Université Catholique de Louvain
Faculté des Sciences Économiques

Econometric Analysis of Financial Count Data and Portfolio Choice: A Dynamic Approach

Thèse présentée pour l'obtention du grade de

Docteur en Sciences Economiques

Erick W. RENGIFO

Promoteur:	L. BAUWENS
Jury:	P. GIOT
	B. LEHMANN
	O. SCAILLET
	L. SIMAR

To Ada and Elizabeth, with love

Acknowledgements

This thesis represents a personal achievement of an old dream. It presents the results of some years of research. I would like to thank the invaluable physical and spiritual support of my lovely wife Ada and my daughter Elizabeth. The constant impulse of my mother Mirtha and my grandmother Carmen and, the higher compromise that my sisters Rita and Lucero make me feel as the oldest brother.

I would like to thank my thesis supervisor, professor Luc Bauwens, who trusted me and gave me the opportunity to do my PhD. I also would like to thank him because he not only guided me since the first moment, but through his constant encouragement kept me on the track. His suggestions and ideas made this thesis richer, both in content and in form. Thanks to Andreas Heinen with whom I have worked a large part of this thesis. I enjoyed working with him, I learned a lot and I really hope to continue doing joint research in the future. Special thanks to Jeroen Rombouts, the co-author of one of the chapter of this thesis with whom we plan to bust the portfolio allocation literature.

I deserve special gratitude to all the members of my thesis committee for their useful comments and suggestions. Thanks to Pierre Giot, Bruce Lehmann, Olivier Scaillet and Léopold Simar.

Finally, I would like to thank all the people of CORE and IRES who have been working here during my stay. I really appreciate the friendship and interesting comments on the work that I was doing. On these lines I want to keep the memory of Antonio, Alfonso, Diego, Fausto, Genaro, Giulio, Helena and Malika. I also want to remember good friends from my master program: Antonio, Alberto, Juan Carlos, among others. I own them the encourage to follow my lines of research.

Introduction

This thesis contributes to the econometric literature in two ways. Firstly, it introduces a new multivariate count model that presents advances in several aspects relative to the existing literature. Our multivariate time series count model can deal with issues of discreteness, overdispersion (variance greater than the mean) and both cross- and serial correlation, all at the same time. We follow a fully parametric approach and specify a marginal distribution for the counts where, conditionally on past observations the means follow a vector autoregressive process (VAR). This enables to attain improved inference on coefficients of exogenous regressors relative to the static Poisson regression, while modelling the serial correlation in a flexible way. The method is also innovative in the use of copulas, which is a relatively new tool in economics and which is designed to build a correlation structure between variables with given marginal distributions. This makes it possible to model the contemporaneous correlation between individual series in a very flexible way. Secondly, this thesis introduces a new approach to estimate the multivariate reduced rank regressions when the normality assumption is not satisfied. We propose to use the copula tool to generate multivariate distributions and, we show that this method can be applied in multivariate settings when a set of K_1 dependent variables is believed to depend on a set of K_2 explanatory variables, including when data have different distributions.

In terms of financial literature, this thesis provides two contributions. Firstly, with our multivariate count model we analyze diverse market microstructure issues about the submission of different types of orders. With this model, we can fully take into account the interactions between submissions of the various types of orders, which represents an advantage with respect to univariate models such as the autoregressive conditional duration model of Engle and Russell (1998). Secondly, it contributes to portfolio research proposing a new dynamic optimal portfolio allocation model in a Value-at-Risk setup. This model allows for time varying skewness and kurtosis of portfolio distributions and the model parameters are estimated by weighted maximum likelihood in an increasing window setup. This last property allows us to have more accurate portfolio recommendations in terms of the amount to invest in the risk-free interest rate and in the risky portfolio.

This thesis is based on four original papers presented in separate chapters divided in two parts. Part I presents the multivariate count model and the reduced rank regression in non-Gaussian contexts and shows their usefulness in microstructure research. Part II is devoted to the dynamic optimal portfolio selection in a VaR framework. In each of these parts there is and introductory chapter (Chapters 1 and 5) that presents the generalities and the existing literature related with the respective topics.

In Chapter 2 we introduce the multivariate autoregressive conditional double Poisson model (MDACP) to deal with discreteness, overdispersion and both auto- and cross-correlation, arising with multivariate counts. We model counts with a double Poisson marginal distribution and assume that conditionally on past observations the means follow a vector autoregression. We resort to copulas to introduce contemporaneous correlation. We use the model to study the impact of sector and stock specific news on the comovements in the number of trades per unit of time of the most important US department stores traded on the New York Stock Exchange (NYSE). We show that the market leaders inside a specific sector are related, in terms of their informational contents, to their size measured by market capitalization.

Chapter 3 contains an empirical analysis of the trading activity in an open automated auction market. We study how liquidity supply, informational factors, and price volatility impact on order submission and cancellation decisions, and discuss the results in the light of predictions implied by theoretical models of financial market microstructure. Using time series of reconstructed limit order books we identify latent factors which should explain, according to hypotheses put forth by microstructure theory, future trading activity. We test those hypotheses by employing the econometric model introduced in Chapter 2 for the analysis of multivariate count processes.

In Chapter 4 we propose a new procedure to perform reduced rank regressions (RRR) in non-Gaussian contexts, based on multivariate dispersion models. Reduced-rank multivariate dispersion models (RR-MDM) generalize RRR to a very large class of distributions, which include continuous distributions like the normal, Gamma, inverse Gaussian, and discrete distributions like the Poisson and the negative binomial. A multivariate distribution is created with the help of the Gaussian copula and estimation is performed using maximum likelihood. We show how this method can be amended to deal with the case of discrete data. We perform Monte Carlo simulations and show that our estimator is more efficient than the traditional Gaussian RRR estimator. In the framework of MDM we introduce a procedure analogous to canonical correlations, which takes into account the distribution of the data. We present a related microstructure application on the number of trades of five US department stores traded on the New York Stock Exchange during the year 1999. We are interested in the common factor underlying the trading activity behavior of the assets, i.e. the sector specific news. This analysis is helpful to identify leaders in a given sector from the point of view of dissemination of sector-based information.

Finally, in Chapter 6 we propose a dynamic portfolio selection model that maximizes expected returns subject to a Value-at-Risk constraint. The model allows for time-varying skewness and kurtosis of portfolio distributions. Estimation of the model parameters is done by weighted maximum likelihood in an increasing window setup. We determine the best daily investment recommendations in terms of percentage to borrow or lend and the optimal weights of the assets in the risky portfolio. Two empirical applications illustrate in an out-of-sample context which models are preferred from a statistical and economic point of view.

Contents

Part I

Dynamic count data models: application to financial market microstructure

Chapter 1

Market microstructure

This chapter presents a general overview of theoretical and empirical approaches to market microstructure and, the most important time series econometric models created to understand and test these theories and hypotheses. Without aiming to be exhaustive, this overview is important to put in perspective and to have a clear understanding of the work presented in Chapters 2, 3 and 4.

The remainder of the chapter is organized as follows. Section 1.1 presents basic definitions of market microstructure. Section 1.2 discusses some theoretical models of market microstructure. Section 1.3 describes the autoregressive conditional duration model, discusses some existing literature on count models and describes the intensity and ordered probit models. Section 1.4 presents a glossary of the most important definitions used all over the chapter.

The basic references for Sections 1.1 and 1.2 are given by O'Hara (1995), Madhavan (2000), Bauwens and Giot (2001) and Hasbrouck (2004). For Section 1.3, Engle and Russell (1998) for the autoregressive conditional duration model, Cox and Isham (1980), Cameron and Trivedi (1998) and Bauwens and Hautsch (2003) for the intensity model, and Hausman, Lo, and Mackinlay (1992) for the ordered probit model. I refer the reader to those references on which I base my exposition for more details about the concepts presented in this chapter.

1.1 Basic definitions

Market microstructure studies the processes by which investors' demands and supplies are translated into transactions of assets under explicit trading rules. This literature studies how the price-setting rules evolve in a given trading mechanism, in order to understand how these mechanisms affect the price formation process. Market microstructure is important for determining the transparency of the markets as well as for market regulation purposes and for the design of new trading mechanisms. However, its importance goes beyond these topics. It has key implications for the study of corporate finance, asset pricing and international finance. It is also related to the field of investments in that it studies the variety of frictions present in asset prices that make them to differ from the expected asset value.

In this context, quickly evolving technological, structural and regulatory changes affecting the securities worldwide, the globalization and inter-market competition, represent nowadays challenges for further research in this field of study. Indeed, this field has grown tremendously in the academic literature since 1990.

Next in this section, I present the basic notions of markets (1.1.1), types of orders (1.1.2), price settings (1.1.3) and the concept of market liquidity (1.1.4).

1.1.1 Markets

In general, markets are categorized as being price driven (also called dealership markets), order driven (also called double auction markets or order book markets) or hybrid markets (combining features from price driven and order driven markets).

A price driven market is conducted by one or several market makers who buy and sell the assets. They can be exchange officials (such as the specialist at the NYSE) or a trader working for a bank (NASDAQ, FOREX). The market maker provides liquidity to the market by posting quotes at the bid and ask sides of the market. The market maker holds a given inventory position that results from the difference of his buying and selling. This inventory position is a source of potential risk for him. For example, if a market maker holds a large inventory when the asset prices fall, he incurs a loss. To reward the market maker for this potential risk, the exchange usually grants him with some advantages: the spread (the ask is higher than the bid) and an informational advantage (market makers usually have more information regarding the existing orders than the traders).

In an order driven market no market maker is involved in the trading process. The orders are directly entered into the electronic order book maintained by the exchange. Trades occur when orders match. Although no market maker is present, a spread also exists. In this type of market the limit orders are executed based on a strict price-time priority. This ensures that the most favorable trades are always executed first. Examples of this kind of markets are the Paris Bourse, the Frankfurt XETRA trading system and the Toronto Stock Exchange, among others.

1.1.2 Type of orders

The types of orders varies according to each specific exchange. Section 1.4 presents most of them. However, the basic types of orders are the following:

- **Market orders**, orders to buy or sell a given number of shares at the current market price.

- **Limit orders**, orders to buy or sell a given number of shares at some pre-specified price. Any non-immediately executed limit order is held in the market maker's book or in the limit order book until the prices moves to the designated level.

- **Cancellations**: not executed limit orders are cancelled according to a given execution condition. If there is no execution condition, limit orders stay in the system until the maximum validity, established for each particular exchange, is reached. For example, in Frankfurt XETRA trading system the maximum validity is one year, after which the limit orders are immediately cancelled.

1.1.3 Economic mechanisms for price settings

Traditional economic theory states that prices in competitive markets are determined by the intersection of the supply and demand curves. Even though this is true in equilibrium, this theory is not able to determine how this equilibrium is attained. In this respect, economic theory has two lines of thought to explain the mechanics of the price formation process. The first line of thought, found for example in the rational expectations literature, argues that since many economic studies involve the analysis of equilibrium, what matters for many economic questions are the properties of the equilibrium prices. Accordingly, the trading mechanism has no relevance at all for the determination of the equilibrium price, i.e. whatever the trading mechanism employed, the same equilibrium will arise. This is particularly troubling for markets in which the traders have different information, for example.

The second line of thought is that the price formation process could be captured by a general representation of a Walrasian auctioneer, who aggregates traders' supplies and demands in order to obtain the equilibrium price. In this simple trading mechanism, the Walrasian auctioneer aggregates the supplies and demands and then announces the first potential trading price. Given this price, traders revise their supplies and demands, which leads to a revision of the prices by the auctioneer. The process iterates until the total excess demand is zero and the equilibrium price is obtained. Just in this final case, trades among buyers and sellers occur.[1]

The question with the last line of thought is whether this in fact captures the real process by which the prices are formed. In practice, this mechanism is rarely encountered (the London gold "fixing" is the most important example), and there are a lot of markets that differ from it. Moreover, as Demsetz (1968) argued, trade may involve some costs. These costs could be explicit, such as the charges of a particular market, or implicit, reflecting costs related with the immediate execution of the trade. These costs are referred as price for immediacy and arise because trading has a time dimension, not considered by the Walrasian auction. In particular, at any given moment in time, the number of shares to sell may not be equal to the number of shares to buy. If this happens and if traders want to trade immediately, the imbalance of trade would make it impossible to find a market-clearing price at that specific time. Demsetz (1968) argued that this lack of equilibrium could be offset by paying for immediacy. If there are traders who want to buy immediately at some price and there are traders who want to sell but not at that given time, then the former would increase the price in order to induce more sellers to trade now. Similarly if there is an imbalance of traders wanting to sell now and no traders wanting to buy, the former should decrease the prices in order to convince the buyers

[1] This kind of trading mechanism is present in the pre-opening procedures of stocks markets such as the New York Stock Exchange, the Paris Bourse and Frankfurt's XETRA trading system. Moreover, this kind of trading mechanism is also used for some infrequently traded assets for which the continuous double auction mechanism is not viable.

to trade now. This results in two prices characterizing the equilibrium, and not one. While a trader willing to wait could trade at the single price found in the Walrasian framework, trades occurring immediately would not share this outcome. Then, the price depends on one wanting to buy or sell immediately and not just in the willingness to trade.

Moreover, one can conceive that the specific structure of the market could influence the trading price. Since the spread depends on the number of traders, characteristics such as volume could affect the price of immediacy and thus the market price. Finally, if trading involves more than simply matching the supply and the demand then the trading mechanism must also play an important role.

Even though the Walrasian mechanism fails to consider important practical aspects of trading, it is nevertheless a useful point of departure for modelling, and is frequently used as a basis for computing the efficiency of a set of trades.

Actually, some of the most common mechanisms are the following:

- **Bargaining**, when there are only two agents. In this case trade arises when one side can make a credibly take-it or leave-it offer. If a counter-offers arises, we observe sequential bargaining.

- **Auction**, when there is one seller and many potential buyers.

- **Call market**, when there are many buyers and many sellers at a single time.

- **Continuous markets**, when trades can potentially occur at any time. Continuous security markets are frequently categorized as dealership (price driven) or double auction (order driven) markets.

In practice, most of the security markets are hybrid and, although continuous markets have been more widespread in the past two decades, there are also a quite large number of periodic call markets.

1.1.4 Market liquidity

Liquidity is usually viewed as a summary attribute of an asset. Even though its definition is not universally accepted, it must satisfy certain widely accepted qualities. Firstly, liquidity is like the static concept of price-elasticity: By how much will an order move the price? Secondly, liquidity should have time and cost dimensions: How much will it cost to trade? How long will it take to trade? The cost dimension is related to the possibility to trade a large volume of shares without disrupting the price process, i.e. to trade at a "fair" price. This dimension is related with the volume traded. Its measure is complicate due to the fact that the fair price notion is vague. The time dimension is related with the possibility to trade at any time and its absence introduces the immediacy costs. In conclusion, we can say that in a liquid market, you can trade a large amount in a relatively short time without moving the price very much and any price perturbations caused by the trade quickly die out. Often, a common definition of liquidity is given in terms of depth, breadth and resilience.[2]

1.2 Theoretical models

The market microstructure theoretical framework is based on two kinds of models: inventory models and the information-based models. The former considers the market maker as the key participant of the trading process seen as a matching process in which the market maker must use prices to balance the supply and demand across time. His inventory plays a crucial role. In the information-based models the trading process is seen as a game involving traders with asymmetric information with respect to the assets' true value. Central to this approach is the learning problem confronting market participants.

In this section I present these two types of models (1.2.1 and 1.2.2) and I present briefly the so-called limit order models (1.2.3).

1.2.1 Inventory models

This section presents some important aspects about Garman (1976) and Amihud and Mendelson (1980) inventory models. It also considers some aspects of the Stoll (1978) and Ho and Stoll (1981) models which investigate explicitly the dealer's optimization problem. The literature related to inventory models include, among others, Stoll (1976), Cohen, Kalman, Maier, Schwartz, and Whitcomb (1981), Ho and Macris (1984), O'Hara and Oldfield (1986), Madhavan and Smidt (1991), Madhavan and Smidt (1993), Hasbrouck and Sofianos (1993) and Reiss and Werner (1998).

Garman (1976) considers the market maker as a smoother of intertemporal order imbalance. He suggests that a dealer is needed because buyers and sellers do not arrive synchronously. The market maker is assumed to be a monopolist whose prices reflect his monopoly power. The market maker is confronted to a succession of buy and sell orders that arrive randomly in continuous time. The arrival processes are considered as independent stochastic Poisson processes with stationary arrival rate functions $\lambda_B(p)$ and $\lambda_A(p)$, for the bid and ask sides respectively. In this setting, the market maker's objective is to maximize expected profit per unit of time, subject to the avoidance of bankruptcy and failure to provide liquidity. In order to achieve his objective, the market maker sets different bid and ask prices.

If the dealer posts a single price p, then the arrival intensities $\lambda_A(p)$ and $\lambda_B(p)$ are monotone increasing and monotone decreasing, respectively. These arrival rate functions describe the dynamic supply and demand curves. In equilibrium there exists a single equilibrium price p^*. However, in a multiple price world, the dealers post an ask price, P_A and a bid price, P_B, at which buyers and sellers trade. Accordingly, the condition of equal arrival rates is $\lambda_A(P_A) = \lambda_B(P_B)$ and the dealer earns the spread $P_A - P_B$ on each buyer-seller pair. Suppose

[2] Kyle (1985) defines the liquidity based on three other measures: tightness (bid and ask spread), depth (amount of one-sided volume that can be absorbed by the market without causing a revision of the bid-ask quotes) and resiliency (speed of return to the equilibrium).

that, subject to equal arrival rates, the dealer sets $P_B < p^* < P_A$. Then, by setting a wide spread, he decreases the number of traders per time unit, he increases his revenue per buyer-seller pair and thus, he can earn a positive revenue per unit of time.
As usual in monopoly models, this pricing strategy results in volume at the optimal prices being less than would occur with a competitive market. Now, define the dealer's inventory of stock at time t as:

$$I_s(t) = I_s(0) + N_B(t) - N_A(t) \tag{1.2.1}$$

where $I_s(0)$ is the dealer's initial inventory, $N_B(t)$ is the cumulative number of trades at the bid (dealer buys and customers sell) until time t and $N_A(t)$ is the cumulative number of trades at the ask side (dealer sells and customers buy). There is a similar expression for the dealers holding of cash:

$$I_c(t) = I_c(0) + P_A N_A(t) - P_B N_B(t) \tag{1.2.2}$$

where $I_c(0)$ is the dealer's initial units of cash. The key constraint is that dealer holdings of stock and cash cannot drop below zero (bankruptcy). Note that if $\lambda_A(P_A) = \lambda_B(P_B)$, holdings of stock follow a zero-drift random walk and cash holdings follow a positive-drift random walk. As Garman noted, if $\lambda_A(P_A) = \lambda_B(P_B)$, the dealer is eventually ruined with probability one: a zero-drift random-walk will eventually hit any finite barrier with probability one. Moreover, with some realistic parameter values, the expected time to ruin is a matter of days. To see this result observe that the dealer fails if he runs out of inventory or cash. Then, since inventories follow a random walk, sooner or later a sequence of trades will force either his stock position or his cash position to their boundary. When this happens, the process meets an absorbing barrier and failure occurs. Then, if the dealer sees an inventory barrier approaching, he simply wishes that the barrier is not hit.

The intuition behind this model is related to the "inventory control principle". Accordingly, the essential mechanism is that dealers change their bid and ask quotes in order to avoid unexpected imbalance of buy and sell orders, in the direction of restoring their inventories to a preferred position.

Amihud and Mendelson (1980) presented a model in which the risk neutral dealer maximizes expected profits per unit of time. In their model the bid and ask prices are functions of the inventory level, allowing the dealers to reset their prices through time according to their inventory evolution. An important assumption is that the inventory is bounded from above and below by some exogenous parameters. This removes the possibility that the dealer can "run out" of inventory and so removes the possibility of failure present in Garman's analysis.

The key results of Amihud and Mendelson (1980) model are:

- As the dealer's inventory increases, he lowers both bid and ask prices. If his inventory falls, he raises both prices, i.e. when his inventory holdings departs from his preferred position, he moves his prices to bring his position back.

- There is a positive spread that is increasing in distance from the preferred position.
- The midquote is not always equal to the true value and price fluctuations associated with inventory control are transient. Moreover, there are no manipulative strategies.

Since the dealer is assumed to be risk neutral and monopolist, in both Garman (1976) and Amihud and Mendelson (1980), the spread results from market power. However, in Amihud and Mendelson (1980) if the dealer faces competition the spread falls to zero. Consequently the spread plays no role in the viability of the market but acts essentially as a transaction cost.

Another approach of this kind of models is given by the works of Stoll (1978), and Ho and Stoll (1981) who explicitly model the dealers optimization problem. Central to this approach are the uncertainties in the order flow, which can result in inventory problems for the market maker or execution problems for the traders. In these models the market makers are risk-averse and the spread compensates them for deviating from their optimal inventory when they supply immediacy. When the market maker has large (low) inventory, he quotes low (high) bid and ask prices to motivate buy (sell) orders but he does not change the spread. The inventory effect is also referred as a transitory component due to the link between the bid-ask prices and the inventory of the market maker. Once, the inventory reverts to its original size, the bid and ask prices revert to their previous levels.

1.2.2 Asymmetric information models

The basic idea of the asymmetric information models is that some agents have superior private information. This kind of models provides insights about the price formation process. They are based on the idea that the outputs of the trading process are signals that show diverse private information of market participants. These models allow for the analysis of strategic behavior among informed and uninformed traders. Related literature is found for example in Grossman (1976) and Grossman and Stiglitz (1980).

Asymmetric information models assume that the traders' benefit from stock holdings is the resale value or terminal liquidating dividend that is the same for all holders. These models also assume that there exist private value components driven by diversification or risk exposure needs. Public information is usually referred to the common knowledge of the unconditional distribution of the terminal security value and the distribution of types of agents. The most important updates to the public information set are market data, such as the bid-ask and the prices and volumes of trades. Private information is assumed to be under the form of perfect knowledge of the terminal value of the stock. Most of the asymmetric information models in microstructure examine market features subject to a single source of uncertainty that is completely revealed at the end of the trading.

Glosten and Milgrom (1985) introduce the sequential trade models where traders arrive at the market singly, sequentially, and independently. O'Hara (1995) introduces the strategic trade models which deal with a single informed agent who can trade at multiple times. This fact implies that there is no need for the single trader to take into account the effect of his actions on subsequent decisions of others. However, as Kyle (1985) notes, a trader who revisits the market must make such calculations involving strategic considerations. These last type of models are known as continuous auction. The essential feature here is that a trade reveals the agents' private information. Rational and competitive market makers set their bid and ask quotes accordingly. All else equal, more extreme information asymmetries lead to a larger spread. Empirical implications of these models try to explain the spread and trade-impact effects.

Continuing with the exposition, I present the generalities of the most popular models that deal with asymmetric information. In order to get a clear idea, I divide the exposition in two parts: in the first, I detail the sequential models and in the second I focus on the strategic trade models of asymmetric information.

Sequential models

Jack Treynor, alias Walter Bagehot (Bagehot (1971)), defined the informed traders as people who posses non-public or private information about the true value of an asset while liquidity traders trade for exogenous or liquidity needs (portfolio rebalancing) rather than on information and they are willing to pay for immediacy. He assumes that informed traders always gain and that the market maker possesses the same information as liquidity traders. Thus, market makers loose when trading with informed traders and gain when trading with liquidity traders. The spread exists due to the trade-off between profits from liquidity traders and losses from informed traders.

Copeland and Galai (1983) were the first who attempted to formalize the concept of information costs. They developed a one-period model of the market maker's pricing problem given that some fraction of traders have superior information. They include two different approaches to the determination of the bid-ask spread. In the first approach the market maker fixes the bid and ask prices to maximize expected profit. In the second approach, the bid and ask prices are viewed as call and put options provided by the market makers to the traders. Even though the last approach is untractable to determine the dealer's maximization problem, they demonstrate that the volatility of the underlying value is an important determinant of the spread. They also show that higher priced stocks have lower percentage spreads. The contribution of Copeland and Galai (1983) to the literature is mainly based on the idea that dealers' order flows must include information-based trades. Moreover, their probabilistic structure is an important contribution of the model.[3] The main implications of the model are:

- As competition between market makers increases, the spread decreases.
- As the probability of informed traders (π_I) increases, the monopoly and competitive spreads converge.
- The spreads are positive in presence of informed traders even if market makers are risk neutral, there is perfect competition and no inventory effect.

[3] The market maker knows that any given trade comes from an informed trader with probability π_I and from and uninformed trader with probability $(1 - \pi_I)$.

Even though this model provides an interesting characterization of the bid-ask spread, it is a static one-period model. The private information is fully revealed after the trade because the decision problem is simply to balance gains and losses. In this sense, this model is similar to the inventory control models. This similarity disappears once dynamic considerations are introduced. With asymmetric information, the order flow is not exogenous to the dealers' problem and consequently trading itself contains information. Moreover, the continued trading of informed traders provides the potential for other uninformed traders to infer the underlying information. Glosten and Milgrom (1985) model is the first one that uses the concept of trades as "signals" of information, introducing the concept of the information content of trades.

Their model is a simple sequential trade model in which traders are confronted with informed and uninformed traders. They face potential losses from trading with informed traders who are the only ones having private information. The market maker wants to protect himself against these possible losses. The Glosten and Milgrom (1985) model specifies that the quoted bid and ask prices are set equal to the dealer's conditional expectation of the asset value given the types of trades submitted by the traders. The bid price equals the conditional expected value of the asset given that the trader wishes to sell the asset to the dealer. Due to the possibility of trading with an informed trader, the quoted bid price takes account of the possibility that bad news have occurred. The opposite is true for the quoted ask price: because informed traders only buy when they know something good about the asset, the quoted ask price must account for this possibility. In both cases, the market maker assumes that the incoming order conveys some signal of possible private information. The main implications of this model are:

- The spread is caused by adverse selection and it is considered as the market maker's compensation for facing this kind of risk in the order flow.
- The spread decreases with trading and converges to the true value once the information is revealed. This corresponds to a strong form of efficiency in the long run. It also implies that the bid and ask prices include the information given by the trades, as opposed to the transitory component predicted by inventory models.
- The transaction prices form a martingale: $E[P_{t+1}|\mathcal{F}_t] = P_t$, where P_t is the asset price and $\mathcal{F}_t$ is the information set at time t. This represents a semi strong efficiency in the short run.

