelihood score.

Proposition 5. *If there is an unknown switch in the data-generating constant of the conditional variance equation of a Gaussian GARCH(1,1) model, as specified in equation (5.3.15), and the model is estimated on the entire series, then, under Assumptions 2 and 4 the condition*

$$\mathbb{E}_{(i)} \hat{\lambda} = \mathbb{E}_{(i)} (\hat{\alpha} + \hat{\beta}) = 1$$

must hold in both segments i , up to terms that vanish with growing length of the segments.

Proof. Take expectations of (5.3.17) conditional on the start value within segments, applying Assumption 2:

$$\mathbb{E}_{(i)} h_t - \mathbb{E}_{(i)} \bar{h} = \mathbb{E}_{(i)} (\hat{\alpha} \varepsilon_{t-1}^2) - \mathbb{E}_{(i)} (\hat{\alpha} \bar{\varepsilon}^2) + \mathbb{E}_{(i)} (\hat{\beta} h_{t-1}) - \mathbb{E}_{(i)} (\hat{\beta} \bar{h}).$$

Use Assumption 4 and the distribution assumption (5.2.2):

$$\mathbb{E}_{(i)} h_t - \mathbb{E}_{(i)} \bar{h} = \mathbb{E}_{(i)} \hat{\alpha} \mathbb{E}_{(i)} h_{t-1} - \mathbb{E}_{(i)} (\hat{\alpha} \bar{\varepsilon}^2) + \mathbb{E}_{(i)} \hat{\beta} \mathbb{E}_{(i)} h_{t-1} - \mathbb{E}_{(i)} (\hat{\beta} \bar{h}) + o(1)_N.$$

Plug in \bar{h} and $\bar{\varepsilon}^2$ from Lemma 3, use Lemma 1:

$$\begin{aligned}
& \mathbb{E}h^{(i)} - \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N - N_1}{N}\mathbb{E}h^{(2)} + O(\lambda^t) + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1} \\
&= \mathbb{E}_{(i)}\hat{\alpha}\mathbb{E}h^{(i)} + O(\lambda^{t-1}) \\
&\quad - \mathbb{E}_{(i)}\left[\hat{\alpha}\left(\frac{N_1}{N}\mathbb{E}h^{(1)} + \frac{N - N_1}{N}\mathbb{E}h^{(2)} + \frac{C_1}{N}\sum_{t=1}^{N_1}\lambda^t\right.\right. \\
&\quad \left.\left.+ \frac{C_2}{N}\sum_{t=N_1+1}^N\lambda^{t-N_1} + \frac{1}{N}\sum_{t=1}^{N_1}y_{t-1} + \frac{1}{N}\sum_{t=N_1+1}^Ny_{t-1}\right)\right] \\
&\quad + \mathbb{E}_{(i)}\hat{\beta}\mathbb{E}h^{(i)} + O(\lambda^{t-1}) \\
&\quad - \mathbb{E}_{(i)}\left[\hat{\beta}\left(\frac{N_1}{N}\mathbb{E}h^{(1)} + \frac{N - N_1}{N}\mathbb{E}h^{(2)} + \frac{C_1}{N}\sum_{t=1}^{N_1}\lambda^t\right.\right. \\
&\quad \left.\left.+ \frac{C_2}{N}\sum_{t=N_1+1}^N\lambda^{t-N_1} + \frac{1}{N}\sum_{t=1}^{N_1}x_{t-1} + \frac{1}{N}\sum_{t=N_1+1}^Nx_{t-1}\right)\right] + o(1)_N, \quad (5.3.19)
\end{aligned}$$

where

$$C_1 = \left| h_0 - \frac{\omega_1}{1 - \lambda} \right|, \quad c_2 = \left| h_{N_1} - \frac{\omega_2}{1 - \lambda} \right|.$$

Except for the sums of the x_t and y_t , all the terms in the expressions in parentheses (5.3.19) are deterministic. From Assumption 4 I also have, for example,

$$\text{cov}(\hat{\alpha}, \varepsilon_{t-1}^2) = \text{cov}(\hat{\alpha}, \mathbb{E}_{(i)}h_{t-1}) + \text{cov}(\hat{\alpha}, y_{t-1}) = o(1)_N,$$

and therefore

$$\begin{aligned}
\frac{1}{N}\mathbb{E}_{(i)}(\hat{\alpha}\sum_{t=1}^{N_1}y_{t-1}) &= \frac{1}{N}\mathbb{E}_{(i)}\hat{\alpha}\mathbb{E}_{(i)}\sum_{t=1}^{N_1}y_{t-1} + \frac{1}{N}\text{cov}(\hat{\alpha}, \sum_{t=1}^{N_1}y_{t-1}) \\
&= o(1)_N + \frac{1}{N}\sum_{t=1}^{N_1}o(1)_N = o(1)_N.
\end{aligned}$$

and analogously for

$$\frac{1}{N}\mathbb{E}_{(i)}(\hat{\alpha}\sum_{t=N_1+1}^Ny_{t-1}), \quad \frac{1}{N}\mathbb{E}_{(i)}(\hat{\beta}\sum_{t=1}^{N_1}x_{t-1}), \quad \frac{1}{N}\mathbb{E}_{(i)}(\hat{\beta}\sum_{t=N_1+1}^Nx_{t-1}).$$

Plugging into (5.3.19) and arranging terms, I arrive at

$$\begin{aligned}
& \mathbb{E}h^{(i)} - \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N - N_1}{N}\mathbb{E}h^{(2)} \\
&= \mathbb{E}_{(i)}\left(\hat{\alpha} + \hat{\beta}\right)\left[\mathbb{E}h^{(i)} - \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N - N_1}{N}\mathbb{E}h^{(2)}\right] \\
&\quad + O(\lambda^{t-1}) + O(1/N) + o(1)_N + o(1)_{N_1} + o(1)_{N-N_1}
\end{aligned} \quad (5.3.20)$$

That is, with growing N_1 and growing $N - N_1$, the expected value of the sum of the estimators of the autoregressive parameters conditional on the start value in each segment i must fulfil

$$\mathbb{E}_{(i)}(\hat{\alpha} + \hat{\beta}) = 1$$

up to vanishing terms, in order to satisfy condition (5.3.20). Observe that the difference

$$\mathbb{E}h^{(i)} - \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N - N_1}{N}\mathbb{E}h^{(2)} \neq 0$$

in both segments i if $\omega_1 \neq \omega_2$. Thus, condition (5.3.20) is not trivial. \square

c) Simulations

I will explore the almost-integration effect in GARCH(1,1) models in numerical simulations. First, I will investigate the case of a single switch in ω , as treated in Sections 5.3.a) and 5.3.b). There, the existence of two (or more) different means of volatility through which a single estimation hyperplane is laid is identified as the cause of almost-integration. From equation (5.2.4) I therefore expect changes in α or β to have a similar effect. Hence, I will consider global GARCH(1,1) estimates of synthetic series constructed in three segments of length 1400 in four mean reversion scenarios:

Table 5.1: GARCH(1,1) segment parameters of artificial series.

segment	1	2	3	average
length	1400	1400	1400	
Scenario 1.				
ω	1e-5	1e-5	2.5e-5	1.5e-5
α	0.10	0.10	0.10	0.10
β	0.75	0.75	0.75	0.75
Scenario 2.				
ω	1e-5	1e-5	1e-5	1e-5
α	0.10	0.10	0.10	0.10
β	0.65	0.80	0.65	0.70
Scenario 3.				
ω	1e-5	1e-5	1e-5	1e-5
α	0.65	0.80	0.65	0.70
β	0.10	0.10	0.10	0.10
Scenario 4.				
ω	1e-5	2e-5	1e-5	1.3e-5
α	0.10	0.05	0.10	0.083
β	0.65	0.80	0.65	0.70

1. ω switches at the point 2800 from 1e-5 to 2.5e-5 while $\alpha = 0.10$ and $\beta = 0.75$ are constant. In terms of annualized standard deviations this is a jump from 13 to 20 per cent volatility.

2. β switches from 0.65 to 0.80 and back, $\alpha \leq \beta$ holds. $\omega = 1\text{e-}5$ and $\alpha = 0.10$ are constant. This corresponds to changes between 10 and 16 per cent volatility.
3. α switches from 0.65 to 0.80 and back, $\alpha \geq \beta$ holds. $\omega = 1\text{e-}5$ and $\beta = 0.10$ are constant.
4. All parameters change: ω from $1\text{e-}5$ to $2\text{e-}5$ and back, α from 0.10 to 0.05 and back, and β from 0.65 to 0.80 and back. The annualized volatility switches between 10 and 18 per cent.

Table 5.1 shows the specification of the parameters in the four scenarios over the three segments.

I generated 10,000 series for each scenario. On each series I estimated a global Gaussian GARCH(1,1) model with constant mean return using maximum likelihood as described in Section 5.2.c). Figure 5.4 shows the histograms of the estimations of ω , α , β , and λ for each scenario. Table 5.2 presents the moments statistics.

Table 5.2: Moments statistics of the estimates of ω , α , β , and λ from the GARCH(1,1) estimation of 10,000 artificial series for every scenario according to Table 5.1.

	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
	Scenario 1.				Scenario 2.			
mean	3e-6	0.0794	0.8926	0.9720	2e-6	0.0765	0.8925	0.9690
std.dev.	1e-6	0.0208	0.0325	0.0123	9e-7	0.0223	0.0366	0.0149
skewness	0.5005	-0.2548	0.0299	-0.5020	0.4957	-0.1554	-0.0559	-0.5195
kurtosis	3.4224	3.1788	3.0779	3.4609	3.2266	2.8505	2.8331	3.3147
	Scenario 3.				Scenario 4.			
mean	1e-5	0.6844	0.1022	0.7867	4e-7	0.0402	0.9545	0.9947
std.dev.	6e-7	0.0510	0.0204	0.0488	2e-7	0.0121	0.0149	0.0029
skewness	0.0357	0.9084	0.1094	0.9214	1.5536	0.9214	-1.0281	-1.5523
kurtosis	2.9608	4.5097	3.0380	4.5514	6.3984	3.8801	4.2242	6.4092

The global estimations in Scenarios 1, 2, and 4 differ widely from the average of the parameters in every scenario. Here, λ is estimated close to one, regardless of the different data generating parameters. In Scenario 3 where α and β take the values of Scenario 2 in reverse order, the effect is not observed. In Scenarios 1, 2, and 4, the global estimate of β is close to or above 0.90 and that of α is less than 0.10.

In Scenarios 1, 2, and 4 the estimation of ω was much lower than the segment's average. Intuitively, this is not very surprising as the high $\hat{\lambda}$ takes much of the variation of the series. The estimator $\hat{\sigma}^2$ of $\mathbb{E}\varepsilon_t^2$ is given by the series (and $\hat{\mu}$) and

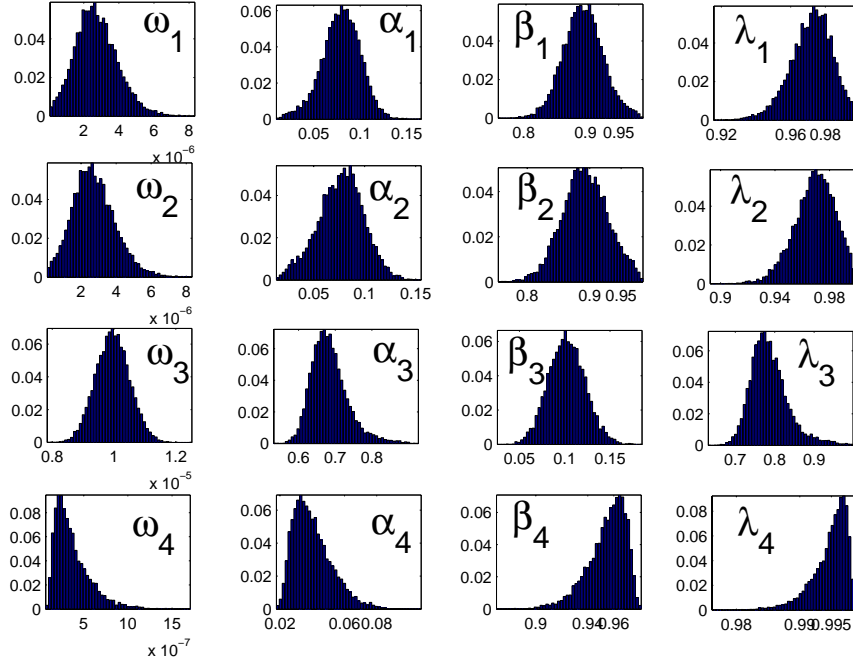


Figure 5.4: Histograms of the GARCH(1,1) estimations of ω , α , β , and λ of 10,000 artificial series for each scenario, constructed according to Table 5.1. The subscripts denote the scenarios.

as $\mathbb{E}\varepsilon_t^2 = \mathbb{E}h_t = \omega/(1 - \lambda)$ in the estimated model, a high estimation of λ must be compensated by a low estimation of ω .

The conclusion from the simulations is that the almost-integration effect can be reproduced easily with switches in ω and β . It is not reproduced with switches in α . Also, the simulations show that the overestimation of λ must be compensated by an underestimation of ω .

4 Estimation of the Short Scale in Stock Volatility

The GARCH model implies correlation structures for the series ε_t^2 :

$$\mathbb{E}\varepsilon_t^2\varepsilon_s^2 = \mathbb{E}\eta_t^2\eta_s^2h_th_s = \mathbb{E}h_th_s, \quad \eta_t \sim \mathcal{N}(0, 1),$$

and for the residual $\nu_t = \varepsilon_t^2 - h_t$:

$$\mathbb{E}\nu_t\nu_s = \mathbb{E}(\eta_t^2 - 1)(\eta_s^2 - 1)h_th_s = 0.$$

I will extract the long time scale by estimating GARCH(1,1) with constant mean return on the $\hat{\varepsilon}_t(\hat{\mu})$ -series, thereby obtaining the $\hat{h}_t(\hat{\omega}, \hat{\alpha}, \hat{\beta})$ -series, and calculate

the residual $\hat{\nu}_t = \hat{\varepsilon}_t^2 - \hat{h}_t$. If there is a second time scale in the $\hat{\varepsilon}_t^2$ apart from the long scale in the \hat{h}_t , it will be visible in the $\hat{\nu}_t$'s.

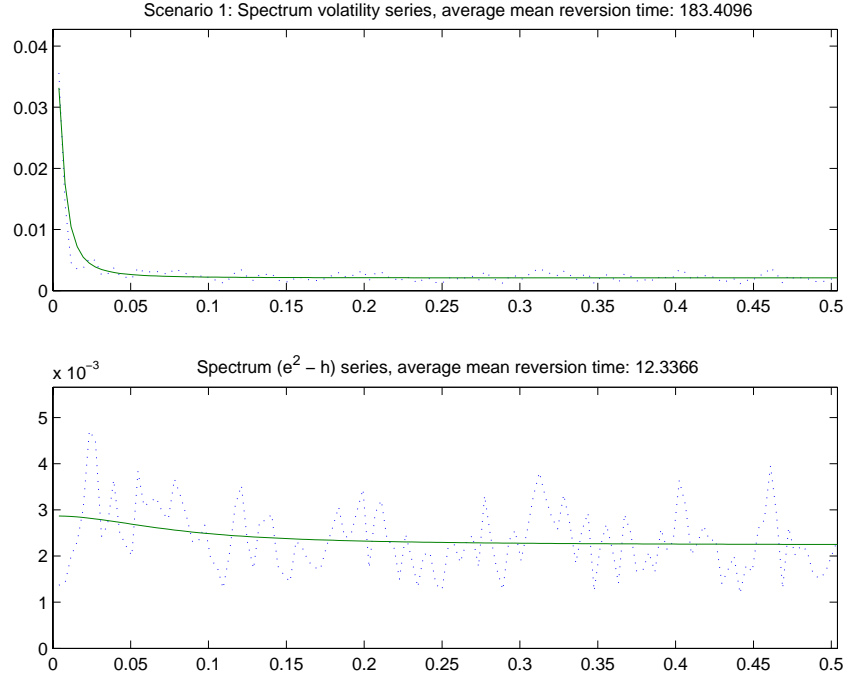


Figure 5.5: Upper graph: Estimation of the power spectrum (dotted line) of the volatility process $\hat{\varepsilon}_t^2$ constructed according to Scenario 1 from Section 5.3.c) and nonlinear least squares fit of a Lorentzian spectrum (solid line). The estimate of the average mean reversion time is computed as $1/c$ from the Lorentzian. Lower graph: same analysis for the residuals $\hat{\nu}_t = \hat{\varepsilon}_t^2 - \hat{h}_t$ after a GARCH(1,1) estimation of the series above. Here we see that the GARCH(1,1) estimation indeed “peels off” the long time scale and the short time scale in the residuals $\hat{\nu}$ is revealed.

To estimate the short scale I use the averaged periodogram, which is model-independent. The averaged periodogram is estimated by subsampling with a Tukey-Hanning window of 256 points length allowing for 64 points overlap. A Lorentzian spectrum model

$$h(w) = a + b/(c^2 + w^2), \quad (5.4.21)$$

is fitted to the periodogram. w denotes the frequencies and (a, b, c) are parameters. The average mean reversion time is estimated by $1/c$. The parameterization of the Lorentzian is motivated in the Appendix.

For the series ε_t^2 and ν_t there is no explicit parameterization of the Lorentzian (5.4.21) in terms of the parameters of the discrete GARCH(1,1) model available. As I will estimate the mean reversion time from the Lorentzian for both real and synthetic data, I will establish the correspondence between the mean reversion

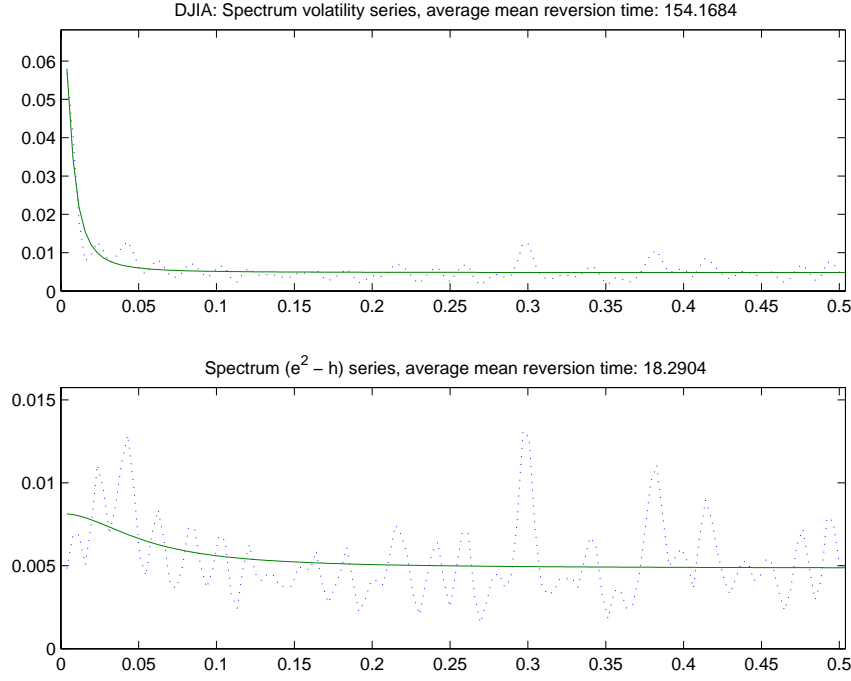


Figure 5.6: Upper graph: Estimation of the power spectrum (dotted line) of the volatility process $\hat{\varepsilon}_t^2$ of the Dow Jones series and nonlinear least squares fit of a Lorentzian spectrum (solid line). The estimate of the average mean reversion time is computed as $1/c$ from the Lorentzian. Lower graph: same analysis for the residuals $\hat{\nu}_t = \hat{\varepsilon}_t - \hat{h}_t$ after a GARCH(1,1) estimation of the series above. Clearly two distinct time scales can be observed in the Dow Jones series, a slower one of the magnitude of about 154 (51) days and a faster one of the magnitude of 18 (6) days. (The numbers in brackets are the time scales according to equation (5.4.22).)

time $1/c$ from the Lorentzian spectral model and $1/(1 - \lambda)$ from the GARCH model heuristically. I generated 10,000 synthetic GARCH(1,1) series of 5000 points length ranging from $\lambda = 0.75$ to $\lambda = 0.99$ and estimated $1/c$ by a nonlinear least squares fit of the Lorentzian to the estimate of the power spectrum. (In particular, I set $\alpha \equiv 0.10$ and let β go through the interval $[0.65, 0.89]$ while $\omega \equiv 1e-5$.) The relation from a linear regression of the resulting series of $\{1/\hat{c}_i\}_i$ on the logs of $\{1/(1 - \lambda_i)\}_i$ was obtained as

$$\frac{1}{\hat{c}} = -86.74 + 61.20 \log \left(\frac{1}{1 - \lambda} \right), \quad R^2 = 0.93. \quad (5.4.22)$$

Note that the $\{1/(1 - \lambda_i)\}_i$ series was known from the construction of the artificial series.