In the basic sequential trade model, the trade quantity can only be one unit. Trades in real markets, of course, occur in varying quantities. Easley and O'Hara (1987) extended the model of Glosten and Milgrom (1985) by allowing for different trade sizes. Their model features event uncertainty and two possible order sizes. The market maker posts one set of bid and ask quotes for small trades and another set for large trades. They concluded that there should be a size dependence on the spread, as market makers should think of large size orders as coming from informed traders.

Easley and O'Hara (1992) make another important extension to the original model of Glosten and Milgrom (1985). They focus in the role of time in the trading process. They argue that time among trades (durations) also conveys information. In their framework a long duration means that no new information (either good or bad) is arriving to the market. Then, the probability of dealing with an informed trader must be smaller than when a short duration is observed. As a consequence, with a low probability of dealing with an informed trader, and because the market makers update their beliefs constantly, the quoted spread decreases. Their model has several important consequences:

- Time is endogenous to the price process. Time between trades should be analyzed in order to incorporate the information contained in nontrading intervals.
- The sequence of prices matters and is informative.
- Volume incorporates valuable information for the market maker. Excess volume (with respect to what is usually observed, or normal volume) is indicative of possible arrival of informed traders.
- The release of news leads to an increase in the trading intensity and this should imply more frequent revisions of the bid-ask prices posted by the market makers.

As noted by Bauwens and Giot (2001), the models of Glosten and Milgrom (1985), Easley and O'Hara (1987) and Easley and O'Hara (1992) are closely related to Bayesian learning models. In a multi-period framework, the quoting strategy of the market makers based on the possibility of facing informed traders, gives rise to a Bayesian updating behavior, with the market makers learning from the sequence of trades. Then, over time, the quoted prices converge to the expected value of the asset given the informed traders' information set.

Strategic trade models

Kyle (1985) introduces a model with a single informed trader who behaves strategically. In the multi-period version of the model, the trader returns to the market in order to minimize the price impact of his trade, distributing his orders over time (order splitting). Kyle (1985) considers that risk neutral informed traders can strategically exploit private information to maximize their profits. Informed traders take account of their trades' effect on the market price. When the informed trader is an informational monopolist, the trader can control the flow of information so that the price path that emerges has constant volatility. When there are multiple informed traders, this control is not great and informed trading causes prices to reflect information sooner. This fact raises the important question of exactly how quickly this price adjustment occurs. If the price adjustment is quite sensitive to the number of informed traders, then market prices may reach full-information efficient levels quite quickly. A concomitant effect will be that the return to information becomes small, leaving little incentive for traders to expend resources to gather new information. In this case, the intriguing results of Kyle's model on the role of volume, depth and price behavior may no longer hold.

This issue of price adjustment with multiple informed traders is addressed by Holden and Subrahmanyam (1992) and by Foster and Viswanathan (1993). Both papers employ variants of Kyle's model in which the number of informed traders is allowed to vary. Foster and

Viswanathan (1993) extend Kyle's model in a potentially important way by allowing the random variables to be elliptically distributed. In this framework, they address several interesting questions, including the impact of multiple informed traders on price behavior. The Holden and Subrahmanyam (1992) paper retains the original Kyle structure, but solves the difference equation system when there are multiple informed traders.

Useful extensions of the Kyle (1985) model include: Admati and Pfleiderer (1988), Foster and Viswanathan (1990), Subrahmanyam (1991b), Subrahmanyam (1991a), Holden and Subrahmanyam (1994), Foster and Viswanathan (1995) and Back (1992).

1.2.3 Limit order models

This section follows Hasbrouck (2004). The models presented in Sections 1.2.1 and 1.2.2 are viewed in settings where market makers set the bid and ask quotes and customers arrive and trade against these quotes. The models in this section examine limit orders and markets organized around them.

Limit orders and dealer quotes

A limit order is defined as an order that specifies quantity and price, and is subject to execution risk. This uncertainty is an important aspect of the traders' decisions. Basically, the buy (sell) limit order is functionally the same as a dealer's bid (ask). Accordingly, it implies that customer limit orders compete with dealer quotes. However, there are two main differences that we must consider. Firstly, the dealers have the ability to condition on size of the incoming order. When they face a large market order they execute the volume at the posted quote but they have some discretion on the remaining part. This is not the case for the limit order traders and this causes them to be less aggressive in terms of the price of their orders. Secondly, customers and dealers have different objectives. Customers usually place their orders for reasons of hedging or long-term portfolio objectives. Dealers set the ask so that, if hit, they can make a profit by quickly buying at a lower price. Moreover, from the customers' point of view, placing limit orders implies the existence of the execution risk, which must be considered.

In the following I present a brief explanation of the main ideas underlying the existing literature that considers those differences between limit orders and dealer quotes. I follow Hasbrouck (2004) and for a detailed exposition I refer the reader to that excellent text.

Limit order placement when faced with incoming orders of varying size The main reference here is the Glosten (1994) model and the most prominent empirical application is the one developed by Sandas (2001). The framework of these models is based on the existence of two types of traders: the market order traders, who trade due to liquidity needs or private information and, the limit order traders who supply liquidity to the market. Moreover, the limit order traders are assumed to be risk-neutral liquidity suppliers who are subject to a zero-expected profit condition.

Limit order models with dealers Seppi (1997) proposes a limit order model that includes the active participation of dealers. These dealers can condition their trades on the total size of the incoming order but they must yield to customer orders at the same price.

Bidding and offering with uncertain execution These models assume that the dealers are in their optimum portfolio and analyze how their quote setting behavior is affected when their posted bid and ask is hit. The idea is that these quotes must be set in such a way that they compensate them for this loss of utility. The relevant paper of this kind of research is Stoll (1978).

Limit order submission strategies This literature differs from the previous one in that it considers an agent who is not at his optimum. In this situation, he should decide among doing nothing, trading with a market order or trading with a limit order. The basic literature here is given by Cohen, Kalman, Maier, Schwartz, and Whitcomb (1981).

Limit order models of choice and strategy These models analyze many realistic trading strategies that are usually multi-period strategies and that involve order revisions. For example, an institutional investor (assumed to be an uninformed trader) facing a deadline, starts by placing limit orders away from the best quotes. As the deadline approaches, the trader revises the orders, pricing them more aggressively, i.e. placing his limit orders closer to the best quotes. Finally, if nothing is executed near the deadline, the trader submits a market order. Angel (1994) and Harris (1998) model these strategies, and Bertsimas and Lo (1998) consider order-splitting strategies of institutional investors.

Dynamic equilibrium models These models belong to the general equilibrium type models. They assume that the quotes that a given agent faces in a given point in time depend on past actions of other agents that face similar problems. Even though these models are stylized ones, they arrive at useful empirical predictions. Here, I focus on Parlour (1998) and Foucault (1999) due to their intensive use in Chapter 3.

Foucault (1999)

Foucault (1999) proposes a discrete time model that allows to do comparative static analysis. It is best viewed as a cross-sectional model rather than a dynamic one. Let $t = 1, \ldots, T$ where T is the terminal payoff date. The underlying value of the security is $v_t = v_0 + \sum_{i=1}^{t} \varepsilon_t$, where the ε_t are i.i.d. Bernoulli, taking values of $\pm\sigma$ with equal probability. The model assumes that at the beginning of every period t, there is 0.5 probability that $t = T$. This implies that the solution is identical in every period, simplifying the analysis. The trader's reservation price at time t, R_t is given by:

$$R_t = v_t + y_t. \tag{1.2.3}$$

where y_t is the idiosyncratic shock that takes the value $(+M, -M)$ with probabilities ϕ and

$1-\phi$, independent of the value process (v_t). This idiosyncratic shock does not reflect private information, it arises from portfolio or liquidity considerations that are not explicitly modelled. This shock drives the direction of the agent's desired trade.

The book is represented at each point in time by the ask and bid prices, A_t and B_t respectively. If there is an empty book it is represented by $A_t = \infty$ and $B_t = -\infty$. It is assumed that the traders have common knowledge on v_t and y_t. The trader can submit a buy or sell market order or he can place a buy or sell limit order. In this framework the execution risk of the limit orders are endogenous in that at time t the trader knows the distribution of v_{t+1} and the distribution of the characteristics for the time $t+1$ trader. This enables him to derive the execution probability for any given limit order. He has not the opportunity to revise his order once it has been submitted and his limit orders are valid for only one period.

If a trade is executed at price P, a seller has utility $U(y_t) = P - v_T - y_t$ and a buyer has utility $U(y_t) = v_T + y_t - P$.

The execution risk arises for example if we have $v_t = v_0$ and if we have two traders tying to trade in the same direction, i.e. if $y_t = +M$ and $y_{t+1} = +M$. Another situation that can be depicted from this model, by setting $\sigma \neq 0$ with $\epsilon_{t+1} = -\sigma$, is the well known "winner's curse". In this case the probability of execution increases but the gain from the trade decreases, i.e. there is a risk to be "picked off" after a "public" information event. Finally, by setting $\epsilon_{t+1} = +\sigma$, the execution probability decreases because the market has moved away. In this situation what is expected is that the trader starts chasing the market by submitting more aggressive limit orders.

The empirical implications of the model are: firstly, if there is a large spread, traders would prefer to send limit orders. Secondly, if the fundamental risk of the security (σ) increases, then the "picked-off" risk increases and traders prefer to submit less aggressive limit orders. With this, market orders become more expensive favoring the submission of limit orders but decreasing their probability of execution.

Parlour (1998)

Parlour's model assumes that there exist at each point in time two discrete prices, the ask (A) and the bid (B). The distance between them is one tick. At this prices there are dealers who are willing to buy an infinite amount at B and sell an infinite amount at A. This is a model of queuing and quantities rather than a model of prices, i.e. the executions in the book respect the time priority.

This is a two days period model. Trade can only occur at times $t = 0, \ldots, T$ on day 1 and clearing occurs at the end of the day. On the second day the non-random payoff per share is realized. In this framework, the agents have preferences $U(C_1, C_2, \alpha) = C_1 + \alpha C_2$ where α is a continuous random variable distributed on the interval $(\underline{\alpha}, \bar{\alpha})$ where $0 < \underline{\alpha} < 1 < \bar{\alpha}$. That is, there is uncertainty and heterogeneity across agents in their relative valuations of C_1 and C_2. Agents also differ in their endowments. Variation in α is the sole source of randomness in the model.

At each time t, a trader arrives. Using a market order, with probability π_B, the arriving trader is a potential buyer of one unit. With probability π_S, the arriving trader has one unit, and is a potential seller. With probability $1 - \pi_B - \pi_S$, the trader is neither a buyer nor a seller.

The main empirical implications of this model are: firstly, when the depth on the same-side is large, the arriving trader is more likely to use a market order because his probability of non execution is high ("crowding out effect"). Secondly, when the depth on the opposite-side is large, the arriving trader is more likely to use a limit order because the probability that opposite side traders use market orders is high and thus, his execution probability is also high.

Table 1.1 summarizes the microstructure models presented in this section.

Table 1.1: **Microstructure models.**

Inventory models
• The dealer is considered as a smoother of intertemporal order imbalance. • This literature is related to the "inventory control principle": dealers change their ask and bid quotes in order to avoid unexpected imbalances. • The spread compensates the dealers for deviating from their optimal inventory. • References: Garman (1976), Stoll (1976), Stoll (1978),Amihud and Mendelson (1980), Ho and Stoll (1981), Cohen, Kalman, Maier, Schwartz, and Whitcomb (1981), Ho and Macris (1984), O'Hara and Oldfield (1986), Madhavan and Smidt (1991), Madhavan and Smidt (1993), Hasbrouck and Sofianos (1993) and Reiss and Werner (1998).

Asymmetric information models
• The basic idea is that some agents have superior private information. • The outputs of the trading process are signals that show diverse private information of market participants. • These models allow for the analysis of strategic behavior between informed and uninformed traders. • This kind of models can be subdivided into sequential models and strategic trade models. • References: Bagehot (1971), Copeland and Galai (1983), Glosten and Milgrom (1985) Easley and O'Hara (1992), for the sequential models. Kyle (1985), Admati and Pfleiderer (1988), Foster and Viswanathan (1990), Subrahmanyam (1991b), Subrahmanyam (1991a), Holden and Subrahmanyam (1992), Back (1992), Foster and Viswanathan (1993), Holden and Subrahmanyam (1994) and Foster and Viswanathan (1995) for the strategic trade models.

Limit order models
• These models examine limit orders and markets organized around them. • They analyze the differences between limit orders and dealer quotes. • This kind of models considers that the execution risk is an important aspect of the limit order trader's decisions. • This kind of models studies several related aspects: limit order placement when faced with incoming orders of varying size, limit order models with dealers, bidding and offering with uncertain execution, limit order submission strategies and dynamic equilibrium models. • References: Stoll (1978), Cohen, Kalman, Maier, Schwartz, and Whitcomb (1981), Glosten (1994), Angel (1994), Seppi (1997), Harris (1998), Bertsimas and Lo (1998), Parlour (1998), Foucault (1999) and Sandas (2001).

1.3 Econometric models

The occurrences of financial market events, such as trades and quotes, can be viewed as point processes whose realizations consist of point events in time. In financial econometrics, point processes are very important in order to model the market activity on a tick-by-tick level.

Let t denote the calendar time and let $\{t_i^s\}_{i\in\{1,2,\dots\}}$, $s = 1, \dots, S$ be S sequences of non-negative random variables on some probability space $(\Omega, \mathcal{F}, \mathcal{P})$ associated with random arrival times $0 \leq t_i^s \leq t_{i+1}^s$. Then, the sequence $\{t_i^s\}$ represent a S-dimensional point process on $[0, \infty)$. A model for financial point processes can be specified in three ways:

1. Via the durations,

2. Via the intensity function,

3. Via the counts in any interval of time.

In principle, one specification implies the other two, but the analytical derivation is rarely possible. Hence it is a matter of simplicity and convenience which one is chosen. However each approach has some advantages or disadvantages with respect to the others.

In this section I present some widely used econometric models in the empirical microstructure literature: the autoregressive conditional duration (ACD) model of Engle and Russell (1998), intensity models, count models and the ordered probit model. The advantages and disadvantages of each of these models with respect to the others are presented in the respective sections. These models are widely used for testing the practical implications of microstructure theories. They are competitors but also complements of the model I will present and use in Chapters 2 and 3.

1.3.1 Autoregressive conditional duration model

The baseline for this section can be found in Engle and Russell (1998), Bauwens and Giot (2000), and Fernandes and Grammig (2003).

The autoregressive conditional duration (ACD) model (Engle and Russell (1998)) is a framework for modelling intertemporally correlated event arrival times.

Consider a stochastic process made of a sequence of arrival times $(t_0, t_1, \dots, t_n)$ with $0 < t_0 < t_1 < \dots < t_n$. The duration is simply defined as the interval between two arrival times: $x_i = t_i - t_{i-1}$. The conditional expectation of the i-th duration is given by:

$$E(x_i|\mathcal{F}_{i-1}) = \psi_i(x_i|\mathcal{F}_{i-1}; \theta) \equiv \psi_i, \tag{1.3.1}$$

where $\mathcal{F}_{i-1}$ represents the conditioning information set generated by the durations preceding x_i and θ is a vector of parameters. The ACD model of Engle and Russell (1998) specifies that the observed duration is a parametrization of equation (1.3.1) such that:

$$x_i = \psi_i \varepsilon_i. \tag{1.3.2}$$

where $\{\varepsilon_i\}$ is a sequence of positive i.i.d. random variables with $E(\varepsilon_i) = 1$, $var(\varepsilon_i) = \sigma^2$, such that $E(x_i|\mathcal{F}_{i-1}) = \psi_i$.

An ACD(1,1) model specifies an autoregressive model for the expected conditional durations:

$$\psi_i = \omega + \alpha x_{i-j} + \beta\psi_{i-j} \tag{1.3.3}$$

with the following constraints on the parameters: $\omega > 0$, $\beta \geq 0$, $\alpha > 0$ and $\alpha + \beta < 1$. The last ensures the existence of the unconditional mean of the duration, the others ensure the positivity of the conditional durations. The model accounts for the clustering of the durations: small durations are more likely to be followed by small durations than by large ones (and likewise for long durations).

The conditional mean is the conditional duration ($E[x_i|\mathcal{F}_{i-1}] = \psi_i$). The conditional variance is $Var[x_i|\mathcal{F}_{i-1}] = \sigma^2\psi_i^2$. Then, the conditional dispersion ratio $\frac{Var[x_i|\mathcal{F}_{i-1}]}{(E[x_i|\mathcal{F}_{i-1}])^2} = \sigma^2$ is constant. As a result, conditional over-, equi- and underdispersion is allowed. Moreover and unless $\alpha = \beta = 0$,

$$\frac{Var[x_i]}{(E[x_i])^2} > \frac{Var[x_i|\mathcal{F}_{i-1}]}{(E[x_i|\mathcal{F}_{i-1}])^2} \tag{1.3.4}$$

i.e. the unconditional dispersion ratio is larger than the conditional one: clustering increases the dispersion.

The unconditional mean and variance of the ACD(1,1) model are:

$$E(x_i) = \mu_x = \frac{\omega}{1 - (\alpha + \beta)} \tag{1.3.5}$$

$$Var(x_i) = \sigma_x^2 = \mu_x^2\sigma^2\frac{1 - 2\alpha\beta - \beta^2}{1 - (\alpha + \beta)^2 - \alpha^2\sigma^2} \tag{1.3.6}$$

provided that the denominator is positive, i.e. $(\alpha + \beta)^2 + \alpha^2\sigma^2 \leq 1$. The autocorrelation function (ACF) is given by:

$$\rho_1 = \frac{\alpha\,(1 - \beta^2 - \alpha\,\beta)}{1 - 2\,\alpha\,\beta - \beta^2} \tag{1.3.7}$$

and

$$\rho_n = (\alpha + \beta)\rho_{n-1} \quad for \;\; n \geq 2. \tag{1.3.8}$$

Model (1.3.3) can also be formulated as an ARMA(p,q) model for durations. Letting $\eta_i = x_i - \psi_i$ which is a martingale difference sequence by construction, the duration process can be expressed as:

$$x_i = \omega + \sum_{j=0}^{max(p,q)} (\alpha_j + \beta_j)x_{i-j} - \sum_{j=0}^{q} \beta_j\eta_{i-j} + \eta_i, \tag{1.3.9}$$

which is an ARMA(p,q) process with non-Gaussian innovations.

Forecasting of waiting times can be computed directly from this representation using conventional ARMA analytics. Thus it is simple to compute analytically the expected waiting time until the $(i+k)$-th transaction occurs. If all roots of the associated polynomial are less than the unity, then the duration process is mean reverting and the impact of a given duration on future expected durations dies out exponentially. Since the transactions to be analyzed occur within seconds of each other, the persistence of shocks is very limited in calendar time unless the roots are very close to unity.

For maximum likelihood estimation, and for simulating the model, we need an assumption on the distribution of ε_i. Many possible choices are available (the number of parameters to estimate are given in parenthesis):

- Exponential (0)
- Weibull (1) (includes Exponential)
- Gamma (1) (includes Exponential)
- Burr (2) (includes Weibull)
- Generalized Gamma (2) (includes Gamma and Weibull)
- Fisher (2)
- Lognormal (2)
- Pareto (2)

Several extensions of the original model of Engle and Russell (1998) have been proposed. Bauwens and Veredas (2004), Lunde (1999) and Grammig and Maurer (2000) propose duration models that accommodate more flexible hazard rate functions. Bauwens and Giot (2000) put forward the logarithmic ACD model in order to have a more suitable framework for testing market microstructure hypothesis avoiding some of the parameter restrictions of the original model. Fernandes and Grammig (2003) propose a very flexible family of augmented ACD processes that nests most of the existing ACD models. Moreover, there are other approaches to specify the dynamics of durations, conditionally on the past durations and exogenous variables. Ghysels, Gouriéroux, and Jasiak (2004) developed the stochastic volatility duration (SVD) model, which is designed for take into account both mean and variance dynamics in financial durations. Another single factor model, the stochastic conditional duration (SCD) model, was developed by Bauwens and Veredas (2004). For a comparative evaluation of these models, see Bauwens, Giot, Grammig, and Veredas (2004).

In empirical studies of market microstructure, the autoregressive conditional duration (ACD) model, introduced by Engle and Russell (1998), and many of its extensions, have been used widely to test theories with tick-by-tick data in a univariate framework. These models are designed specifically to deal with irregularly-spaced nature of financial time series. However, extensions to more than one series have proven to be very difficult. The difficulty comes from the very nature of the data, which are by definition not aligned in time, i.e. the times at which an event of any type happens are random. Engle and Lunde (2003) suggest a model for the bivariate case, but the specification is not symmetric in the two processes. These kind of multivariate extensions of the ACD model result in competing risk models. In these approaches one models the time until the occurrence of one of the individual processes and treat all the others (non-observed) processes as right-censored. However, this approach implies some information loss when successive points of one process occur without intervening points of the other processes. For this reason, such an approach is not appropriate for a complete modelling of multivariate point processes.

1.3.2 Intensity models

This section follows Cox and Isham (1980) and Bauwens and Hautsch (2003). As usual, I refer the reader to those references for an extended presentation of the ideas developed in this section.

The basis of this kind of models relies on the concept of the intensity function. This function is defined as the instantaneous rate of occurrence given the process history and observable factors. It allows to account for events that occur at any point in time, defining the instantaneous event arrival time rate and thus, it could be interpreted as a continuous time measure for economic activity. In order to present a mathematical definition of the intensity function I present some preliminary definitions.

Having defined a point process, let $N^s(t) := \sum_{i\geq 1} \mathbb{1}_{\{t_i^s \leq t\}}$ and $\breve{N}^s(t) := \sum_{i\geq 1} \mathbb{1}_{\{t_i^s < t\}}$ denote the right-continuous and, respectively, left-continuous counting functions associated with the s-type events. Correspondingly, $N(t)$ and $\breve{N}(t)$ are the right-continuous and, respectively, left-continuous counting functions of the pooled process, i.e. $N(t)$ counts the number of points until and including t and $\breve{N}(t)$ counts the number of events before t. Let n denote the number of points in the pooled process, which pools and orders the arrival times of all single processes. Let n^s be the number of s-type elements in the sample. Moreover, define $x_i^s := t_i^s - t_{i-1}^s$ (with $t_0 := 0$) as the interval of time (or duration) between two successive points associated with the s-th process and call $\{x_i^s\}_{i\in\{1,\ldots,n^s\}}$ the duration process associated with $\{t_i^s\}_{i\in\{1,\ldots,n^s\}}$. Finally, $x(t) := t - t^s_{\breve{N}(t)}$ is called the backward recurrence time at t, i.e. it is the time elapsed since the previous point, and it is a left-continuous function that grows linearly through time with discrete jumps back to zero after each arrival time t_i.

In the following, it is assumed that for small positive Δ,

$$Pr[N(t+\Delta) - N(t) > 1|\mathcal{F}_t] = o(\Delta), \tag{1.3.10}$$

where $o(\Delta)$ denotes a remainder term with the property $o(\Delta)/\Delta \to 0$ as $\Delta \to 0$. This property assumes that the pooled process is orderly, and thus, excludes the possibility of multiple occurrences simultaneously. This is a restrictive assumption but simplifies the probability theory of point processes.

Definition 1.3.1. *Let $N^s(t)$ be the s-type component of a S-dimensional point process on* $[0, \infty)$ *that is adapted to some history $\mathcal{F}_t$ and assume that $\lambda^s(t|\mathcal{F}_t)$ is a positive process with sample paths that are left-continuous and have right-hand limits. Then, the process*

$$\lambda^s(t|\mathcal{F}_t) := \lim_{\Delta \downarrow 0} \frac{1}{\Delta} E[N^s(t+\Delta) - N^s(t)|\mathcal{F}_t], \quad \lambda^s(t|\mathcal{F}_t) > 0, \ \ \forall t, \tag{1.3.11}$$

is called the $\mathcal{F}_t$-intensity process of the counting process $N^s(t)$.

Under assumption (1.3.10), the intensity function is alternatively written as

$$\lambda^s(t|\mathcal{F}_t) := \lim_{\Delta \downarrow 0} \frac{1}{\Delta} Pr[N^s(t+\Delta) - N^s(t) > 0|\mathcal{F}_t], \tag{1.3.12}$$

which can be associated, roughly speaking, with the conditional probability per unit time to observe an s-type event in the next instant, given the conditioning information.

Continuing with this section devoted to intensity models, I present some special types of processes in order to establish some basic ideas. I start with the Poisson process that is the usual starting point for point processes that, even though being very restrictive for applied work, constitutes the basis for many theoretical approaches. From this simple process, usually called homogenous Poisson process, I discuss some of the existing literature that deal with more complex processes called non-homogenous Poisson processes.

The Poisson process This is the simplest point process with no memory, i.e. the probability of finding a point in $(t, t+\Delta)$ does not depend on whether there have been relatively few or relatively many points just before t, or indeed on whether there is a point exactly at t, then:

$$Pr[N(t+\Delta) - N(t) = 1|\mathcal{F}_t] = Pr[N(t+\Delta) - N(t) = 1], \tag{1.3.13}$$

A homogeneous Poisson process of rate λ occurs if events occur independently with constant probability equal to Δ times the length of the interval. The number of events in disjoint time intervals are independent, and the distribution of events in each interval of unit length is $Poisson[\lambda]$. Formally, as the length of the interval Δ tends to zero:

$$Pr[N(t+\Delta) - N(t) = 1] = \lambda\Delta + o(\Delta). \tag{1.3.14}$$

Since the Poisson process is simple, i.e. the occurrence of several events simultaneously is excluded, it satisfies:

$$Pr[N(t+\Delta) - N(t) > 1] = o(\Delta), \tag{1.3.15}$$

and therefore

$$Pr[N(t+\Delta) - N(t) = 0] = 1 - \lambda\Delta + o(\Delta). \tag{1.3.16}$$

An essential element in (1.3.14)-(1.3.16) is that λ is constant, i.e. it does not depend on t. However, for some purposes we can be interested to replace this assumption by λ being a function of time λ_t, the other assumptions remaining the same. This is the so-called non-homogeneous Poisson process. A closely related possibility is that there is an observed explanatory variable $z(t)$, and that the rate of the Poisson process at time t is a function of $z(t)$.

Two key results are as follows. First, let X be a random variable measuring the duration from the origin to the first point. Then the probability density function, f_X, and survivor function $\mathcal{F}_X$ of X are

$$f_X(x) = \lambda e^{-\lambda x}, \;\; \mathcal{F}_X(x) = Pr(X > x) = e^{-\lambda x} \quad (x > 0), \tag{1.3.17}$$

i.e. X is exponentially distributed with parameter λ and mean $1/\lambda$.