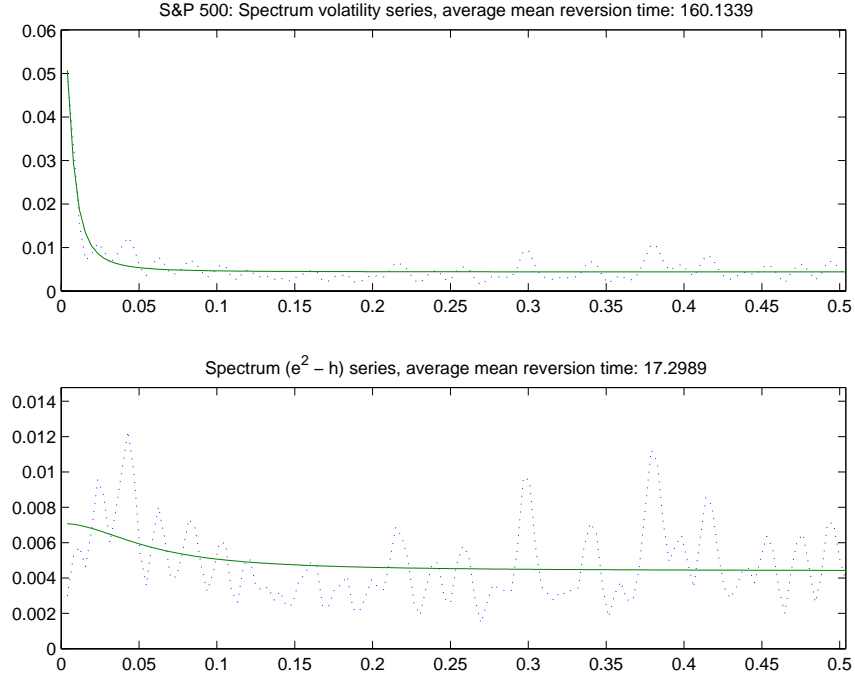


Figure 5.7: Upper graph: Estimation of the power spectrum (dotted line) of the volatility process $\hat{\varepsilon}_t^2$ of the S&P500 series and nonlinear least squares fit of a Lorentzian spectrum (solid line). The estimate of the average mean reversion time is computed as $1/c$ from the Lorentzian. Lower graph: same analysis for the residuals $\hat{\nu}_t = \hat{\varepsilon}_t - \hat{h}_t$ after a GARCH(1,1) estimation of the series above. The two distinct time scales compare to those of the Dow Jones series in Figure 5.6.

a) Synthetic Data

I estimate the power spectra for series built according to Scenario 1 in Section 5.3.c). Here, a time scale of 7 days is superposed by a time scale of the complete length of the series as there is a single switch in ω . The power spectra of the residuals exhibit a short time scale as shown in Figure 5.5 for a typical realization. The GARCH(1,1) maximum likelihood estimates of this realization were $(\hat{\omega}, \hat{\alpha}, \hat{\beta}) = (8e-7, 0.0384, 0.9556)$. This is a long time scale of $1/(1 - \hat{\lambda}) = 167$ days according to the parameter estimates, 183 days according to the Lorentzian model and $1/(1 - \lambda) = 82$ days according to (5.4.22). A maximum likelihood estimation of a GARCH(1,1) model for the residuals $\hat{\nu}_t$ could not detect the second, short time scale but essentially repeated the estimates of the GARCH(1,1) model for the volatility series $\hat{\varepsilon}_t^2$.

The power spectrum estimation reveals a short scale in the residuals that is of the magnitude of $1/\hat{c} = 12$ days or $1/(1 - \lambda) = 5$ days according to (5.4.22). This compares to the data-generating short scale of $1/(1 - 0.85) \approx 7$ days.

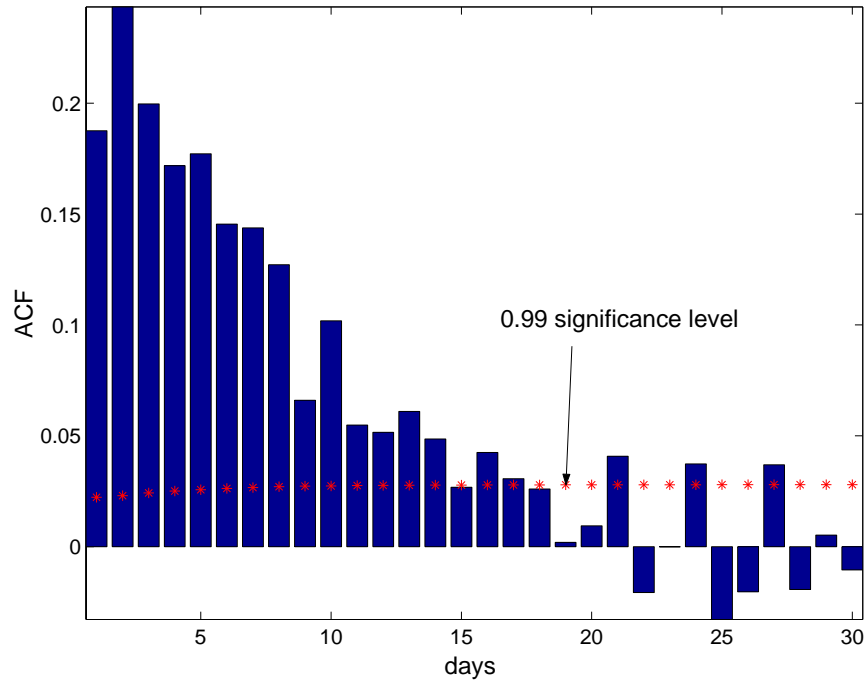


Figure 5.8: Sample autocorrelation function of the series $\hat{\nu}_t = \hat{\varepsilon}_t^2 - \hat{h}_t$ of the Dow Jones series. Up to the lag of 14 days, the estimates are clearly significant and the median lag is 6 days. The sample autocorrelation function of the $\hat{\nu}_t$ series of the S&P500 looks essentially the same.

b) Market Data

Figure 5.6 shows the spectra of the volatility series $\hat{\varepsilon}_t^2$ of the Dow Jones series (above) and of the residual $\hat{\nu}_t$ (below). Two distinct time scales can be observed, a longer one of about 154 days or $1/(1 - \lambda) \approx 51$ days according to (5.4.22), and a faster scale of about 18 days or $1/(1 - \lambda) \approx 6$ days. Figure 5.7 shows the spectra of the volatility (above) and the residual (below) of the S&P500 series. Again, two time scales can be observed and their magnitudes compare closely to those of the Dow Jones.

For inference statistics, I turn to the estimation of the autocorrelation function, which is equivalent to the estimation of the power spectrum of stationary processes by the Wiener-Khinchine theorem (e.g., Priestley 1981). The results for the Dow Jones and for the S&P500 series look essentially the same, so that I will report only the estimation for the Dow Jones in Figure 5.8.

An alternative way to arrive at the short scale is to solve the changepoint detection problem for (G)ARCH models, as discussed in Kokoszka and Leipus (1999) and (2000) or Andreou and Ghysels (2001). Local GARCH estimations on the obtained segmentation may capture the short run dynamics. However, it suffices to miss a single changepoint and the estimation will not reveal the short

scale. Periodogram estimation is more robust as it does not have to find the changepoints.

I conclude that by estimating GARCH(1,1) and computing the residual $\hat{\nu}_t = \hat{\varepsilon}_t^2 - \hat{h}_t$ using the estimated GARCH parameters, the long time scale can be eliminated from the data. Spectral analysis is capable of measuring the short time scale left in the residual.

5 Summary and Conclusion

Changes in GARCH(1,1) parameters that are not accounted for in global estimations lead to an estimated persistence that is much higher than the average persistence within the regimes. I show that for switches in the constant ω of the conditional variance equation of the GARCH(1,1) model, the sum of the estimated autoregressive parameters $\hat{\lambda} = \hat{\alpha} + \hat{\beta}$ must be close to one. In simulations I obtain global estimates close to integration for parameter changes in ω and β within realistic ranges for stock-price volatility. It is not necessary to have a certain underlying stochastic structure that drives the changes but a single deterministic changepoint is sufficient.

The switches induce different volatility means in each segment that the global GARCH(1,1) estimation has to capture. Thereby, the long time scale of the parameter switches dominates the parameter estimation and masks the short correlation structure that governs the process within regimes. For the daily volatility of the Dow Jones and for the S&P500 for the sample Jan 2, 1985 to Jan 2, 2001, I obtain an induced mean reversion time of 73 days and of 114 days, respectively.

The short time scale within regimes can be uncovered in the GARCH(1,1) residual $\hat{\nu}_t = \hat{\varepsilon}_t^2 - \hat{h}_t$, where \hat{h}_t is the estimated conditional volatility. By periodogram estimation of synthetic data I recover the short time scale that I inserted in the data. Applying this method to the daily Dow Jones and S&P500 volatility, I find a short time scale of the magnitude of 5 to 10 days. I conclude that at least two overlaying time scales are present in the considered series.

In summary, global GARCH(1,1) estimations that do not take parameter changes in ω and β into account will only capture the time scale of these parameter changes. To obtain the short-run, proper GARCH dynamics of the process without having to find the changepoints, I propose periodogram estimation of the residual $\hat{\nu}_t = \hat{\varepsilon}_t^2 - \hat{h}_t$.

In general, the findings confirm the existence of mean reversion in volatility and show that it undergoes different regimes. The long memory that was ubiquitously found in volatility is most likely caused by the regime changes. Therefore, the question is no longer what determines the long memory of volatility, but

what causes the regime changes, and what determines the short memory within regimes.

The methods presented here cannot identify economic causes of regime switches, as discussed in Chapter 4. I will return to that question in Chapter 7, where the influence of foreign exchange interventions by the Japanese central bank on the mean reversion regime of the yen/dollar exchange rate is investigated. Also, the restriction to the GARCH(1,1) model may be critical. This is examined in the following chapter.

Chapter 6

Generalization to GARCH(p,q)

1 Multiple Scales and Higher Order ARMA and GARCH

It is a fair conjecture that multiple time scales, or multiple correlation structures, might be captured by higher order GARCH models.

This idea is motivated by the fact that for ARMA models, this problem has a very nice analysis. When two ARMA(1,1) models (driven by the same white noise) are aggregated, the result is an ARMA(2,2) model. The two different correlation structures implied by the two ARMA(1,1) processes can be found by solving for the roots of the characteristic equation of the ARMA(2,2) model.

As GARCH models have a well known ARMA representation, it is natural to assume that it will work out the same in this case. Unfortunately, it turns out however that the underlying GARCH(1,1) models cannot easily be identified by solving for the roots of the lag polynomial.

In the following sections, I will show that different scales can be recovered in ARMA aggregations. For this purpose, it will be shown that the concept of the mean reversion time presented in Section 5.2.b) also applies to the ARMA(1,1) case. I will compare this to GARCH aggregations. In particular, I will discuss the roots of the GARCH lag polynomial. There are a number of differences between ARMA and GARCH that are helpful to keep in mind.

I will generalize the proof that unknown changes in the parameter regime lead to almost-integration to models of orders up to GARCH(2,2). This will be illustrated by simulations in different mean reversion environments.

2 Measuring Mean Reversion in ARMA(1,1)

Consider an ARMA(1,1) model driven by a white noise process η_t , for example $\eta_t \sim \mathcal{N}(0, \sigma)$ (Box and Jenkins 1976).

$$x_t + \phi x_{t-1} = c + \eta_t + \theta \eta_{t-1}, \quad (6.2.1)$$

where $\phi \in (-1, 1)$, $c, \theta \in \mathbb{R}$, $x_0 \in \mathbb{R}$.

ARMA models are pure reduced form models. A priori, they have no economic interpretation whatsoever. They represent, however, a very general class of time series processes. The Wold Decomposition Theorem (for example, Priestley 1981, p. 755), states that *every* stationary process can be represented as an ARMA(∞, ∞) process. In other words, any *structural* explanatory model that has an immediate economic interpretation also has an ARMA representation, provided that the model is stationary. This ARMA representation then has no economic interpretation but the data it generates have exactly the same properties as those of the structural model. As ARMA models are much easier to estimate than many structural models, they are frequently used when the objective is not to uncover correlations between economic magnitudes but when a good fit to the data and an accurate forecast are sufficient. ARMA models can have an immediate interpretation when the process under investigation is structurally self-referring, as in population dynamics, for example.

The expected value of model (6.2.1) is given by

$$\mathbb{E}x = \frac{c}{1 + \phi}. \quad (6.2.2)$$

To find a measure of the mean reversion time, we can proceed exactly as in Section 5.2.b). Take expectations of the process at time $t + k$ with respect to the information at time t , $k \geq 2$.

$$\mathbb{E}_t x_{t+k} + \phi \mathbb{E}_t x_{t+k-1} = c. \quad (6.2.3)$$

Denote by ξ_{t+k} the distance of this conditional expectation from the unconditional mean $\mathbb{E}x$:

$$\mathbb{E}_t x_{t+k} = \mathbb{E}x + \xi_{t+k}. \quad (6.2.4)$$

From (6.2.3) and (6.2.4), we have

$$\mathbb{E}x + \xi_{t+k} + \phi(\mathbb{E}x + \xi_{t+k-1}) = c,$$

and as $\mathbb{E}x = c - \phi \mathbb{E}x$ from (6.2.2),

$$\xi_{t+k} = -\phi \xi_{t+k-1}.$$

Same as in Section 5.2.b), in the continuous limit we have

$$d\xi_t = -\phi\xi_t dt$$

and thus

$$\xi_t = \xi_0 e^{-\phi t}$$

such that the e -folding time

$$\left(t_e : \frac{\xi_{t_e}}{\xi_0} = e^{-1}\right)$$

is given by

$$t_e = \frac{1}{1 + \phi},$$

which I will use as a measure of the mean reversion time.

3 Aggregation of ARMA(1,1) Models

Consider two ARMA(1,1) processes driven by the same white noise process η_t but with different parameters:

$$\begin{aligned} x_t + \phi_1 x_{t-1} &= c_1 + \eta_t + \theta_1 \eta_{t-1}, \\ y_t + \phi_2 y_{t-1} &= c_2 + \eta_t + \theta_2 \eta_{t-1} \end{aligned} \quad (6.3.5)$$

and define the aggregate

$$z_t = x_t + y_t.$$

Then, z_t clearly contains two different time scales, $1/(1 + \phi_1)$ and $1/(1 + \phi_2)$ from the two ARMA(1,1) models.

The solutions to (6.3.5) are given by

$$\begin{aligned} x_t &= \frac{1 + \theta_1 L}{1 + \phi_1 L} \eta_t + \frac{1}{1 + \phi_1} c_1, \\ y_t &= \frac{1 + \theta_2 L}{1 + \phi_2 L} \eta_t + \frac{1}{1 + \phi_2} c_2, \end{aligned}$$

where L is the lag operator. Thus,

$$\begin{aligned} z_t &= \left(\frac{1 + \theta_1 L}{1 + \phi_1 L} + \frac{1 + \theta_2 L}{1 + \phi_2 L} \right) \eta_t + \tilde{c} \\ &= \frac{(1 + \theta_1 L)(1 + \phi_2 L) + (1 + \theta_2 L)(1 + \phi_1 L)}{(1 + \phi_1 L)(1 + \phi_2 L)} \eta_t + \tilde{c}, \end{aligned} \quad (6.3.6)$$

where $\tilde{c} = [(c_1(1 + \phi_2) + c_2(1 + \phi_1))/((1 + \phi_1)(1 + \phi_2))]$. The aggregate z_t therefore has the representation

$$(1 + \phi_1 L)(1 + \phi_2 L)z_t = (1 + \theta_1 L)(1 + \phi_2 L)\eta_t + (1 + \theta_2 L)(1 + \phi_1 L)\eta_t + c_1(1 + \phi_2) + c_2(1 + \phi_1),$$

or

$$z_t + (\phi_1 + \phi_2)z_{t-1} + \phi_1\phi_2z_{t-2} = 2\eta_t + (\theta_1 + \theta_2 + \phi_1 + \phi_2)\eta_{t-1} + (\theta_1\phi_2 + \theta_2\phi_1)\eta_{t-2} + c_1(1 + \phi_2) + c_2(1 + \phi_1),$$

which is an ARMA(2,2) model. The roots of the characteristic equation, or lag polynomial, of z_t are given by

$$(1 + \phi_1 L)(1 + \phi_2 L) = L^2 + \frac{\phi_1}{\phi_2}L + \frac{1}{\phi_2} = 0.$$

As the polynomial is given in its decomposition into linear factors, it can be seen immediately that the roots are given by

$$L_1 = -\frac{1}{\phi_1} \text{ and } L_2 = -\frac{1}{\phi_2}.$$

The time scales $1/(1 + \phi_1)$ and $1/(1 + \phi_2)$ can thus easily be obtained by estimating an ARMA(2,2) model

$$z_t + \phi'_1 z_{t-1} + \phi'_2 z_{t-2} = c' + \eta_t + \theta'_1 \eta_{t-1} + \theta'_2 \eta_{t-2}$$

and calculating the roots of the characteristic polynomial

$$1 + \hat{\phi}'_1 L + \hat{\phi}'_2 L^2 = 0.$$

The argument readily generalizes to the aggregation of n ARMA(1,1) models. The n roots of the lag polynomial of the resulting ARMA(n, n) process are the $-1/\phi_i$, $i = 1, \dots, n$. Let $\Phi(L)$ be the root polynomial corresponding to the ARMA(k, k) model that resulted from the aggregation of k ARMA(1,1) models. Assume that the k roots are the $-1/\phi_i$, $i = 1, \dots, k$. From (6.3.6) it is clear that by adding another ARMA(1,1) component to the model, the lag polynomial of the resulting ARMA($k+1, k+1$) model is $\Phi(L)(1 + \phi_{k+1}L)$ which adds the root $-1/\phi_{k+1}$ to the k assumed to be known. Thus the assertion holds for $k + 1$.