Further, the above definition of the Poisson process implies that, starting from an arbitrary time origin, subsequent points are at times

$$X_1,\; X_1 + X_2,\; X_1 + X_2 + X_3, \ldots, \tag{1.3.18}$$

where the random variable X_i are i.i.d. with the exponential distribution (1.3.17). Note from (1.3.18) that the r-th point after the origin occurs at time $T_r = X_1 + \cdots + X_r$ and that this has a gamma distribution with density

$$\frac{\lambda(\lambda t)^{r-1} e^{-\lambda t}}{(r-1)!}. \tag{1.3.19}$$

Note that T_r is the waiting time (duration) for observing r events and that for $r > 0$, $\{N(t) < r\}$ is equivalent to $\{T_r > t\}$.

As a conclusion, the durations $\{x_i\}_{i \in \{1,2,\ldots\}}$ of the homogeneous Poisson process are i.i.d. exponential random variables with parameter λ and standard deviation $1/\lambda$ (equidispersed).

Another way to generalize the Poisson process is to relax the assumption of an exponential distribution for the durations, i.e. that the durations $\{x_i\}_{i \in \{1,2,\ldots\}}$ are i.i.d. but not exponentially distributed: e.g. they can follow gamma, lognormal, Weibull or other continuous distributions for non-negative random variables. This type of process is called a renewal process. For such processes, the intensity is not constant between two points: it depends on the backward recurrence time $x(t)$. Positive (negative) duration dependence corresponds to an increasing (decreasing) hazard function.

For the second key property of the Poisson process, consider the number $N(a, b)$ of points in a fixed interval $(a, b]$; slightly more general, consider an arbitrary set A on the time axis and the number of points $N(A)$ therein. It can be shown that $N(a, b)$ has a Poisson distribution with mean $\lambda(b - a)$ and that $N(A)$ has a Poisson distribution with mean $\lambda|A|$, where $|A|$ is the length (Lebesgue measure) of the set A. It follows from the strong independence properties of the Poisson process that the distribution of $N(a, b)$ depends only on $b - a$.
Thus, $N(t)$, the number of events until t, has a Poisson distribution with parameter $t\lambda$ such

that $E[N(t)] = t\lambda = Var[N(t)]$. Again, one can generalize the Poisson process by assuming a more general distribution such as the negative binomial, the double Poisson and so on.

An important result in the point process literature is that under fairly weak regularity conditions cited in Theorem 16 of Brémaud (1981), the integrated intensity function

$$\Lambda^s(t^s_{i-1}, t^s_i) := \int_{t^s_{i-1}}^{t^s_i} \lambda^s(u|\mathcal{F}_u)du \tag{1.3.20}$$

follows and i.i.d. standard exponential process, i.e.

$$\Lambda^s(t^s_{i-1}, t^s_i) \sim i.i.d.\ Exp(1). \tag{1.3.21}$$

The Poisson property of the integrated intensity process holds if $\int_0^\infty \lambda^s(u|\mathcal{F}_u)du = \infty$. The interpretation of this condition is that the point process does not "die" at some point in time, i.e. it is assumed a zero probability of no occurrences of more points after some point in time. This is realistic for financial point processes, for example trades and quotes do not stop under usual market conditions.

Other non-homogeneous Poisson process have been developed in the literature. Grandell (1976) proposed a type of non-homogeneous Poisson process called the doubly stochastic Poisson process. Bauwens and Hautsch (2003) proposed the latent factor intensity (LFI) model. This model assumes that the conditional intensity function given the observable history of the process is itself stochastic and follows an autoregressive process. The intensity is parameterized in terms of two components, a univariate latent one, and an observable one which is driven by the history of the process and can be specified univariately or multivariately. In the latter case, the latent factor corresponds to a common component that captures the impact of a general factor that influences all individual process components. In the context of financial markets, the latent factor may be economically interpreted as a variable representing the information flow that cannot be observed directly but influences the general activity of the markets (and hence the intensity of the process). In this sense, the LFI approach combines the idea of latent factor models arising from the mixture-of-distribution hypothesis (see Clark (1973)) with the concept of dynamic intensity processes. It can be seen as the counterpart of the stochastic volatility (SV) model (Taylor (1982)) or the stochastic conditional duration (SCD) model (Bauwens and Veredas (2004)). However, while in the SV or SCD model, the process dynamics are completely driven by the dynamics of the latent component, the LFI model is based on observation driven and latent dynamics. Hence, using the terminology of Cox (1981), the LFI model combines the idea of observation driven models and parameter driven models. In this sense, the latent component models unobserved dynamic heterogeneity which is not captured by the observation driven part of the model. Then, two limit cases emerge naturally: one when the latent factor is irrelevant and the intensity is completely described by observation driven dynamics, and the other when the observable components are not relevant and the latent factor completely dominates. Hence, econometrically, the existence of a latent factor can be interpreted as an indication that the observation driven part of the model is not able to completely capture the dynamics in the data. Bauwens and Hautsch (2003) present a dynamic extension of this kind of processes and introduced the use of the latent dynamic factor. They illustrate alternative parameterizations of the observation driven component based on autoregressive conditional intensity (ACI) specifications (Russell (1999)), as well as Hawkes types models (Hawkes (1971)). Based on simulation studies, they showed that the proposed model provides a flexible tool to capture the joint dynamics of multivariate point processes. Since the latent component has to be integrated out, the model is estimated by simulated maximum likelihood based upon efficient importance sampling techniques proposed by Richard (1999).

Among others, the intensity approach has been used by Russell (1999), Kawasaki (2002), Bowsher (2002) and Bauwens and Hautsch (2003). The intensity approach is convenient for multivariate specifications and time-varying covariates but it is not straightforward for computation of forecasts.

1.3.3 Time series count models

Many different approaches have been proposed to model time series count data. Good reviews can be found both in Cameron and Trivedi (1998), Chapter 7 and in MacDonald and Zucchini (1997), Chapter 1. In this section I briefly introduce the most used count models in time series and present their strengths and weaknesses.

Markov chains are one way of dealing with count data in time series. The method consists of defining transition probabilities between all the possible values that the count variable can take and determining, in the same way as in usual time series analysis, the appropriate order for the series. This method is only reasonable, when there are very few possible values that the observations can take. A prominent area of application for Markov chains is binary data. As soon as the number of values that the dependent variable takes gets too large, these models lose tractability.

Discrete Autoregressive Moving Average (DARMA) models introduced by Jacobs and Lewis (1978a), Jacobs and Lewis (1978b) and Jacobs and Lewis (1983), are models for time series count data with properties similar to those of ARMA processes found in traditional time series analysis. They are probabilistic mixtures of discrete i.i.d. random variables with suitably chosen marginal distribution. The simplest example is the DARMA(1,0) model with:

$$y_t = u_t y_{t-1} + (1 - u_t)\varepsilon_t, \tag{1.3.22}$$

where u_t is a binary mixing random variable that takes the value 1 with probability ρ and 0 with probability $(1 - \rho)$. This model implies that:

$$Pr[y_t = y_{t-1}] = \rho, \tag{1.3.23}$$

$$Pr[y_t = \varepsilon_t] = 1 - \rho, . \tag{1.3.24}$$

Clearly, the autocorrelation at lag k is ρ^k, as in the AR(1) case, but only positive correlations are possible. Generalizations can be made to the DARMA(p,q) with correlation structures equal to the standard linear ARMA(p,q) models, although with greater restrictions on the permissible range of correlation structures. The major restriction of DARMA models is that for high serial correlation the data will be characterized by a series of runs of a single value. It is for this reason that this kind of models is rarely used. Another problem associated with these models seems to be the difficulty of estimating them. Moreover, the model is only applied to a time series which can take at most three values.

The integer-valued Autoregressive Moving Average (INARMA) is a generalization of the linear ARMA models for continuous data but specifies y_t to be the sum of a count random variable whose value is determined by its past and an iid innovation ε_t. This model resembles the linear model $y_t = \rho y_{t-1} + \varepsilon_t$, for example, but it explicitly models y_t as a count. Different distributions of ε_t lead to different marginal distributions for y_t. The model is interesting in that it has the same serial correlation structure as ARMA models for continuous data. This models where introduced by McKensie (1986) and Al-Osh and Alzaid (1987) and extended to the regression case by Brännäs (1995).

To understand how the model works, I present the simple INAR(1) model. Let $\mathbf{Y}^{t-1} = (y_{t-1}, y_{t-2}, \ldots, y_0)$. The INAR(1) process is given by:

$$y_t = \rho \circ y_{t-1} + \varepsilon_t, \quad 0 \leq \rho < 1, \tag{1.3.25}$$

where ε_t is an iid latent count variable independent of $\mathbf{Y}^{t-1}$. The symbol $\circ$ denotes the binomial thinning operator. Then, $\rho \circ y_{t-1}$ is the realized value of a binomial random variable with y_{t-1} trials and probability ρ of success of each trial. More formally $\rho \circ y_{t-1} = \sum_{j=1}^{y_{t-1}} u_j$ where u_j is a sequence of iid binary random variables that take value 1 with probability ρ and value 0 with probability $1 - \rho$.

The conditional expectation of y_t (conditional on y_{t-1}) is:

$$\mu_{t \mid t-1} = E[y_t \mid y_{t-1}] = \rho y_{t-1} + E[\varepsilon_t], \tag{1.3.26}$$

and the conditional variance (using standard results on the binomial with y_{t-1} trials) is:

$$\sigma^2_{t \mid t-1} = Var[y_t \mid y_{t-1}] = \rho(1-\rho) y_{t-1} + Var[\varepsilon_t]. \tag{1.3.27}$$

It can be shown using the law of iterated expectations that the unconditional distribution of y_t has:

$$\mu = E[y_t] = \frac{E[\varepsilon_t]}{1-\rho}, \tag{1.3.28}$$

and

$$\sigma^2 = Var[y_t] = \frac{\rho E[\varepsilon_t] + Var[\varepsilon_t]}{1-\rho^2}, \tag{1.3.29}$$

The parameter ρ is analogous to the coefficient on the lagged value in an AR(1) model. Moreover, this model has the same autocorrelations as the AR(1) model of traditional time series analysis, which makes it its discrete counterpart. This family of models has been generalized to include integer valued ARMA processes as well as to incorporate exogenous regressors. However, the problem with this type of models is the difficulty in estimating them. Many models have been proposed and the emphasis was put more on their stochastic properties than on how to estimate them.

The hidden Markov models advocated by MacDonald and Zucchini (1997) are an extension of the basic Markov chains models, in which various regimes characterizing the possible values of the mean are identified. It is then assumed that the transition from one regime to another is governed by a Markov chain.

Let m be the possible regimes and C_t, $t = 1, \ldots, T$, a Markov chain on the state-space $\{1, 2, \ldots, m\}$. Thus, $C_t = j$ if at time t we are in regime j. I consider the simplest case where C_t is an irreducible homogeneous Markov chain, with transition probabilities given by:

$$\gamma_{ij} = Pr[C_t = j \mid C_{t-1} = i], \quad i, j = 1, \ldots, m, \tag{1.3.30}$$

which are time invariant. It is assumed that exists an unique strictly positive stationary distribution

$$\delta_j = Pr[C_t = j], \quad j = 1, \ldots, m, \tag{1.3.31}$$

where the δ_j are a function of γ_{ij}. For example, for the Poisson hidden Markov model, it is assumed that the count data y_t in each regime are Poisson distributed, with mean parameters that varies with exogenous variables and the regime.

One of the problems of this approach is that there is no accepted way of determining the appropriate order for the Markov chain. Whereas in some cases there is a natural interpretation for what might constitute a suitable regime, in most applications, and in particular in the applications considered in this thesis, this is not the case. Another problem is that the number of parameters to be estimated can get big, especially when the number of regimes is large. Finally, the results are, in most cases, not very easy to interpret.

State-space models specify the conditional distribution of y_t to depend on stochastic parameters that evolve according to a specified distribution whose parameters are determined by some regressors $\mathbf{X}^t$ and $\mathbf{Y}^{(t-1)}$. Harvey and Fernandes (1989) used state-space models with conjugate prior distributions. In this context, counts are modelled as a Poisson distribution whose mean itself is drawn from a gamma distribution. Their starting point is a Poisson regression model, where y_t conditional on μ_t is Poisson $P[\mu_t]$ distributed, so

$$f(y_t \mid \mu_t) = \frac{e^{-\mu_t}\mu_t^{y_t}}{y_t!}. \tag{1.3.32}$$

In this case the mean parameter μ_t is modelled to evolve stochastically over time with distribution determined by past values of y_t. A convenient choice of the distribution is the gamma

$$f(\mu_t \mid a_{t|t-1}, b_{t|t-1}) = \frac{e^{-b\mu_t}\mu_t^{a-1}}{\Gamma(a)b^{-a}}, \quad a_{t|t-1} > 0, \;\; b_{t|t-1} > 0, \tag{1.3.33}$$

where a and b are evaluated at $a = a_{t|t-1} = \omega a_{t-1}$ and $b = b_{t|t-1} = \omega b_{t-1}$ and $0 < \omega \leq 1$. The marginal density of y_t given $Y^{(t-1)}$, obtaining by computing:

$$f(y_t \mid Y^{(t-1)}) = \int_0^\infty f(y_t \mid \mu_t) f(\mu_t \mid Y^{(t-1)}) d\mu_t, \tag{1.3.34}$$

is negative binomial with parameters $a_{t|t-1}$ and $b_{t|t-1}$. Estimation of ω and the parameters of μ_t is done by maximum likelihood and the Kalman filter is used to update recursively $a_{t|t-1}$ and $b_{t|t-1}$.

Generalized linear models. The count models presented until here provide an explicit model for dependence of y_t on past outcomes. Zeger (1988) proposed a model which extends the generalized linear models and introduces a latent multiplicative autoregressive term ε_t with unit expectation, variance σ^2 and autocovariances $Cov[\varepsilon_t, \varepsilon_{t-k}] = \sigma^2 \rho_{k\varepsilon}$ where $\rho_{k\varepsilon} = Corr[\varepsilon_t, \varepsilon_{t-k}]$ is the autocorrelation function.

Again, I use the Poisson distribution to present this model. The dependent variable y_t is specified conditionally on ε_t, to be independent over t with mean and variance $\lambda_t \varepsilon_t$ where λ_t is a function of regressors via

$$\lambda_t = \exp(x_t^{'} \beta), \tag{1.3.35}$$

and ε_t is an unobservable latent variable ($\varepsilon_t > 0$). The moments conditional on observation of ε_t are

$$E[y_t \mid \lambda_t, \varepsilon_t] = \lambda_t \varepsilon_t, \tag{1.3.36}$$

and

$$Var[y_t \mid \lambda_t, \varepsilon_t] = \lambda_t \varepsilon_t, \tag{1.3.37}$$

In this sense the model is equidispersed. However, when we consider the marginal distribution of y_t, marginal with respect to ε_t but still conditional on λ_t, it is overdispersed:

$$E[y_t \mid \lambda_t] = \lambda_t, \tag{1.3.38}$$

and

$$Var[y_t \mid \lambda_t] = \lambda_t + \sigma^2 \lambda_t^2. \tag{1.3.39}$$

In order to estimate this model by maximum likelihood, one would need to specify a density for $y_t \mid \varepsilon_t$ and a density for ε_t. In most cases, no closed-form solution would be available. Instead, a quasilikelihood method is adopted, which only requires the knowledge of mean, variance and autocovariances of y_t. This method can also be viewed as a count data analog of the Cochrane-Orcutt method for regression models in which all the serial correlation is assumed to come from the error term. The method has been applied to sudden infant death syndrome by Campbell (1994). Brännäs and Johansson (1994) study the Zeger (1988) model in detail. They observe that the Poisson pseudo-maximum likelihood estimation (PMLE) is still consistent and that yields similar estimates and efficiency to estimates using weighted least squares as proposed by Zeger (1988), once appropriate correction for serial correlation is made to the Poisson PMLE standard errors.

While conceptually this method is quite close to what is proposed in this thesis, there are nonetheless important differences in that it is fundamentally a static model with a correction for autocorrelation in the same sense as generalized least squares (GLS) are, whereas the model in this thesis is an explicitly dynamic one. The interest is not limited to getting correct inference about the parameters on the exogenous variables but also lies in adequately capturing the dynamics of the system. In order to achieve this, a more parametric approach is taken, which, among other things, allows forecasting.

The autoregressive conditional Poisson model. This model was introduced by Heinen (2003). In this univariate count model the Poisson distribution is replaced by the double Poisson distribution of Efron (1986) in a regression context. The advantages of using this distribution are that it can be both under- and overdispersed, depending on whether the dispersion parameter is smaller or larger than 1. The main advantages of this model are that it is flexible, parsimonious and easy to estimate using maximum likelihood. Results are easy to interpret and standard hypothesis tests are available. In addition, given that the autocorrelation and the density are modelled explicitly, the model is well suited for both point and density forecasts, which can be of interest in many applications. In Chapter 2 we develop the multivariate generalization of this model, hence we do not present it here in more detail.

1.3.4 Ordered probit models

The ordered probit model is a technique frequently used in cross-sectional studies when a dependent variable takes only finite number of values possessing a natural ordering. Heuristically, ordered probit analysis is a generalization of the linear regression model to cases where the dependent variable is discrete. The ordered probit model have been used by Hausman, Lo, and Mackinlay (1992) to deal with price discreteness. With the ordered probit they capture the impact of explanatory variables on price changes while also accounting for discreteness and irregular trade times. Ranaldo (2003) and Pascual and Veredas (2004) use it to test different microstructure hypotheses. For the mathematical development of this section I follow Hausman, Lo, and Mackinlay (1992).

Let y_t^* be an unobservable continuous random variable such that:

$$y_t^* = X_t^{'}\beta + \varepsilon_t. \tag{1.3.40}$$

where X_t is a (q x 1) vector of predetermined variables that determine the conditional mean of y_t^*, and ε_t is the independent but not identically distributed error such that $E[\varepsilon_t \mid X_t] = 0$.

The essence of the order probit model is the assumption that the observed random variable y_t, such as observed prices changes (Hausman, Lo, and Mackinlay (1992)) or order aggressiveness (Ranaldo (2003)), is related to the continuous variable y_t^* in the following manner:

$$y_t = \begin{cases} s_1 & \text{if } \; -\infty < y_t^* \leq \alpha_1, \\ s_2 & \text{if } \; \alpha_1 < y_t^* \leq \alpha_2, \\ \vdots & \vdots \\ s_m & \text{if } \; \alpha_{m-1} < y_t^* \leq \infty. \end{cases} \tag{1.3.41}$$

where the right hand side of equation (1.3.41) is a partition of the state space S^* of y_t^*, and the s_j's are the discrete values that make up the state space S of y_t. The motivation for the order probit specification is to uncover the mapping between S^* and S as a function of the predetermined variables.

The conditional distribution of the observed variable y_t, conditioned on X_t, is determined by the partition of boundaries and the distribution of ε_t. For example, the conditional distribution is:

$$P(y_t = s_i \mid X_t) = P(X_t^{'}\beta + \varepsilon_t \; \in \; A_i \mid X_t) \tag{1.3.42}$$

where A_i represents the partition of the state space S^* written explicitly in the right hand side of equation (1.3.41). For example, $A_1 = (-\infty, \alpha_1]$, $A_2 = (\alpha_1, \alpha_2]$, and so on. Hence, assuming a $N(0, \sigma^2)$ distribution for ε_t,

$$P(y_t = s_i \mid X_t) = \begin{cases} P(X_t^{'}\beta + \varepsilon_t \leq \alpha_1 \mid X_t) = \Phi(\frac{\alpha_1 - X_t^{'}\beta}{\sigma}) & \text{if } \; i = 1 \\ P(\alpha_{i-1} < X_t^{'}\beta + \varepsilon_t \leq \alpha_i \mid X_t) = \Phi(\frac{\alpha_1 - X_t^{'}\beta}{\sigma}) - \Phi(\frac{\alpha_{i-1} - X_t^{'}\beta}{\sigma}) & \text{if } \; 1 < i < m \\ P(\alpha_{m-1} < X_t^{'}\beta + \varepsilon_t \mid X_t) = 1 - \Phi(\frac{\alpha_{m-i} - X_t^{'}\beta}{\sigma}) & \text{if } \; i = m \end{cases} . \tag{1.3.43}$$

where we define $\Phi(.)$ as the standard normal cumulative distribution function. In order to identify the model it is often used the following normalization: $\beta = [1, \tilde{\beta}]'$. It is also usually to assume one of the $\alpha_i = 0$. In both cases $\sigma^2 = 1$.

To better understand the ordered probit model, observe that the probability of any particular observed variable is determined by where the conditional mean lies relative to the partition boundaries. Therefore, for a given conditional mean $X_t^{'}\beta$, shifting the boundaries will alter the probabilities of observing each state. In fact, shifting the boundaries appropriately, the ordered probit model can fit any arbitrary multinomial distribution. This implies that the assumption of normality underlying the order probit plays no special role in determining the probabilities of states.

Given the partition boundaries, a higher conditional mean $X_t^{'}\beta$ implies a higher probability of observing a more extreme positive state. Of course, the label of states is arbitrary, but the order probit model makes use of a natural ordering of states. The regressors allow to separate the effects of various economic factors that influence the likelihood of one state versus the other. For example, suppose that a large positive value of X_1 usually implies a large negative observed price change and vice versa. Then the order probit coefficient β_1 will be negative in sign and large in magnitude (relative to σ).

In empirical studies of market microstructure, this model has been used to test theories with tick-by-tick data in a univariate framework. See for example Ranaldo (2003) and Pascual and Veredas (2004). However, with these approaches the possibility of auto- and cross-correlation among different market events cannot be analyzed. This is a serious limitation of this model if one wants to analyze not only univariate dynamics but also the multivariate dependence among different market events.

1.4 Glossary

This section presents some terminology that is going to be used in the development of this part of the thesis. It does not intend to be exhaustive but it aims to serve as a good tool to understand some terms used later.

1. **Ask price**: Quoted price at which a market maker is committed to sell an asset. This price is usually valid for a certain volume (depth). In an order book market, the price of the lowest limit sell order is called the ask price.

2. **Bid price**: Quoted price at which a market maker is committed to buy an asset. This price is usually valid for a certain volume (depth). In an order book market, the price of the highest limit buy order is called the bid price.

3. **Block orders**: Orders to buy or sell large block of shares, i.e. large trades with a volume in excess of 10,000 shares. Normally, these trades are prearranged in the up-stairs market by a block trader, who is in charge of locating potential counterparties to the large order.

4. **Breadth**: The market has many participants.

5. **Deep market**:A market is deep if its prices are close to the fair price for large trades (see tight markets).

6. **Depth**: Volume for which the bid-ask prices quoted by a market maker are guaranteed.

7. **Depth (liquidity)**: A market is considered "depth" in the liquidity sense, if there is a large incremental quantity available for sale above the current market price or if there

is a large incremental quantity that is sought (by a buyer or buyers) below the current market price.

8. **Duration**: Time between two market events, which can be for example two trades (trade duration) or two quotes (quote duration).

9. **Execution conditions**: Limit orders and market orders may be restricted with regard to execution, validity and trading properties.

10. **Fill-or-kill (FOK) order**: An execution condition that establishes that the order should be executed immediately and fully. Otherwise it will be deleted without being inserted in the system.

11. **Immediate-or-cancel (IOC) order**: An execution condition that establishes that the order should be executed immediately and fully- or at least as fully as possible. Every part that is not executed immediately will not be inserted into the order book.

12. **Inside Spread**: Price difference between the best bid-ask prices quoted by the market makers (these best quotes can be posted by different market makers) in a price driven market, or price difference between the price of the lowest limit sell order and the price of the highest buy order in an order driven market. See also the definition of spread.

13. **Limit order**: Order to buy or sell an asset at a given price and within a given time interval. A limit buy order specifies the maximum price at which the asset can be bought, a limit sell order specifies the minimum price at which the asset can be sold.

14. **Making a market**: In a price driven market, a market maker makes the market for a given asset, meaning that he has the obligation (usually enforced by the exchange) to quote firm bid and ask prices for this asset.

15. **Market maker**: Designated person (usually employed by the exchange or by banks affiliated with the exchange) who has the obligation to quote firm bid-ask prices for a given asset. These bid-ask prices are valid up to a given number of shares (depth). The market maker buys the asset at the bid price and sell the asset at the ask price.

16. **Market order**: Order to buy or sell an asset at the best available price. This order is executed immediately, i.e. upon the receipt of the order by a market maker or by the centralized order book system.

17. **Market-on-close orders**: Orders which are executed near the end of the trading session. They require special trading procedures.

18. **Market transparency**: Ability of the market participants to observe information about the trading process.

19. **Matching of orders**: Orders are said to be matched when a buy order (sell order) is crossed with a sell order (buy order), i.e. when a trade occurs between a buyer and a seller.

20. **Odd-lot orders**: Orders to trade a number of shares that are not multiple of 100 (the usual round lot size).

21. **Order book**: Complete collection of limit buy and sell orders entered by traders. In an order driven market, the order book is usually managed by a centralized computer. In a price driven market or hybrid system, the order book is maintained by the market maker.

22. **Over-the-counter security system**: Trading system where traders deal the securities by phone.

23. **Preferencing**: In a price driven market, preferencing occurs when traders assign their orders to specific market makers, i.e. their orders are routed to market makers on the basis of pre-arrangements.

24. **Resilience**. Price impacts caused by the trading are small and quickly die out, i.e. it is related to the speed of the prices to return to the equilibrium.

25. **Spread**: Price difference between the quoted bid-ask prices of a market maker. See also the definition of inside spread.

26. **Stop orders**: Orders that are inserted in the system as a market or limit order, when the indicative price reaches a certain stop limit.

27. **Tight market**: A market is tight if its prices are close to the fair market values for small trades.

28. **Trading mechanism**: Set of rules governing the exchange of financial assets (stocks, derivatives) or foreign currencies in a market.

29. **Up-stairs trade**: trade that is routed to the so-called upstairs market of NYSE, where block trades are prearranged by a block trader.

Chapter 2

Dynamic count data models

This chapter reports on research done jointly with Andréas Heinen.

2.1 Introduction

In empirical studies of market microstructure, the autoregressive conditional duration (ACD) model, introduced by Engle and Russell (1998), has been used widely to test theories with tick-by-tick data in a univariate framework. This model is designed specifically to deal with the irregularly-spaced nature of financial events. However, extensions to more than one series have proven to be very difficult. The difficulty comes from the very nature of the data, which are by definition not aligned in time, i.e. the times at which an event of any type happens are random. Engle and Lunde (2003) suggest a model for the bivariate case, but the specification is not symmetric in the two processes. They analyze jointly the duration between successive trades and the duration between a trade and the next quote arrival. This is done in the framework of competing risks. Spierdijk, Nijman, and van Soest (2002) model bivariate durations using a univariate model for the duration between the arrival of all events, regardless of their type, and a probit specification which determines the type of event that occurred. These models become intractable when the number of series is greater than two.

In this chapter we suggest working with counts instead of durations, especially when there is more than two series. Any duration series can be transformed into a series of counts by choosing an appropriate interval, which depends on the applications at hand, and counting the number of events that occur in each interval. The loss of information from considering counts is largely compensated for by the possibility of flexibly modelling interactions between several series. Moreover, most applications involve relatively rare events, which makes the use of the normal distribution questionable. Thus, modelling this type of series requires one to deal explicitly with the discreteness of the data as well as its time series properties and correlation. Neglecting any of these characteristics would lead to potentially serious misspecification.

We introduce a new multivariate model for time series count data. The multivariate autoregressive conditional double Poisson model (MDACP) makes it possible to deal with issues of discreteness, over- and underdispersion (variance greater or smaller than the mean) and both cross- and serial correlation. This chapter constitutes a multivariate extension to the univariate time series of counts model developed in Heinen (2003). We take a fully parametric approach where the counts have the double Poisson distribution proposed by Efron (1986) and their mean, conditional on past observations, is autoregressive. In order to introduce contemporaneous correlation we use a multivariate normal copula. This copula is very flexible, since it can accommodate positive and negative correlation, something that is impossible in most existing multivariate count distributions. The models are estimated using maximum likelihood, which makes the usual tests available. In this framework autocorrelation of the count processes can be tested with a straightforward likelihood ratio test, whose simplicity is in sharp contrast with test procedures in the latent variable time series count model of Zeger (1988). We apply a two-stage estimation procedure developed in Patton (2002), which consists in estimating first the marginal models and then the copula, taking the parameters of the marginal models as given. This considerably eases estimation of the model. In order to capture the dynamic interactions between the series we model the conditional mean as a VARMA-type structure, focusing our attention to the (1,1) case, motivated largely by considerations of parsimony.

It is well documented in market microstructure literature that the trading process conveys information. The key element in this literature is the existence of informed traders that trade on private information. This private information can be of two kinds, one linked directly with one asset (stock-specific) and other related to a given industry or sector (sector-wide). According to Admati and Pfleiderer (1988) and Easley and O'Hara (1992) frequent trading implies that news is arriving to the market. Thus a higher number of trades in a given time interval is a signal for the arrival of news. How much sector-specific information a stock trading activity contains has important implications from the point of view of identifying sectorial leaders.[1] For example, assume that exist two related stocks S_1 and S_2. It could be interesting for a S_1-holder to observe the traders of the other stock since there is a positive probability that S_2-traders are trading based on sector-wide private information and thus, their trades may be indicative of relevant information.

Spierdijk, Nijman, and van Soest (2002) study this question using a duration-based approach for pairs of assets. As a feasible alternative to multivariate duration models, we apply the MDACP to the study of sector and stock specific news of the most important US department stores traded on the New York Stock Exchange during the year 1999. We model the dynamics of the number of transactions of all stocks simultaneously using an intuitive and parsimonious factor structure, whereby the conditional mean of every series depends on one lag of itself, one lag of the count and one factor of the cross-section of lagged counts. We show that the assets that contain more sector information correspond to assets with larger market capitalizations. This is in contradiction with the findings of Spierdijk, Nijman, and van Soest (2002), who find that it is the most frequently traded stocks that contain most sector-specific information.

The chapter is organized as follows. Section 2.2 introduces the multivariate autoregressive double Poisson model and shows how we use copulas in the present context. Section 2.3

[1]Sectorial leaders are the assets that have a large impact on trades of related assets before some sector-wide information is revealed to the marked.

presents the empirical application. Section 2.4 concludes and presents some ideas for future research. Finally Section 2.5 presents a mathematical appendix.

2.2 The multivariate autoregressive double Poisson model

In this section we discuss the way in which we use copulas and the continuous extension argument to generate a multivariate discrete distribution. Then we present the conditional distribution and the conditional mean of the Multivariate autoregressive double Poisson (MDACP). Next we summarize the features of our model and establish its properties.

2.2.1 A general multivariate model using copulas

In order to generate richer patterns of contemporaneous correlations, we resort to copulas. Copulas provide a very general way of introducing dependence among several series with known marginals. Copula theory goes back to the work of Sklar (1959), who showed that a joint distribution can be decomposed into its K marginal distributions and a copula, that describes the dependence between the variables. This theorem provides an easy way to form valid multivariate distributions from known marginals that need not be necessarily the same distributions, i.e. it is possible to use normal, Student-t or any other marginal, combine them with a copula and get a suitable joint distribution, which reflects the kind of dependence present in the series. A more detailed account of copulas can be found in Joe (1997) and in Nelsen (1999).

Let $H(y_1, \ldots, y_K)$ be a continuous K-variate cumulative distribution function with univariate margins $F_i(y_i)$, $i = 1, \ldots, K$, where $F_i(y_i) = H(\infty, \ldots, y_i, \ldots, \infty)$. According to Sklar (1959), there exists a function C, called copula, mapping $[0,1]^K$ into $[0,1]$, such that:

$$H(y_1, \ldots, y_K) = C(F_1(y_1), \ldots, F_K(y_K)) \ . \tag{2.2.1}$$

The joint density function is given by the product of the marginals and the copula density:

$$\frac{\partial H(y_1, \ldots, y_K)}{\partial y_1 \ldots \partial y_K} = \prod_{i=1}^{K} f_i(y_i) \frac{\partial C(F_1(y_1), \ldots, F_K(y_K))}{\partial F_1(y_1) \ldots \partial F_K(y_K)} \ . \tag{2.2.2}$$

With this we can define the copula of a multivariate distribution with Uniform $[0,1]$ margins as:

$$C(z_1, \ldots, z_K) = H(F_1^{-1}(z_1), \ldots, F_K^{-1}(z_K)) \ , \tag{2.2.3}$$

where $z_i = F_i(y_i)$, for $i = 1, \ldots, K$.

As we can see with the use of the copulas we are able to map the univariate marginal distributions of K random variables, each supported in the $[0,1]$ interval, to their K-variate distribution, supported on $[0,1]^K$, something that holds, no matter what the dependence among the variables is (including if there is none).

Most of the literature on copulas is concerned with the bivariate case. However, we are trying to specify a general type of multivariate count model, not limited to the bivariate case. Whereas there are many alternative formulations in the bivariate case, the number of possibilities for multi-parameter multivariate copulas is rather limited. We choose to work with the most intuitive one, which is arguably the Gaussian copula, obtained by the inversion method (based on Sklar (1959)). This is a K-dimensional copula such that:

$$C(z_1, \ldots, z_K; \Sigma) = \Phi^K(\Phi^{-1}(z_1), \ldots, \Phi^{-1}(z_K); \Sigma) \ , \tag{2.2.4}$$

and its density is given by,

$$c(z_1, \ldots, z_K; \Sigma) = \mid \Sigma \mid^{-1/2} \exp\left(\frac{1}{2}(q'(I_K - \Sigma^{-1})q\right) \ , \tag{2.2.5}$$

where Φ^K is the K-dimensional standard normal multivariate distribution function, Φ^{-1} is the inverse of the standard univariate normal distribution function and $q = (q_1, \ldots, q_K)'$ is a vector of normal scores $q_i = \Phi^{-1}(z_i)$, $i = 1, \ldots, K$. Furthermore, it can be seen that if $Y_1, \ldots, Y_K$ are mutually independent, the matrix Σ is equal to the identity matrix I_K and the copula density is then equal to 1.

In the present chapter we are using a discrete marginal, the double Poisson, whose support is the set of integers, instead of continuous ones, which are defined for real values. If the marginal distributions functions are all continuous then C is unique. However when the marginal distributions are discrete, this is no longer the case and the copula is only uniquely identified on $\overset{K}{\underset{i=1}{\otimes}} Range(F_i)$, a K-dimensional set, which is the Cartesian product of the range of all marginals. Moreover, a crucial assumption, which underlies the use of copulas, is that the marginal models are well specified and that the probability integral transformation (PIT) of the variables under their marginal distribution is distributed uniformly on the $[0, 1]$ interval. The problem with discrete distributions is that the Probability Integral Transformation Theorem (PITT) of Fisher (1932) does not apply, and the uniformity assumption does not hold, regardless of the quality of the specification of the marginal model. The PITT states that if Y is a continuous variable, with cumulative distribution F, then

$$Z = F(Y), \tag{2.2.6}$$

is uniformly distributed on $[0, 1]$.

We use the continuous extension argument proposed by Denuit and Lambert (2004) to overcome these difficulties and apply copulas with discrete marginals.[2] The main idea of the continuous extension is to create a new random variable Y^* by adding to a discrete variable Y a continuous variable U valued in $[0, 1]$, independent of Y, with a strictly increasing cdf, sharing no parameter with the distribution of Y, such as the Uniform $[0, 1]$ for instance:

$$Y^* = Y + (U - 1). \tag{2.2.7}$$

[2]Machado and Santos Silva (2003) use this extension in order to work out the theoretical properties of a quantile estimator for discrete data.

The continuous extension does not alter the concordance between pairs of random variables; intuitively, two random variables Y_1 and Y_2 are concordant, if large values of Y_1 are associated with large values of Y_2. Concordance is an important concept, since it underlies many measures of association between random variables, such as Kendall's tau for instance. It is easy to see that the continuous extension does not affect concordance, since $Y_1^* > Y_2^* \iff Y_1 > Y_2$.

Using the continuous extension, we state a discrete analog of the PITT. If Y is a discrete random variable with domain χ, in $\mathbf{N}$, such that $f_y = P(Y = y),\ y \in \chi$, continued by U, then

$$Z^* = F^*(Y^*) = F^*(Y + (U-1)) = F\left([Y^*]\right) + f_{[Y^*]+1}U = F(Y-1) + f_yU \tag{2.2.8}$$

is uniformly distributed on $[0, 1]$, and $[Y]$ denotes the integer part of Y.

In this chapter, we use the continuous extension of the probability integral transformation in order to test the correct specification of the marginal models. If the marginal models are well-specified, then Z^*, the PIT of the series under the estimated distribution and after the continuous extension, is uniformly distributed. We also use Z^* as an argument in the copula, since, provided that the marginal model is well specified, this ensures that the conditions for use of a copula are met.

One remark needs to be made concerning the use of the continuous extension in the present context. In a sense the lack of identifiability of the copula outside of the range of the cumulative distribution of the marginal model is less acute in the time-varying distribution case, as the number of points at which the copula is observed increases, relative to the static case. In order to illustrate this point, let's consider the case of Bernoulli variables, which are in a sense, the 'most discrete' possible random variables. The problem we describe is the same with the Poisson or the double Poisson distribution. We consider the Bernoulli variables Y_i, for $i = 1, \ldots, K$, whose cumulative density functions F_i can only take 3 possible values:

$$Z_{i,t} = F_i(Y_{i,t}) = \begin{cases} 0 & \text{if } y_{i,t} \leq 0 \\ p_i & \text{if } 0 < y_{i,t} < 1 \\ 1 & \text{otherwise} \end{cases}$$

The copula is then only identified on the set $S = \bigotimes_{i=1}^{K} \{0, p_i, 1\}$. Therefore it is impossible to distinguish two copulas which have the same values on S, but are different on $[0, 1]^n \bigcap \overline{S}$. In the case where the distributions are time-varying, we have:

$$F_{i,t}(Y_{i,t}) = \begin{cases} 0 & \text{if } y_{i,t} \leq 0 \\ p_{i,t} & \text{if } 0 < y_{i,t} < 1 \\ 1 & \text{otherwise} \end{cases}$$

The copula is now identified on the set $\bigcup_{t=1}^{T} \bigotimes_{i=1}^{K} \{0, p_{i,t}, 1\}$, which is obviously much larger a set than S. Nonetheless, it remains true in the time-varying case, that the non-corrected Z-statistic is not uniformly distributed, which, alone, justifies the use of the continuous extension.

2.2.2 The conditional distribution and the conditional mean

In order to extend the autoregressive conditional double Poisson model to a $(K \times 1)$ vector of counts N_t, we build a VARMA-type system for the conditional mean. In a first step, we assume that conditionally on the past, the different series are uncorrelated. This means that there is no contemporaneous correlation and that all the dependence between the series is assumed to be captured by the conditional mean. Even though the Poisson distribution with autoregressive means is the natural starting point for counts, one of its characteristics is that the mean is equal to the variance, a property referred to as equidispersion. However, by modelling the mean as an autoregressive process, we generate overdispersion even in the simple Poisson case.

In some cases one may want to break the link between overdispersion and serial correlation. It is quite probable that the overdispersion in the data is not attributable solely to the autocorrelation, but also to other factors, for instance unobserved heterogeneity. It is also imaginable that the amount of overdispersion in the data is less than the overdispersion resulting from the autocorrelation, in which case an underdispersed marginal distribution might be appropriate. In order to account for these possibilities we consider the double Poisson distribution introduced by Efron (1986) in the regression context, which is a natural extension of the Poisson model and allows one to break the equality between conditional mean and variance. This density is obtained as an exponential combination with parameter ϕ of the Poisson density of the observation y with mean μ and of the Poisson with mean equal to the observation y, which can be thought of as the likelihood function taken at its maximum value.
The density of the double Poisson is:

$$f(y, \mu, \phi) = c(\mu, \phi) \left(\phi^{\frac{1}{2}} e^{-\phi\mu}\right) \left(\frac{e^{-y} y^y}{y!}\right) \left(\frac{e^{\mu}}{y}\right)^{\phi y} \tag{2.2.9}$$

where $c(\mu, \phi)$ is such that the probabilities add up to 1. Efron (1986) shows that the value of the multiplicative constant $c(\mu, \phi)$, correctly normalized varies little and suggests the following approximation for this constant:

$$\frac{1}{c(\mu, \phi)} \approx 1 + \frac{1-\phi}{12\mu\phi}\left(1 + \frac{1}{\mu\phi}\right) \tag{2.2.10}$$

Furthermore, he suggests maximizing the likelihood while leaving out the highly nonlinear multiplicative constant in order to find the parameter estimates and using the correction factor when making probability statements using the density.

The advantages of using this distribution are that it can be both under- and overdispersed, depending on whether the parameter ϕ is larger or smaller than 1. We write the model as:

$$N_{i,t} \mid \mathcal{F}_{t-1} \sim DP(\mu_{i,t}, \phi_i) \ , \ \forall i = 1, \dots, K. \tag{2.2.11}$$

where $\mathcal{F}_{t-1}$ designates the past of all series in the system up to time $t-1$.[3] With the double

[3]It is shown in Efron (1986) (Fact 2) that the mean of the Double Poisson is μ and that the variance is

Poisson, the conditional variance is equal to:

$$V[N_{i,t} \mid \mathcal{F}_{t-1}] = \sigma^2_{i,t} = \frac{\mu_{i,t}}{\phi_i} \tag{2.2.12}$$

The coefficient ϕ_i of the conditional distribution is a parameter of interest, as values different from 1 represent departures from the Poisson distribution. The double Poisson generalizes the Poisson in the sense of allowing more flexible dispersion patterns.

The vector of conditional means, μ_t, is assumed to follow a VARMA-type process:

$$E[N_t \mid \mathcal{F}_{t-1}] = \mu_t = \omega + \sum_{j=1}^{p} A_j N_{t-j} + \sum_{j=1}^{q} B_j \mu_{t-j} \tag{2.2.13}$$

For reasons of simplicity, in most of the ensuing discussion, we focus on the most simple $(1,1)$ case and for notational simplicity, we denote $A = \sum_{j=1}^{p} A_j$, $B = \sum_{j=1}^{q} B_j$, and drop the index whenever there is no ambiguity.

In most empirical applications, especially when the number of series analyzed jointly is large, some additional restrictions might have to be imposed on A and B.

In systems with large K, which could be found, for instance when analyzing a large group of stocks like the constituents of an index, the full approach would not be feasible, as the number of parameters would get too large. If we assume that A and B are of full rank, the number of parameters that has to be estimated in this model would be $2K^2 + K$. In situations where this is not an option, we propose to impose some additional structure on the process of the conditional mean. The most interesting structure is the reduced rank and own effect model where

$$\mu_t = \omega + (diag(\alpha_i) + \gamma\delta')N_{t-1} + diag(\beta_i)\mu_{t-1} \ . \tag{2.2.14}$$

In this formulation it is assumed that for every series the conditional mean depends on one lag of itself, one lag of the count and r factors of the cross-section of lagged counts. In this context, the matrix $A = diag(\alpha_i) + \gamma\delta'$ where γ and δ are $(K \times r)$ matrices and $diag(\alpha_i)$ is a $(K \times K)$ diagonal matrix that accounts for the own effect. The matrix $B = diag(\beta_i)$ accounts for the effect of the lagged mean count on the same current mean. This is adapted to large systems, where imposing a reduced rank is necessary for practical reasons, but there is reason to believe that each series own past has explanatory power beyond the factor structure.

Moreover, in some cases one might want to assume that the dynamics of all the series under consideration is common, and that one factor explains the dynamics of the whole system. This is obtained as a special case of equation (2.2.14) when $\alpha_i = 0$, i.e. $A = \gamma\ \delta^{'}$, with $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_K)^{'}$ and $\delta = (1, \delta_2, \ldots, \delta_K)^{'}$, and we impose the normalization $\delta_1 = 1$ in order to identify the model. This means that we have an autoregressive process:

$$\mu_t = \omega + \gamma\delta' N_{t-1} + diag(\beta_i)\mu_{t-1} \tag{2.2.15}$$

approximately equal to $\frac{\mu}{\phi}$. Efron (1986) shows that this approximation is highly accurate, and we will use it in our more general specifications.

It is easy to show that the MDACP is stationary as long as the roots of the sum of the autoregressive coefficient matrices are within the unit circle, or equivalently, the eigenvalues of $(I - A - B)$ lie within the unit circle. In that case, the unconditional mean of the MDACP(p,q) is identical to the one of a VARMA process:

$$E[N_t] = \mu = (I - A - B)^{-1}\omega. \tag{2.2.16}$$

In the appendix we present two properties about the unconditional variance and the autocorrelation of the MDACP with an ARMA(1,1) structure.

2.2.3 Inference

Having dealt with the problems due to the discreteness and the time-varying nature of the marginal density in earlier sections, we proceed with the estimation of the model. The joint density of the counts in the double Poisson case with the Gaussian copula is:

$$h(N_{1,t}, \ldots, N_{K,t}, \theta, \Sigma) = \prod_{i=1}^{K} f_{DP}(N_{i,t}, \mu_{i,t}, \phi_i) \cdot c(q_t; \Sigma), \tag{2.2.17}$$

where $f_{DP}(N_{i,t}, \mu_{i,t}, \phi_i)$ denotes the double Poisson density as a function of the observation $N_{i,t}$, the conditional mean $\mu_{i,t}$ and the dispersion parameter ϕ_i. The function c denotes the copula density of a multivariate normal $N(0, \Sigma)$, and $\theta = (\omega, vec(A), vec(B))$.
The $q_{i,t}$, gathered in the vector q_t are the normal quantiles of the $z_{i,t}$:

$$q_t = (\Phi^{-1}(z_{1,t}), \ldots, \Phi^{-1}(z_{K,t}))', \tag{2.2.18}$$

where the $z_{i,t}$ are the PIT of the continued count data, under the marginal densities:

$$z_{i,t} = F^*(N^*_{i,t}) = F(N_{i,t} - 1) + f(N_{i,t}) * U_{i,t}, \tag{2.2.19}$$

$N^*_{i,t}$ is the continued version of the original count data $N_{i,t}$:

$$N^*_{i,t} = N_{i,t} + (U_{i,t} - 1). \tag{2.2.20}$$

Finally $U_{i,t}$ is a uniform random variable on $[0, 1]$.

Taking the log of the joint density (equation (2.2.17)) we obtain the conditional log likelihood function,

$$logL = \sum_{t=1}^{T} \left[\sum_{i=1}^{K} log(f_{DP}(N_{i,t}; \mu^*_{i,t}, \phi_i)) + log(c(q_t; \Sigma)) \right] \tag{2.2.21}$$

where T denotes the number of observations.

To estimate the model parameters we adopt a two-step procedure proposed by Patton (2002). In the first step we maximize the first part of the log-likelihood (2.2.21) which, written in detail, is given by

$$\sum_{t=1}^{T}\sum_{i=1}^{K}\left(\frac{1}{2}log\phi_i - \mu^*_{i,t}\phi_i - N_{i,t} + (1-\phi_i)log(N_{i,t}^{N_{i,t}}) + N_{i,t}\phi_i(1+log\mu^*_{i,t}) - log(N_{i,t}!)\right). \tag{2.2.22}$$

Since we employ the multivariate Gaussian copula, the second estimation step does not require any numerical optimization. The maximum likelihood estimate of the variance-covariance matrix Σ is simply the sample variance-covariance matrix,

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}\hat{q}_t\hat{q}_t^{'} . \tag{2.2.23}$$

Patton (2002) shows the asymptotic normality for the two-stage estimator based on the work of Newey and McFadden (1994).[4] The two-stage estimates of the copula parameters are less asymptotically efficient than the full information estimates. To take into consideration this fact, Patton adjusts the two-stage estimator taking a single iteration of the Newton-Raphson algorithm starting from the two-stage parameter estimate. With this modification the updated estimates attain the minimum asymptotic variance bound.

Specification tests can be conducted based on the usual likelihood statistics, but conveniently also by analyzing the properties of the "Pearson residuals", which are defined as $\epsilon_{i,t} = \frac{N_{i,t}-\mu_{i,t}}{\sigma_{i,t}}$. When a model is correctly specified, the estimated Pearson residuals should have an empirical variance close to one and exhibit no significant autocorrelation. Besides these methods for specification testing, one can also use the sequence of probability integral transforms as discussed by Diebold, Gunther, and Tay (1998). As outlined above, it is important to realize that correct specification of the density in the marginal models is crucial to the specification of the copula, as any mistake would have as a consequence the fact that the uniformity assumption is violated, which would invalidate the use of copulas. Then, if a model is correctly specified, the sequence of probability integral transforms $\{z_{i,t}\}$ must be uniformly distributed and iid. Straightforward tests for iid uniformity are readily available.

2.3 Sector- and stock-specific news

Much of the microstructure literature is based on the existence of asymmetric information and consequently of two types of traders: the uninformed who trade for liquidity reasons and informed traders who possess superior information. This superior information can be macroeconomic, sector- or stock-specific information. Through the trading process this information is disseminated to the public, therefore trading conveys information. According to Admati and Pfleiderer (1988) and Easley and O'Hara (1992) frequent trading implies that news is arriving to the market. Thus a higher number of trades in a given time interval is a signal of the arrival of news.

The trading activity of one asset does not only convey information about that specific

[4]Asymptotic theory of the one-step and two-step estimation is also investigated in Genest, Ghoudi, and Rivest (1995) for the well-specified case and in Cebrián, Denuit, and Scaillet (2004) for the misspecified case.

asset, it can also contain information about the whole sector this asset belongs to. In order to model comovement in trading activity within a sector, Spierdijk, Nijman, and van Soest (2002) propose a duration model for the trading intensities of pairs of stocks of department stores. Their model consists of a univariate duration model for the pooled trades of two stocks and a probit specification which determines in which stock a transaction took place. They classify stocks according to how much sector-wide information they contain, based on a series of ratios of the sample variance of the conditional intensity of the pooled and univariate ACD models for each pair of stocks. In recent years, the focus of empirical microstructure has shifted from the study of an individual asset to the analysis of the cross-sectional interactions amongst stocks. Hasbrouck and Seppi (2001) document the existence of commonalities in order flow that are responsible for about two thirds of the commonalities in returns, using principal components analysis and canonical correlations on the stocks of the Dow Jones Industrial Average.

We analyze the same data as Spierdijk, Nijman, and van Soest (2002), but the MDACP allows us to take into account the interaction amongst all stocks simultaneously as in Hasbrouck and Seppi (2001), which is helpful for the purpose of identifying leaders from the point of view of dissemination of sectorial information, while at the same time modelling the dynamics in a general framework.

2.3.1 Data

We are working with the five most important US department stores traded on the New York Stock Exchange during the year 1999: May Department Stores (MAY), Federated Department Stores (FD), J.C. Penney Company, Inc (JCP), Dillar's INC (DDS) and Saks Inc (SKS). We work with the number of trades in 5-minute intervals. The data we use was taken from the Trades and Quotes (TAQ) data set, produced by the New York Stock Exchange (NYSE). This data set contains every trade and quote posted on the NYSE, the American Stock Exchange and the NASDAQ National Market System for all securities listed on NYSE. We first remove any trade that occurred with non-standard correction or G127 codes (both of these are fields in the trades data base on the TAQ CD's), such as trades that were cancelled, trades that were recorded out of time sequence, and trades that were called for delivery of the stock at some later date. All trades that were recorded to have occurred before 9:45 AM or after 4 PM (the official close of trading) were removed. The reason for starting at 9:45 instead of 9:30 AM, the official opening time, is to make sure that none of the opening transactions were accidentally included in the sample, or that there would not be artificially low numbers of events at the start of the day, due to the fact that part of the first interval was taking place before the opening transaction. This could have biased estimates of intradaily seasonality.

The sample period goes from January 2nd 1999 to December 30th 1999. This means that the sample covers 252 trading days, that yield $18,900$ observations, as there are 75 5-minute intervals every day between 9:45 AM and 4 PM. The descriptive statistics are given in Table 2.1.