4 Aggregation of GARCH(1,1) Models

Consider the conditional variance equation of the GARCH(1,1) model

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}.$$

Add $\varepsilon_t^2 - \beta\varepsilon_{t-1}^2$ on both sides:

$$\varepsilon_t^2 - \alpha\varepsilon_{t-1}^2 - \beta\varepsilon_{t-1}^2 = \omega + \varepsilon_t^2 - h_t - \beta(\varepsilon_{t-1}^2 - h_{t-1}).$$

Define $\nu_t := \varepsilon_t^2 - h_t$, just as in Section 5.4, where it is also shown that ν_t is white noise. Then, the GARCH(1,1) model has the representation

$$(1 - \alpha L - \beta L)\varepsilon_t^2 = \omega + (1 - \beta L)\nu_t \quad (6.4.7)$$

which is an ARMA(1,1) model. Notice however that in order to obtain the ARMA representation, ε_t^2 is added on both sides. Therefore, the ARMA representation puts no constraints on ε_t^2 , it does not define it. The only constraint, $\mathbb{E}_t \varepsilon_t^2 = h_t$, comes from the distribution assumption (5.2.2).

If two GARCH(1,1) processes are to be aggregated, we cannot just add two models of type (6.4.7). Say we add two processes $\varepsilon_{1,t}^2$ and $\varepsilon_{2,t}^2$. Then, we will also have two different ν 's, and the analysis of Section 6.3 will not apply. Even if we neglect this difference, assume $\nu_{1,t} \approx \nu_{2,t} \approx \nu_t$, and proceed as in Section 6.3, the aggregate is a genuine ARMA(2,2) model, that is, it does place constraints on ε_t^2 . It has the form

$$\begin{aligned} h_t &= \frac{\omega_1}{2}(1 - \alpha_2 - \beta_2) + \frac{\omega_2}{2}(1 - \alpha_1 - \beta_1) \\ &\quad + \frac{1}{2}\varepsilon_t^2 - \frac{\beta_1 + \beta_2}{2}\varepsilon_{t-1}^2 + \frac{\beta_1\beta_2 - \alpha_1\alpha_2}{2}\varepsilon_{t-2}^2 \\ &\quad + \frac{\alpha_1 + \alpha_2 + 2(\beta_1 + \beta_2)}{2}h_{t-1} + \frac{\alpha_1\beta_2 + 2\beta_1\beta_2 + \alpha_2\beta_1}{2}h_{t-2}. \end{aligned}$$

That is, the contemporaneous term ε_t^2 appears in the variance process, which is thus not conditional on \mathcal{F}_{t-1} anymore. This is not a GARCH model.

Another natural way to aggregate two GARCH(1,1) processes is by formulating the GARCH model as

$$\varepsilon_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, h_t),$$

$$h_t = h_t^{(1)} + h_t^{(2)},$$

$$h_t^{(1)} = \omega_1 + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}^{(1)},$$

$$h_t^{(2)} = \omega_2 + \alpha_2\varepsilon_{t-1}^2 + \beta_2 h_{t-1}^{(2)}.$$

Then,

$$h_t^{(1)} = \frac{\omega_1}{1 - \beta_1} + \frac{\alpha_1}{1 - \beta_1 L} \varepsilon_{t-1}^2,$$

$$h_t^{(2)} = \frac{\omega_2}{1 - \beta_2} + \frac{\alpha_2}{1 - \beta_2 L} \varepsilon_{t-1}^2,$$

such that

$$\begin{aligned} h_t &= \frac{\omega_1}{1-\beta_1} + \frac{\omega_2}{1-\beta_2} + \left(\frac{\alpha_1}{1-\beta_1 L} + \frac{\alpha_2}{1-\beta_2 L} \right) \varepsilon_{t-1}^2, \\ &= \frac{\omega_1(1-\beta_2) + \omega_2(1-\beta_1)}{(1-\beta_1)(1-\beta_2)} + \frac{\alpha_1(1-\beta_2 L) + \alpha_2(1-\beta_1 L)}{(1-\beta_1 L)(1-\beta_2 L)} \varepsilon_{t-1}^2. \end{aligned}$$

Therefore, the variance process has the representation

$$\begin{aligned} (1-\beta_1 L)(1-\beta_2 L)h_t &= w_1(1-\beta_2) + w_2(1-\beta_1) \\ &\quad + \alpha_1(1-\beta_2 L)\varepsilon_{t-1}^2 + \alpha_2(1-\beta_1 L)\varepsilon_{t-1}^2, \end{aligned}$$

or

$$\begin{aligned} h_t &= \tilde{\omega} + \alpha_1 \alpha_2 \varepsilon_{t-1}^2 - (\alpha_1 \beta_2 + \alpha_2 \beta_1) \varepsilon_{t-2}^2 \\ &\quad + (\beta_1 + \beta_2) h_{t-1} - \beta_1 \beta_2 h_{t-2}, \end{aligned}$$

where $\tilde{\omega} = \omega_1(1-\beta_2) + \omega_2(1-\beta_1)$.

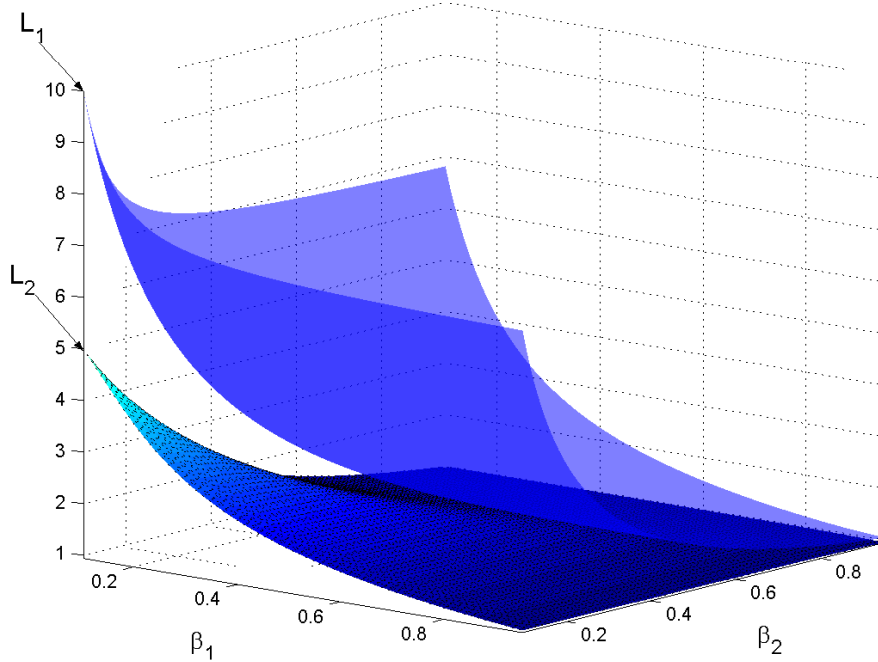


Figure 6.1: Plot of (6.4.9) when $\alpha_1 \equiv \alpha_2 \equiv 0.05$ and β_1 and β_2 vary between 0.10 and 0.95. In the region of interest, where $\beta_{1,2} \approx 0.85$ and $\beta_{2,1} \approx 0.94$ such that the time scales of the two GARCH(1,1) components are 10 and 100 days, the roots hardly discriminate between the persistence parameters.

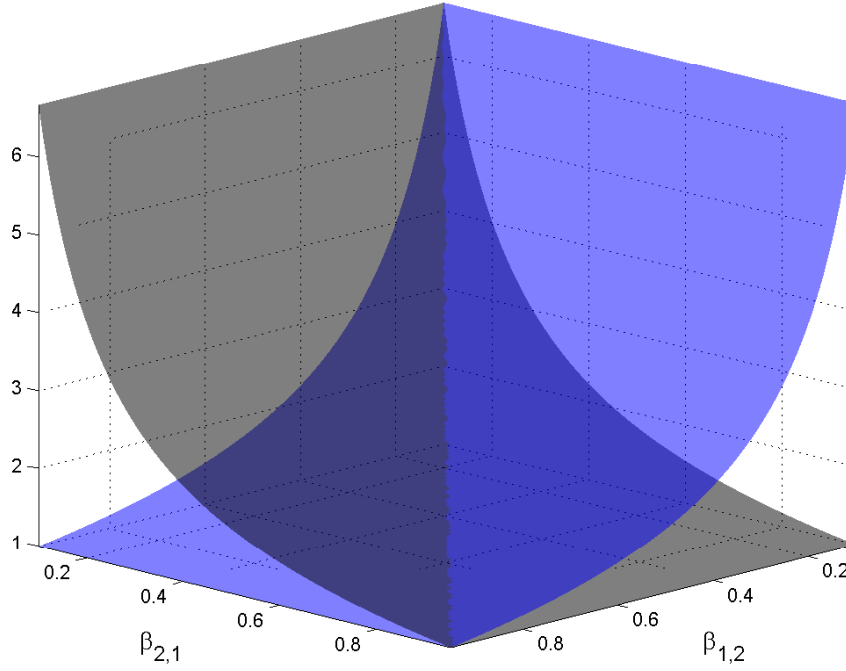


Figure 6.2: Plot of the roots $L_1 = 1/\phi_1$ and $L_2 = 1/\phi_2$ of the lag polynomial of an ARMA(2,2) process that was generated by aggregating two ARMA(1,1) processes. Unlike in the case of the GARCH aggregation, the two roots separate the two persistence parameters of the ARMA(1,1) components at all parameter values, except, of course, where the parameters are the same.

This is a venerable GARCH(2,2) process, but what are the roots of the characteristic equation? Write the GARCH(2,2) process as

$$h_t = \tilde{\omega} + a_1 \varepsilon_{t-1}^2 + a_2 \varepsilon_{t-2}^2 + b_1 h_{t-1} + b_2 h_{t-2},$$

where

$$\begin{aligned} a_1 &= \alpha_1 + \alpha_2, \\ a_2 &= -\alpha_1 \beta_2 - \alpha_2 \beta_1, \\ b_1 &= \beta_1 + \beta_2, \\ b_2 &= -\beta_1 \beta_2. \end{aligned}$$

Then, the characteristic equation is given by

$$1 - (a_1 + b_1)L - (a_2 + b_2)L^2 = 0, \quad (6.4.8)$$

as can be seen when the process is written in ARMA form

$$\begin{aligned} \varepsilon_t^2 - (a_1 + b_1)\varepsilon_{t-1}^2 - (a_2 + b_2)\varepsilon_{t-2}^2 \\ = \nu_t - b_1\nu_{t-1} - b_2\nu_{t-2}. \end{aligned}$$

As can be seen by substitution, the self-suggesting conjectures $1/(\alpha_1 + \beta_1)$ and $1/(\alpha_2 + \beta_2)$ are not the roots of (6.4.8). Instead, conventional solving leads to the quite cumbersome expression

$$L_{1,2} = \frac{1}{2} \left(\frac{\alpha_1 + \alpha_2 + \beta_1 + \beta_2}{\alpha_1\beta_2 + \alpha_2\beta_1 + \beta_1\beta_2} \pm \frac{\alpha_1 + \alpha_2 + \beta_1 + \beta_2 - 2\sqrt{\alpha_1\beta_2 + \alpha_2\beta_1 + \beta_1\beta_2}}{2(\alpha_1\beta_2 + \alpha_2\beta_1 + \beta_1\beta_2)} \right). \quad (6.4.9)$$

Figure 6.1 shows the L_1 and L_2 when $\alpha_1 = \alpha_2 = 0.05$ and β_1 and β_2 vary between 0.10 and 0.95. The question of interest is how sensitive the root surfaces are to different persistence parameters of the underlying GARCH(1,1) components. In the interesting region, where $\beta_{1,2} \approx 0.85$ and $\beta_{2,1} \approx 0.94$ such that the time scales of the two GARCH(1,1) components are 10 and 100 days, the root surfaces flatten out and show little sensitivity. This indicates that the power of GARCH(2,2) models to discriminate different time scales is quite poor.

Figure 6.2 shows the roots of the ARMA(2,2) process that was built from two ARMA(1,1) for comparison. These are simple plots of $L_1 = 1/\phi_1$ and $L_2 = 1/\phi_2$. In the region of interest, where $\phi_{1,2} \approx 0.90$ and $\phi_{2,1} \approx 0.99$, the model clearly discriminates the scales, the surfaces remain sensitive to the persistence parameters.

So far, I have considered *spatial* aggregation. In the following section, I will discuss the question whether higher order GARCH models can distinguish scales when the aggregation is *temporal*, that is, there are unknown changepoints in the data. It turns out that the almost-integration effect described in Chapter 5 fully generalizes to GARCH(2,2).

5 Unknown Parameter Regime Changes and Global GARCH(p, q) Modelling

We will consider the Gaussian GARCH(p, q) specifications with $p = 1, 2$ and $q = 1, 2$. The conditional variance equations are thus given by

$$h_t = \omega + a_1^* h_{t-1} + a_2^* h_{t-2}, \quad (6.5.10)$$

where

$$a_1^* = \alpha_1 \eta_{t-1}^2 + \beta_1, \\ a_2^* = \begin{cases} 0, & \text{for GARCH(1,1),} \\ \alpha_2 \eta_{t-2}^2, & \text{for GARCH(1,2),} \\ \beta_2, & \text{for GARCH(2,1),} \\ \alpha_2 \eta_{t-2}^2 + \beta_2, & \text{for GARCH(2,2).} \end{cases}$$

The persistence of the volatility process h_t is governed by the sum of the autoregressive parameters (Engle and Patton 2001, Nelson 1990),

$$\lambda := \sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i, \quad (6.5.11)$$

where $\lambda = 1$ is equivalent to integration of order one.

Denote by $\mathbb{E}_0 h_t = \mathbb{E}(h_t | \mathcal{F}_0)$ the expected value of h_t when the process $\{h_t\}$ is finite, $t \in \{1, \dots, N\}$ and $h_{-1}, h_0 \in \mathbb{R}$ being start values. Take expectations of (6.5.10):

$$\mathbb{E}_0 h_t = \omega + a_1 \mathbb{E}_0 h_{t-1} + a_2 \mathbb{E}_0 h_{t-2}, \quad (6.5.12)$$

where

$$a_1 = \alpha_1 + \beta_1, \quad a_2 = \begin{cases} 0, & \text{for GARCH}(1,1), \\ \alpha_2, & \text{for GARCH}(1,2), \\ \beta_2, & \text{for GARCH}(2,1), \\ \alpha_2 + \beta_2, & \text{for GARCH}(2,2), \end{cases} \quad (6.5.13)$$

as $\mathbb{E}_0 \eta_t^2 = 1 \forall t$ and the η_t are independent.

Lemma 6. *Consider the Gaussian GARCH(p, q) model with $p = 1, 2$ and $q = 1, 2$. Denote the start values of the volatility process $\{h_t\}$ by $h_0, h_{-1} \in \mathbb{R}$. Then, the expectation of h_t conditional on the start values is given by*

$$\begin{aligned} \mathbb{E}_0 h_t &= \frac{1}{2d} (\gamma_1^{t+1} (h_0 - \gamma_2 h_{-1}) - \gamma_2^{t+1} (\gamma_1 h_{-1} - h_0)) \\ &\quad + \frac{\omega}{2d} \left(\gamma_1 \frac{1 - \gamma_1^t}{1 - \gamma_1} - \gamma_2 \frac{1 - \gamma_2^t}{1 - \gamma_2} \right), \end{aligned} \quad (6.5.14)$$

where

$$\gamma_1 = \frac{a_1}{2} + \sqrt{\frac{a_1^2}{4} + a_2}, \quad \gamma_2 = \frac{a_1}{2} - \sqrt{\frac{a_1^2}{4} + a_2},$$

and $d = \sqrt{a_1^2/4 + a_2}$.

Proof. For simplicity, denote $y_t := \mathbb{E}_0 h_t$. Then, (6.5.12) reads

$$y_t = \omega + a_1 y_{t-1} + a_2 y_{t-2}.$$

This is a linear second-order difference equation with constant coefficients given by (6.5.13). Vectorize the process:

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \omega \\ 0 \end{bmatrix}, \quad (6.5.15)$$

and denote $Y_t := (y_t, y_{t-1})'$ and $\Omega := (\omega, 0)'$, and

$$A := \begin{bmatrix} a_1 & a_2 \\ 1 & 0 \end{bmatrix}.$$

The eigenvalue decomposition of A yields

$$A = S \Gamma S^{-1},$$

with

$$S = \begin{bmatrix} \gamma_1 & \gamma_2 \\ 1 & 1 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{bmatrix}, \quad S^{-1} = \frac{1}{2d} \begin{bmatrix} 1 & -\gamma_2 \\ -1 & \gamma_1 \end{bmatrix}.$$

Pre-multiply (6.5.15) by S^{-1} :

$$\begin{aligned} S^{-1} Y_t &= S^{-1} A Y_{t-1} + S^{-1} \Omega \\ &= \Gamma S^{-1} Y_{t-1} + S^{-1} \Omega \end{aligned}$$

and denote $U_t := S^{-1} Y_t$ and $\tilde{\Omega} = S^{-1} \Omega$, yielding

$$\begin{aligned} U_t &= \Gamma U_{t-1} + \tilde{\Omega}, \\ &= \Gamma^t U_0 + \sum_{i=0}^{t-1} \Gamma^i \tilde{\Omega}, \\ &= \begin{bmatrix} \gamma_1^t & 0 \\ 0 & \gamma_2^t \end{bmatrix} U_0 + \begin{bmatrix} \frac{1-\gamma_1^t}{1-\gamma_1} & 0 \\ 0 & \frac{1-\gamma_2^t}{1-\gamma_2} \end{bmatrix} \tilde{\Omega}. \end{aligned} \tag{6.5.16}$$

From the definition of $U_t = S^{-1} Y_t$, we have

$$U_t = \frac{1}{2d} \begin{bmatrix} 1 & -\gamma_2 \\ -1 & \gamma_1 \end{bmatrix} \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \frac{1}{2d} \begin{bmatrix} y_t - \gamma_2 y_{t-1} \\ -y_t + \gamma_1 y_{t-1} \end{bmatrix},$$

and hence, from (6.5.16),

$$U_t = \frac{1}{2d} \begin{bmatrix} \gamma_1^t (y_0 - \gamma_2 y_{-1}) \\ \gamma_2^t (-y_0 + \gamma_1 y_{-1}) \end{bmatrix} + \frac{\omega}{2d} \begin{bmatrix} \frac{1-\gamma_1^t}{1-\gamma_1} \\ -\frac{1-\gamma_2^t}{1-\gamma_2} \end{bmatrix}.$$

We are interested in the first entry in $Y_t = S U_t$:

$$Y_t = \frac{1}{2d} \begin{bmatrix} \gamma_1 & \gamma_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \gamma_1^t (y_0 - \gamma_2 y_{-1}) \\ \gamma_2^t (-y_0 + \gamma_1 y_{-1}) \end{bmatrix} + \frac{\omega}{2d} \begin{bmatrix} \gamma_1 & \gamma_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1-\gamma_1^t}{1-\gamma_1} \\ -\frac{1-\gamma_2^t}{1-\gamma_2} \end{bmatrix},$$

which is

$$y_t = \frac{1}{2d} (\gamma_1^{t+1}(y_0 - \gamma_2 y_{-1}) - \gamma_2^{t+1}(-y_0 + \gamma_1 y_{-1})) \\ + \frac{\omega}{2d} \left(\gamma_1 \frac{1 - \gamma_1^t}{1 - \gamma_1} - \gamma_2 \frac{1 - \gamma_2^t}{1 - \gamma_2} \right).$$

As $y_0 = \mathbb{E}_0 h_0 = h_0$ and $y_{-1} = \mathbb{E}_0 h_{-1} = h_{-1}$, the statement is proven. \square

From the stationarity assumption $a_1 + a_2 < 1$, we have that $|\gamma_{1,2}| < 1$. For $t \rightarrow \infty$, the expectation conditional on the start values converges to the expectation with respect to the stationary measure, and

$$\mathbb{E}h = \lim_{t \rightarrow \infty} y_t = \frac{\omega}{2d} \left(\frac{\gamma_1}{1 - \gamma_1} - \frac{\gamma_2}{1 - \gamma_2} \right) = \frac{\omega}{1 - a_1 - a_2}, \quad (6.5.17)$$

a well-known result for GARCH models.