The means of the series are relatively small, which makes the use of a continuous distribution

Table 2.1: **Descriptive statistics**

	DDS	FD	JCP	MAY	SKS
No.trades	55,399	100,928	108,392	90,881	59,725
Mean	2.93	5.34	5.73	4.81	3.16
Median	2.00	5.00	5.00	4.00	3.00
Std.Dev.	2.57	3.56	3.89	3.04	2.84
Dispersion	2.25	2.38	2.64	1.92	2.55
Maximum	37	35	38	22	32
Minimum	0	0	0	0	0
$Q(20)$	11,560	15,504	34,482	8,531.7	33,679

This table presents the descriptive statistics for the number of trades of the Dillar's INC (DDS), Federated Department Stores (FD), J.C. Penney Company, Inc (JCP), May Department Stores (MAY) and Saks Inc (SKS). The sample period goes from January 2nd 1999 to December 30th 1999. The number of observations is $18,900$. $Q(20)$ is the Ljung-Box Q-statistic of order 20 on the series. The dispersion refers to the ratio of the variance to the mean.

like the normal problematic. As can be seen, the data exhibits significant overdispersion (the variance is greater than the mean), which can be due to autocorrelation or to overdispersion in the marginal distribution. The presence of overdispersion is confirmed by looking at the histogram of the data in Figure 2.1, which shows that, whereas the probability mass is fairly concentrated around the mean, there exist large outliers. There is significant autocorrelation in each series, as can be seen from the Ljung-Box Q-statistic shown here at order 20.

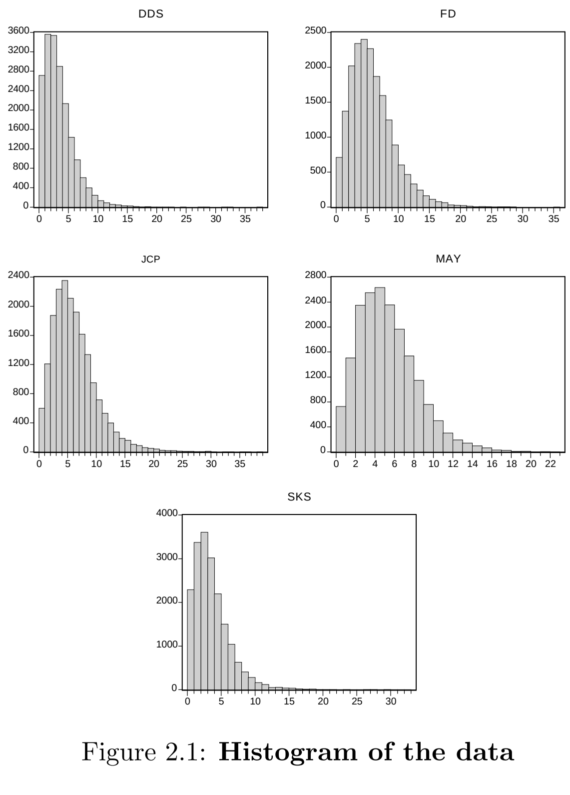


Figure 2.1: **Histogram of the data**

Table 2.2 presents the empirical correlation matrix among the five series. Figure 2.2 shows the auto- and cross-correlations of the vector of market-events, up to 375 5-minute intervals, which corresponds to 5 trading days. A very striking pattern of seasonality can be observed.

Clearly looking only at contemporaneous correlations does not reveal the full picture, there is a very significant and systematic link across time between the various trading events. The correlations move from positive to negative in a systematic way, which seems to be due to the presence of diurnal seasonality of the U-shape type (see Figure 2.3), which is commonly found in time series based on high-frequency data. The models we present consider the seasonality present in the data. We eliminate this seasonality pattern by the use of 30-minute dummies.

Table 2.2: **Correlation matrix of the counts**

	DDS	FD	JCP	MAY	SKS
DDS	1.00				
FD	0.27	1.00			
JCP	0.24	0.29	1.00		
MAY	0.25	0.30	0.31	1.00	
SKS	0.12	0.10	0.15	0.12	1.00

This table presents the correlation matrix among the five series we analyze. The series are: Dillar's INC (DDS), Federated Department Stores (FD), J.C. Penney Company, Inc (JCP), May Department Stores (MAY) and Saks Inc (SKS). The sample period goes from January 2nd 1999 to December 30th 1999.

2.3.2 Empirical results

In this subsection we discuss the estimates of two different specifications of the model, one based on the idea of a common factor and the second based on a mean structure based on a common factor, a series-specific lagged term in the moving average part and a diagonal autoregressive part. An obvious disadvantage of using these two factor-type specifications is that we do not consider cross effects, i.e. the impact that the lagged counts of a given asset could have on the others. However, this kind of formulation is very helpful when the number of variables increases and the number of parameters to estimate grows exponentially, making the estimation procedure untractable. In order to fit the dispersion we use the double Poisson distribution and we model seasonality using a series of half hourly dummy variables.

The results are shown in Table 2.3. The estimates of the MDACP are maximum likelihood estimates and the asymptotic variances are estimated using the "sandwich" estimator. The eigenvalues of $A + B$ are smaller than 1, which means that the estimated model seems to be stationary. However, some eigenvalues are very close to one, indicating the possibility of non stationarity of the model. A likelihood ratio test shows that the seasonality variables, i.e. the 30-minutes dummies (the estimates are not shown), are jointly significant. The coefficients on the seasonality shown in Figure 2.3 exhibit the well-documented U-shape, which means that there is more activity at the beginning and end of the trading day and less at lunch time. The dispersion parameter ϕ of the double Poisson is also very significantly smaller than 1 (which corresponds to the Poisson case). This means that the Poisson distribution is strongly rejected and that we have a much better model for the conditional distribution. Furthermore, if the model is well specified, the Pearson residuals have unitary variance and no significant auto- or cross-correlation left. The reported variances in Table 2.3 are close to 1.

An assumption underlying the use of copulas is the uniformity of Z, the PIT of the observations under the marginal distributions. If the density from the model is accurate, these values will be uniformly distributed and will have no significant autocorrelation left. In order to assess how close the distribution of the Z variable is to a uniform, we show quantiles of Z plotted against quantiles of the uniform distribution. The closer the plot is to a 45-degree line, the closer the distribution is to a uniform. Visual inspection of the Q-Q plots of the Z statistic of the factor plus own effect model in Figure 2.4 reveals that indeed the distribution

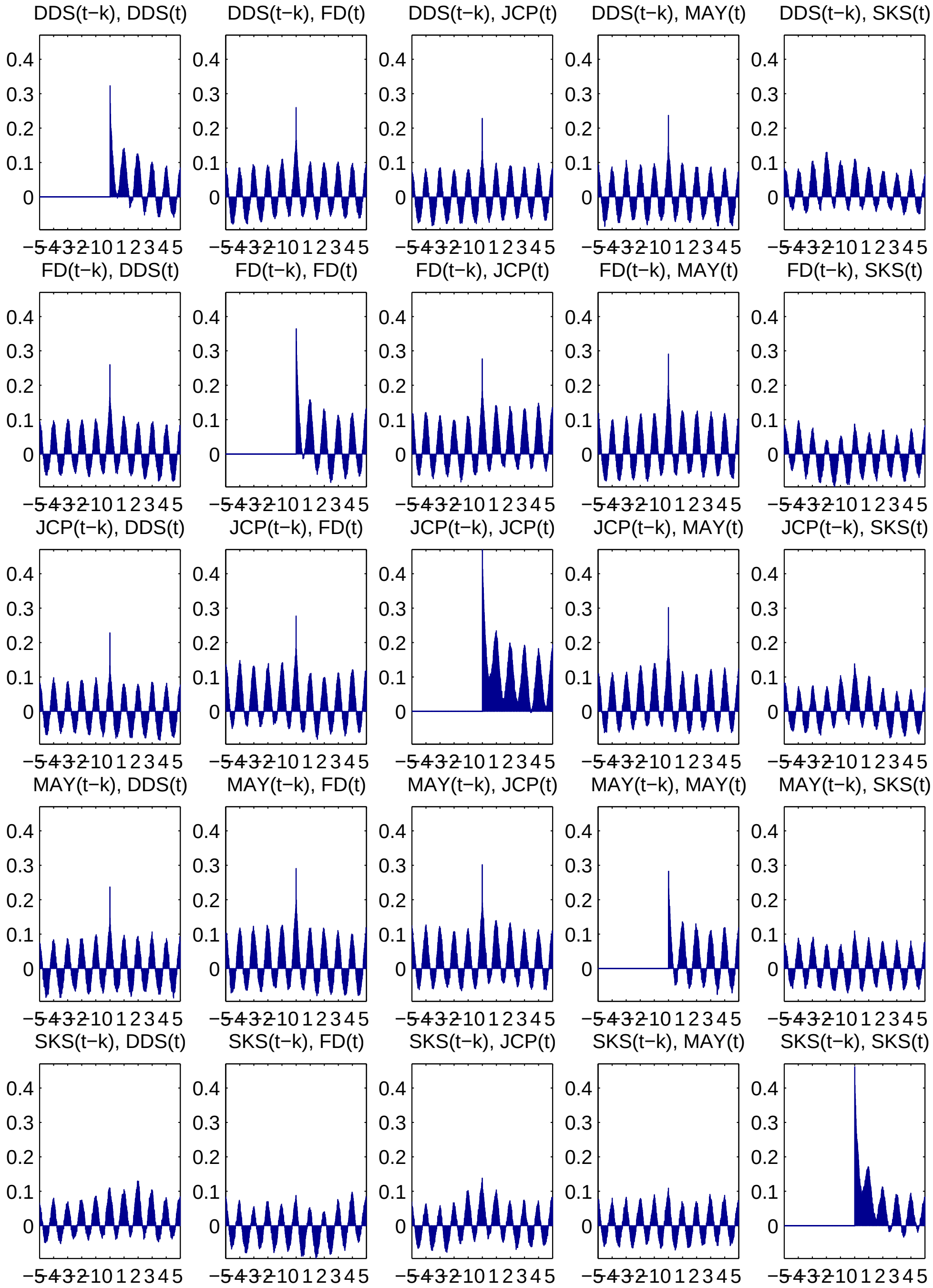


Figure 2.2: **Auto- and cross-correlogram of the data**

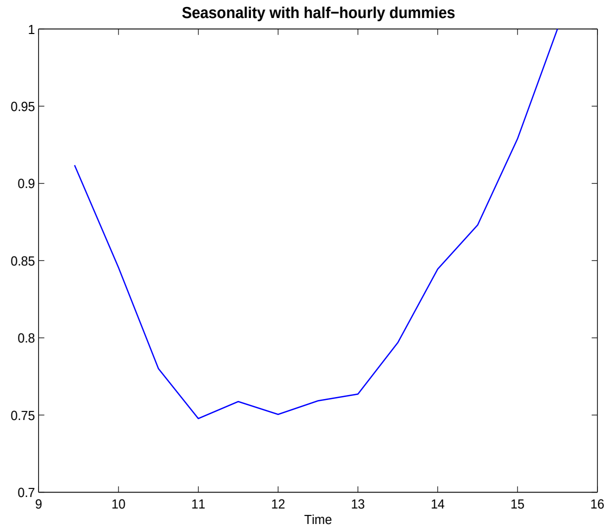


Figure 2.3: **Seasonality plot of JCP trades**

is well specified: the Z is uniformly distributed on [0,1], since the Q-Q plots nearly coincide with the 45-degree line. This means that with the use of the double Poisson we satisfy the uniformity assumption, which is the theoretical basis for using copulas. The same results hold for the factor only model. The autocorrelations of the Z statistic, shown in Figure 2.5 are essentially not significant, which indicates that the dynamics of the series is well accounted for. The autocorrelations of the Pearson residuals of the series (not shown) confirm that there is no more seasonal pattern left and the correlations are well below significance. This however is not the case for the factor only model, for which there still remains autocorrelation in the residuals.

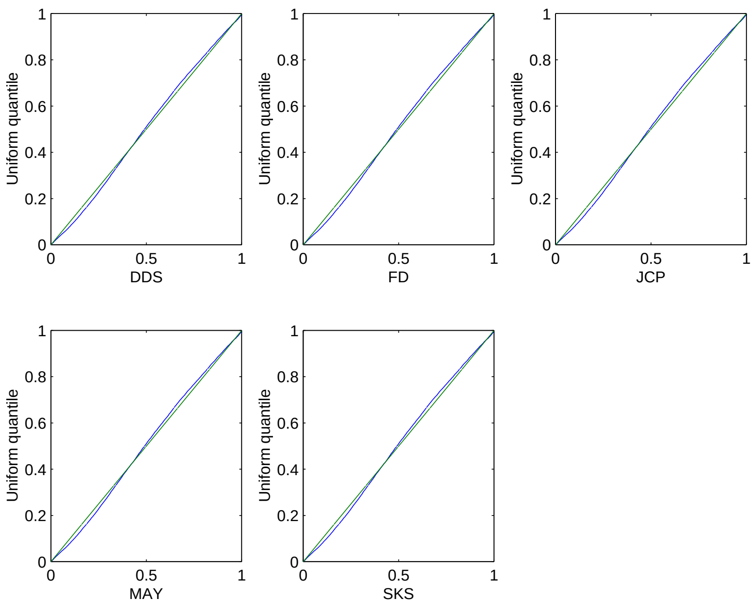


Figure 2.4: **Quantile Plots of the Z statistics of the MDACP model with factor and own effect**

In order to model the contemporaneous correlations we estimate a multivariate normal copula. As this model is somewhat involved in terms of the number of parameters, we use the

Table 2.3: **Maximum Likelihood Estimates of the MDACP models.**

θ	Single factor model					Single factor and own effect				
	DDS	FD	JCP	MAY	SKS	DDS	FD	JCP	MAY	SKS
ω_i	0.325	0.773	0.098	0.600	0.043	0.136	0.331	0.216	0.275	0.094
	(8.73)	(12.62)	(5.47)	(11.40)	(4.04)	(7.97)	(11.98)	(10.30)	(10.67)	(7.46)
α_{1i}						0.137				
						(25.80)				
α_{2i}							0.151			
							(17.51)			
α_{3i}								0.178		
								(36.50)		
α_{4i}									0.108	
									(12.19)	
α_{5i}										0.161
										(33.73)
γ	0.119	0.217	0.213	0.124	0.100	0.022	0.050	0.008	0.043	0.011
	(14.01)	(14.57)	(17.50)	(12.97)	(15.97)	(3.22)	(3.68)	(1.838)	(3.87)	(3.43)
δ	0.250	0.247	0.375	0.218	0.335	0.250	0.382	0.122	0.461	0.072
		(15.44)	(18.90)	(13.20)	(17.39)		(3.32)	(2.89)	(3.20)	(1.77)
β	0.694	0.664	0.790	0.761	0.839	0.811	0.777	0.814	0.820	0.825
	(35.26)	(38.06)	(106.25)	(48.28)	(111.66)	(110.05)	(97.78)	(155.75)	(103.36)	(143.00)
ϕ	0.504	0.496	0.514	0.571	0.498	0.546	0.542	0.575	0.600	0.584
	(94.09)	(96.24)	(95.95)	(95.58)	(99.52)	(92.78)	(98.59)	(101.26)	(97.26)	(95.21)
Eigenval	0.99	0.68	0.71	0.83	0.77	0.94	0.97	0.99	0.99	0.93
$Var(\epsilon_{i,t})$	1.00	0.99	1.00	0.97	1.04	0.97	0.98	0.99	0.97	0.98
LogL	-219,072					-214,463				

The table presents the Maximum Likelihood estimates of the multivariate autoregresive conditional double Poisson (MDACP) models on counts based on data of the 5 most important retail department stores: Dillar's INC (DDS), Federated Department Stores (FD), J.C. Penney Company, Inc (JCP), May Department Stores (MAY) and Saks Inc (SKS). The sample period goes from January 2nd 1999 to December 30th 1999. These models consider the seasonality presented in the data and eliminate it by the use of 30 minutes dummies. The t-statistics are presented in parenthesis. We impose the normalisation $\delta_1 = 0.25$ in order to identify the model. $\epsilon_{i,t} = \frac{N_{i,t}-\mu_{i,t}}{\sigma_{i,t}}$ are the Pearson residuals from the model. The equation of the factor only model is:
$\mu_t = \omega + \gamma\delta' N_{t-1} + diag(\beta_i)\mu_{t-1}$, and of the model with a factor and an own effect:
$\mu_t = \omega + (diag(\alpha_i) + \gamma\delta')N_{t-1} + diag(\beta_i)\mu_{t-1}$
If a model is well specified, the Pearson residuals have variance one.

two-step procedure of Patton (2002). Table 2.4 shows the copula correlation matrix Σ, which is responsible for the part of the contemporaneous and lagged cross-correlation which does not go through the time-varying mean.

To see the influence on the factor of each of the assets involved we just need to take a look at the vector of factor weights (the δ's). According to this the ranking of sectorial influence in the factor only model (left panel of Table 2.3) is JCP, SKS, DDS, FD and MAY. These results are closely related to Spierdijk, Nijman, and van Soest (2002) who find that the assets that contain more sectorial information are, in descending order, JCP, FD, SKS, DDS and MAY. As they mention, this ranking is related to the average number of transactions (see Table 2.1). This amounts to saying that the stocks with most sectorial information are the most frequently

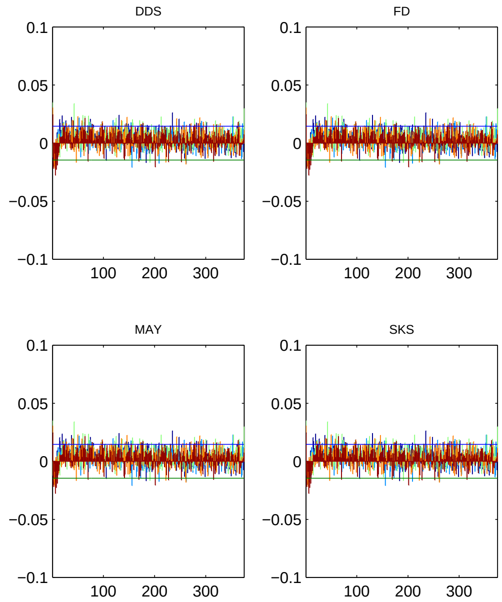


Figure 2.5: **Autocorrelation of the Z statistics of the MDACP model with factor and own effect**

traded ones. However, if instead we rely on the intuitive idea that every stock's past trading activity plays a special role for that asset, in addition to an effect through a common factor, we find quite a different result. For the results in the right side of Table 2.3, we find a quite different ranking: MAY, FD, DDS, JCP and SKS. This ranking no longer matches the ranking based on the average trading activity of the asset, but it is instead highly related with the market capitalizations of the stocks. Indeed, ranking the most important US department stores by their size we have: MAY, FD, JCP, DDS and SKS.[5]

Based on these results, we can conclude that indeed, within a sector two kinds of information matter for traders: stock specific information, related to the series-specific autocorrelation coefficients (the $\alpha_{i,i}$'s) and sector specific news, captured by the common factor (the δ's and γ's). Unlike traders trying to benefit from a stock-specific information, a trader with sector-specific information who is trying to conceal it, has the choice of which asset to trade in. He should naturally chose the asset with the least amount of sectorial information in its trading activity, as this would allow him to hide his private information without impacting the market too much. Based on our system of five department stores, trading in SKS and JCP could be appropriate. Of course, to obtain more general results, one would need to incorporate all stocks of that sector and not only the biggest ones as we do.

The comparison of our results with the ones of duration-based models suggests that taking into consideration all the assets simultaneously does make a difference. We are able to capture cross-sectional interactions with an intuitive factor-structure, commonly used in finance since the CAPM and also used more recently in the context of liquidity and order flow by Hasbrouck and Seppi (2001). This advantage of our multivariate specification over bivariate duration models compensates for the loss of information due to the aggregation from durations to counts. We have estimated our models for different time intervals (10 and 15 minutes) and we obtain

[5]Their market-cap in millions of US Dollars were: 11,226, 8,945, 3,538, 1,647, and 1,612 respectively.

Table 2.4: **Correlation Matrix of the q estimated by the MDACP model.**

	$COPULA - MDACP4S$				
	DDS	FD	JCP	MAY	SKS
DDS	1.00				
FD	0.16	1.00			
JCP	0.17	0.18	1.00		
MAY	0.15	0.17	0.20	1.00	
SKS	0.02	0.02	0.04	0.03	1.00

The table presents the correlation matrix of q, based on the probability integral transformation, Z, of the continued count data under the marginal densities estimated using the single factor plus own effect model by the two-step procedure. Recall that $q_i = \Phi^{-1}(z_i^*)$, where Φ^{-1} is the inverse of the standard univariate normal distribution function. The series are: Dillar's INC (DDS), Federated Department Stores (FD), J.C. Penney Company, Inc (JCP), May Department Stores (MAY) and Saks Inc (SKS). The sample period goes from January 2nd 1999 to December 30th 1999.

the same results (available upon request). This robustness over time aggregation and the accordance of our results with economic intuition increases our confidence in the findings.

2.4 Conclusions and future work

In this chapter we introduce new models for the analysis of multivariate time series of count data with many possible specifications. These models have proved to be flexible and easy to estimate. We discuss how to adapt copulas to the case of time series of counts and show that the multivariate autoregressive conditional double Poisson model (MDACP) can accommodate many features of multivariate count data, such as discreteness, over- and underdispersion (variance greater and smaller than the mean) and both auto- and cross-correlation. Hypothesis testing in this context is straightforward, because all the usual likelihood-based tests can be applied. Another important advantage of this model is that it can accommodate both positive and negative correlation among variables, which most multivariate count models cannot do, and this is shown to be important in our financial application.

As a feasible alternative to multivariate duration models, the model is applied to the study of sector and stock specific news related to the comovements in the number of trades per unit of time of the most important US department stocks traded on the New York Stock Exchange. We show that the informational leaders inside a specific sector are related to their size measured by their market capitalization rather than to their trading activity.

We advocate the use of the multivariate autoregressive conditional double Poisson model for the study of multivariate point processes in finance, when the number of variables considered simultaneously exceeds two and looking at durations becomes too difficult. Plans for further research include evaluating the forecasting ability of these models, both in terms of point and density forecasts and we leave more empirical applications for further work with more detailed tick-by-tick data sets. In Chapter 3 we develop a detailed analysis of the trading activity of the Xetra trading system using the MDACP model.

Related to the single factor model, Chapter 4 presents an alternative way to do this type of analysis. We present a procedure to improve the principal components analysis (PCA) when non-Gaussian data is used. We use this improved PCA in order to determine the common factor underlying the behavior of the stocks and to compare these results with the ones obtained in this chapter using the MDACP model.

2.5 Appendix

Proposition 2.5.1 (Unconditional variance of the MDACP(1,1) Model)**.** *The unconditional variance of the MDACP(1,1) model, when the conditional mean is given by (2.2.13), is equal to:*

$$vec(Var[N_t]) = \left(I_{K^2} + \left(\left(I_{K^2} - (A+B)\otimes(A+B)^{'}\right)^{-1} \cdot (A \otimes A^{'})\right)\right) \cdot vec(\Omega)\ , \tag{2.5.1}$$

where $\Omega = E\left(Var[N_t \mid \mathcal{F}_{t-1}]\right)$*. Under small dispersion asymptotics of Jorgensen (1987),* $\Omega \simeq V^{\frac{1}{2}}\Sigma V^{\frac{1}{2}}$*, where* Σ *is the copula covariance and* V *is the variance of the marginal models:* $V = diag(\frac{\mu_i}{\phi_i})$*.*

This is a multivariate extension of Proposition 3.2 of Heinen (2003). The variance is equal to the ratio of the mean to the dispersion parameter, the covariances are zero and therefore the variance-covariance matrix is diagonal. It can be seen that the variance-covariance of the counts is the product of a term reflecting the autoregressive part of the model, a term capturing the variance of the marginal models and a copula term responsible for the part of the contemporaneous cross-correlation which does not go through the time-varying mean.

Proof of Proposition 2.5.1. Upon substitution of the mean equation in the autoregressive intensity, one obtains:

$$\mu_t - \mu = A(N_{t-1} - \mu) + B(\mu_{t-1} - \mu) \tag{2.5.2}$$

$$\mu_t - \mu = A(N_{t-1} - \mu_{t-1}) + (A+B)(\mu_{t-1} - \mu) \tag{2.5.3}$$

Squaring and taking expectations gives:

$$V[\mu_t] = AE\left[(N_{t-1} - \mu_{t-1})(N_{t-1} - \mu_{t-1})^{'}\right]A + (A+B)V[\mu_{t-1}](A+B)^{'} \tag{2.5.4}$$

Using the law of iterated expectations and denoting $\Omega = V\left[N_t \mid \mathcal{F}_{t-1}\right]$, one gets:

$$V[\mu_t] = A\Omega A + (A+B)V[\mu_{t-1}](A+B)^{'} \tag{2.5.5}$$

Vectorialising and collecting terms, one gets:

$$vec(V[\mu_t]) = \left(I_{K^2} - (A+B) \otimes (A+B)^{'}\right)^{-1} \cdot (A \otimes A^{'}) \cdot vec(\Omega) \tag{2.5.6}$$

Now, applying the following property on conditional variance

$$V[y] = E_x \left[V_{y|x}(y|x)\right] + V_x \left[E_{y|x}(y|x)\right] \tag{2.5.7}$$

to the counts and vectorialising, one obtains:

$$vec(V\left[N_t\right]) = vec(\Omega) + vec(V[\mu_t]) \tag{2.5.8}$$

Again using the law of iterated expectations, substituting the conditional variance σ_t for its expression, then making use of the previous result, and after finally collecting terms, one gets the announced result.

$$vec(V[N_t]) = \left(I_{K^2} + \left(\left(I_{K^2} - (A+B) \otimes (A+B)^{'}\right)^{-1} \cdot (A \otimes A^{'}\right)\right) \cdot vec(\Omega) \tag{2.5.9}$$

Based on Song (2000), and on tail area approximations (Jorgensen (1997), we can approximate the Pearson residual as follows:

$$F(N_{i,t}, \mu_{i,t}, \phi) \simeq \Phi\left(\frac{N_{i,t} - \mu_{i,t}}{\sqrt{\frac{\mu_{i,t}}{\phi_i}}}\right) , \tag{2.5.10}$$

Equivalently, we have:

$$q_{i,t} \equiv \Phi^{-1}(F(N_{i,t}, \mu_{i,t}, \phi)) \simeq \frac{N_{i,t} - \mu_{i,t}}{\sqrt{\frac{\mu_{i,t}}{\phi_i}}} \equiv \epsilon_{i,t} , \tag{2.5.11}$$

Therefore we can approximate the variance-covariance of the Pearson residuals with the copula covariance:

$$\Sigma = Cov(q_t) \simeq Cov\left(\epsilon_{i,t}\right) \tag{2.5.12}$$

Now the average conditional variance-covariance matrix Ω can be obtained simply from Σ as:

$$\Omega \simeq V^{\frac{1}{2}} \Sigma V^{\frac{1}{2}} \tag{2.5.13}$$

□

Proposition 2.5.2 (Autocovariance of the MDACP(1,1) Model)**.** *The autocovariance of the MDACP(1,1) model, when the conditional mean is given by 2.2.13, is equal to:*

$$vec(Cov[N_t, N_{t-s}]) = \left[I \otimes A^{-1}\left((A+B)^s - B(A+B)\right)\right] \cdot vec\left(V[N_t] - \Omega\right) \tag{2.5.14}$$

where Ω *and* $V[N_t]$ *are as defined in Proposition 2.5.1.*

Proof of Proposition 2.5.2. As a consequence of the martingale property, deviations between the time t value of the dependent variable and the conditional mean are independent from the information set at time t. Therefore:

$$E[(N_t - \mu_t)(\mu_{t-s} - \mu)^{'}] = 0 \qquad \forall\, s \geq 0 \tag{2.5.15}$$

By distributing $N_t - \mu_t$, one gets:

$$Cov[N_t, \mu_{t-s}] = Cov[\mu_t, \mu_{t-s}] \qquad \forall\ s \geq 0 \tag{2.5.16}$$

By the same "non-anticipation" condition as used above, it must be true that:

$$E[(N_t - \mu_t)(N_{t-s} - \mu)^{'}] = 0 \qquad \forall\ s \geq 0 \tag{2.5.17}$$

Again, distributing $N_t - \mu_t$, one gets:

$$Cov[N_t, N_{t-s}] = Cov[\mu_t, N_{t-s}] \qquad \forall\ s \geq 0 \tag{2.5.18}$$

Now,

$$\begin{aligned} Cov[\mu_t, \mu_{t-s}] &= ACov[N_t, \mu_{t-s+1}] + BCov[\mu_t, \mu_{t-s}] \\ &= (A+B)Cov[\mu_t, \mu_{t-s}] \\ &= (A+B)^s V[\mu_t] \end{aligned} \tag{2.5.19}$$

The first line was obtained by replacing μ_t by its expression, the second line by making use of 2.5.16, the last line follows from iterating line two.

$$Cov[\mu_t, \mu_{t-s+1}] = ACov[\mu_t, N_{t-s}] + BCov[\mu_t, \mu_{t-s}] \tag{2.5.20}$$

Rearranging and making use of 2.5.18, one gets:

$$\begin{aligned} ACov[N_t, N_{t-s}] &= Cov[\mu_t, \mu_{t-s+1}] - BCov[\mu_t, \mu_{t-s}] \\ &= ((A+B)^s - B(A+B))\, V[\mu_t] \end{aligned} \tag{2.5.21}$$

Under the condition that A is invertible, which is not an innocuous assumption, as it excludes the pure factor model, we get after vectorialising:

$$vec(Cov[N_t, N_{t-s}]) = \left[I \otimes \left(A^{-1}(A+B)^s - A^{-1}B(A+B)\right)\right] vec(V[\mu_t]) \tag{2.5.22}$$

After substituting in 2.5.9, we get:

$$vec(Cov[N_t, N_{t-s}]) = \left[I \otimes A^{-1} \left((A+B)^s - B(A+B) \right) \right] \cdot \left(I_{K^2} + \left(\left(I_{K^2} - (A+B) \otimes (A+B)^{'} \right)^{-1} \cdot (A \otimes A^{'} \right) \right) \cdot vec(\Omega) \tag{2.5.23}$$

□

Chapter 3

Trading activity and liquidity in a pure limit order book: An empirical analysis

This chapter reports on research done jointly with Joachim Grammig and Andréas Heinen.

3.1 Introduction

The most important stock markets of continental Europe are organized as electronic open limit order book markets.[1] Unlike traditional stock markets, most prominently the New York Stock Exchange, no specialist is responsible for managing liquidity supply and demand. Whether or not a trader asking for immediate execution of an order has to incur a volume dependent price adjustment depends on the state of the open limit order book, which consists of previously submitted, non executed buy and sell orders. The arrival of new information induces traders to cancel, revise and (re)submit limit and market orders which implies that the open limit order book is permanently in flux. The resiliency of such a market design is crucial both for the operator of the trading venue and the agents participating in the trading process. Microstructure theory has put forth a variety of hypotheses about how information events affect liquidity supply and demand in open limit order book markets. The availability of detailed transaction data makes it possible to test these predictions, assess market resiliency, and draw conclusions for market design.

This paper uses data from the Xetra system, a pure limit order book market which operates at several exchanges in continental Europe, to test hypotheses and empirically assess predictions of microstructure models. We identify liquidity and informational factors describing the state of the limit order book and show how these factors, as well as volatility and liquidity demand,

[1] The largest of these markets is Euronext, the joint venture of the Amsterdam, Brussels and Paris stock exchanges, with a trading volume of 890 billions euro (in stocks) during the first two quarters of 2004, followed by the German stock exchange/Xetra (490 billions euro) and the Swiss SWX/Virt-X trading platform (170 billions euro). Trading volume at the London Stock Exchange during this period amounted to 660 billions euro.

affect future trading activity and market resiliency. For these purposes we use the dynamic model for multivariate time series of counts introduced by Heinen and Rengifo (2003) presented in Chapter 2.

It is important to note that this is not the first study that deals with those issues. Related work has focussed on whether a trader chooses a market or limit order, and how market conditions affect these choices (see, e.g., Biais, Hillion, and Spatt (1995), Griffiths, Turnbull, and White (2000), and Ranaldo (2003)). Sandas (2001) uses Swedish order book data and estimates a version of the celebrated Glosten (1994) limit order book model. Pascual and Veredas (2004) analyze the limit order book information of the Spanish Stock Exchange. Degryse, de Jong, Ravenswaaij, and Wuyts (2003) analyze the resiliency of a pure limit order market by investigating the order flow around aggressive orders using data from Paris Bourse. The present paper links and contributes to the literature in the following ways. As in Biais, Hillion, and Spatt (1995) we study in detail the trading process in an electronic limit order market. Following their approach we categorize limit orders according to their aggressiveness and study the interdependence of the order submission, execution and cancellation processes. Additionally, we distinguish less aggressive limit orders in terms of their relative position in the limit order book with respect to the best quotes. We show that this constitutes an improvement over the categories proposed in Biais, Hillion, and Spatt (1995) as the analysis of the disaggregated order categories provides new insights into the trading process. The detailed analysis is possible since we can exploit the information of a complete record of submission/cancellation/execution events (referred to as "market events") of different types of orders over a three month period. The market events we are particularly interested in are market order entries, limit and market order submissions and cancellations. Using these data and implementing the trading rules of the electronic market, we are able to reconstruct the prevailing order book at any point in time. No hidden orders were allowed during the sample period which implies that market participants and econometricians have an unobstructed (ex post) view of the entire order book.

The main empirical results can be summarized as follows. As predicted by theoretical models of financial market microstructure (Foucault (1999), Handa and Schwartz (1996)) we find that larger spreads reduce the relative importance of market order trading compared to limit order submissions. Consistent with Parlour's (1998) theoretical model, depth at the best quotes stimulates the submission of aggressive limit orders on the same side of the market, as limit order traders strive for price priority. On the other hand, larger depth on the opposite side of the market reduces the aggressiveness of own-side limit orders.

As Beltran-Lopez, Giot, and Grammig (2004), we show that two factors (extracted by a principal components analysis) explain a considerable fraction of the variation of market liquidity. Consistent with hypotheses derived from the theoretical analyzes in Foucault (1999) and Handa, Schwartz, and Tiwari (2003), and the empirical results of Biais, Hillion, and Spatt (1995) we find that the first two extracted principal components, identified as "market liquidity" and "informational" factor, respectively, can predict future trading activity. If the informational factor indicates a "bad news" state, aggressive limit and market sell order trading increases while buyer activity decreases. We also find that order aggressiveness is reduced and cancellation activity increases when price volatility is high. Evidence for market resiliency in this automated auction market is provided by the result that an increase in liquidity demand induces an increase in limit order submission activity. Furthermore, we show that cancellations do matter in the sense that they carry information for predicting future market activity and liquidity supply.

The methodological challenge when modelling financial transactions data is the irregular spacing of the multivariate time series data (see Hasbrouck (1999) for a useful discussion). The count data methodology employed in the present paper avoids the caveats of discrete choice models (see e.g. Ranaldo (2003)), in which time series aspects cannot adequately be taken into account, and the limitations of financial duration models for which it is difficult to formulate multivariate specifications (see e.g. Bauwens and Hautsch (2003), Engle and Lunde (2003) and Russell (1999)).

The remainder of the chapter is organized as follows. Section 3.2 describes the market structure. Section 3.3 presents the data and Section 3.4 discusses the empirical results. Section 3.5 concludes and provides an outlook for future research.

3.2 Market structure

We use data from the automated auction system Xetra. After its introduction at the Frankfurt Stock Exchange (FSE) in 1997, Xetra has become the main trading venue for German blue chip stocks. The Xetra system is also the trading platform of the Dublin and Vienna stock exchanges as well as the European Energy Exchange. The Xetra system operates as a pure electronic order book market. The computerized trading protocol keeps track of the entries, cancellations, revisions, executions and expirations of market and limit orders.

For blue chip stocks there are no dedicated market makers, like the specialists at the New York Stock Exchange (NYSE) or the Japanese saitori. For some small capitalized stocks listed in Xetra there may exist so-called designated sponsors - typically large banks - who are required to provide a minimum liquidity level by simultaneously submitting competing buy and sell limit orders.

Xetra/FSE does face some local competition for order flow. The FSE maintains a parallel floor trading system, which bears some similarities with the NYSE. Furthermore, like in the US, some regional exchanges participate in the hunt for liquidity. However, due to the success of the Xetra system, the FSE floor, previously the main trading venue for German blue chip stocks, as well as the regional exchanges became less important. The same holds true for the regional exchanges. Initially, Xetra trading hours at the FSE extended from 8.30 a.m to 5.00 p.m. central European time (CET). From September 20, 1999 the trading hours were shifted to 9.00 a.m to 5.30 p.m.. The trading day begins and ends with call auctions and is interrupted by another call auction which is conducted at 12.00 p.m.. The regular, continuous trading process is organized as a double auction mechanism with automatic matching of orders based on price and time priority.[2]

Five other Xetra features should be noted.

- Assets are denominated in euros, with a decimal system, which implies a small minimum tick size (1 euro-cent).
- Unlike at Paris Bourse, market orders exceeding the volume at the best quote are allowed to "walk up the book". At Paris Bourse the volume of a market order in excess of the depth at the best quote is converted into a limit order at that price entering the opposite side order book. However, in Xetra, market orders are guaranteed immediate full execution, at the cost of incurring a higher price impact on the trades.
- Dual capacity trading is allowed, i.e. traders can act on behalf of customers (agent) or as principal on behalf of the same institution (proprietary).
- Until March 2001 no block trading facility (like the upstairs market at the NYSE) was available.
- Before 2002, and during the time interval from which our data is taken, only round lot order sizes could be filled during continuous trading hours. A Xetra round lot was defined as a multiple of 100 shares. Execution of odd-lot parts of an order - this is an integer valued fraction of one hundred shares - was possible only during call auctions.

Besides these technical details, the trading design entails some features which render our sample of Xetra data (described in the next section) particularly appropriate for our empirical analysis. First, the Xetra system displays not only best quotes, but the contents of the whole limit order book. This is a considerable difference compared to other systems like the Paris Bourse's CAC system, where only the five best orders are displayed. Second, hidden limit orders (or iceberg orders) were not allowed until a recent change in the Xetra trading rules that permitted them.[3] As a result, the transparency of liquidity supply offered by the system was quite unprecedented. On the other hand, Xetra trading is completely anonymous, i.e. the Xetra order book does not reveal the identity of the traders submitting market or limit orders.[4]

3.3 Data

The dataset used for our study contains complete information about Xetra market events, that is all entries, cancellations, revisions, expirations, partial-fills and full-fills of market and limit orders that occurred between August 2, 1999 and October 29, 1999. Due to the considerable amount of data and processing time, we had to restrict the number of assets. Market events

[2] Bauwens and Giot (2001) provide a complete description of an order book market and Biais, Hillion, and Spatt (1999) describe the opening auction mechanism employed in an order book market and corresponding trading strategies.

[3] Biais, Hillion, and Spatt (1995) show that the possibility of hiding part of the volume of a limit order leads to all sorts of specific trading behavior, for example submitting orders to "test" the depth at the best quote for hidden volume.

[4] Further information about the organization of the Xetra trading process and a description of the trading rules that applied to our sample period is provided in Deutsche Börse AG (1999).

were extracted for three blue chip stocks, Daimler Chrysler (DCX), Deutsche Telekom (DTE) and SAP. At the end of the sample period their combined market capitalization represented 30.4 percent of the German blue chip index DAX 30. The three blue-chip stocks under study are traded at several important exchanges. Daimler-Chrysler shares are traded at the NYSE, the London Stock Exchange (LSE), the Swiss Stock Exchange, Euronext, the Tokyo Stock Exchange (TSE) and at most German regional exchanges. SAP is traded at the NYSE and at the Swiss Stock Exchange. Deutsche Telekom is traded at the NYSE and at the TSE. The stocks are also traded on the FSE floor trading system, but this accounts for less than 5% of daily trading volume in those shares. Trading volume at the NYSE accounts for about 20% of daily trading volume in those stocks. As the prices for the three stocks remained above 30 euros during the sample period, the tick size of 0.01 euros is less than 0.05% of the stock price. Hence, we should not observe any impact of the minimum tick size on prices or liquidity. Starting from the initial state of the order book, we track each change in the order book implied by entry, partial or full fill, cancellation and expiration of market and limit orders and perform a real time reconstruction of the order books. For this purpose we implement the rules of the Xetra trading protocol outlined in Deutsche Börse AG (1999) in the reconstruction program. From the resulting real-time sequences of order books, snapshots were taken at one minute intervals during continuous trading hours.

Following Biais, Hillion, and Spatt (1995) we classify market and limit orders in terms of aggressiveness:

- Category 1: Large market orders, orders that walk up or down the book (BMO-agg and SMO-agg).

- Category 2: Market orders, orders that consume all the volume available at the best quote (BMO-inter and SMO-inter).

- Category 3: Small market orders, orders that consume part of the depth at the best quote (BMO-small and SMO-small).

- Category 4: Aggressive limit orders, orders submitted inside the best quotes (BLO-inside and SLO-inside).

- Category 5: Limit orders submitted at the best quote (BMO-at and SMO-at).

- Category 6: Limit orders submitted outside the best quotes, orders that are below (above) the bid (ask). (BMO-outside and SMO-outside).

- Category 7: Cancellations. (BCANC and SCANC)

Moreover, we break up categories 6 and 7 according to their relative position in terms of the number of quotes away from the best quote:

- Limit Orders submitted within the first two quotes away from the best quotes (BLO-outside-1-2 and SLO-outside-1-2).

- Limit Orders submitted within the third and fifth quotes away from the best quotes (BLO-outside-3-5 and SLO-outside-3-5).
- Limit Orders submitted outside the best quotes beyond the fifth quote from the inside market (BLO-outside-5+ and SLO-outside-5+).
- Cancellations of standing limit orders at, or one or two quotes away from the best quotes (BCANC-0-2 and SCANC-0-2).
- Cancellations of standing limit orders between the third and the fifth quotes away from the best quotes (BCANC-3-5 and SCANC-3-5).
- Cancellations of standing limit orders beyond the fifth quote away from the best quotes (BCANC-5+ and SCANC-5+).

For our empirical analysis we then count the submission/cancellation events in the different categories during each one minute interval of the sample. The resulting multivariate sequence of counts provides the input for the econometric model used in the next section. To avoid dealing with the change in trading times, and given the large number of observations, we restrict the whole sample to observations between August 20 to September 20, 1999. The data therefore contain information about 21 trading days with 510 one-minute intervals per day giving a total of 10730 one minute intervals. Due to space limitations we only report the results for Daimler-Chrysler (DCX).[5] Sample statistics are presented in Table 3.1 where the main characteristics of the data can be read. The large number of marketable limit orders (MLO) compared to "true" market orders is remarkable. A MLO is a limit order which is submitted at a price which makes it immediately executable. In this respect it is indistinguishable from a "true" market order. However, MLOs differ from market orders in that the submitter specifies a limit of how much the order can walk up the book. Hence, a MLO might be immediately, but not necessarily completely filled. The non-executed volume of the MLO then enters the book.[6] In our empirical analysis we therefore treat the either completely or partially filled parts of a MLO just like a market order. When, for the sake of brevity, we refer in the following to "market orders" what we precisely mean is "true market orders and completely/partially filled marketable limit orders". The number of buy (sell) limit orders is 3.35 (4.7) times larger than the number of market orders. As one can see from Table 3.1, the sample means of the counts series are very small and all series are overdispersed (the sample variance is greater than the sample mean). This has implications for the appropriate statistical specification.

Table 3.2 presents the descriptive statistics for Daimler-Chrysler (DCX) in which the limit orders submitted outside the best quotes have been further disaggregated according to their relative position to the inside market, as well as descriptive statistics on cancellations, also categorized relative to the best quotes.

Figure 3.1 depicts the intraday seasonality in the series of market event counts. Neither buy nor sell market order counts reflect the often reported U-shape of intra-day financial series.

[5] The results obtained with the other two assets confirm the findings. These results are available upon request.
[6] MLOs therefore share some properties with Paris Bourse market orders.

Table 3.1: **Descriptive statistics for market event one-minute counts**

	Obs.	Mean	Std. Dev.	Disp.	Max.	Q(60)
BUY ORDERS	**52712**	**4.91**	**4.37**	**3.89**	**68**	**37817**
Category 1 (BMO-agg)	**3494**	**0.33**	**0.71**	**1.53**	**22**	**7888**
- True Market Orders	898	0.08	0.36	1.51	18	872
- Marketable Limit Orders	2596	0.24	0.56	1.28	6	7397
Category 2 (BMO-inter)	**3369**	**0.31**	**0.64**	**1.32**	**6**	**1629**
- True Market Orders	18	0.01	0.04	1.00	1	64
- Marketable Limit Orders	3351	0.31	0.64	1.33	6	1627
Category 3 (BMO-small)	**5250**	**0.49**	**0.81**	**1.33**	**7**	**11106**
- True Market Orders	2564	0.24	0.54	1.22	6	8990
- Marketable Limit Orders	2686	0.25	0.55	1.23	5	1344
Buy Market Orders (BMO)	**12113**	**1.13**	**1.46**	**1.89**	**29**	**22759**
Category 4 (BLO-inside)	18312	1.71	1.85	2.00	17	21309
Category 5 (BLO-at)	11411	1.06	1.33	1.68	18	14313
Category 6 (BLO-outside)	10876	1.01	1.28	1.62	11	8657
Buy Limit Orders (BLO)	**40599**	**3.78**	**3.35**	**2.96**	**39**	**33304**
Cancellations (BCANC)	**20534**	**1.91**	**2.03**	**2.15**	**18**	**13623**
SELL ORDERS	**43163**	**4.02**	**3.92**	**3.82**	**38**	**20498**
Category 1 (SMO-agg)	**2263**	**0.21**	**0.53**	**1.36**	**6**	**1442**
- True Market Orders	524	0.05	0.23	1.12	3	472
- Marketable Limit Orders	1739	0.16	0.45	1.25	5	1125
Category 2 (SMO-inter)	**3077**	**0.29**	**0.63**	**1.38**	**8**	**2602**
- True Market Orders	94	0.01	0.11	1.33	5	305
- Marketable Limit Orders	2983	0.28	0.62	1.36	7	2551
Category 3 (SMO-small)	**2241**	**0.21**	**0.52**	**1.32**	**10**	**833**
- True Market Orders	892	0.08	0.31	1.14	5	362
- Marketable Limit Orders	1349	0.13	0.40	1.24	8	426
Sell Market Orders (SMO)	**7581**	**0.71**	**1.15**	**1.86**	**15**	**5331**
Category 4 (SLO-inside)	15012	1.34	1.68	2.00	13	11184
Category 5 (SLO-at)	10166	0.95	1.30	1.78	23	8660
Category 6 (SLO-outside)	10404	0.97	1.25	1.62	11	6738
Sell Limit Orders (SLO)	**35582**	**3.32**	**3.14**	**2.97**	**38**	**21272**
Cancellations (SCANC)	**20010**	**1.86**	**2.09**	**2.34**	**29**	**11379**

This table presents the descriptive statistics of the one-minute Daimler-Chrysler count series of market events. Category 1 orders are market orders that walk up the book. Category 2 orders are market orders which consume all (but not more than) the volume available at the best quote. Category 3 orders are market orders that consume part of the depth at the best quote. Category 4 orders are aggressive limit orders, i.e. orders submitted inside the best quotes. Category 5 orders are limit orders submitted at the best quote. Category 6 orders are limit orders outside the best quotes, i.e. below (above) the bid (ask). Q(60) reports the Ljung-Box Q-statistic computed with 60 lagged autocorrelations. The Disp. column reports the ratio of sample variance to sample mean.

Table 3.2: **Descriptive statistics for disaggregated market event one-minute counts (least aggressive limit orders and cancellations)**

	Obs	Mean	Std. Dev.	Disp.	Max.	Q(60)
Category 6 buy orders (BLO-outside)	**10876**	**1.01**	**1.28**	**1.62**	**11**	**8657**
- BLO-outside-1-2	4322	0.40	0.75	1.41	7	2513
- BLO-outside-3-5	3702	0.35	0.66	1.28	7	3375
- BLO-outside-5+	2852	0.27	0.61	1.40	9	1929
Buy cancellations: (BCANC)	**20534**	**1.70**	**1.86**	**2.04**	**16**	**12715**
- BCANC-0-2	8518	0.79	1.12	1.58	9	4748.5
- BCANC-3-5	6306	0.59	0.90	1.39	8	4350
- BCANC-5+	5710	0.53	0.96	1.72	11	4948
Category 6 sell orders (SLO-outside)	**10404**	**0.97**	**1.25**	**1.62**	**11**	**6738**
- SLO-outside-1-2	4286	0.40	0.77	1.49	8	2642
- SLO-outside-3-5	3479	0.32	0.62	1.19	5	2023
- SLO-outside-5+	2639	0.25	0.58	1.36	7	1723
Sell cancellations: (SCANC)	**20010**	**1.66**	**1.94**	**2.27**	**29**	**9799**
- SCANC-0-2	8139	0.76	1.09	1.58	9	4714.6
- SCANC-3-5	6219	0.58	0.89	1.36	9	3689
- SCANC-5+	5652	0.53	0.11	2.33	26	4227

This table presents the descriptive statistics of the one-minute Daimler-Chrysler count series of market events. Buy and sell orders of category 6 (limit orders submitted outside the best quotes) have been disaggregated according to their relative position to the best quotes. BLO-outside-1-2 and SLO-outside-1-2 count the number of buy and sell orders submitted one or two quotes away from the best quotes. The categories BLO-outside-3-5, SLO-outside-3-5, BLO-outside-5+ and SLO-outside-5+ are defined accordingly. A disaggregation of the buy and sell cancellations (BCANC and SCANC) is conducted accordingly: BCANC-0-2 and SCANC-0-2 denote cancellations of standing limit orders at, or one or two quotes away from the best quotes. The categories BCANC-3-5, SCANC-3-5, BCANC-5+ and SCANC-5+ are defined accordingly. Q(60) reports the Ljung-Box Q-statistic computed with 60 autocorrelations

There is a small increase in the number of counts at about 2.30 p.m. CET which most likely corresponds to the NYSE opening. The number of buy limit orders is large early in the morning, but decays quite fast. Limit orders at both sides of the book behave similarly in that we observe an increase in trading activity in the afternoon at the same time as the market order activity increases. We observe a similar diurnal pattern in the cancellation series.

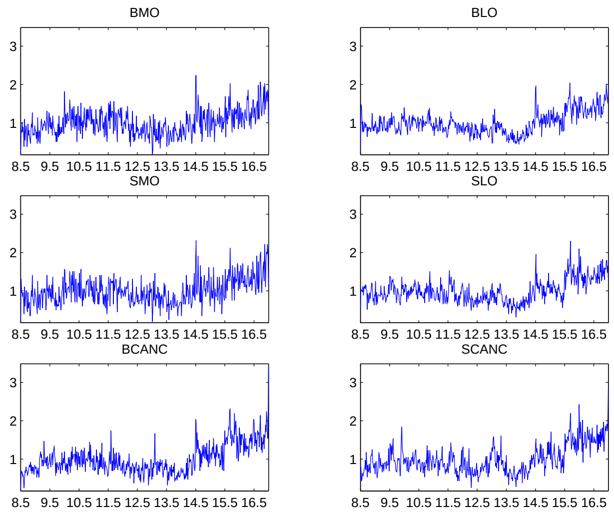


Figure 3.1: **Seasonality in market event count series**

The figure depicts the daily seasonality of the aggregated market event counts for Daimler-Chrysler. Xetra trading hours at the FSE extended from 8.30 a.m to 5.00 p.m. CET. BMO denote buy market orders, SMO market orders, BLO buy limit orders, SLO sell limit orders, BCANC buy cancellations and SCANC sell cancellations.

Figure 3.2 presents two-day auto- and cross-correlograms of the aggregated count series for Daimler-Chrysler (DCX). We consider buy market orders (BMO), sell market orders (SMO), buy limit orders (BLO), sell limit orders (SLO), buy cancellations (BCANC) and sell cancellations (SCANC). Observing the autocorrelations one can see that all series of counts show persistence. Moreover, a diurnal seasonality of the U-type can be observed in some of the auto- or cross-correlograms. A visual inspection of the cross correlations between market buys and market sells reveals that they are almost symmetric. This implies that the tendency of market buys at time t to follow market sells at time $t-k$ is almost the same as the tendency of market sells to follow market buys. This indicates that the informational effects, found by Hasbrouck (1999), are not detectable in our data.

3.4 Empirical results

In order to model the dynamics of the multivariate series of counts of order submissions and cancellations within one minute intervals, we adopt the multivariate autoregressive conditional double Poisson (MDACP) modelling framework introduced by Heinen and Rengifo (2003) and presented in Chapter 2. A detailed exposition of this model can be found in that chapter.