Lemma 7. *In the Gaussian GARCH(p,q) model with $p = 1, 2$ and $q = 1, 2$, the relation*

$$\mathbb{E}_0 h_t = \mathbb{E}h + O(\gamma_1^{t+1}) + O(\gamma_2^{t+1})$$

holds for $t \in \{1, \dots, N\}$.

Proof. From Lemma 6 and equation (6.5.17), we have that

$$|\mathbb{E}_0 h_t - \mathbb{E}h| = \left| \frac{1}{2d} (\gamma_1^{t+1}(h_0 - \gamma_2 h_{-1}) - \gamma_2^{t+1}(\gamma_1 h_{-1} - h_0)) \right. \\ \left. + \frac{\omega}{2d} \left(\frac{\gamma_2^{t+1}}{1 - \gamma_2} - \frac{\gamma_1^{t+1}}{1 - \gamma_1} \right) \right| \\ = O(\gamma_1^{t+1}) + O(\gamma_2^{t+1}),$$

remembering that $|\gamma_{1,2}| < 1$ and $d = \sqrt{\frac{a_1^2}{4} + a_2}$ positive and bounded. \square

Assumption 8. *I will assume that the processes $\{h_t\}$ and $\{\varepsilon_t\}$ are observable without measurement error, or at least with a measurement error that is independent of the parameter estimates $(\hat{\mu}, \hat{\omega}, \hat{a}_1, \hat{a}_2)$.*

Again, this assumption is unrealistic. The process h_t is not observable and in real estimation problems h_t is estimated by $\hat{h}_t(\hat{\omega}, \hat{a}_1, \hat{a}_2)$ and ε_t by $\hat{\varepsilon}_t(\hat{\mu})$. I make the same conjecture as in Chapter 5 that if I can show that almost-integration would occur if h_t were observable, it will also occur when I have less information. I will also make the assumption again that the measurement is error free instead of introducing an error term that is independent of the parameter estimates and vanishes with growing segment size.

Now, let $\{h_t\}$ be generated by

$$h_t = \begin{cases} \omega_1 + a_1^* h_{t-1} + a_2^* h_{t-2}, & t \in \{1, \dots, N_1\}, \\ \omega_2 + a_1^* h_{t-1} + a_2^* h_{t-2}, & t \in \{N_1 + 1, \dots, N\}. \end{cases} \quad (6.5.18)$$

This segmentation is not known to the econometrician and the estimated model equation is

$$h_t = \hat{\omega} + \hat{a}_1^* h_{t-1} + \hat{a}_2^* h_{t-2} \quad (6.5.19)$$

One might argue that having exact measurement of h_t , it would be an obvious approach to just back out the parameters, thereby finding that there was a jump in ω . However, as I am interested in the case of h_t being unobservable, where this cannot be done, I will nevertheless proceed with the estimation of (6.5.19).

Subtract the mean from (6.5.19):

$$h_t - \bar{h} = \hat{\alpha}_1(\varepsilon_{t-1}^2 - \bar{\varepsilon}^2) + \hat{\alpha}_2(\varepsilon_{t-2}^2 - \bar{\varepsilon}^2) + \hat{\beta}_1(h_{t-1} - \bar{h}) + \hat{\beta}_2(h_{t-2} - \bar{h}), \quad (6.5.20)$$

where $\hat{\alpha}_2$ or $\hat{\beta}_2$ or both might be zero, depending on the GARCH(p,q) specification estimated. If the segmentation were known, the econometrician would insert a term for the difference in ω .

Let $\mathbb{E}_{(i)} h_t$ denote the expected values with respect to the start value in segment i , where i is 1 for $t \in \{-1, 0, 1, \dots, N_1\}$ and i is 2 for $t \in \{N_1 + 1, \dots, N\}$. In other words,

$$\begin{aligned} \mathbb{E}_{(1)} h_t &= \mathbb{E}(h_t | \mathcal{F}_0) \\ \mathbb{E}_{(2)} h_t &= \mathbb{E}(h_t | \mathcal{F}_{N_1}). \end{aligned}$$

Lemma 9. *Let h_t be generated according to (6.5.18). Let $\mathbb{E}h^{(i)} = \omega_i/(1 - a_1 - a_2)$ denote the expected value with respect to the stationary measure of a process h_t generated by (ω_i, a_1, a_2) without parameter changes. Then,*

$$\begin{aligned} \bar{h} &= \frac{N_1}{N} \mathbb{E}h^{(1)} + \frac{N - N_1}{N} \mathbb{E}h^{(2)} + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1}. \\ \bar{\varepsilon}^2 &= \frac{N_1}{N} \mathbb{E}h^{(1)} + \frac{N - N_1}{N} \mathbb{E}h^{(2)} + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1}. \end{aligned} \quad (6.5.21)$$

Proof. Write $h_t = \mathbb{E}_{(i)} h_t + x_t$, x_t being the deviation from the expectation conditional on the start values in the segments, so that

$$\begin{aligned} \frac{1}{N_1} \sum_{t=1}^{N_1} x_t &= o(1)_{N_1} \\ \frac{1}{N - N_1} \sum_{t=N_1+1}^N x_t &= o(1)_{N-N_1}. \end{aligned}$$

Then, from this and Lemma 7 we obtain

$$\begin{aligned}
\bar{h} &= \frac{1}{N} \sum_{t=1}^N h_t, \\
&= \frac{1}{N} \sum_{t=1}^{N_1} \mathbb{E}_{(1)} h_t + \frac{1}{N} \sum_{t=N_1+1}^N \mathbb{E}_{(2)} h_t + \frac{1}{N_1} \sum_{t=1}^{N_1} x_t + \frac{1}{N - N_1} \sum_{t=N_1+1}^N x_t, \\
&= \frac{1}{N} \sum_{t=1}^{N_1} \mathbb{E} h^{(1)} + \frac{1}{N} \sum_{t=N_1+1}^N \mathbb{E} h^{(2)} + o(1)_{N_1} + o(1)_{N-N_1} + \frac{1}{N} \sum_{t=1}^{N_1} O(\gamma_1^{t+1}) \\
&\quad + \frac{1}{N} \sum_{t=1}^{N_1} O(\gamma_2^{t+1}) + \frac{1}{N} \sum_{t=N_1+1}^N O(\gamma_1^{t+1-N_1}) + \frac{1}{N} \sum_{t=N_1+1}^N O(\gamma_2^{t+1-N_1}), \\
&= \frac{N_1}{N} \mathbb{E} h^{(1)} + \frac{N - N_1}{N} \mathbb{E} h^{(2)} + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1}.
\end{aligned}$$

In the same manner, write $\varepsilon_t^2 = \mathbb{E}_{(i)} \varepsilon_t^2 + y_t = \mathbb{E}_{(i)} h_t + y_t$ by the distribution assumption (5.2.2). Then,

$$\begin{aligned}
\overline{\varepsilon^2} &= \frac{1}{N} \sum_{t=1}^N \varepsilon_t^2, \\
&= \frac{N_1}{N} \mathbb{E} h^{(1)} + \frac{N - N_1}{N} \mathbb{E} h^{(2)} + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1} \quad \square
\end{aligned}$$

Concerning the estimators, I make the same assumption as in the GARCH(1,1) case.

Assumption 10. *The influence of a single realisation of the processes ε_t^2 and h_t on the estimators $\hat{\alpha}_j$ and $\hat{\beta}_j$ vanishes with growing sample size:*

$$\begin{aligned}
\text{cov}(\hat{\alpha}_j, \varepsilon_t^2) &= o(1)_N \quad \forall t, j \\
\text{cov}(\hat{\beta}_j, h_t) &= o(1)_N \quad \forall t, j.
\end{aligned}$$

Proposition 11. *If there is an unknown switch in the data-generating constant of the conditional variance equation of a Gaussian GARCH(p,q) model with $p = 1, 2$ and $q = 1, 2$, as specified in equation (6.5.18), and the model is estimated on the entire series, then, under Assumption 8 the condition*

$$\mathbb{E}_{(i)}(\hat{a}_1 + \hat{a}_2) = 1$$

must hold in both segments i , up to terms that vanish with growing length of the segments.

Proof. Take expectations of (6.5.20) conditional on the start values within segments:

$$\begin{aligned}\mathbb{E}_{(i)}h_t - \mathbb{E}_{(i)}\bar{h} &= \mathbb{E}_{(i)}(\hat{\alpha}_1 h_{t-1}) - \mathbb{E}_{(i)}(\hat{\alpha}_1 \bar{\varepsilon}^2) \\ &\quad + \mathbb{E}_{(i)}(\hat{\alpha}_2 h_{t-2}) - \mathbb{E}_{(i)}(\hat{\alpha}_2 \bar{\varepsilon}^2) \\ &\quad + \mathbb{E}_{(i)}(\hat{\beta}_1 h_{t-1}) - \mathbb{E}_{(i)}(\hat{\beta}_1 \hat{h}) \\ &\quad + \mathbb{E}_{(i)}(\hat{\beta}_2 h_{t-2}) - \mathbb{E}_{(i)}(\hat{\beta}_2 \bar{h})\end{aligned}$$

Use Assumption 10 and the distribution assumption (5.2.2)

$$\begin{aligned}\mathbb{E}_{(i)}h_t - \mathbb{E}_{(i)}\bar{h} &= \mathbb{E}_{(i)}\hat{\alpha}_1 \mathbb{E}_{(i)}h_{t-1} - \mathbb{E}_{(i)}(\hat{\alpha}_1 \bar{\varepsilon}^2) \\ &\quad + \mathbb{E}_{(i)}\hat{\alpha}_2 \mathbb{E}_{(i)}h_{t-2} - \mathbb{E}_{(i)}(\hat{\alpha}_2 \bar{\varepsilon}^2) \\ &\quad + \mathbb{E}_{(i)}\hat{\beta}_1 \mathbb{E}_{(i)}h_{t-1} - \mathbb{E}_{(i)}(\hat{\beta}_1 \hat{h}) \\ &\quad + \mathbb{E}_{(i)}\hat{\beta}_2 \mathbb{E}_{(i)}h_{t-2} - \mathbb{E}_{(i)}(\hat{\beta}_2 \bar{h}) + o(1)_N.\end{aligned}$$

Plug in \bar{h} and $\bar{\varepsilon}^2$ from Lemma 9, use Lemma 7

$$\begin{aligned}\mathbb{E}h^{(i)} &- \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N - N_1}{N}\mathbb{E}h^{(2)} \\ &\quad + O(\gamma_1^{t+1-N_i}) + O(\gamma_2^{t+1-N_i}) + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1} \\ &= \mathbb{E}_{(i)}\hat{\alpha}_1 \mathbb{E}h^{(i)} + O(\gamma_1^{t-N_i}) + O(\gamma_2^{t-N_i}) \\ &\quad - \mathbb{E}_{(i)} \left[\hat{\alpha}_1 \left(\frac{N_1}{N}\mathbb{E}h^{(1)} + \frac{N - N_1}{N}\mathbb{E}h^{(2)} + \frac{C_1^{(1)}}{N} \sum_{t=1}^{N_1} \gamma_1^t + \frac{C_2^{(1)}}{N} \sum_{t=1}^{N_1} \gamma_2^t \right. \right. \\ &\quad \left. \left. + \frac{C_1^{(2)}}{N} \sum_{t=N_1+1}^N \gamma_1^{t-N_1} + \frac{C_2^{(2)}}{N} \sum_{t=N_1+1}^N \gamma_2^{t-N_1} + \frac{1}{N} \sum_{t=1}^{N_1} y_{t-1} + \frac{1}{N} \sum_{t=N_1+1}^N y_{t-1} \right) \right] \\ &\quad + \mathbb{E}_{(i)}\hat{\alpha}_2 \mathbb{E}h^{(i)} + O(\gamma_1^{t-1-N_i}) + O(\gamma_2^{t-1-N_i}) \\ &\quad - \mathbb{E}_{(i)} \left[\hat{\alpha}_2 \left(\frac{N_1}{N}\mathbb{E}h^{(1)} + \frac{N - N_1}{N}\mathbb{E}h^{(2)} + \frac{C_1^{(1)}}{N} \sum_{t=1}^{N_1} \gamma_1^{t-1} + \frac{C_2^{(1)}}{N} \sum_{t=1}^{N_1} \gamma_2^{t-1} \right. \right. \\ &\quad \left. \left. + \frac{C_1^{(2)}}{N} \sum_{t=N_1+1}^N \gamma_1^{t-1-N_1} + \frac{C_2^{(2)}}{N} \sum_{t=N_1+1}^N \gamma_2^{t-1-N_1} + \frac{1}{N} \sum_{t=1}^{N_1} y_{t-2} + \frac{1}{N} \sum_{t=N_1+1}^N y_{t-2} \right) \right]\end{aligned}$$

$$\begin{aligned}
& + \mathbb{E}_{(i)} \hat{\beta}_1 \mathbb{E} h^{(i)} + O(\gamma_1^{t-N_i}) + O(\gamma_2^{t-N_i}) \\
& - \mathbb{E}_{(i)} \left[\hat{\beta}_1 \left(\frac{N_1}{N} \mathbb{E} h^{(1)} + \frac{N - N_1}{N} \mathbb{E} h^{(2)} + \frac{C_1^{(1)}}{N} \sum_{t=1}^{N_1} \gamma_1^t + \frac{C_2^{(1)}}{N} \sum_{t=1}^{N_1} \gamma_2^t \right. \right. \\
& \quad \left. \left. + \frac{C_1^{(2)}}{N} \sum_{t=N_1+1}^N \gamma_1^{t-N_1} + \frac{C_2^{(2)}}{N} \sum_{t=N_1+1}^N \gamma_2^{t-N_1} + \frac{1}{N} \sum_{t=1}^{N_1} x_{t-1} + \frac{1}{N} \sum_{t=N_1+1}^N x_{t-1} \right) \right] \\
& + \mathbb{E}_{(i)} \hat{\beta}_2 \mathbb{E} h^{(i)} + O(\gamma_1^{t-1-N_i}) + O(\gamma_2^{t-1-N_i}) \\
& - \mathbb{E}_{(i)} \left[\hat{\beta}_2 \left(\frac{N_1}{N} \mathbb{E} h^{(1)} + \frac{N - N_1}{N} \mathbb{E} h^{(2)} + \frac{C_1^{(1)}}{N} \sum_{t=1}^{N_1} \gamma_1^{t-1} + \frac{C_2^{(1)}}{N} \sum_{t=1}^{N_1} \gamma_2^{t-1} \right. \right. \\
& \quad \left. \left. + \frac{C_1^{(2)}}{N} \sum_{t=N_1+1}^N \gamma_1^{t-1-N_1} + \frac{C_2^{(2)}}{N} \sum_{t=N_1+1}^N \gamma_2^{t-1-N_1} + \frac{1}{N} \sum_{t=1}^{N_1} x_{t-2} + \frac{1}{N} \sum_{t=N_1+1}^N x_{t-2} \right) \right] \\
& + o(1)_N, \tag{6.5.22}
\end{aligned}$$

where $N_0 = 0$ and

$$\begin{aligned}
C_1^{(1)} &= \frac{1}{\gamma_1 - \gamma_2} \left[h_0 - \gamma_2 h_{-1} - \frac{\omega_1}{1 - \gamma_1} \right] \\
C_2^{(1)} &= \frac{1}{\gamma_1 - \gamma_2} \left[h_0 - \gamma_1 h_{-1} + \frac{\omega_1}{1 - \gamma_2} \right] \\
C_1^{(2)} &= \frac{1}{\gamma_1 - \gamma_2} \left[h_{N_1} - \gamma_2 h_{N_1-1} - \frac{\omega_2}{1 - \gamma_1} \right] \\
C_2^{(2)} &= \frac{1}{\gamma_1 - \gamma_2} \left[h_{N_1} - \gamma_1 h_{N_1-1} + \frac{\omega_2}{1 - \gamma_2} \right].
\end{aligned}$$

Only the sums of the x_t and y_t are non-deterministic in the expressions in parentheses in (6.5.22). Exactly as for the GARCH(1,1) case, Assumption 10 is used to decouple the expectations of these products, for example,

$$\frac{1}{N} \mathbb{E}_{(i)} \left(\hat{\alpha} \sum_{t=1}^{N_1} y_t \right) = \frac{1}{N} \mathbb{E}_{(i)} \hat{\alpha} \mathbb{E}_{(i)} \sum_{t=1}^{N_1} y_t + \frac{1}{N} \sum_{t=1}^{N_1} \text{cov}(\hat{\alpha}, y_t),$$

$$\text{cov}(\hat{\alpha}, y_t) = \text{cov}(\hat{\alpha}, \mathbb{E}_{(i)} h_t) + \text{cov}(\hat{\alpha}, y_t) = \text{cov}(\hat{\alpha}, \varepsilon_t^2) = o(1)_N.$$

Table 6.1: Values of the autoregressive parameters of the data generating GARCH(2,2) processes in the three mean reversion environments.

mean reversion:	slow	fast	medium
α_1	0.05	0.05	0.05
β_1	0.40	0.10	0.30
α_2	0.10	0.05	0.10
β_2	0.40	0.10	0.40
λ	0.95	0.30	0.85
mr-time	20	3	7

Plugging into (6.5.22) and arranging terms, I obtain

$$\begin{aligned}
& \mathbb{E}h^{(i)} - \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N - N_1}{N}\mathbb{E}h^{(2)} \\
&= \mathbb{E}_{(i)}(\hat{a}_1 + \hat{a}_2) \left[\mathbb{E}h^{(i)} - \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N - N_1}{N}\mathbb{E}h^{(2)} \right] \\
&+ o(1)_N + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1} + O(\gamma_1^{t-1-N_i}) + O(\gamma_2^{t-1-N_i}).
\end{aligned} \tag{6.5.23}$$

That is, with growing N_1 and growing $N - N_1$, the expectation of the sum of the estimators of the autoregressive coefficients conditional on the start values within segments is

$$\mathbb{E}_{(i)}(\hat{a}_1 + \hat{a}_2) = 1$$

up to vanishing terms, in order to satisfy condition (5.3.20). Observe that the difference

$$\mathbb{E}h^{(i)} - \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N - N_1}{N}\mathbb{E}h^{(2)} \neq 0$$

for both segments i if $\omega_1 \neq \omega_2$. Thus, condition (5.3.20) is not trivial. \square

The fact that condition (6.5.23) is not trivial is the main difference to the stationary GARCH analysis without parameter switches. In that case, the condition reads zero equals zero with respect to the stationary measure.

6 Simulations

I will exert simulation experiments in three mean reversion environments, a fast, medium, and slow mean reversion. The process considered is GARCH(2,2). Table 6.1 shows the autoregressive parameters of the processes in the three environments.