Besides the VARMA dynamics in equation (2.2.13) we allow predetermined variables observed at $t-1$, and collected in a vector X_{t-1}, to impact on the conditional mean $E(N_{i,t} \mid \mathcal{F}_{t-1})$ of the one-minute submission/cancelation count. The predetermined variables are derived from

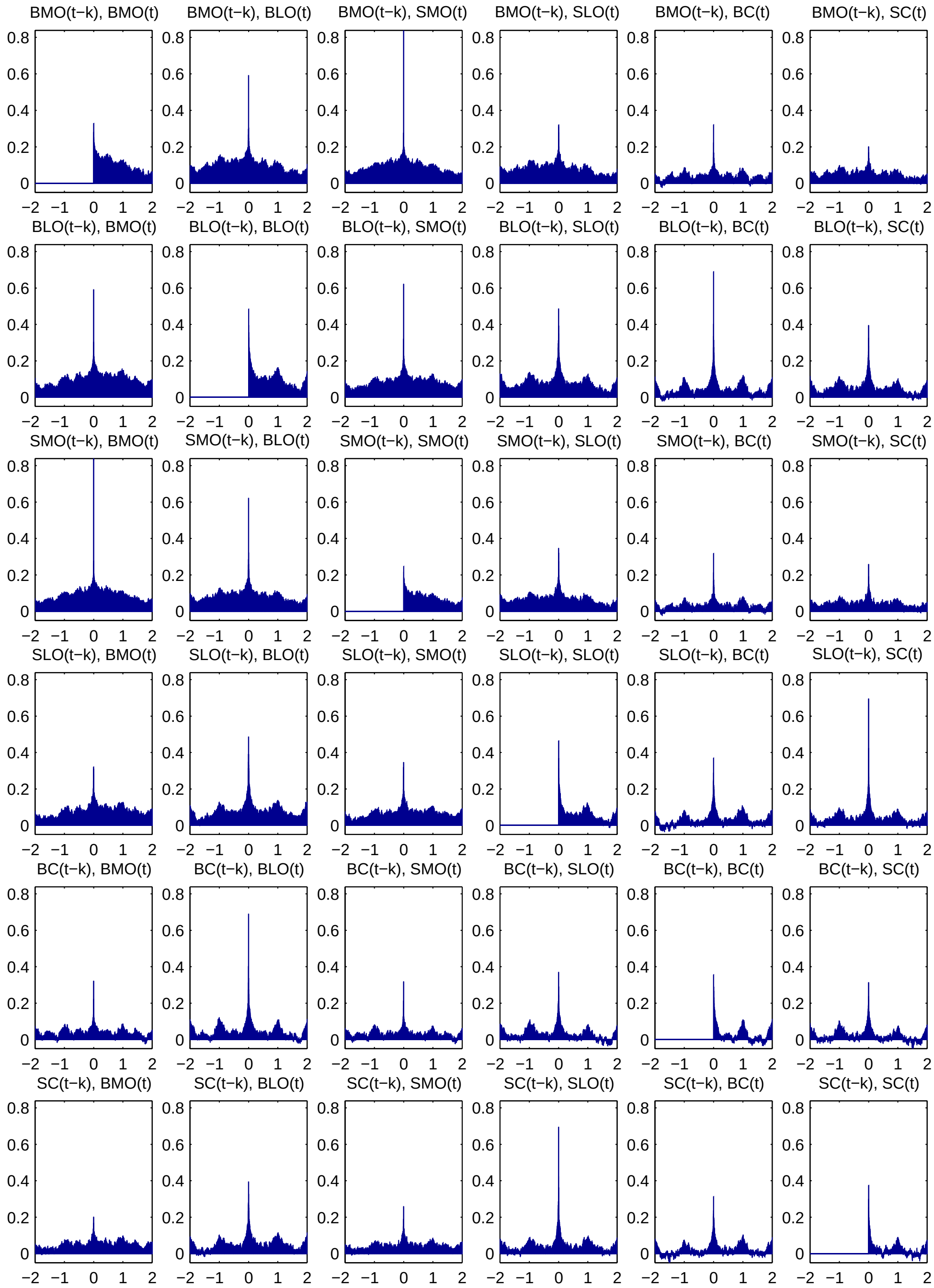


Figure 3.2: **Cross-correlation of market events**

The figure depicts two days auto- and cross-correlograms of the aggregated market event counts for Daimler-Chrysler. BMO denotes buy market orders, SMO sell market orders, BLO buy limit orders, SLO sell limit orders, BC buy cancellations, and SC sell cancellations.

models of market microstructure and include liquidity and informational indicators that can be extracted from the order book information and transaction data (e.g. inside spread, depth and volatility). Furthermore, to account for intra-day seasonality (or "diurnality") of the count sequences, we include a trigonometric spline function in the conditional mean equation. This method has been advocated and successfully applied by Andersen and Bollerslev (1997) to account for diurnality in volatility models. Including both predetermined variables and the seasonality model, the conditional distribution of $N_{i,t}$ in equation (2.2.11) becomes

$$N_{i,t} \mid \mathcal{F}_{t-1} \sim DP(\mu^*_{t,i}\ , \phi_i)\ ,\ \forall i = 1, \ldots, K. \tag{3.4.1}$$

where

$$\mu^*_{t,i} = \mu_{t,i} \exp\left(X'_{t-1}\gamma_i + \sum_{p=1,2}\left(\psi_{c,p} \cos \frac{2\pi p\, Re[t,N]}{N} + \psi_{s,p} \sin \frac{2\pi p\, Re[t,N]}{N}\right)\right) \tag{3.4.2}$$

The first term in the exponent accounts for the effect of the predetermined variables X_{t-1} on the conditional mean, where γ_i is a parameter vector. The second term is the trigonometric spline function, where $Re[t,N]$ is the remainder of the integer division of t by N, the number of one-minute periods in a trading session. The coefficients $\psi_{c,p}$ and $\psi_{s,q}$ are parameters to be estimated.

3.4.1 Parameter estimates and specification tests

Estimation and test results are reported in Tables 3.4, 3.5, 3.6, 3.8, 3.9 and 3.10.

- Table 3.4 contains the results for an MDACP model with six endogenous count variables: buy market orders (BMO), buy limit orders (BLO), sell market orders (SMO), sell limit orders (SLO), buy order cancellations (BCANC) and sell order cancellations (SCANC). This specification (henceforth referred to as the aggregated model) is useful to test several predictions of theoretical microstructure.
- Tables 3.5 (bid side) and 3.6 (ask side) report the estimation results for a disaggregated MDACP system, where order counts are classified, according to aggressiveness, into the six categories described in Section 3.3.
- Table 3.8 presents the results of a bivariate MDACP model for buy and sell market orders in which lagged cancellation counts enter as predetermined variables.
- Table 3.9 reports the results of an MDACP model which focuses on the counts of the three limit order categories (LO-inside, LO-at, and LO-outside) and that also uses lagged cancellation counts as predetermined variables.
- To obtain the results reported in Table 3.10 we estimated an MDACP model which is based on a finer categorization of limit orders outside the best bid as described in Section 3.3.

In all tables we report the estimates of the autoregressive parameters (β), the parameters of the lagged counts (α), the parameters which determine the impact of the predetermined variables on the expected number of counts (γ), and the dispersion parameters (ϕ). Significant (at 5 % and 10 %) parameter estimates are printed in boldface and with a star, respectively. The last rows of the estimation result tables report the empirical variance of the Pearson residuals. Because of space limitations we do not present the estimates of the seasonality parameters. Instead, we report the p-value of the Wald statistic ($W(\psi's = 0)$) for a test of the joint significance of the seasonality parameters. Under the null hypothesis the test statistic is distributed as chi-squared with four degrees of freedom. Except for two cases, the Wald statistic is highly significant, underlining the necessity to account for diurnality in the count sequences. We have outlined in Chapter 2, Section 2.2 that a correctly specified model implies that the Pearson residuals have unitary variance and exhibit no significant autocorrelation. Inspecting the estimated variances of the Pearson residuals in the result tables and the sample autocorrelogram of the Pearson residuals (aggregated system) in Figure 3.3 we find no evidence of specification problems.[7] Following the suggestions of Diebold, Gunther, and Tay (1998) we also employed graphical tools to check for uniformity and serial dependence in the probability integral transform (PIT) sequences. The visual inspections did not point to specification problems, as the Q-Q plots almost coincide with the 45-degree line and the empirical autocorrelograms of the PIT sequences do not indicate serial correlation.

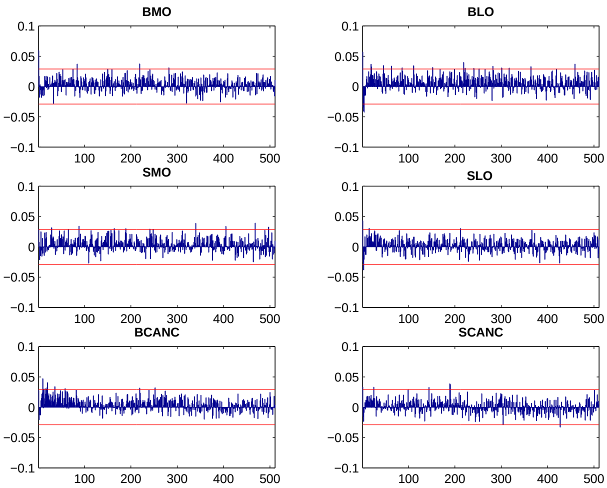


Figure 3.3: **Autocorrelogram of the Pearson residuals - Aggregated MDACP system**

The figures depict one-day (510 one-minute intervals) autocorrelograms of the Pearson residuals of the aggregated MDACP system (estimation results in table 3.4). BMO denote buy market orders, SMO sell market orders, BLO buy limit orders, SLO sell limit orders, BCANC buy cancellations, and SCANC sell cancellations.

The estimation results indicate a clear rejection of the Poisson assumption as all estimated dispersion coefficients are significantly different from one. The distributions are either over- or underdispersed, supporting the use of the double Poisson distribution.

[7] To save space and since the results are qualitatively identical we do not present the autocorrelograms of all models.

Table 3.3: **Contemporaneous dependence of market events for the aggregated MDACP system**

	BMO	BLO	SMO	SLO	BCANC	SCANC
BMO	1.000					
BLO	0.100	1.000				
SMO	-0.025	0.187	1.000			
SLO	0.171	0.195	0.159	1.000		
BCANC	0.177	0.574	0.103	0.193	1.000	
SCANC	0.119	0.202	0.196	0.574	0.191	1.000

For the MDACP approach we use a Gaussian copula to account for contemporaneous dependence in the market event count sequences. This implies that the degree of contemporaneous dependence can be measured by computing the correlation matrix of the quantile vector $q_t = (\Phi^{-1}(z_{1,t}), \ldots, \Phi^{-1}(z_{K,t}))'$, where Φ^{-1} denotes the quantile function of the standard normal distribution, K the number of count series, and $z_{i,t}$ the sequence of probability integral transforms of the ith continuous extension of the count data series (see 2.2 for details). The table reports this estimated correlation matrix for an aggregated MDACP system which uses buy and sell market orders (BMO and SMO), limit orders (BLO and SLO) as well as cancellation counts (BCANC and SCANC) as dependent variables.

Table 3.3 reports the estimated contemporaneous correlation matrix of the quantile vector q_t implied by the aggregated MDACP system. The appendix on Section 2.5 shows that this correlation measures the part of the contemporaneous and lagged cross-correlation which does not go through the time-varying mean. With a single exception, all correlations are positive and especially the own-side correlations of limit order submissions and cancellations are considerable. This indicates that an increase in trading activity generally involves all types of market events, but that the same side dependence is stronger. The market sell and buy order events are negatively correlated (although, slightly different from zero). This last result shows the capability of the normal copula to capture negative dependence among variables.

3.4.2 Liquidity supply, volatility and order submission activity

Inside Spread and depth, and trading activity

Theoretical models put forth by Handa and Schwartz (1996) and Foucault (1999) hypothesize that large spreads reduce the proportion of market orders relative to limit orders in the total order flow. The explanation is that a larger spread implies a higher price of immediacy. This makes market orders less attractive than limit orders which receive a higher premium for providing liquidity. Griffiths, Turnbull, and White (2000) and Ranaldo (2003) have provided empirical evidence for these predictions. The estimation results for the aggregated MDACP system (Table 3.4) indicate that an increase of the inside spread exerts a negative effect on all six order categories and cancellations, thus inducing a general slowdown in trading activity. Except for the effect of the inside spread on buy limit orders, all the estimates are significant at 5% or at 10%. Moreover, in line with theory, the impact on market orders is considerably stronger than the effect on limit orders. The estimation results for the disaggregated system (Tables 3.5 and 3.6) lead to the same conclusions. The empirical analysis thus confirms the theoretical prediction that the proportion of market orders relative to limit orders decreases when large spreads prevail.

In the models proposed by Parlour (1998) and Handa, Schwartz, and Tiwari (2003) the volume (depth) at the best quotes is related to the execution probability of limit orders at the respective side of the book, which in turn affects trading activity. More precisely, it is predicted that when the execution probability of a limit order is low, traders on the respective side of the market act more aggressively when striving for price-time priority. A large volume at the best quote (at the bid side, say) will induce bid-side traders to act aggressively by submitting more market orders or limit orders inside the best quotes. On the other hand, when the depth at the **opposite** side of the market is large, **own side** order aggressiveness is expected to decrease. This is a mechanical consequence of the previous result. Coming back to the example, large volume at the bid-side, which induces bid-side traders to submit more aggressive buy limit orders, increases the probability of execution of limit orders at the ask side relative to market orders, thereby decreasing aggressiveness on the opposite side. The empirical evidence for these hypotheses obtained from the estimation of the aggregated MDACP system is mixed.[8] Table 3.4 shows that volume at the best quotes (denoted BIDVOL and ASKVOL) exerts a positive effect on all components of the order flow. The effect of BIDVOL is not significant for its own side[9] and significant for the opposite side. The variable ASKVOL has a significant impact in its own side and on the buy market orders (BMO) but its impact is not significant for the buy limit orders (BLO). Thus, larger volume at the best quotes does not only have a positive effect on own side trading activity, but also on the opposite side. While the own side effect is in line with the theoretical predictions outlined above, the opposite side effect is clearly not. The estimation results of the disaggregated MDACP system presented in Tables 3.5 and 3.6 are more in accordance with the theoretical predictions. As hypothesized, the empirical results confirm that traders on the respective side of the market act more aggressively when the volume at the best quote is large. For example, when depth at the bid is large, traders are expected to submit more buy limit orders inside the best quotes (1.46) and less buy limit orders at the best quotes (-1.82). As predicted, volume at the bid exerts a positive effect on the expected number of buy market orders of the most aggressive categories (BMO-agg and BMO-inter). The ask side results are quite similar. The opposite side effects are now also in accordance with the theoretical predictions. For example in Table 3.6, an increase of the volume at the best bid decreases the expected number of most aggressive sell market orders (-3.37 and -1.86 for SMO-agg and SMO-inter, respectively). In other words, own-side order aggressiveness tends to decrease when opposite-side depth at the best quote increases, as hypothesized.

Beyond the inside market: Liquidity and informational factors, and trading activity

Beltran-Lopez, Giot, and Grammig (2004) propose to employ principal components analysis

[8]The volume is measured in unit of shares, this explains the small values of the parameter estimates.
[9]Significant at 13%.

Table 3.4: **Estimation results for an aggregated MDACP system**

	BMO	BLO	SMO	SLO	BCANC	SCANC
ω	**0.032**	**0.270**	**0.052**	**0.348**	**0.115**	**0.159**
α_{BMO}	**0.086**	**0.051**	0.002	**0.168**	-0.019*	-0.017
α_{BLO}	**0.023**	**0.178**	0.004	**0.041**	**0.102**	0.012*
α_{SMO}	0.011	**0.266**	**0.087**	**0.048**	-0.001	0.015
α_{SLO}	-0.006*	-0.008	**0.034**	**0.180**	**-0.015**	**0.121**
α_{BCANC}	**-0.016**	0.023*	-0.001	**0.027**	**0.104**	0.009
α_{SCANC}	0.004	**0.064**	**-0.011**	**0.054**	**0.044**	**0.086**
β	**0.848**	**0.647**	**0.709**	**0.541**	**0.660**	**0.600**
γ_{SPREAD}	**-1.019**	-0.134	**-1.736**	**-0.341**	-0.300*	-0.391*
$\gamma_{\text{BIDVOL}} * 10^{-5}$	0.632	0.396	**1.015**	**0.506**	0.065*	0.657*
$\gamma_{\text{ASKVOL}} * 10^{-5}$	**0.726**	0.295	**1.650**	**0.601**	0.410	0.412
γ_{VOLAT}	**-0.643**	0.160	-0.492	**0.380**	**1.475**	**0.573**
ϕ	**0.713**	**0.525**	**0.752**	**0.511**	**0.614**	**0.604**
$W(\psi's = 0)$	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Var(ϵ_t)	1.02	1.00	1.05	1.01	1.00	1.03
Log likelihood	-14661.9	-23535.2	-12049.8	-22717.2	-18410.6	-18254.2

The table reports Maximum Likelihood estimates of a MDACP system. The dependent variables are the one-minute counts of buy market orders (BMO), buy limit orders (BLO), sell market orders (SMO), sell limit orders (SLO), buy cancellations (BCANC) and sell cancellations (SCANC). The mean equations are specified as

$$\mu^*_{t,i} = \mu_{t,i} \exp \left(X_{t-1}\gamma_i + \textstyle\sum_{p=1,2} \psi_{c,p} \cos \frac{2\pi p \, Re[t,N]}{N} + \psi_{s,p} \sin \frac{2\pi p \, Re[t,N]}{N} \right),$$

where $\mu_{t,i} = \omega_i + \sum_{j=1}^{6} \alpha_{i,j} N_{t-1,j} + \beta_i \mu_{t-1,i}$, for $t = 1, \ldots, 10731$. $Re[t, N]$ denotes the remainder of the integer division of t by N, the number of periods in a trading session. X_{t-1} collects the vector of predetermined variables, the inside spread (SPREAD), the volume at the best bid (BIDVOL), the volume at the best ask (ASKVOL) and volatility measured by the standard deviation of the last 5 minutes midquote returns (VOLAT). ϕ is the dispersion parameter of the Double Poisson. Parameters significant at the 5% and 10% level are printed boldface and accompanied with a star, respectively. For ϕ the null hypothesis is that the parameter is one, for β and the α parameters the null is that the true parameter is zero. The $W(\psi's = 0)$ row reports the p-values for a Wald-test of the hypothesis that the seasonality parameters $\psi_{s,1}$, $\psi_{s,2}$, $\psi_{c,1}$ and $\psi_{c,2}$ are jointly zero. $Var(\varepsilon_t)$ is the variance of the Pearson residual which should be close to one for a correctly specified model.

Table 3.5: **Estimation results for a disaggregated MDACP system - Bid side**

	BMO-agg	BMO-inter	BMO-small	BLO-inside	BLO-at	BLO-outside
ω	**0.008**	**0.050**	**0.007**	**0.150**	**0.045**	**0.060**
$\alpha_{\text{BMO-agg}}$	**0.056**	0.009	**0.011**	**0.062**	0.019	-0.004
$\alpha_{\text{BMO-inter}}$	0.008*	**0.047**	-3.13E-4	-0.011	0.001	-0.003
$\alpha_{\text{BMO-small}}$	**0.022**	0.011*	**0.035**	**0.066**	0.011	0.001
$\alpha_{\text{BLO-inside}}$	**0.009**	**0.038**	0.003*	**0.110**	**0.018**	0.011
$\alpha_{\text{BLO-at}}$	**0.011**	**0.011**	**0.006**	**0.052**	**0.091**	**0.031**
$\alpha_{\text{BLO-outside}}$	-0.001	0.001	-0.001	**0.024**	**0.058**	**0.138**
$\alpha_{\text{SMO-agg}}$	0.001	-0.014	**-0.011**	**0.092**	**0.158**	**0.110**
$\alpha_{\text{SMO-inter}}$	**0.017**	0.002	-0.004	**0.104**	**0.070**	**0.051**
$\alpha_{\text{SMO-small}}$	0.001	0.015	-0.003	0.020	**0.034**	0.010
$\alpha_{\text{SLO-inside}}$	-0.003	0.004	0.004*	**0.027**	-0.001	0.011
$\alpha_{\text{SLO-at}}$	-0.001	-0.003	0.001	-0.015*	**0.017**	0.012*
$\alpha_{\text{SLO-outside}}$	-0.003	0.007*	-0.002	-0.003	-0.001	**0.015**
β	**0.807**	**0.579**	**0.921**	**0.656**	**0.721**	**0.689**
γ_{BFACT1}	0.000	**0.010**	0.004	**0.005**	**0.010**	**-0.009**
γ_{BFACT3}	**0.039**	0.011	0.001	0.018*	0.013	-0.014
γ_{SFACT1}	**0.018**	0.004	**0.010**	**0.005**	**0.005**	-0.003
γ_{SFACT3}	**-0.008**	-0.004	-0.006	-0.012	0.003	0.005
$\gamma_{\text{DIFFSLOPE}}$	0.009	**-0.023**	0.002	**-0.016**	-0.008	**0.014**
γ_{SPREAD}	0.150	**-2.549**	**-0.809**	0.086	-0.500*	-0.404*
$\gamma_{\text{BIDVOL}} * 10^{-5}$	0.980*	**2.200**	**-1.700**	**1.460**	**-1.820**	0.270
$\gamma_{\text{ASKVOL}} * 10^{-5}$	-0.333	**-3.210**	**2.870**	0.265	0.266	-0.024
γ_{VOLAT}	**-0.622**	-0.591	-0.263	**-0.153**	**0.174**	0.257*
ϕ	**1.177**	**1.146**	**1.004**	**0.652**	**0.790**	**0.762**
$W(\psi's = 0)$	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Var(ϵ_t)	1.06	1.04	1.02	0.98	1.00	1.01
Log likelihood	-7318.7	-7425.4	-9486.6	-17499.5	-14001.5	-13891.0

The table reports Maximum Likelihood estimates of a MDACP system. The dependent variables are the one-minute counts of category 1 orders (BMO-agg/SMO-agg, market orders that walk up or down the book), category 2 orders (BMO-inter/SMO-inter, market orders which consume all volume available at the best quote), category 3 orders (BMO-small/SMO-small, market orders which consume part of the depth at the best quote), category 4 orders (BLO-inside/SLO-inside, limit orders submitted inside the best quotes), category 5 orders (BLO-at/SLO-at, limit orders submitted at the best quote), category 6 orders (BLO-outside/SMO-outside, limit orders outside the best quotes) and cancellations (BCANC/SCANC). The table reports the results of the bid side equations. The mean equations are specified as equation (3.4.2).
The predetermined variables are the inside spread (SPREAD), the volume at the best bid (BIDVOL), the volume at the best ask (ASKVOL) and volatility measured by the standard deviation of the last 5 minutes midquote returns (VOLAT). BFACT1 (SFACT1) denotes the first factor extracted by PCA at the bid (ask) side, DIFFSLOPE is the difference of the absolute values of the second factors (informational factor). BFACT3 and SFACT3 denote the third factor for the bid (ask) side. ϕ is the dispersion parameter of the Double Poisson. Parameters significant at the 5% and 10% level are printed boldface and accompanied with a star, respectively. For ϕ the null hypothesis is that the parameter is one, for β and the α parameters the null is that the true parameter is zero.The $W(\psi's = 0)$ row reports the p-values for a Wald-test of the hypothesis that the seasonality parameters $\psi_{s,1}$, $\psi_{s,2}$, $\psi_{c,1}$ and $\psi_{c,2}$ are jointly zero. $Var(\varepsilon_t)$ is the variance of the Pearson residual which should be close to one for a correctly specified model

Table 3.6: **Estimation results for a disaggregated MDACP system - Ask side**

	SMO-agg	SMO-inter	SMO-small	SLO-inside	SLO-at	SLO-outside
ω	**0.017**	**0.027**	**0.024**	**0.151**	**0.057**	**0.121**
$\alpha_{\text{BMO-agg}}$	**-0.014**	-0.008	-0.002	0.010	**0.069**	**0.046**
$\alpha_{\text{BMO-inter}}$	**-0.010**	**-0.016**	**-0.011**	**0.051**	**0.082**	**0.067**
$\alpha_{\text{BMO-small}}$	**0.009**	0.008*	0.004	0.024*	**0.020**	0.017
$\alpha_{\text{BLO-inside}}$	0.003	**0.011**	0.001	**0.049**	**0.016**	**0.026**
$\alpha_{\text{BLO-at}}$	**0.005**	**0.009**	**-0.004**	0.009	0.011*	**0.022**
$\alpha_{\text{BLO-outside}}$	-0.002	-0.004*	0.003*	-0.001	0.002	**0.022**
$\alpha_{\text{SMO-agg}}$	**0.071**	0.004	**0.016**	**0.094**	-0.005	**0.060**
$\alpha_{\text{SMO-inter}}$	**0.022**	**0.035**	0.008*	0.014	-0.003	-0.007
$\alpha_{\text{SMO-small}}$	0.009*	-0.002	**0.032**	-0.005	0.022	0.018
$\alpha_{\text{SLO-inside}}$	**0.013**	**0.036**	**0.012**	**0.132**	0.015*	**0.025**
$\alpha_{\text{SLO-at}}$	**0.009**	**0.007**	**0.007**	**0.053**	**0.103**	**0.052**
$\alpha_{\text{SLO-outside}}$	0.002	**0.006**	**0.006**	**0.040**	**0.037**	**0.123**
β	**0.682**	**0.632**	**0.677**	**0.559**	**0.676**	**0.512**
γ_{BFACT1}	**0.008**	-0.003	**0.035**	0.000	0.001	-0.002
$\gamma_{\text{DIFFSLOPE}}$	0.006	**0.037**	0.000	**0.022**	0.006	**-0.019**
γ_{BFACT3}	0.004	-0.023	**0.072**	-0.013	0.010	**-0.025**
γ_{SFACT1}	**0.009**	**0.014**	0.000	**0.007**	**0.011**	**-0.014**
γ_{SFACT3}	-0.026*	0.001	**0.049**	0.005	-0.016	0.007
γ_{SPREAD}	**-1.349**	**-2.467**	**-1.902**	-0.456*	-0.274	-0.378
$\gamma_{\text{BIDVOL}} * 10^{-5}$	**-3.370**	**-1.865**	**3.860**	**0.582**	0.524	0.168
$\gamma_{\text{ASKVOL}} * 10^{-5}$	**1.390**	**1.070**	**1.310**	**1.670**	**-0.953**	-0.196
γ_{VOLAT}	**-0.493**	**-0.334**	-0.142	**-0.120**	**0.431**	0.487*
ϕ	**1.388**	**1.195**	**1.402**	**0.652**	**0.773**	**0.775**
$W(\psi's = 0)$	(0.00)	(0.00)	(0.01)	(0.01)	(0.71)	(0.00)
Var(ϵ_t)	1.07	1.04	1.09	1.01	1.03	1.02
Log likelihood	-5516.6	-6952.0	-5497.3	-16220.2	-13507.1	-13613.9

The table reports Maximum Likelihood estimates of a MDACP system. The dependent variables are the one-minute counts of category 1 orders (BMO-agg/SMO-agg, market orders that walk up or down the book), category 2 orders (BMO-inter/SMO-inter, market orders which consume all volume available at the best quote), category 3 orders (BMO-small/SMO-small, orders are market orders that consume part of the depth at the best quote), category 4 orders (BLO-inside/SLO-inside, limit orders submitted inside the best quotes), category 5 orders (BLO-at/SLO-at, limit orders submitted at the best quote), category 6 orders (BLO-outside/SMO-outside, limit orders outside the best quotes) and cancellations (BCANC/SCANC). The table reports the results of the ask side equations. The mean equations are specified as equation (3.4.2).
The vector of predetermined variables are the inside spread (SPREAD), the volume at the best bid (BIDVOL), the volume at the best ask (ASKVOL) and volatility measured by the standard deviation of the last 5 minutes midquote returns (VOLAT). BFACT1 (SFACT1) denotes the first factor (liquidity factor) extracted by PCA at the bid (ask) side, DIFFSLOPE is the difference of the absolute values of the second factors (informational factor). BFACT3 and SFACT3 denote the third factor for the bid (ask) side. ϕ is the dispersion parameter of the Double Poisson. Parameters significant at the 5% and 10% level are printed boldface and accompanied with a star, respectively. For ϕ the null hypothesis is that the parameter is one, for β and the α parameters the null is that the true parameter is zero.The $W(\psi's = 0)$ row reports the p-values for a Wald-test of the hypothesis that the seasonality parameters $\psi_{s,1}$, $\psi_{s,2}$, $\psi_{c,1}$ and $\psi_{c,2}$ are jointly zero. $Var(\varepsilon_t)$ is the variance of the Pearson residual which should be close to one for a correctly specified model.