For each environment, I generate 10×10000 GARCH(2,2) series of length 5000. At $t = 2500$ each of the time series has a changepoint of the GARCH intercept ω . During the first 2500 observations, the process always has the mean 12.5 percent annualized volatility. For the slow mean reversion environment,

this corresponds to $\omega = 3.125\text{e-}6$. For the fast environment, it corresponds to $\omega = 4.375\text{e-}5$, and for the medium environment $\omega = 9.375\text{e-}6$. In general, ω is determined by

$$\sqrt{250 \frac{\omega}{1 - \alpha_1 - \alpha_2 - \beta_1 - \beta_2}} = 0.125.$$

Table 6.2: Ten jump sizes of the intercept of the data generating GARCH(2,2) processes in the three mean reversion environments.

experiment	1	2	3	4	5	6	7	8	9	10
slow mean reversion										
ω	3.1e-6	3.9e-6	4.7e-6	5.4e-6	6.2e-6	6.9e-6	7.7e-6	8.5e-6	9.2e-6	1e-5
$\sigma = \sqrt{\frac{250\omega}{0.05}}$	0.125	0.139	0.153	0.165	0.176	0.186	0.196	0.206	0.215	0.224
$\sigma - 0.125$	0	0.014	0.028	0.040	0.051	0.061	0.071	0.081	0.090	0.099
fast mean reversion										
ω	4.4e-5	5.1e-5	5.9e-5	6.7e-5	7.4e-5	8.2e-5	8.9e-5	9.7e-5	1e-4	1.1e-4
$\sigma = \sqrt{\frac{250\omega}{0.70}}$	0.125	0.135	0.145	0.154	0.163	0.171	0.179	0.186	0.193	0.200
$\sigma - 0.125$	0	0.010	0.020	0.030	0.038	0.046	0.054	0.061	0.068	0.075
medium mean reversion										
ω	9.4e-6	1.2e-5	1.4e-5	1.6e-5	1.9e-5	2.1e-5	2.3e-5	2.5e-5	2.8e-5	3e-5
$\sigma = \sqrt{\frac{250\omega}{0.15}}$	0.125	0.139	0.153	0.165	0.176	0.186	0.196	0.206	0.215	0.224
$\sigma - 0.125$	0	0.014	0.028	0.040	0.051	0.061	0.071	0.081	0.090	0.099

During the second 2500 observations, the second segment of each series, the mean of the volatility process is higher. In 10 experiments I use different magnitudes of the jumps. For each of these magnitudes, I generate 10000 series and estimate GARCH(2,2) on it.

Table 6.2 shows the values of ω in the second segment in the 10 experiments, the corresponding mean in the second segment (in the first segment, it is always 0.125), and the jump size over 12.5 percent.

Figures 6.3 through 6.5 show the mean and the two-sided 95 percent quantile of $\hat{\lambda}$ for each of the three mean reversion environments as a function of the jump size. That is, for each experiment (= each jump size) 10,000 series are generated and GARCH(2,2) is estimated on each series. This gives a sample distribution of $\hat{\lambda}$ for each experiment. The figures plot mean and two-sided 95 percent quantile for each experiment. The label for each experiment is the difference of the annualized standard deviation in the second segment of each time series from 12.5 percent, the annualized deviation in the first segment of each time series.

The almost-integration effect is clearly visible in GARCH(2,2) estimations and it is monotonously increasing with the jump size. From Section 5.3.a) this is intuitively clear. If the jump size is small, the different volatility means through which the estimation hyperplane must be laid are close together. The point clusters belonging to the two segments may have a much larger range than the distance of their means and as both have the same slopes, the ranges and slopes

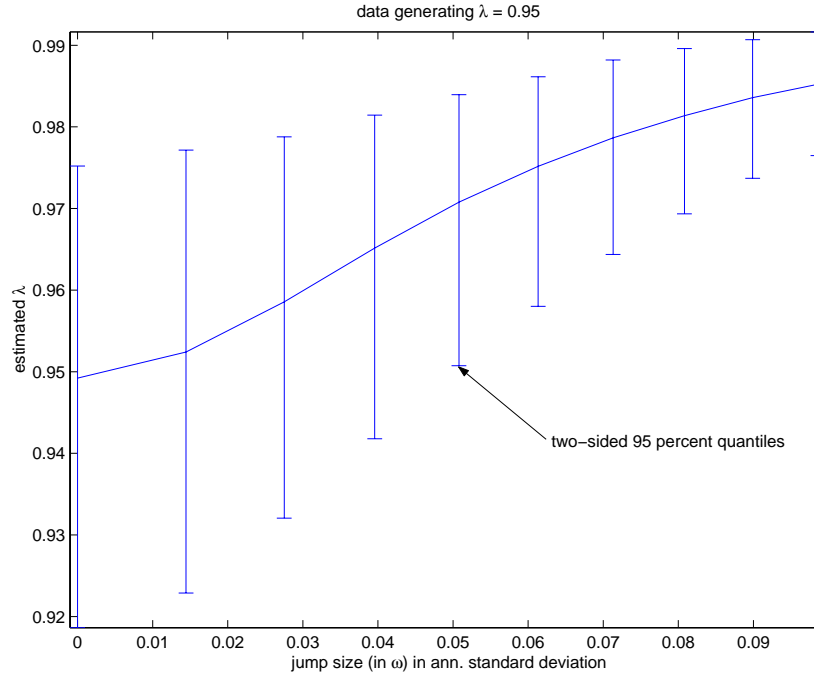


Figure 6.3: Plot of the mean and two-sided 95 percent quantiles for each experiment in the slow mean reversion environment according to Table 6.2. The almost-integration effect is clearly visible, the estimated mean of $\hat{\lambda}$ increases towards one with increasing jump size.

of the clusters may dominate the estimation. The result is that the estimation captures the true in-segment dynamics fairly well. Only when the means grow further apart, that is, when the jump size increases, the distance in the means dominates over the ranges of the clusters and the almost-integration effect takes over.

Figures 6.3 through 6.5 make also clear that the variance of the estimates of λ decreases with growing jump size. This is also caused by the geometry of the problem: The larger the distance in the means, the more dominant is the almost-integration effect and the less slack there is for $\hat{\lambda}$ to take any value other than almost one.

How precise are the estimations? For this problem, the first experiment in each mean reversion environment is the most interesting. Here, the volatility in the second segment is also 12.5 percent, that is, there is no real changepoint. In the slow and medium mean reversion environments, the estimation appears not very biased, the means of the $\hat{\lambda}$'s are at 0.95 and 0.85. In the fast mean reverting environment, the estimation is upward biased, the mean is at 0.35 instead of 0.30. This is probably a start value effect, as I set $\hat{\lambda}_0 = 0.35$. Also, for some experiments in this environment the 95 percent quantiles are huge, for the 2 percent jump for instance, they cover almost the whole parameter subspace.

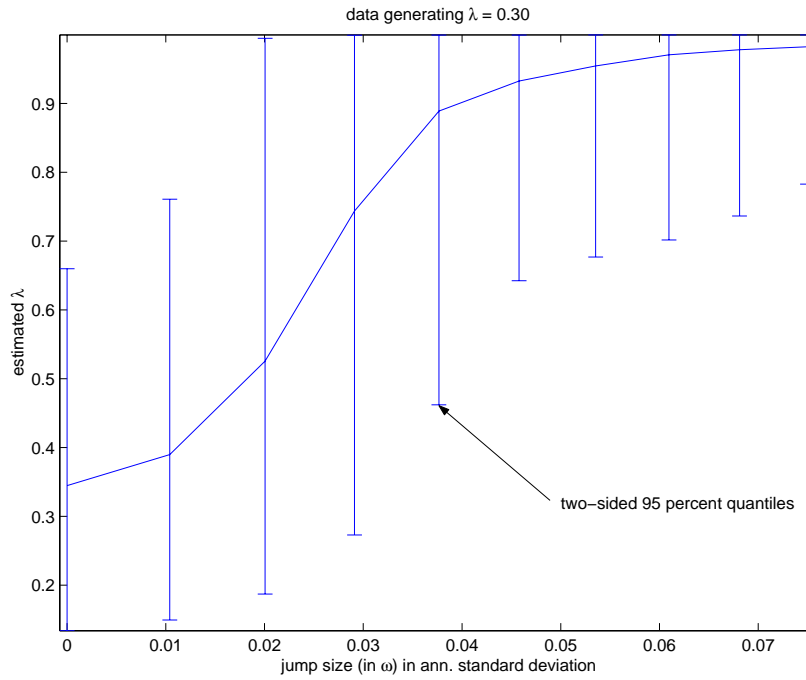


Figure 6.4: Plot of the mean and two-sided 95 percent quantiles for each experiment in the fast mean reversion environment according to Table 6.2. Again, the almost-integration effect is clearly visible.

The estimations were done using two types of code, own code written in C++ using the quasi-Newton methods from the “Numerical Recipes” (Press et al. 2002), and the other code written in EViews 3.1. The differences between the results were negligible and the EViews results are reported.

The upshot is that the less persistent the data generating process, the more sensitive is the estimation to the choice of the start value. GARCH estimates of very low persistence processes are not very precise and tend to stick to the start value of $\hat{\lambda}$. In general, the higher the persistence, the more precise is the estimation, as can be seen from the decreasing 95 percent quantiles as λ grows larger over the three mean reversion environments. Even though, the almost-integration clearly takes over, it causes a substantial distortion for jumps of one percent annualized volatility already. Jump sizes between 5 and 6 percent place λ safely above 0.95, regardless of the in-segment persistence.

7 Conclusion

Higher order GARCH models, at least up to order GARCH(2,2) are not the remedy for the problem of multiple scales. In the case of spatial aggregation, the

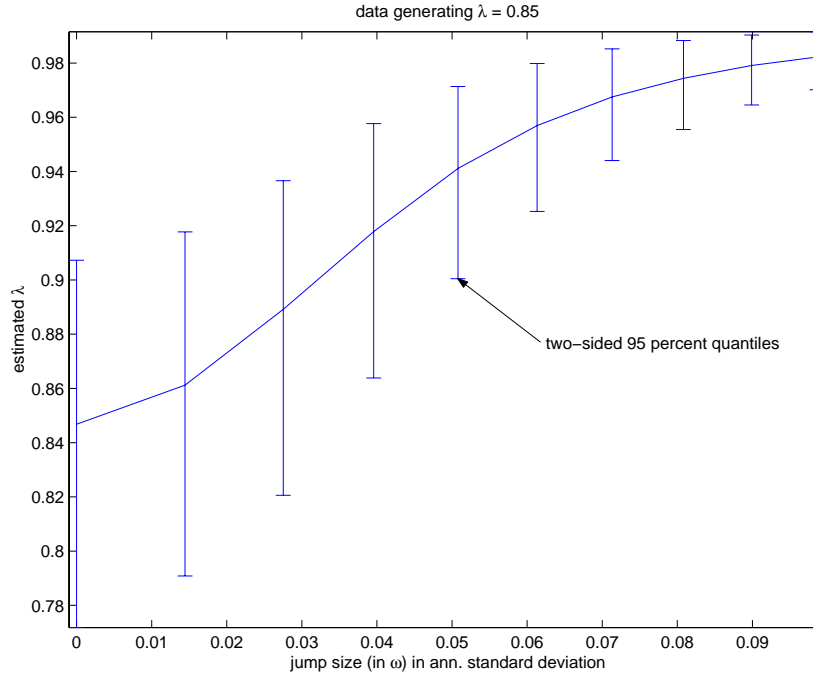


Figure 6.5: Plot of the mean and two-sided 95 percent quantiles for each experiment in the medium mean reversion environment according to Table 6.2. Again, the almost-integration effect is clear.

roots of the characteristic equation of GARCH(2,2) do not have the nice properties of their ARMA(2,2) counterparts and cannot discriminate cleanly between the scales of the component processes.

In the case of temporal aggregation, that is, unknown points where the GARCH(2,2) parameters change, this chapter showed that the arguments presented in Chapter 5 for the GARCH(1,1) case fully generalize to GARCH(2,2). In other words, changes in the parameter regime push the sum of the estimates of the autoregressive parameters close to one, indicating high persistence. This is regardless of the data generating persistence.

Simulations showed that the magnitude of the distortion that the almost-integration effect causes depends on the magnitude of the jump in the volatility mean implied by the parameter change. The larger the jump, the more dominant is the effect and the less related are the parameter estimates to the data generating persistence.

Another observation from the simulations is that the higher the data generating persistence, the more accurate the GARCH(2,2) estimations are. For the fast mean reverting environment, the estimation errors were huge and there was a strong bias toward the start value of $\hat{\lambda}$.

Chapter 7

The Impact of Japanese Foreign Exchange Intervention on Level and Volatility of the Yen/Dollar Exchange Rate

The combined weak economy and strong yen has brought Japan into the dilemma that the strong export sector is curtailed in its contribution to much needed growth. Interventions on the yen/dollar market help the export but have a negative effect on the domestic capital markets, where interest rates are at a record low. The Japanese authorities have used ‘sterilized’ interventions as a remedy to the dilemma. Sterilized interventions are neutralized by domestic open market operations with an opposite sign, so that the total monetary base remains unchanged. As intervention dates are potential candidates for changepoints in volatility, I will consider the volatility of the yen/dollar exchange rate and newly released daily data of the Japanese foreign exchange interventions. I will show that accounting for interventions largely reduces the estimated mean reversion time of the yen/dollar exchange rate.

1 Sterilized Intervention and Volatility

Japan’s economy has suffered from a long lasting depression ever since the real estate bubble burst in the mid 1990s. Coinciding with the economic slump is a surprisingly strong yen. As the export sector is one of the remaining driving forces of growth, the Japanese authorities have a clear interest in not letting the yen grow too strong especially against the dollar. This inclines them towards interventions on the yen/dollar exchange markets, selling yen against dollars.

With interest rates approaching zero at home however, it is counterproductive to further increase the monetary base, as it would usually be the case when yen

is sold by the central bank against dollars. Therefore these interventions have been ‘sterilized’, that is, while creating money for the interventions, the Bank of Japan has sold titles against yen on the domestic market. Thereby the total effect on the monetary base of the yen was neutralized.

Central bank interventions are prime candidates for causing changes in volatility regimes. When they come as a surprise, they signal a shift in the attitude of the central bank towards the development of the exchange rate.¹ Mostly, interventions are concerned with the level of the rate and not its volatility. When the interventions are sustained however, market participants expect the exchange rate to revert to the bliss point set by the authorities. That is, they expect lower volatility. When the interventions are discretionary and arrive somewhat arbitrarily, the opposite effect might be achieved. The uncertainty about the conduct of monetary policy increases and with it the volatility, as agents do not know what to expect.

I will investigate the effect of the interventions of the Japanese authority using a GARCH approach with exogenous variables as suggested by Baillie and Bollerslev (1989) and also using the changepoint detector for ARCH processes suggested by Kokoszka and Leipus (2000). Earlier investigations used changes in monthly reserves of the central bank or sporadic press releases as a proxy for interventions. Recently, the Bank of Japan has published time series of its daily intervention activities reaching back through 1991. The investigation is based on Hillebrand and Schnabl (2003).

2 The Discussion of the Japanese Interventions in the Literature

The effectiveness of sterilized interventions is highly controversial ever since the Working Group on Exchange Market Intervention set in by the G7 devised sterilized interventions and suggested the concept in its report (Jurgensen 1983).

The proponents of sterilized interventions suggest that foreign and domestic capital assets are imperfect substitutes and that interventions shift the relative supplies even when sterilized. This would change the relative returns, leading to restructuring of portfolios. Important contributions in this direction are Rogoff (1984) and Dominguez and Frankel (1993).

The opponents of sterilized interventions argue that these interventions leave the domestic money supply unchanged. Without any effect on the domestic interest rate, no restructuring of portfolios is triggered. The only effect can

¹In Japan, it is actually the Ministry of Finance that decides over interventions, the Bank of Japan acts simply as an agent for the ministry, according to the Foreign Exchange and Trade Law, Article 7(3).

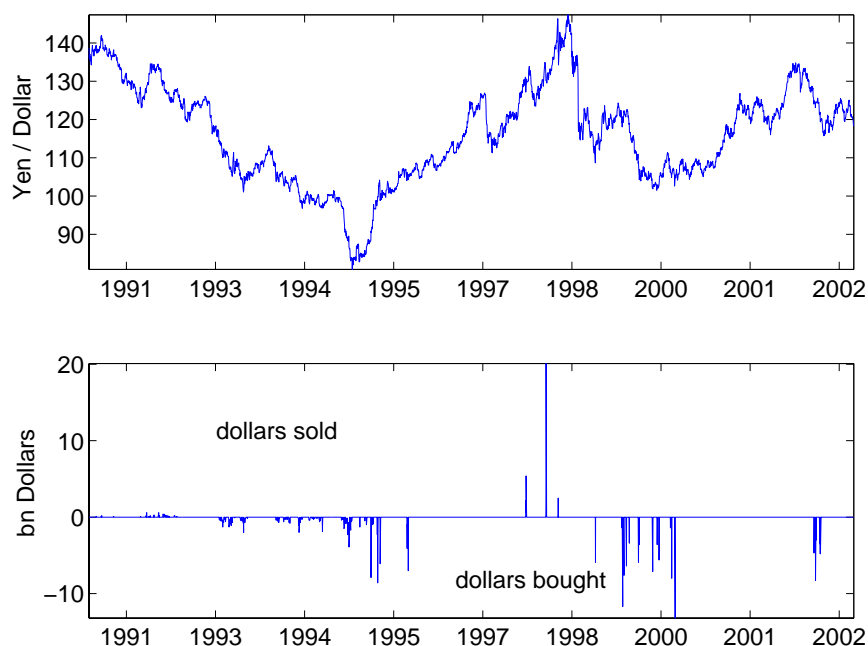


Figure 7.1: The upper panel shows the yen/dollar exchange rate in the period January 1, 1991 through December 31, 2002. The lower panel shows the interventions carried out by the Japanese central bank during the same period. The numbers were converted from yen into billion dollars.

therefore arise from the sheer volume of additional supply or demand in the market that the interventions amount to. These volumes however are much too small in comparison to the total market turnover to have any lasting effect. These are the main points put forward in Galati and Melick (1999) and in Dominguez (1998).

The influence of sterilized interventions on exchange rate volatility is discussed likewise. As outlined in the introduction, fully credible interventions that push the exchange rate back to its target corridor should clearly reduce volatility. This view is proposed in Dominguez (1998). On the other hand, sporadic and surprising interventions may raise uncertainty and increase volatility. Galati and Melick (1999) find for the period 1987 through 1991 that Japanese interventions increased the volatility of the yen/dollar rate.

The sustained strong yen after eight years of massive interventions by the Japanese authorities is clearly a case in the latter point as illustrated by Figure 7.1. Figure 7.2 relates interventions to the development of one-year volatility. It is apparent that there is quite some correlation between interventions and volatility. The model considered in this thesis will make this correlation more precise. As with every econometric model, it will not be possible to clearly

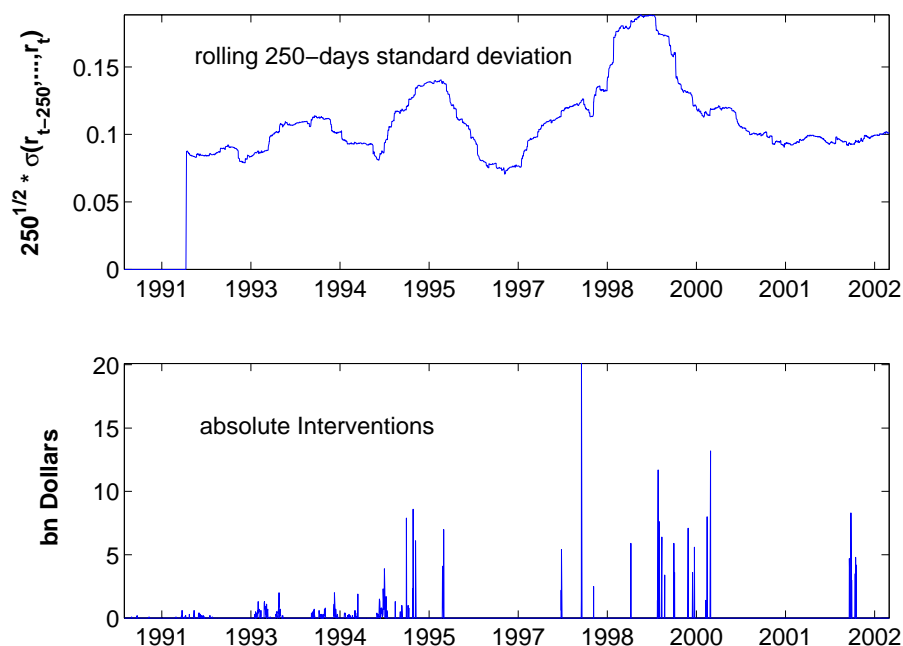


Figure 7.2: The upper panel shows a rolling 250 days standard deviation of the returns of the yen/dollar exchange rate. The lower panel shows the absolute magnitude of interventions carried out by the Japanese central bank during the same period. The numbers were converted from yen into billion dollars.

identify the causal direction, but another interesting aspect will be clarified: The interventions coincide with changepoints in the volatility data and taking these into account largely reduces the measured mean reversion.

Figures 7.3 and 7.4 show the Japanese and Federal Reserve interventions in the yen/dollar relation and compare it with the development of central bank reserves in the respective currencies. From a phenomenological point of view, the immense holdings of foreign assets by Japanese investors result in sustained appreciation pressure against the yen. As investors sell their assets abroad or merely convert the returns into yen, demand for the yen soars.

As the appreciation curtails the export sector, the Japanese authorities react by intervening on the yen/dollar market by selling yen and buying dollars. The result is a shift of dollar holdings from the Japanese private sector to the Japanese public sector. In comparison to US reserves, for example, the Japanese reserves are very large and consist for the better part of intervention volume (the intervention series of the Bank of Japan adds up to about 210 billion dollars bought).

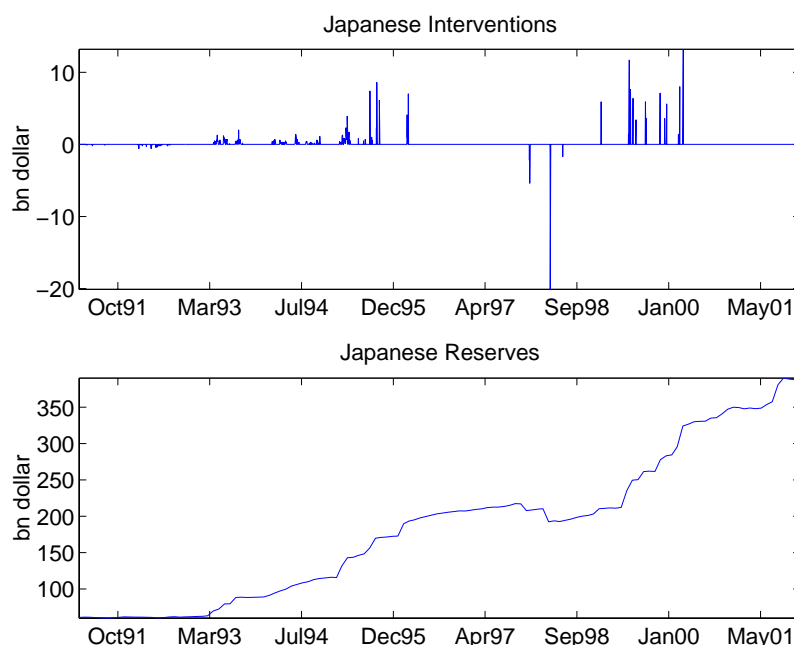


Figure 7.3: The upper panel shows the Japanese interventions in the yen/dollar market between 1991 and 2001, measured in billion US dollar. Yen sales (dollar purchases) have a positive and yen purchases (dollar sales) have a negative sign. The lower panel shows the development of dollar reserves of the Bank of Japan.

3 Data

The sample period is January 1, 1991 through December 31, 2002. The yen/dollar exchange rate was obtained from Datastream (BBI series). The foreign exchange rate intervention data are provided by the Japanese Ministry of Finance. The intervention series are published in billion yen and separately for transactions in the yen/dollar, yen/euro, and yen/other markets. Interventions in the yen/euro and other markets are very small and rare. I consider only interventions in the yen/dollar market and convert all amounts into trillion dollars.

The U.S. foreign exchange intervention data are provided by the Federal Reserve Board separately for the dollar/yen, dollar/euro (dollar/deutschemark), and dollar/others markets. Again, only the dollar/yen data are considered.

Table 7.1 shows some summary statistics for the intervention series. The Japanese foreign exchange interventions are largely focused on the yen/dollar market (97.2 percent), whereas the Federal Reserve also intervened on the deutsche-mark/dollar market (there have been no interventions of the Federal Reserve after the introduction of the euro). 48.7 percent of the US interventions are exerted on the yen/dollar market.

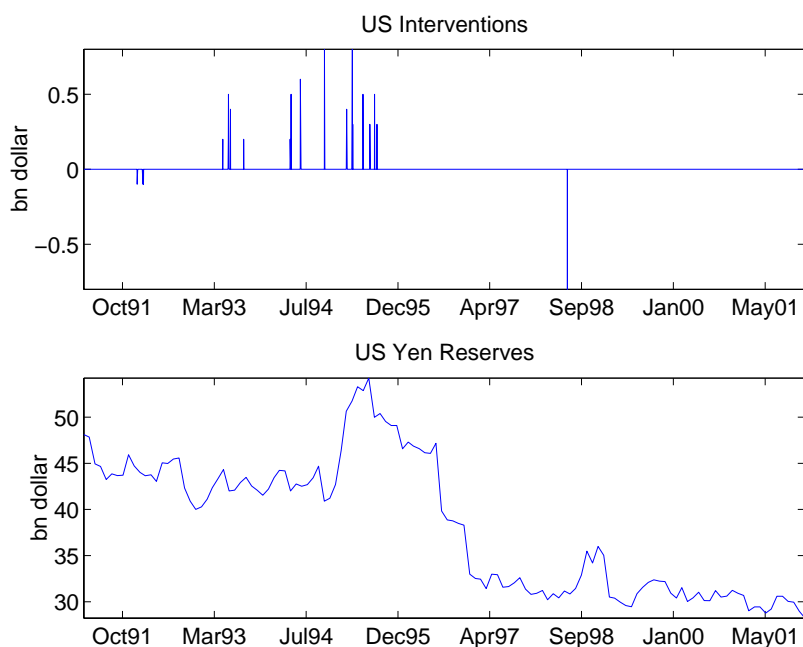


Figure 7.4: The upper panel shows the interventions of the Federal Reserve in the yen/dollar market between 1991 and 2001, measured in billion dollars. Yen sales (dollar purchases) have a positive and yen purchases (dollar sales) have a negative sign. The lower panel shows the development of dollar reserves of the Federal Reserve.

Japan is much more frequently intervening both in terms of times and volumes, about ten times as often and with a total 270 billion dollars, compared to eight billion dollars US interventions. All 22 US intervention days coincide with Japanese interventions, so that it is clear that US interventions were mostly carried out to support the Japanese interventions.

To infer the targeted bliss point from these numbers is difficult, in particular because the interventions were not that successful in the long run, as can be seen from Figure 7.1 and as will be shown in the next Section. There is no clear corridor discernible. It is well conceivable that the bliss point changes over time. There is not even a need to define a bliss point, the interventions of the Bank of Japan may be based on macroeconomic data, in particular export figures, and not on the exchange rate itself.

To control for the influence of other markets' volatilities, in particular the stock market, I include the Dow Jones Industrial Average and the Nikkei, both obtained from Datastream.

Table 7.1: Summary statistics of the Japanese and Federal Reserve interventions in the yen/dollar market. Interventions against all currencies are reported in parentheses.

	Bank of Japan	Federal Reserve
total intervention days	208 (208)	22 (36)
total transaction volume in bn dollars	273.6 (281.4)	8.4 (17.2)
percentage of interventions in the yen/dollar market	97.2%	48.7%
number of days with dollar purchases (yen sales)	175 (175)	18 (30)
total intervention volume in bn dollar	236.0	7.3
mean intervention volume in bn dollar	1.349	0.408
number of days with dollar sales (yen purchases)	33 (33)	4 (6)
total intervention volume in bn dollar	37.6	1.0
mean intervention volume in bn dollar	1.139	0.258

4 A GARCH(p,q) Model with Interventions as Exogenous Variables

As a benchmark, I estimated a simple GARCH(1,1) model with constant mean returns on the series of returns from the yen/dollar exchange rate. These are the gains from converting one dollar into yen on one day and converting it back into dollars on the next day.

The estimated GARCH(1,1) model with constant mean return for the sample period is

$$\begin{aligned}
 r_t &= -9\text{e-}4 + \varepsilon_t, \\
 \varepsilon_t | \mathcal{F}_{t-1} &\sim \mathcal{N}(0, h_t), \\
 h_t &= 8\text{e-}7 + 0.0411\varepsilon_{t-1}^2 + 0.9427h_{t-1}.
 \end{aligned} \tag{7.4.1}$$

The estimated sum of the autoregressive parameters is $\hat{\lambda} = \hat{\alpha} + \hat{\beta} = 0.9838$ so that the estimated mean reversion time is

$$\frac{1}{1 - \hat{\lambda}} \approx 62 \text{ days.}$$

Baillie and Bolerslev (1989) suggested the introduction of explanatory variables into both mean and conditional variance equation. I will use a consolidated intervention variable that contains the Japanese as well as the Federal Reserve interventions in the yen/dollar market. As the Federal Reserve merely supported the Japanese interventions, there is no need to consider them separately. Instead, they would introduce multicollinearity bias. To control for the influence of the stock market, I also include returns on two leading indicators, the Dow Jones Industrial Average and the Nikkei 300.

The result is the following GARCH specification

$$\begin{aligned} r_t &= b_0 + b_1 I_{t-1} + b_2 r(\text{Nikkei})_{t-1} + b_3 r(\text{Dow})_{t-1} + \varepsilon_t, \\ \varepsilon_t | \mathcal{F}_{t-1} &\sim \mathcal{N}(0, h_t), \\ h_t &= \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} + \gamma_1 |I_t| + \gamma_2 r(\text{Nikkei})_t^2 + \gamma_3 r(\text{Dow})_t^2. \end{aligned} \quad (7.4.2)$$

The returns are influenced by the interventions I (measured in trillion dollars) and the returns on the U.S. and Japanese stock markets. The conditional volatility is influenced by the absolute magnitude of the interventions and the volatilities of the U.S. and Japanese stock markets. A deviation of the exchange rate from the target increases the probability of an intervention. I am interested in the effects of interventions on the exchange rate and therefore the intervention variable I is lagged by one period to avoid simultaneity bias. In the conditional variance equation however, chances are slim that a high volatility triggers an intervention and therefore the contemporary value of I is inserted. The absolute value is used to avoid negative variances.

A Bayes Information Criterion and Akaike Information Criterion search for the best model order among the class $p \in \{1, \dots, 4\}$ and $q \in \{1, \dots, 4\}$ favored the GARCH(4,3) specification.

5 Estimation Results

The estimation results of model (7.4.2) are reported in Table 7.2. The sum of the estimates of the autoregressive coefficients is

$$\hat{\lambda} = \sum_{j=1}^4 \hat{\alpha}_j + \sum_{j=1}^3 \hat{\beta}_j \approx 0.831,$$

corresponding to a mean reversion time of

$$\frac{1}{1 - \hat{\lambda}} \approx 6 \text{ days.}$$

As there are exogenous variables in the conditional variance equation, the derivation of the mean reversion time in Section 5.2.b) is no longer valid, so that this is only a heuristic.

Also, one might argue that the long scale is not taken out by the interventions but by the long scale in the volatility of the stock indices. To control for this, I repeated the AIC and BIC search for a GARCH specification without any stock market data, neither in the variance nor in the mean. Only the lagged intervention was included in the mean and the absolute intervention in the conditional

variance. The model favored by both criteria was GARCH(3,3). The estimated sum of the autoregressive parameters was $\hat{\lambda} = 0.887$, corresponding to about 9 days of mean reversion time. In this case, the derivation of the mean reversion time in Section 5.2.b) can be argued to be valid as it is a fair assumption that the expected value of the absolute interventions is a constant c adding to the intercept of the conditional variance equation without changing its autoregressive dynamics.

Table 7.2: Estimation of model (7.4.2) for the sample period January 1, 1991 through December 31, 2002.

coefficient		standard error	z statistic	probability
b_0	-2E-5	0.0001	-0.173	0.863
b_1 (I_{t-1})	-0.0871	0.164	-0.532	0.595
b_2 (Nikkei $_{t-1}$)	-0.016	0.011	-1.511	0.131
b_3 (Dow $_{t-1}$)	0.038***	0.013	2.976	0.003
ω	4E-6***	1E-6	2.783	0.005
α_1	0.048*	0.028	1.672	0.095
α_2	0.027*	0.015	1.785	0.074
α_3	0.064***	0.018	3.596	0.0003
α_4	-0.005	0.025	-0.191	0.848
β_1	0.093*	0.049	1.882	0.06
β_2	-0.071*	0.039	-1.791	0.073
β_3	0.675***	0.062	10.954	0.000
γ_1 ($ I_t $)	0.016***	0.005	2.949	0.003
γ_2 (Nikkei 2)	0.007**	0.004	1.984	0.047
γ_3 (Dow 2)	0.016**	0.007	2.460	0.014

These findings compare to the short scale found in Section 5.4. The results indicate that the long scale is superimposed by the timing of the interventions. When using changepoint detection methods, I therefore expect to see a correspondence between the changepoints and the dates of the interventions.

It can also be seen from Table 7.2 that the interventions on the whole were not that successful. The influence on the returns of the exchange rate is short lived as the coefficient of the lagged intervention is insignificant (the coefficient is indeed significant for contemporary interventions). Unfortunately, the coefficient in the variance equation is highly significant. This shows that interventions increased volatility in the yen/dollar market. The U.S. stock market has a significant influence on returns and volatility in the yen/dollar market, interestingly more so than the Japanese stock market.

6 Changepoint Detection

I will apply a changepoint detector for ARCH models proposed by Kokoszka and Leipus (2000). Let k^* denote a single changepoint in a series generated by a standard GARCH(1,1) model with constant mean return. That is, at k^*

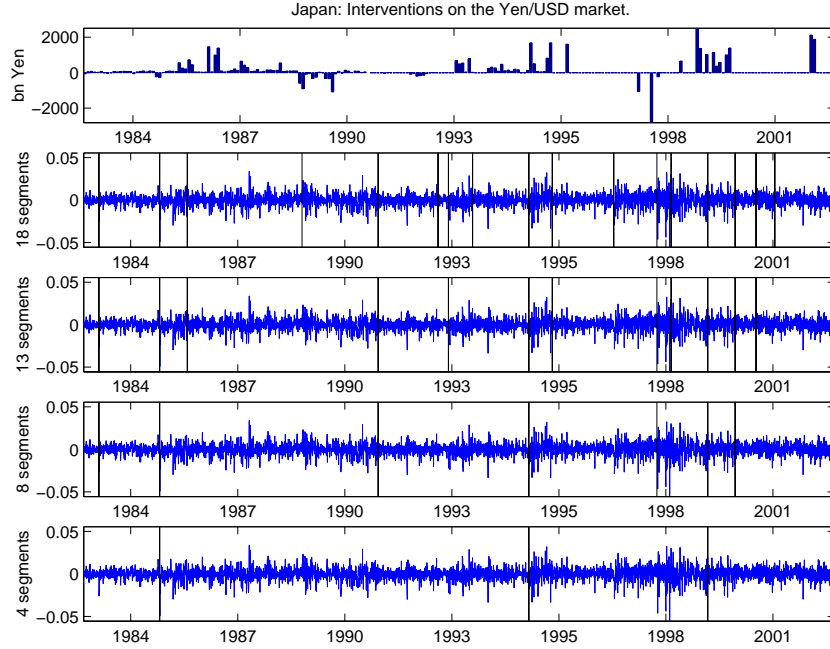


Figure 7.5: Comparison of estimated changepoints in the volatility of the daily yen/dollar returns with monthly interventions by the Japanese authorities. The first panel shows the interventions, the lower panels show the segmentations after the fifth, fourth, third, and second application of the changepoint detector (7.6.3).

the data generating parameter vector changes from $\theta_1 = (\mu_1, \omega_1, \alpha_1, \beta_1)$ to $\theta_2 = (\mu_2, \omega_2, \alpha_2, \beta_2)$. The changepoint detector is the estimator of k^* defined by

$$\hat{k} = \min \left\{ k : |R_k| = \max_{1 \leq j \leq n} |R_j| \right\}, \quad (7.6.3)$$

where

$$R_k = \frac{k(n-k)}{n^2} \left(\frac{1}{k} \sum_{j=1}^k r_j^2 - \frac{1}{n-k} \sum_{j=k+1}^n r_j^2 \right). \quad (7.6.4)$$

Kokoszka and Leipus (2000) show that this estimator is consistent and converges in probability to the true changepoint k^* with rate $1/n$.

I approach the multi-changepoint problem of finding a segmentation of the yen/dollar ex-change rate series as a sequential single-changepoint problem. In other words, first I take the whole exchange rate series and apply the detector. This results in an estimate of a single changepoint in the series and thus two subseries. Then I apply the detector to both subseries, resulting in a changepoint estimate for each subseries. I obtain three estimated changepoints and four subseries, and so forth.

I extended the sample period to October 11, 1983 through December 31, 2002 and repeated the detection step five times, theoretically resulting in 32 segments. However, I stopped subdividing a segment when its length was either less than 250 points or when in the next segmentation step a new segment of less than 50 points would have been cut off. Therefore, the total number of segments at the last step is 18.

Figure 7.5 reports the results of the changepoint detector. The first panel shows the interventions of the Bank of Japan in billion yen, aggregated to monthly observations. For the period before 1991, I use monthly changes in reserves as a proxy for the interventions. The next panels show the segmentations after the fifth, fourth, third, and second step.

A synopsis of the interventions and the changepoint estimates cannot be more than a correlation study. It does not provide any evidence of causality. For example, in the second panel from the bottom a changepoint appears in 1997 that coincides with the largest sales of yen by the Bank of Japan in our sample. Rather than interventions causing the changepoint, both changepoint and intervention are caused by the Asian crisis that was unfolding at the time.

The coincidence between changepoints and interventions is nevertheless remarkable. The first changepoint in the entire series is the one occurring in late 1994 in the bottom panel. It falls into a period of intense intervention. The two next changepoints depicted in the bottom panel also accompany interventions. Eight out of twelve changepoints in the middle panel occur within or immediately after interventions, nine out of 17 in the second panel from the top.

An estimation of the GARCH(1,1) model with constant mean return on the segmentation of the third step, that is, the second panel from the bottom of Figure 7.5 reveals a short autocorrelation structure. The average mean reversion time measured on the segments is

$$\frac{1}{1 - \hat{\lambda}} \approx \frac{1}{1 - 0.915} \approx 12 \text{ days.}$$

This corroborates the results of Table 7.2. The estimates for the segments are given in Table 7.3.

7 Conclusion

The Japanese authorities, that is, the Ministry of Finance and the Bank of Japan, are trying to stabilize the economy and especially the export sector by massive so-called ‘sterilized’ interventions. Sterilized interventions in foreign exchange markets are neutralized by domestic open market operations with opposite sign.

Table 7.3: Estimation of a GARCH(1,1) model with constant mean return on each of the segments detected by (7.6.3) in the third step.

Segment	I	II	III	IV	V	VI	VII	VIII	average
Observations	97	390	1367	961	816	329	175	645	
$\hat{\omega}$	0.018	0.0005	0.0015	0.0006	0.0001	0.0022	0.002	0.0005	0.0009
$\hat{\alpha}$	-0.1090	0.032	0.077	0.045	0.019	0.12	0.024	0.054	0.0511
$\hat{\beta}$	0.615	0.903	0.795	0.888	0.97	0.805	0.828	0.893	0.864
$\hat{\lambda}$	0.506	0.935	0.872	0.933	0.989	0.925	0.852	0.947	0.915

The analysis presented here shows that the interventions have a clear effect on the volatility of the yen/dollar exchange rate. The changepoint detector study shows that many of the estimated changepoints fall into periods of interventions.

A GARCH model with interventions as exogenous variables in mean and conditional variance reveals that the interventions add a long time scale to the process of the returns of the yen/dollar exchange rate. The absolute magnitude of interventions is highly significant in the variance equation, that is, the interventions increase volatility. The influence on the level of the exchange rate, which is the intended effect, is much more elusive. The estimated mean reversion time of the return series without accounting for interventions is about 60 days. After accounting for interventions, I measure a mean reversion time of 9 days.

Estimating a simple GARCH(1,1) model with constant mean return on the segments found by a changepoint detector for ARCH models, an average in-segment mean reversion of 12 days is estimated. The short scale in the volatility of the yen/dollar exchange rate is very distinct. Furthermore, many of the detected changepoints coincide with periods of interventions of the Japanese authorities.

Chapter 8

Conclusions and Directions

In this thesis, I have considered the property of mean reversion, an aspect that many models of financial prices, returns, and volatility have in common. Mostly, these models are considered as disparate topics and I have attempted to show that there is a benefit from considering them as “Mean Reversion Models of Financial Markets”. Mean reversion appears to be in financial data but it is not trivial to find, at least in prices and returns. In volatility, it is much more obvious.

The first interesting thing to note is that the literatures on mean reversion in stock prices and returns on the one hand and volatility on the other hand are not connected. Papers on the one topic hardly ever mention the other and the models formulated for mean reversion in prices and returns differ largely from the models for mean reversion in volatility.

The reason for this is not ignorance but the fact that one of the dividing lines of the controversy whether models should be formulated according to economic theory (the hypothesis of efficient markets) or according to the features of the data runs right through our topic: Mean reversion in volatility complies with the hypothesis of efficient markets, mean reversion in returns does not.

This is one of the reasons why one can obtain a fair overview of the literature body on mean reversion in prices and returns in the course of a Ph.D. thesis while the same is utterly impossible for the work on volatility and its persistence. Another reason is that mean reversion in prices and returns is statistically elusive. The main point of the seminal papers on the subject was to show that mean reversion in returns leaves no statistical trace and that it is therefore impossible to tell the null hypothesis of a random walk reliably apart from a persistent mean reversion.

The first part of this thesis discussed mean reversion in stock prices and returns. Taking up an idea of Fischer Black, I showed that mean reversion in prices and returns, even though it appears statistically to be marginal, can have serious consequences. It appears to have played a role in the stock market crash of

1987. While the focus of this part of the thesis is on the effects of mean reversion, I discuss some possible causes that were suggested in the literature. Most causal explanations of mean reversion in prices and returns are microeconomic.

The second part of the thesis considered mean reversion in volatility. While the evidence of mean reversion in prices and returns is commonly viewed as ambiguous (although it is clearly measurable with the methods presented here), mean reversion in volatility is a commonly accepted fact. Volatility is seen as one of the closest proxies to the information stream that is relevant to the market. Fluctuations in economic activity are assumed to cause persistence in volatility, the sustained high volatility in the early 1930's, caused by the Great Depression, being an example. But is it high persistence or is it just a period of constantly high volatility? There are few explanations of the exact transmission from fundamental news into market volatility around.

I showed in this thesis that the long memory that is found in virtually all financial data is probably caused by regime changes and that within regimes, the memory of volatility is quite short, that is, mean reversion is fast.¹ If long memory is caused by regime changes, the question naturally arises if economic causes of regime changes can be identified. This question is followed up for the example of exchange rates and foreign exchange interventions in Chapter 7. Another natural question is what determines the short memory within regimes. This question is beyond the scope of this thesis but I will give an outline of possible future research in this direction at the end of this chapter.

1 Summaries

A Mean Reversion Theory of Stock Market Crashes

I use a model that is often employed in the analysis of mean reversion in volatility and apply it to the mean reversion in prices problem. Using daily data of the Dow Jones Industrial Average between 1901 and 2002, I find that mean reversion in prices was a rare but recurring phenomenon in the 20th century. The model can offer inference results for mean reversion on short time horizons only, for longer horizons, the statistical indifference problem mentioned above applies. Mean reversion on short horizons implies mean reversion on long horizons, however, as shown in Chapter 2.

With this model, I study the occurrence of mean reversion around the stock market crash of 1987 in search of evidence for a theory that Fischer Black posed in 1988 as an explanation of the crash.

¹Regime changes are understood as any events that cause the parameter vector of the considered model, in particular GARCH, to change. It is not a political interpretation of regime change, even though these may coincide.

Simply put, the theory states that market participants in 1987 were fooled by a phenomenon of group perception. They saw a series of upward moves (the boom 1982–1987) and inferred that the other investors must have quite low expectations of mean reversion. They could not see, however, that there was a substantial hedge position in the market that protected its owners from a fast mean reversion (which, after a sustained boom, means a sharp downturn). The reason why they could not see it was that the hedge position consisted of synthesized put options, that is, positions in stocks (or futures) and bonds. While so protected, the hedgers could engage in the market as if they expected high returns in the future and thereby profit from the boom while it lasted.

Many investors who only perceived the latter activity got the wrong impression of the market's expected mean reversion and adapted their own, probably much more conservative expectations to these inflated ones. The result was that for quite a while, the market was allowed to operate below its true expected mean reversion velocity and took a large upward swing away from the mean. During the week prior to the crash, however, fundamental news about an array of issues, from the twin deficit in the budget and in the trade balance, new restrictive legislation concerning takeovers, to some shootings in the Persian Gulf, disrupted the markets and caused downside movements of up to ten percent in the course of trading.

These movements triggered the dynamic hedges, which must be adapted constantly in order to be protective, and the large and well-noticed transactions made it plain to the average investor that 1) major players were hedged against a much faster reversion, that is, a falling market, 2) a good deal of the purchases during the boom phase were not motivated by fundamental news but by portfolio insurance and could be regarded as noise, and 3) that the own, adapted mean reversion expectations were too low and better had to be put somewhere into the neighborhood of the conservative *a priori* expectations. The result was that the market had to be put into a position as if the illusion had not happened in the first place and this meant not only a correction in one parameter at that time but a correction over the whole history of the illusion. Mathematically speaking, the crash was the integral of the difference in the two trajectories implied by the two different mean reversion parameters, the “illusioned” and the “disillusioned” one, over the duration of the illusion, which lasted some nine months.

Using a discretized version of an Ornstein-Uhlenbeck process and daily data on the S&P500, I show that the mean reversion parameter estimates jump from obscurity before the crash to high significance afterwards (leaving out the days around the crash itself, which would undoubtedly overestimate the mean reversion). Likelihood ratio tests of the mean reversion model against the null of a random walk confirm this.

The segmentation of boom and exaggeration prior to the crash that is posited

in the report of the Brady-Commission is supported by a changepoint detector study which locates the beginning of the mean reversion illusion at about the change of the years 1986 to 1987.

Simulations of the model with the parameters obtained from estimations on these segments reveal that substantial corrections had quite a high probability. A downward jump of 20 percent or more had a probability of seven percent, a downward jump of ten percent or more had a probability of more than 40 percent.

The stock market crash of 1929, however, cannot be explained by changes in mean reversion expectations. Estimates of the model before and after the crash do not distinguish the mean reversion parameter estimates.

Mean Reversion and Persistence in GARCH(1,1)

I consider mean reversion in volatility using GARCH models. Mean reversion in volatility is a very well documented phenomenon, only the time scales of the correlation structures involved are not agreed upon. There is a broad consensus that volatility has a long scale, or long memory, even models with indefinite memory have been suggested. Nevertheless, there are a couple of accounts of shorter scales in volatility, mainly when non-standard econometric techniques are used or high frequency data are analyzed, or both.

In GARCH models, long time scales in the data are indicated by the sum of the estimates of the autoregressive coefficients, which are close to one in that case. I show that changes in the data generating parameters that occur at unknown points and that are not accounted for in global GARCH(1,1) estimations cause exactly this effect. In other words, even when the data generating persistence is very low, changepoints cause the estimations to indicate long memory. The reason for this huge distortion effect lies solely in the geometry of the problem posed, it has nothing to do with the statistical properties of the estimators.

Simulations confirm this finding. As it is quite likely that over longer periods of time the parameter regimes of financial volatility change, I consider daily data on the Dow Jones Industrial Average and the S&P500 between 1985 and 2001. The hypothesis is that changes in parameter regimes cause the well documented long scale and that as soon as the long scale is filtered out of the data, a short scale appears that corresponds to the fast mean reversion within segments of constant parameters.

Knowing that a global GARCH(1,1) estimation can pick up the long scale only, we can consider a properly defined difference of the measured volatility from the data and the estimated volatility from the GARCH(1,1) model. Spectral estimations of this residual show a short scale of about 6 days for both, the Dow Jones and the S&P500 series. This contrasts to the long scales of about 80–100 days

that are identified by GARCH(1,1). Estimations of the sample autocorrelation functions of these residuals corroborate the findings. In summary, there are two distinct time scales in the data, one of the order of months corresponding to changes in the volatility means and one of the order of days corresponding to the in-segment persistence.

Generalization to GARCH(p,q)

I generalize the analysis of the almost-integration effect to GARCH(2,2), that is, the effect that unknown parameter changes in the data generating structure push the sum of the estimates of the autoregressive parameters of GARCH close to one.

This is a non-trivial step as considerations for ARMA(2,2) models suggest that these models may actually be able to capture two time scales. The argument set forth for GARCH(1,1) however generalizes to GARCH(2,2), such that the almost-integration effect prevents these models from distinguishing different time scales. This phenomenon is also explored in simulation experiments. Besides supporting the main point, these experiments reveal some other interesting finite sample properties of GARCH(2,2) estimations. I find that the higher the data generating persistence, the more precise are the estimations. If the data generating persistence is very low, the start value of the mean reversion parameter causes a large bias.

Japanese Foreign Exchange Interventions and Yen/Dollar Volatility

Finally, I consider the daily exchange rate of the Japanese yen against the U.S. dollar between 1983 and 2002 and investigate it together with newly released data on foreign exchange interventions. The hypothesis is that the interventions mark points where the mean of the volatility of the exchange rate changes, either by causing the change itself or because the interventions occur simultaneously with or shortly after the event that caused the change (like the 1997 Asian crisis).

I use a GARCH model with the interventions series as exogenous variable. Thus accounting for changepoints, the time scale in the exchange rate reduces from 62 days (measured with a simple GARCH(1,1) model) to 6–9 days. Using a changepoint detector to segment the exchange rate series, I find a segmentation that supports this finding: When I estimate a simple GARCH(1,1) model on the obtained segments, the average estimated mean reversion within the segments is 12 days. I emphasize that in this investigation, the short scale is estimated by using traditional econometric techniques only.

2 Future Research

ARMA

The almost integration effect is also dominant for ARMA models. The argument extends in a very straightforward manner. As this is more a topic of time series analysis than of financial markets analysis, I left it out of this thesis. However, as ARMA models are a very general class, I will follow up this argument in a forthcoming paper.

Interpretations of Multiple Scales

The existence of multiple scales in the volatility of financial markets is a more and more recognized fact that still needs economic interpretation. One interpretation suggested in this thesis is that a short scale is generating the data and the long scale is added by discrete macroeconomic events that cause shifts in the volatility mean. There are other possible interpretations. A popular view is that there are different persistences of different fluctuation magnitudes. According to this view, large fluctuations have a short memory (the influence of crashes washes out very quickly) while small fluctuations have a high persistence.

It will be interesting, I think, to try to distinguish these interpretations. One way to go about this is to specify a model in which large and small movements in the volatility process have different persistence coefficients (Threshold-ARCH, or TARARCH, models have this property). Generating synthetic data with it, we will see how the instruments used in this thesis will react to these data. Conversely, synthetic data generated with parameter changes may be analyzed with TARARCH in order to see whether these models can discriminate the different data generating mechanisms. Using both methods on real data may contribute evidence in favor of the one or other interpretation.

This may shed some light on the question how the short time scale in volatility comes about. It would be interesting to formulate a microeconomic model that explains in what way investors implement their investment time horizon into their sales and purchases and thereby into the financial data. This might impose a time scale on stock price volatility.

Changepoint Detection

There are a number of changepoint detectors discussed in the literature. The findings in this thesis suggest that GARCH(1,1) may be used itself for purposes of changepoint detection. Iteratively estimating GARCH(1,1) on a time series,

observation by observation, a changepoint may be set whenever the sum of the estimates of the autoregressive parameters falls within a pre-specified neighborhood of one. It will be a worthwhile investigation to find out whether any statistical statements can be made about changepoints determined this way.

Other changepoint detectors are worth further study. The arguments set out in Section 5.3.a) suggest that the detector used in Chapter 7 is exactly the right number to look at, as it measures the distances in the volatility means.

Secondary Markets Data

In this thesis, only primary markets data, that is, price series, were investigated. As the state of market volatility is more clearly reflected in secondary markets data, that is, option prices, it will be a very interesting subject to search for multiple time scales in options data.

“Structural” Volatility Models

GARCH as well as stochastic volatility models are so-called reduced form models. Their dynamic structure is not designed to reflect economic facts but to capture salient features of the data. Now that we know about the existence of two different scales and that we have competing explanations for it, it will be interesting to explicitly model the time scales, for example by sine/cosine waves of different frequency and try to distinguish the interpretations that way. This would not be a structural model in an economic sense but it would model the time scale more directly.

What Is the Mean Reverting Object?

In this thesis, I treated mean reversion as if it were a phenomenon that is confined to the distinct moment of the return distribution for which I considered it. Is it conceivable, however, that there is only one single mean reverting driver that implies mean reversion in multiple moments? The class of ARCH-in-mean (ARCH-M) models is a step into that direction. It adds the conditional volatility into the conditional mean equation. This question will probably lead closer to the “missing link” between mean reversion in prices and returns on the one side and mean reversion in volatility on the other.

Generalization of the Crash Theory

In the mean reversion theory of stock market crashes presented in this thesis, I made a tacit assumption. When market participants have, on average, a certain level of mean reversion expectations, their actions (sales and purchases) transform these expectations into market prices. In other words, I assume that mean reversion is endogenous. Is that so? The model I have in mind is one of heterogeneous agents with rational expectations, or possibly rational beliefs, of mean reversion. I assume that agents incorporate their expectations and preconceptions into the data, that financial time series are measurements of social constructions. That is, there is no “physical” process that determines the series and that is measured with error. The data generating process largely depends on what the agents think that it is.² It does not matter whether the agents understand how this transformation happens and what it implies, or not. A model like this would generalize the theory beyond the explanation of one single event and might be able to explain why mean reversion seems to occur infrequently but recurrently.

In summary, there are many open questions about mean reversion in financial markets and the subject still promises to be fruitful. In this thesis, I attempted to carefully raise some more questions and I hope that I could also contribute to answer some.

²As an example of this phenomenon consider option prices before the stock market crash of 1987. They behaved more like the commonly used Black-Scholes model than after the crash. Then, the volatility “smirk” appeared, that is, the empirical fact that out-of-the-money put options are higher priced than according to Black-Scholes. This reflects the fact that large downward deviations have a much higher probability than predicted by a Gaussian error as assumed by Black-Scholes, as investors learned in 1987.

Bibliography

- [1] ANDERSEN, T. G. / BOLLERSLEV, T. (1997). *Intraday Periodicity and Volatility Persistence in Financial Markets*. *Journal of Empirical Finance* 4: 115–158.
- [2] ANDREOU, E. / GHYSELS, E. (2001). *Detecting Multiple Breaks in Financial Market Volatility Dynamics*. Cirano working paper series.
- [3] ARNOLD, L. (1973). **Stochastische Differentialgleichungen: Theorie und Anwendung**. Oldenbourg: München.
- [4] BAILLIE, R. / BOLLERSLEV, T. (1989). *The Message in Daily Exchange Rates: A Conditional-Variance Tale*. *Journal of Business and Economic Statistics* 7(3): 297–305.
- [5] BAILLIE, R. T. / DEGENNARO, R. P. (1990). *Stock Returns and Volatility*. *Journal of Financial and Quantitative Analysis* 25(2): 203–214.
- [6] BAILLIE, R. T. / BOLLERSLEV, T. / MIKKELSEN, H. O. (1996). *Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity*. *Journal of Econometrics* 74: 3–30.
- [7] BARSKY, R. B. / DELONG, J. B. (1993). *Why Does the Stock Market Fluctuate?* *Quarterly Journal of Economics* 108(2): 291–311.
- [8] BASU, P. (1993). *Mean Reversion in GNP and Stock Prices: An Adjustment Cost Hypothesis*. *Kyklos* 46(1): 87–104.
- [9] BASU, P. / VINOD, H. D. (1994). *Mean Reversion in Stock Prices: Implications from a Production Based Asset Pricing Model*. *Scandinavian Journal of Economics* 96(1): 51–65.
- [10] BLACK, F. (1976). *Studies of Stock Price Volatility Changes*. American Statistical Association, *Proceedings of the Business and Economic Statistics Section* 177–181.
- [11] BLACK, F. (1988). *An Equilibrium Model of the Crash*. *NBER Macroeconomics Annual*: 269–275.

-
- [12] BLACK, F. (1989). *Mean Reversion and Consumption Smoothing*. NBER Working Paper 2946, April 1989. Also in: *Review of Financial Studies* 1990. 3(1): 107–114.
- [13] BODMER, D. (1996). “*Mean Reversion*” und “*Time-Varying Expected Returns*” in internationalen Aktienmärkten: *Theorie und empirische Evidenz*. Dissertation Nr. 1868, Universität St. Gallen.
- [14] BOLLERSLEV, T. (1986). *Generalized Autoregressive Conditional Heteroskedasticity*. *Journal of Econometrics* 31: 307–327.
- [15] BOLLERSLEV, T. (1987). *A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return*. *Review of Economics and Statistics* 69: 542–547.
- [16] BOLLERSLEV, T. / WOOLDRIDGE, J. M. (1992). *Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances*. *Econometric Reviews* 11(2): 143–172.
- [17] BOLLERSLEV, T. / ENGLE, R. F. (1993). *Common Persistence in Conditional Variances*. *Econometrica* 61(1): 167–186.
- [18] BRADY, N.F. / COTTING, J.C. / KIRBY, R.G. / OPEL, J.R. / STEIN, H.M. / GLAUBER, R.R. (1988). *Report of the Presidential Task Force on Market Mechanisms*. U.S.-Government Printing Office.
- [19] CAMPBELL, J. Y. / MANKIW, N. G. (1987). *Are output fluctuations transitory?* *Quarterly Journal of Economics* 102: 857–880.
- [20] CAMPBELL, J.Y. / LO, A.W. / MACKINLAY, A.C. (1997). **The Econometrics of Financial Markets**. Princeton University Press: New Jersey.
- [21] CECCHETTI, S. G. / LAM, P.-S. / MARK, N. C. (1990). *Mean Reversion in Equilibrium Asset Prices*. *American Economic Review* 80(3): 398–418.
- [22] CHERNOV, M. / GALLANT, A.R. / GHYSELS, E. / TAUCHEN, G. (2002). *Alternative Models for Stock-Price Dynamics*. mimeo.
<http://www.ssc.upenn.edu/~fdiebold/w2002/program.htm>
- [23] CHOI, K. / ZIVOT, E. (2002). *Long Memory and Structural Changes in the Forward Discount: An Empirical Investigation*. mimeo.
<http://faculty.washington.edu/ezivot/ezresearch.htm>.
- [24] CUTLER, D. M. / POTERBA, J. M. / SUMMERS, L. H. (1991). *Speculative Dynamics*. *Review of Economic Studies* 58: 529–546.
- [25] DEBONDT, W. / THALER, R. (1987). *Does the Stock-Market Overreact?* *Journal of Finance* 40: 793–805.

-
- [26] DIEBOLD, F. X. (1986). *Modeling the Persistence of Conditional Variances: A Comment*. *Econometric Reviews* 5(1): 51–56.
- [27] DIEBOLD, F. X. / INOUE, A. (2001). *Long Memory and Regime Switching*. *Journal of Econometrics* 105: 131–159.
- [28] DING, Z. / GRANGER, C. W. J. / ENGLE, R. F. (1993) *A Long Memory Property of Stock Market Returns and a New Model*. *Journal of Empirical Finance* 1: 83–106.
- [29] DING, Z. / GRANGER, C. W. J. (1996). *Modeling Volatility Persistence of Speculative Returns: A New Approach*. *Journal of Econometrics* 73: 185–215.
- [30] DOMINGUEZ, K. / FRANKEL, G. (1993). *Does Foreign Exchange Intervention Work?* Working Paper. Institute for International Economics, Washington.
- [31] DOMINGUEZ, K. (1998). *Central Bank Intervention and Exchange Rate Volatility*. *Journal of International Money and Finance* 17: 161–190.
- [32] DRĂGULESCU, A. A. / YAKOVENKO, V. M. (2002). *Probability Distribution of Returns for a Model with Stochastic Volatility*. mimeo.
<http://lanl.arXiv.org/abs/cond-mat/0203046>.
- [33] DUFFEE, G. R. (1991). *A New Test for Mean Reversion in Stock Prices*. Federal Reserve Board, Finance and Economics Discussion Series No 152.
- [34] ENGLE, R. F. (1982). *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*. *Econometrica* 50(4): 987–1007.
- [35] ENGLE, R. F. / BOLLERSLEV, T. (1986). *Modelling the Persistence of Conditional Variances*. *Econometric Reviews* 5(1): 1–50.
- [36] ENGLE, R. F. / LILIEN, D. M. / ROBINS, R. P. (1987) *Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model*. *Econometrica* 55(2): 391–407.
- [37] ENGLE, R. F. / PATTON, A. J. (2001). *What good is a volatility model?* *Quantitative Finance* 1(2): 237–245.
- [38] FAMA, E. F. (1965). *The Behavior of Stock-Market Prices*. *Journal of Business*, 38(1): 34–105.
- [39] FAMA, E. F. (1970). *Efficient Markets: A Review of Theory and Empirical Work*. *Journal of Finance*, 25: 383–417.
- [40] FAMA, E. F. / FRENCH, K. R. (1988). *Permanent and Temporary Components of Stock Prices*. *Journal of Political Economy* 96: 247–273.

-
- [41] FOUQUE, J. P. / PAPANICOLAOU, G. / SIRCAR, K. R. / SØLNA, K. (2002). *Short time-scale in S&P 500 Volatility*. To appear in the *Journal of Computational Finance*. <http://georgep.stanford.edu/~papanico/pubs.html>
- [42] FISHER, I. (1933). *The Debt-Deflation Theory of Great Depressions*. *Econometrica* 1: 337–357.
- [43] FRANCO, C. / ROUSSIGNOL, M. / ZAKOÏAN, J.-M. (2001). *Conditional Heteroskedasticity Driven by Hidden Markov Chains*. *Journal of Time Series Analysis* 22(2): 197–220.
- [44] GALATI, G. / MELICK, W. (1999). *Perceived Central Bank Intervention and Market Expectations: An Empirical Study of the Yen/Dollar Exchange Rate 1993–1996*. Working Paper No 77. Bank for International Settlement, Basel.
- [45] GALLANT, A. R. / TAUCHEN, G. (2001). *Efficient Method of Moments*. mimeo. <http://www.unc.edu/~arg>.
- [46] GOURIÉROUX, C. (1997). **ARCH Models and Financial Applications**. Springer: New York.
- [47] GRANGER, C. W. J. (1980). *Long Memory Relationships and the Aggregation of Dynamic Models*. *Journal of Econometrics* 14: 227–238.
- [48] GRANGER, C. W. J. / HYUNG, N. (1999). *Occasional Structural Breaks and Long Memory*. UC San Diego Discussion Paper 99-14.
- [49] GRANGER, C. W. J. / TERÄSVIRTA, T. (2001). *A Simple Nonlinear Time Series Model with Misleading Linear Properties*. *Economics Letters* 62: 161–165.
- [50] GRAY, S. F. (1996). *Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process*. *Journal of Financial Economics* 42: 27–62.
- [51] GREENE, W. H. (2000). **Econometric Analysis**. 4th ed., Prentice-Hall: New Jersey.
- [52] GRIFFIN, D. / TVERSKY, A. (1992). *The weighing of evidence and the determinants of confidence*. *Cognitive Psychology* 24: 411–435.
- [53] HAMILTON, J. D. / SUSMEL, R. (1994). *Autoregressive Conditional Heteroskedasticity and Changes in Regime*. *Journal of Econometrics* 64: 307–333.
- [54] HARVEY, A. C. (1976). *Estimating Regression Models with Multiplicative Heteroscedasticity*. *Econometrica* 44(3): 461–465.
- [55] HILLEBRAND, E. / SCHNABL, G. (2003). *Reassessing Japanese Foreign Exchange Intervention: Evidence from GARCH and changepoint detection*. mimeo. <http://math.stanford.edu/~erhil>.

-
- [56] JACKLIN, C. J. / KLEIDON, A. W. / PFLEIDERER, P. (1992). *Underestimation of Portfolio Insurance and the Crash of 1987*. *Review of Financial Studies* 5: 35–64.
- [57] JEGADEESH, N. (1990). *Evidence of Predictable Behavior of Security Returns*. *Journal of Finance* 45(3): 881–898.
- [58] KEYNES, J. M. (1936). **The General Theory of Employment, Interest, and Money**. Macmillan: London.
- [59] KIM, C.-J. / NELSON, C. R. / STARTZ, R. (1998). *Testing for Mean Reversion in Heteroskedastic Data Based on Gibbs-sampling-augmented Randomization*. *Journal of Empirical Finance* 5: 131–154.
- [60] KIM, C.-J. / NELSON, C. R. (1998). *Testing for Mean Reversion in Heteroskedastic Data II: Autoregression Test Based on Gibbs-sampling-augmented Randomization*. *Journal of Empirical Finance* 5: 385–396.
- [61] KIRMAN, A. / TEYSSIERE, G. (2000). *Microeconomic Models for Long Memory in the Volatility of Financial Time Series*. mimeo. <http://members.aol.com/gteyssiere>.
- [62] KOKOSZKA, P. / LEIPUS, R. (1999). *Testing for Parameter Changes in ARCH Models*. *Lithuanian Mathematical Journal* 39(2): 182–195 (in translation).
- [63] KOKOSZKA, P. / LEIPUS, R. (2000). *Changepoint estimation in ARCH models*. *Bernoulli* 6(3): 513–539.
- [64] LAI, T. L. (1995). *Sequential Changepoint Detection in Quality Control and Dynamical Systems*. *Journal of the Royal Statistical Society B* 57(4): 613–658.
- [65] LAMOUREUX, C. G. / LASTRAPES, W. D. (1990). *Persistence in Variance, Structural Change, and the GARCH Model*. *Journal of Business and Economic Statistics* 8(2): 225–234.
- [66] LEBARON, B. (2001). *Stochastic Volatility as a Simple Generator of Apparent Financial Power Laws and Long Memory*. *Quantitative Finance* 1(6): 621–631.
- [67] LO, A.W. / MACKINLAY, A.C. (1988). *Stock Prices do not Follow Random Walks: Evidence from a Simple Specification Test*. *Review of Financial Studies* 1:41–66.
- [68] LOBATO, I. N. / SAVIN, N. E. (1998). *Real and Spurious Long-Memory Properties of Stock-Market Data*. *Journal of Business and Economic Statistics* 16(3): 261–268.
- [69] LUMSDAINE, R. (1996). *Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models*. *Econometrica* 64(3): 575–596.

-
- [70] MANDELBROT, B. (1963). *The Variation of Certain Speculative Prices*. *Journal of Business* 36(4): 394–419.
- [71] MERTON, R.C. (1980). *On Estimating the Expected Return on the Market*. *Journal of Financial Economics* 9: 323–361.
- [72] MIKOSCH, T. / STARICA, C. (2000). *Change of Structure in Financial Time Series, Long Range Dependence and the GARCH Model*. University of Aarhus, Aarhus School of Business, Centre for Analytical Finance, Working Paper No. 58.
- [73] METCALF, G. E. / HASSETT, K.A. (1995). *Investment under Alternative Return Assumptions. Comparing Random Walks and Mean Reversion*. *Journal of Economic Dynamics and Control* 19: 1471–1488.
- [74] MITCHELL, M.L. / NETTER, J.M. (1989). *Triggering the 1987 Stock Market Crash: Antitakeover Provisions in the Proposed House Ways and Means Tax Bill?* *Journal of Financial Economics* 24: 37–68.
- [75] NELSON, D.B. (1990). *ARCH Models as Diffusion Approximations*. *Journal of Econometrics* 45: 7–38.
- [76] NELSON, D. B. (1990a). *Stationarity and Persistence in the GARCH(1,1) Model*. *Econometric Theory* 6: 318–334.
- [77] POTERBA, J.M. / SUMMERS, L.H. (1988). *Mean Reversion in Stock-Prices: Evidence and Implications*. *Journal of Financial Economics* 22: 27–59.
- [78] PRESS, W. H. / TEUKOLSKY, S. A. / VETTERLING, W. T. / FLANNERY, B. P. (2002). **Numerical Recipes in C++**. Cambridge University Press: Cambridge, U.K.
- [79] PRIESTLEY, M. B. (1981). **Spectral Analysis and Time Series**. Academic Press: San Diego.
- [80] ROGOFF, K. (1984). *On the Effects of Sterilized Intervention: An Analysis of Weekly Data*. *Journal of Monetary Economics* 14: 133–150.
- [81] SAKOULIS, G. / ZIVOT, E. (2000). *Time-Variation and Structural Change in the Forward Discount: Implications for the Forward Rate Unbiasedness Hypothesis*. mimeo. <http://faculty.washington.edu/ezivot/ezresearch.htm>.
- [82] SUMMERS, L. H. (1986). *Does the Stock-Market Rationally Reflect Fundamental Values?* *Journal of Finance* 41: 591–601.
- [83] TVERSKY, A. / KAHNEMAN, D. (1981). *The Framing of Decisions and the Psychology of Choice*. *Science* 211: 453–458.
- [84] VASICEK, O. (1977). *An Equilibrium Characterisation of the Term Structure*. *Journal of Financial Economics* 5: 177–188.

-
- [85] WEISS, A. A. (1986). *Asymptotic Theory for ARCH Models: Estimation and Testing*. *Econometric Theory* 2: 107–131.
 - [86] WHITE E.N. (1990). *The Stock Market Boom and Crash of 1929 Revisited*. *Journal of Economic Perspectives* 4(2): 67–83.
 - [87] WHITE, H. (1982). *Maximum Likelihood Estimation of Misspecified Models*. *Econometrica* 50: 1–25.
 - [88] WHITE, H. (2001). **Asymptotic Theory for Econometricians**. Academic Press: San Diego.

The net effect of market transactions of investors buying a stock and simultaneously replicating a put option on it is positive. That is, the purchases are greater than the sales.

This will be shown here for the case of a European put option. According to the Black-Scholes model, the replicating portfolio of a European put on one share of the underlying stock consists of a short position of $|\Delta(t)|$. Δ is the sensitivity of the option to changes in the price of the underlying given by

$$\Delta(t) = \Phi(d_1) - 1 < 0,$$

$$d_1 = \frac{\log \frac{S}{X} + (r + \frac{1}{2}\sigma^2)(T - t)}{\sigma\sqrt{T - t}}.$$

Φ is the cumulative distribution function of the standard normal distribution, S is the stock price, X is the exercise price of the put option, r is the risk-free interest rate, $T - t$ is the time to maturity and σ^2 is the variance of the stock price.

The proceeds from the short position are invested and gain the risk-free interest rate r . Assume that the investor hedges every single stock that he buys. His position $P(t)$ then is (in terms of inventories)

$$P(t) = (S - S\Delta(t)e^{rt}) \cdot n,$$

where n denotes the number of shares. The assertion made here is equivalent to

$$\frac{1}{n}P(t) > 0 \iff S > S\Delta(t)e^{rt}.$$

Now, it is obvious that

$$1 + e^{-rt} > \Phi(d_1),$$

as the exponential function is strictly positive on \mathbb{R} and $\Phi(d_1) \in [0, 1]$ as it is a probability. It follows that

$$1 > (\Phi(d_1) - 1)e^{rt} \implies 1 > \Delta(t)e^{rt}.$$

Multiplying with $S > 0$ proves the assertion.

The expected value of the process solving model (3.3.1) is given by $\vartheta_t = S_0 e^{\mu t}$.

Rewrite (3.3.1) to

$$dS_t = (\mu - \lambda)S_t dt + \lambda\vartheta_t dt + \sigma S_t dW_t,$$

and solve the associated homogeneous equation

$$dX_t = (\mu - \lambda)X_t dt + \sigma X_t dW_t,$$

to obtain $X_t = \exp[(\mu - \lambda - \sigma^2/2)t + \sigma W_t]$. Then the solution to (3.3.1) is given by

$$\begin{aligned} S_t &= X_t \left(S_0 + \int_0^t (X_u)^{-1} \lambda \vartheta_u du \right) \\ &= S_0 \exp \left[\left(\mu - \lambda - \frac{\sigma^2}{2} \right) t + \sigma W_t \right] \left(1 + \lambda \int_0^t \exp \left[\left(\lambda + \frac{\sigma^2}{2} \right) u - \sigma W_u \right] du \right) \end{aligned}$$

Taking expectations, we obtain

$$\begin{aligned} \mathbb{E}S_t &= S_0 e^{(\mu - \lambda - \frac{\sigma^2}{2})t} \left(\mathbb{E}e^{\sigma W_t} + \lambda \int_0^t e^{(\lambda + \frac{\sigma^2}{2})u} \mathbb{E}e^{\sigma(W_t - W_u)} du \right) \\ &= S_0 e^{\mu t - \lambda t} + S_0 e^{\mu t - \lambda t} \lambda \int_0^t e^{\lambda u} du \\ &= S_0 e^{\mu t}. \end{aligned}$$

Model (3.3.2) is a first-order approximation to model (3.3.1).

The mean reversion term in the model (3.3.1) can be rewritten as

$$\lambda \frac{\tilde{\vartheta}_t - S_t}{S_t} dt = \lambda \left(\frac{\tilde{\vartheta}_t}{S_t} - 1 \right) dt.$$

Denote $r := \tilde{\vartheta}_t/S_t - 1$, then

$$1 + r = \frac{\tilde{\vartheta}_t}{S_t}$$

and as $\log(1 + r) \doteq r$ we have a first-order equivalent representation

$$\lambda \frac{\tilde{\vartheta}_t - S_t}{S_t} dt \doteq \lambda \log \frac{\tilde{\vartheta}_t}{S_t} dt = \lambda (\log \tilde{\vartheta}_t - \log S_t) dt.$$

From Ito's Lemma, we have

$$d \log S_t = \frac{dS_t}{S_t} - \frac{\sigma^2}{2} dt.$$

Define $\tilde{\mu} = \mu - \sigma^2/2$ and $\vartheta_t := S_0 \exp(\tilde{\mu} t)$. Then there is a first-order equivalent of the model (3.3.1) given by (3.3.2):

$$\log S_t = \log S_0 + \tilde{\mu} t + \lambda \int_0^t (\log \vartheta_u - \log S_u) du + \sigma W_t. \quad (0.0.1)$$

In this Appendix I will derive the Lorentzian model

$$h(w) = a + b/(c^2 + w^2)$$

for the power spectrum of the GARCH(1,1) process as used in Section 5.4. (a, b, c) are parameters and w denotes the frequencies.

Proposition 12. *The power spectrum of the log of ε_t^2 in the continuous time analogue of a Gaussian GARCH(1,1) model with constant mean return can be represented by the function*

$$h(w) = \frac{\gamma^2}{2\pi} + \frac{\alpha^2}{2\pi} \frac{1}{w^2 + \vartheta^2}$$

where $\vartheta \approx 1 - \alpha - \beta$, γ^2 is the variance of $\log \eta_t^2$, $\eta_t \sim \mathcal{N}(0, 1)$, and w denotes the frequencies.

Proof. Nelson (1990) showed that the discrete GARCH(1,1) model with constant mean return converges with $\Delta t \rightarrow 0$ in distribution to the system of stochastic differential equations

$$\begin{aligned} dY(t) &= \sigma(t)dW_1(t) \\ d\sigma^2(t) &= (\omega - \vartheta\sigma^2(t))dt + \alpha\sigma^2(t)dW_2(t), \end{aligned}$$

where $Y_t = \sum_{i=0}^t (r_i - \mu)$ are the cumulative excess returns, σ_t^2 is the volatility process, ω and α are the discrete GARCH(1,1) parameters and $\vartheta \approx 1 - \alpha - \beta$. $W_1(t)$, $W_2(t)$ are two independent Brownian Motions.

Taking the log of the volatility driver and denoting $V_t = \log \sigma^2(t)$, $f(V_t) = \sqrt{\exp(V_t)}$, and $m = \log(\omega/\vartheta) - \alpha^2/2\vartheta$, I have the first order equivalent

$$\begin{aligned} dY_t &= f(V_t)dW_1(t), \\ dV_t &= \vartheta(m - V_t)dt + \alpha dW_2(t). \end{aligned}$$

V_t is an Ornstein-Uhlenbeck process with solution

$$V_t = m + (V_0 - m)e^{-\vartheta t} + \alpha \int_0^t e^{-\vartheta(t-s)} dW_2(s).$$

As described for example in Arnold (1973), the correlation with respect to the stationary measure is given by

$$\text{cov}(V_s, V_{s+t}) = \frac{\alpha^2}{2\vartheta} e^{-\vartheta t} \text{ for } s \rightarrow \infty \text{ and } t > 0.$$

I discretize the volatility process with $\Delta t = 1$ and obtain

$$Y_t - Y_{t-1} = \varepsilon_t = r_t - \mu = \sqrt{e^{V_t}} \eta_t, \quad \eta_t \sim \mathcal{N}(0, 1).$$

This motivates the transformation

$$x_t := \log \varepsilon_t^2 = V_t + \log \eta_t^2,$$

$\log \eta_t^2$ being White Noise with mean zero and variance γ^2 .

The autocorrelation of x_t is then given by

$$R_x(t) = \frac{\alpha^2}{2\vartheta} e^{-\vartheta t} + \gamma^2 \delta_0(t),$$

where $\delta_0(t)$ is the Dirac-function with unit mass at zero.

According to the Wiener-Khintchine theorem, the power spectrum of the real process x_t has the form

$$\begin{aligned} h(w) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iwt} \left(\frac{\alpha^2}{2\vartheta} e^{-\vartheta t} + \gamma^2 \delta_0(t) \right) dt \\ &= \frac{\gamma^2}{2\pi} + \frac{1}{2\pi} \frac{\alpha^2}{\vartheta} \operatorname{Re} \int_0^{\infty} e^{-(iw+\vartheta)t} dt \\ &= \frac{\gamma^2}{2\pi} + \frac{\alpha^2}{2\pi} \frac{1}{w^2 + \vartheta^2} \end{aligned} \quad \square$$

Simplifying this to $h(w) = a + b/(w^2 + c^2)$, I can recover the e -folding time by $1/c$.