(PCA)[10] for the analysis of commonalities in the limit order book. We adopt their approach to analyze the impact of the order book state beyond the inside market on trading activity. We conduct a PCA based on the reconstructed limit order book. The basic idea is to compute the hypothetical unit price of a market order of volume v if it were executed immediately against the order book at time t. Dividing the unit price by the best quote prevailing at time t yields the relative price impact. In our application the relative price impact is computed for v=3,000 to 40,000 with $1,000$ shares increments. PCA is then employed to summarize the information using a small number of factors (principal components) which are, by construction, uncorrelated.[11] The PCA is conducted separately for buy and sell side of the order book.

Table 3.7 presents the variance shares of the first five principal components and Figure 3.4 depicts the factor loadings of the first three principal components. We focus on the buy side results, the sell side results being quite similar. Table 3.7 shows that the first three factors explain 99% of the total variation of the data. Figure 3.4 shows that the first factor has nearly constant loadings for all volumes v. An increase of this factor, given the positivity of the factor weights, implies that the book is depleted and thus the percentage relative price impact increases.[12] The second factor is negatively related to the price impacts at small volumes, with factor loadings increasing monotonically with v. In other words, an increase in the second factor induces the slope of the price impact curve to become steeper. A steep slope of the book indicates that limit order traders are more cautious and want to protect themselves against information based trading by submitting less aggressive limit orders. The second principal component could therefore be interpreted as an "informational" factor.

Table 3.7: **Principal Components Analysis of the Limit Order Book - Buy side**

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Eigenvalue	32.83	3.90	0.80	0.24	0.09
Variance Share	0.864	0.103	0.021	0.006	0.002
Cumulative Share	0.864	0.967	0.988	0.994	0.996

The principal components analysis is based on relative price impact series. To compute these series we compute the unit price of a market order of volume v if it were executed immediately against the time t order book and divide it by the best quote prevailing at time t. The relative unit price is computed for v=3,000 to v=40,000 with $1,000$ shares increments. The table presents the eigenvalues, percentage of the explained variance and the cumulative explained variance of the first five principal components extracted from these relative price impact series. The analysis is conducted separately for the buy and sell side. The table shows the buy side results.

The extracted principal components can conveniently be used to test hypotheses found in the theoretical papers and previous empirical findings. Tables 3.5 and 3.6 present results of a MDACP model where the first principal component from each market side is used as an explanatory variable (in the result tables denoted as SFACT1 and BFACT1). Biais, Hillion, and Spatt (1995) find that investors provide liquidity to the market when it is valuable and consume liquidity when it is plentiful. Our empirical results seem to support this finding.

[10] For a general description of principal components analysis see Anderson (1984b).

[11] Prior to the PCA the data is standardized by substracting time-of-day specific means and dividing by standard deviation.

[12] We thank Pierre Giot for having pointed out this.

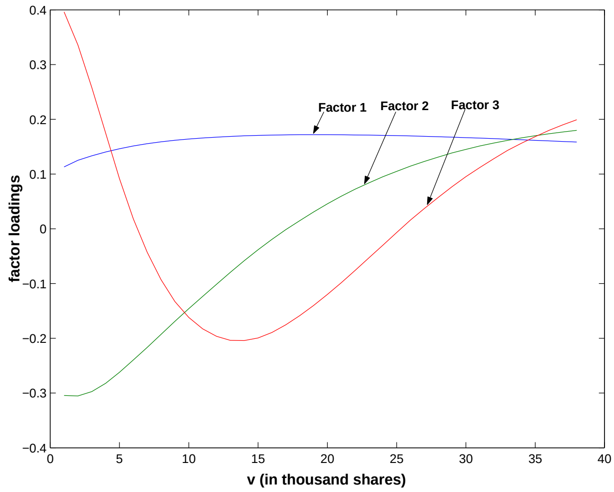


Figure 3.4: **Factor loadings of first three principal components.**

When the first factors increase, the most aggressive own-side limit orders increases, i.e. traders use more aggressive limit orders to replenish the book. For example, in Table 3.5 an increase in the first factor increases BLO-inside and BLO-at (0.005 and 0.010, respectively). Order aggressiveness on the opposite market side is also increased as the price impacts stimulate the submission of opposite-side market orders.[13] In that sense, this factor could be related with the liquidity provision and consumption in the markets: liquidity is offered when it is needed and consumed when the book is filled.

Hall, Hautsch, and Mcculloch (2003) point out that the liquidity effect, which stimulates overall trading activity, has to be distinguished from an informational effect for which the theoretical predictions are quite different. In the theoretical models of Foucault (1999) and Handa, Schwartz, and Tiwari (2003) an imbalance in the order book with a steep buy side and flat sell side order book indicates a bad news state in which prospective buy side traders act cautiously, by submitting buy limit orders away from the best bid, while sellers are expected to submit market orders and aggressive sell limit orders. To test this hypothesis, we construct a convenient indicator by taking the difference of the absolute values of the bid and ask side informational factors extracted by the PCA. This indicator (in the results tables denoted DIFFSLOPE) is positive when the ask side of the limit order book is relatively flat and the bid side of the book relatively steep (thus indicating a bad news state), i.e. with the slope of the ask-side smaller (in absolute terms) than the slope of the bid-side. The disaggregated MDACP specification uses this "bad news" indicator as an explanatory variable. The estimation results in Tables 3.5 and 3.6 show that the bad news indicator induces the bid side to become less aggressive, while the ask side acts more aggressively, which is in accordance with the theoretical predictions. The significance of the estimates are mixed but most of them have the expected direction. For example, an increase of the variable DIFFSLOPE, meaning bad news, decreases the BMO-inter (-0.023) and the BLO-inside (-0.016) (less aggressive buy orders), and increases SMO-inter (0.037) and the SLO-inside (0.022) (more aggressive sell orders).

[13] However, note that there is no significant effect of the first factor on both sides on the market on the opposite market orders (BMO-inter and SMO-inter).

Volatility and order submission activity

Foucault's (1999) theoretical model implies that when volatility increases, limit order traders ask for a higher compensation for the risk of being picked off, i.e. being executed when the market has moved against them. Then, the sell (buy) limit order traders increase (decrease) their reservation prices and market order trading becomes more costly. In equilibrium this results in higher volatility leading to the submission of less aggressive orders. Empirical evidence confirming this prediction was found by Bae, Jang, and Park (2003) and Danielson and Payne (2001). Also, Griffiths, Turnbull, and White (2000) and Ranaldo (2003) report less aggressive trades when volatility temporarily increases. In order to test the hypothesis in the MDACP framework, we measure volatility as the standard deviation of the midquote returns during the last 5 minutes and include it as a predetermined variable (in the result tables, it is denoted VOLAT). The estimation results for the aggregated system in Table 3.4 are in accordance with theoretical prediction. Volatility exerts a negative impact on the most aggressive orders (market orders) and a positive impact on less aggressive orders (limit orders).[14] Moreover, volatility affects cancellations on both sides of the book positively and significantly, which is in line with the prediction that as volatility increases, traders cancel their positions more frequently to avoid being picked off.

The estimation results for the disaggregated system in Tables 3.5 and 3.6 confirm these conclusions and provide a more detailed view. Volatility affects negatively and significantly the submission intensity of the most aggressive market orders (category one) and negatively, but not significantly, the orders of categories two[15] and three. Furthermore, volatility exerts a positive impact on limit orders at (BLO-at and SLO-at) or outside (BLO-outside and SLO-outside) the best quotes (at 10% significance for the last), but has a negative and significant effect on limit orders inside the best quotes (BLO-inside and SLO-inside). These results again confirm the theoretical prediction that order aggressiveness decreases when volatility increases.

Order submission dynamics, cancellations, and market resiliency

The VARMA structure of the MDACP model provides a convenient framework to analyze autoregressive dynamics of order submissions and cancellations in an automated auction market in the spirit of the papers by Biais, Hillion, and Spatt (1995) and Bisière and Kamionka (2000). In the following we exploit this feature in an empirical assessment of market resiliency, particularly with regard to cancellation events.

The estimation results for the aggregated MDACP system in Table 3.4 show that lagged buy (sell) market order counts exert a positive and significant effect on the expected number of sell (buy) limit orders. In other words, when liquidity is consumed by market orders, liquidity suppliers (voluntarily) enter into the market, and new (competitive) limit orders are submitted which replenish the limit order book. These results indicate market resiliency despite the absence of designated market makers. Estimation results for the disaggregated MDACP system

[14] The significance of the estimates are also mixed but they are in the expected direction.

[15] It is significant for SMO-inter in Table 3.6.

Table 3.8: **Estimation results for a bivariate MDACP system of buy and sell market orders with cancellation counts as predetermined variables**

	BMO	SMO
ω	**0.034**	**0.057**
α_{BMO}	**0.086**	0.002
α_{SMO}	**0.020**	0.003
β	**0.848**	**0.698**
$\gamma_{\text{BCANC-0-2}}$	**-0.020**	-0.007
$\gamma_{\text{BCANC-3-5}}$	-0.013*	-0.007
$\gamma_{\text{BCANC-5+}}$	-0.009*	0.010*
$\gamma_{\text{SCANC-0-2}}$	**-0.016**	**-0.031**
$\gamma_{\text{SCANC-3-5}}$	0.014	-0.006
$\gamma_{\text{SCANC-5+}}$	**0.012**	0.004
ϕ	**0.714**	**0.753**
$W(\psi's = 0)$	(0.00)	(0.00)
Var(ϵ_t)	.02	.99
Log likelihood	-14656.7	-12042.91

The table reports Maximum Likelihood estimates of a bivariate MDACP system. The dependent variables are one minute counts of buy (BMO) and sell (SMO) market orders. The mean equations are specified as equation (3.4.2).
The vector of predetermined variables consists of cancellations categorized according to their position away from the best quotes. BCANC-0-2 and SCANC-0-2 denote cancellations of standing limit orders at, or one or two quotes away from the best quotes. The categories BCANC-3-5, SCANC-3-5, BCANC-5+ and SCANC-5+ are defined accordingly. ϕ is the dispersion parameter of the Double Poisson. Parameters significant at the 5% and 10% level are printed boldface and accompanied with a star, respectively. For ϕ the null hypothesis is that the parameter is one, for β and the α parameters the null is that the true parameter is zero. The $W(\psi's = 0)$ row reports the p-values for a Wald-test of the hypothesis that the seasonality parameters $\psi_{s,1}$, $\psi_{s,2}$, $\psi_{c,1}$ and $\psi_{c,2}$ are jointly zero. $Var(\varepsilon_t)$ is the variance of the Pearson residual which should be close to one for a correctly specified model.

lead to the same conclusion as lagged market orders impact positively on all opposite-side limit order categories.[16]

So far, the theoretical literature did not devote a great deal of attention to the role of limit order cancellations in explaining future trading activity. This is surprising, as it seems natural to hypothesize that cancellation events, especially when occurring near the inside market, carry informational content. The estimation results for the aggregated system already provide some empirical evidence for the informational significance of cancellations: Table 3.4 shows that by affecting own-side expected market order submissions negatively (-0.016 and -0.011 for BMO and SMO, respectively), but by exerting a positive impact on own-side limit order submissions (0.023 and 0.054 for BLO and SLO, respectively), cancellations tend to reduce own-side order aggressiveness.

More detailed empirical analyzes provide further evidence for the informational content of cancellation events. First, we estimate a bivariate MDACP model for buy and sell market orders set up to study the effect of cancellation events on market order submissions. The results are reported in Table 3.8. Secondly, we estimate a MDACP model designed to assess the effect

[16] The estimation results of the aggregated and disaggregated MDACP systems (Tables 3.4, 3.5 and 3.6) also provide empirical evidence for the "diagonal effect" identified by Biais, Hillion, and Spatt (1995). The diagonal effect describes the stylized fact that the probability of observing a market event (a market order submission, say), given that the most recent market event was of the same type, is higher than the unconditional probability. The statistically and economically significant effect of the lagged counts on the expected number of counts of the same order category is consistent with the presence of a diagonal effect.

Table 3.9: **Estimation results for a MDACP system of buy and sell limit order categories with cancellation counts as predetermined variables**

Parameters	BLO-inside	BLO-at	BLO-outside	SLO-inside	SLO-at	SLO-outside
ω	**0.133**	**0.041**	**0.057**	**0.137**	**0.048**	**0.109**
$\alpha_{\text{BLO-inside}}$	**0.143**	**0.035**	-0.007	**0.065**	**0.045**	**0.044**
$\alpha_{\text{BLO-at}}$	**0.069**	**0.109**	0.005	0.003	-0.002	**0.015**
$\alpha_{\text{BLO-outside}}$	**0.040**	**0.077**	**0.091**	-0.009	**-0.014**	0.012
$\alpha_{\text{SLO-inside}}$	**0.058**	**0.041**	**0.043**	**0.155**	**0.024**	0.004
$\alpha_{\text{SLO-at}}$	**-0.023**	0.001	-0.002	**0.064**	**0.115**	**0.019**
$\alpha_{\text{SLO-outside}}$	**-0.036**	-0.010	0.010	**0.046**	**0.052**	**0.087**
β	**0.680**	**0.713**	**0.685**	**0.581**	**0.707**	**0.533**
$\gamma_{\text{BCANC-0-2}}$	**-0.037**	**-0.036**	**0.053**	0.012	0.005	0.015
$\gamma_{\text{BCANC-3-5}}$	**-0.021**	**-0.037**	**0.093**	0.009	**0.027**	**0.024**
$\gamma_{\text{BCANC-5+}}$	**-0.027**	0.006	**0.025**	0.011	**0.017**	-0.007
$\gamma_{\text{SCANC-0-2}}$	0.000	-0.006	-0.005	-0.014	**-0.039**	**0.089**
$\gamma_{\text{SCANC-3-5}}$	**0.061**	0.011	0.016	-0.028	**-0.044**	**0.103**
$\gamma_{\text{SCANC-5+}}$	**0.029**	**0.049**	**0.017**	0.007	**0.016**	0.013
ϕ	**0.649**	**0.786**	**0.765**	**0.649**	**0.771**	**0.777**
$W(\psi's = 0)$	(0.00)	(0.00)	(0.01)	(0.01)	(0.94)	(0.00)
Var(ϵ_t)	0.98	1.00	1.02	1.01	1.03	1.03
Log likelihood	-17525.1	-14027.4	-13866.3	-16242.7	-13519.5	-13601.6

The table reports the Maximum Likelihood estimates of a MDACP system. The dependent variables are the one-minute counts of category 4 orders (BLO-at and SLO-at, limit orders submitted inside the best quotes), category 5 orders (BLO-inside and SLO-inside, limit orders submitted at the best quote) and category 6 orders (BLO-outside and SLO-outside, limit orders submitted outside the best quotes). The mean equations are specified as equation (3.4.2).
The vector of predetermined variables consists of cancellations categorized according to their position away from the best quotes. BCANC-0-2 and SCANC-0-2 denote cancellations of standing limit orders at one or two quotes away from the best quotes. The categories BCANC-3-5, SCANC-3-5, BCANC-5+ and SCANC-5+ are defined accordingly. ϕ is the dispersion parameter of the Double Poisson. Parameters significant at the 5% and 10% level are printed boldface and accompanied with a star, respectively. For ϕ the null hypothesis is that the parameter is one, for β and the α parameters the null is that the true parameter is zero. The $W(\psi's = 0)$ row reports the p-values for a Wald-test of the hypothesis that the seasonality parameters $\psi_{s,1}$, $\psi_{s,2}$, $\psi_{c,1}$ and $\psi_{c,2}$ are jointly zero. $Var(\varepsilon_t)$ is the variance of the Pearson residual which should be close to one for a correctly specified model.

Table 3.10: **Estimation results for a MDACP system of buy and sell limit order categories (submitted outside the best quotes) with cancellation counts as predetermined variables**

	BLO-out-1-2	BLO-out-3-5	BLO-out-5+	SLO-out-1-2	SLO-out-3-5	SLO-out-5+
ω	**0.019**	**0.014**	**0.025**	**0.038**	**0.021**	**0.032**
$\alpha_{\text{BLO-outside-1-2}}$	**0.058**	0.005	**0.011**	-0.004	**0.012**	**0.013**
$\alpha_{\text{BLO-outside-3-5}}$	**0.022**	**0.074**	**0.025**	0.011	0.001	**-0.012**
$\alpha_{\text{BLO-outside-5+}}$	0.001	**0.015**	**0.084**	**0.012**	-0.001	-0.002
$\alpha_{\text{SLO-outside-1-2}}$	0.006	**0.011**	0.005	**0.079**	**0.011**	**-0.008**
$\alpha_{\text{SLO-outside-3-5}}$	0.006	0.001	0.006	**0.021**	**0.046**	**0.013**
$\alpha_{\text{SLO-outside-5+}}$	0.003	**0.011**	**-0.013**	**-0.014**	-0.001	**0.087**
β	**0.684**	**0.716**	**0.768**	**0.558**	**0.692**	**0.663**
$\gamma_{\text{BCANC-0-2}}$	**0.046**	**0.023**	**-0.009**	**0.025**	**0.012**	**0.008**
$\gamma_{\text{BCANC-3-5}}$	**0.029**	**0.060**	**-0.009**	**0.012**	**0.014**	**0.009**
$\gamma_{\text{BCANC-5+}}$	**0.007**	-0.005	**0.027**	0.002	-0.001	**0.0097**
$\gamma_{\text{SCANC-0-2}}$	**0.011**	0.005	-0.004	**0.084**	**0.016**	**-0.007**
$\gamma_{\text{SCANC-3-5}}$	**0.014**	0.006	**0.008**	**0.021**	**0.063**	**0.007**
$\gamma_{\text{SCANC-5+}}$	**0.020**	0.000	**0.012**	0.003	-0.003	**0.023**
ϕ	**1.030**	**1.149**	**1.347**	**1.0223**	**1.197**	**1.287**
$W(\psi's = 0)$	(0.00)	(0.02)	(0.01)	(0.01)	(0.00)	(0.00)
$\text{Var}(\epsilon_t)$	1.01	1.04	1.06	1.02	1.02	1.04
Log likelihood	-8615.7	-7683	-6701.9	-8560.2	-7364.3	-6238.8

The table reports the Maximum Likelihood estimates of a MDACP model. The dependent variables are the counts of limit orders submitted outside the best quotes. BLO-outside-1-2 and SLO-outside-1-2 count the number of buy and sell orders submitted one or two quotes away from the best quotes. The categories BLO-outside-3-5, SLO-outside-3-5, BLO-outside-5+ and SLO-outside-5+ are defined accordingly (note that we use *out* instead of *outside* in the upper part of the table, due to space limitations). The mean equations are specified as equation (3.4.2).
The vector of predetermined variables consists of cancellations categorized according to their position away from the best quotes. BCANC-0-2 and SCANC-0-2 denote cancellations of standing limit orders at, or one or two quotes away from the best quotes. The categories BCANC-3-5, SCANC-3-5, BCANC-5+ and SCANC-5+ are defined accordingly. ϕ is the dispersion parameter of the Double Poisson. Parameters significant at the 5% and 10% level are printed boldface and accompanied with a star, respectively. For ϕ the null hypothesis is that the parameter is one, for β and the α parameters the null is that the true parameter is zero. The $W(\psi's = 0)$ row reports the p-values for a Wald-test of the hypothesis that the seasonality parameters $\psi_{s,1}$, $\psi_{s,2}$, $\psi_{c,1}$ and $\psi_{c,2}$ are jointly zero. $Var(\varepsilon_t)$ is the variance of the Pearson residual which should be close to one for a correctly specified model.

of cancellations on limit order submissions (see Table 3.9). For both models we categorize the position of the cancelled limit order counts relative to the best quotes. The estimation results evidence that, as hypothesized, cancellations close to the inside market are informationally the most important events. Observing Table 3.8, these "aggressive" cancellations exert a negative and significant impact on the expected number of own-side market order submissions (BMO (-0.020) and SMO (-0.031)). Furthermore looking at Table 3.9, they decrease the expected number of most aggressive limit orders (those submitted inside and at the best quotes). For example, an increase of the most aggressive buy cancellations (BCANC-0-2), decreases BLO-inside (-0.037) and BLO-at (-0.036). However, aggressive cancellations also exert a positive impact on limit order submissions *outside* the best quotes (BLO-outside (0.053)).[17] This leads to the conclusion that the aggressive cancellations induce limit order traders to act more cautiously and to demand higher liquidity premia.

The estimation results reported in Table 3.10 provide more detailed insights. We estimated an MDACP model where the limit order submission category "outside the best quote" is further

[17] The same pattern is observed when the most aggressive sell cancellations (SCANC-0-2) increases.

disaggregated. The results show that aggressive cancellations (presented as BCANC-0-2 and SCANC-0-2) induce a higher limit order submission activity close to (yet not inside or at) the best quotes. For example, an increase of the most aggressive buy cancellations (BCANC-0-2) impacts positively and significantly to the submission of buy limit orders outside the best quote but inside the best five quotes (BLO-out-1-2 (0.046) and BLO-out-3-5 (0.023)). We conclude that, although cancellations negatively affect liquidity quality (by the negative effect on limit orders which provide the best liquidity quality), it is only reduced, and not erased. This again indicates the market resiliency property of this automated auction market.

3.5 Conclusions and future work

This chapter has presented an empirical analysis of the trading process in an automated auction market. For this purpose we use the dynamic model for multivariate time series of counts introduced by Heinen and Rengifo (2003), presented in Chapter 2. This econometric methodology is tailor-made to account for the various dimensions of the trading process. Compared with alternative empirical strategies which tackle the natural irregular spacing of the transactions data by formulating a duration model or marked point process our approach delivers results that are much easier to communicate. We have tested several hypotheses put forth by market microstructure. The results that we have obtained using the new methodology both confirm previous findings and offer new insights:

- We have found empirical support for hypothesis that larger spreads reduce the relative importance of market order trading compared to limit order submission activity. Furthermore, we have confirmed the hypothesis that increasing depth at the best quotes stimulates the submission of aggressive limit order at the same side of the market while larger depth on the opposite side of the market reduces the aggressiveness of own-side limit orders. Consistent with theoretical predictions we have found that order aggressiveness is reduced when volatility is high.

- Using a principal components analysis of the order book we have obtained the result that one of the extracted factors, identified as the "informational factor", proved to be very successful in predicting the future order submission process. As predicted both by theory and intuition, we found that if the informational factor indicates "bad news" the number of aggressive sell limit and market orders increases while buyer activity decreases.

- The results indicate the important role that cancellations play for predicting future order submission activity. More precisely, we have found that cancellations of aggressive limit orders (standing orders close to the best quotes) generally reduce the trading activity. However, those "aggressive" cancellations increase the submission activity within the first five quotes, again indicating market resiliency.

There are a number of directions for further research along the lines presented here. A potential extension could use the econometric methodology to study cross-security or cross-trading venues differences. For example, it is tempting to conduct a comparative analysis of the trading activity for assets with different ownership and/or market capitalization. Will the encouraging results regarding market resiliency still hold for small caps? Another idea is to compare trading venues which offer different degrees of pre-trade transparency. Will we still obtain those empirical confirmations of theoretical predictions outlined above if the trading process is less transparent, for example if hidden orders are allowed? The presence of hidden orders disguises part of the liquidity, i.e. blurs the "informational" component discussed above.

An extension to issues in international finance is another interesting research direction. For example, Daimler Chrysler is both listed at the NYSE and Xetra/FSE (and other international exchanges, too). As a matter of fact, the DCX globally registered share is traded simultaneously during overlapping trading hours of these international exchanges. In a comparative study one could analyze how the different degrees of pre-trade transparency at Xetra/FSE and NYSE affects the trading process and price discovery for those cross listed stocks. The method presented here is straightforwardly extended to multiple markets, thus offering the possibility to study linkages of international stock markets on a much more detailed level.

Chapter 4

Multivariate reduced rank regression in non-Gaussian contexts

This chapter reports on research done jointly with Andréas Heinen.

4.1 Introduction

Reduced-rank Regression (RRR) is an important tool in multivariate statistical analysis. It provides interpretable results based on a low dimensional view of the data, allowing for parsimonious models. However, in all the classical references on RRR, like in Anderson (1984b), there is an implicit assumption of normality. Moreover, many other techniques of multivariate analysis are also implicitly based on the normality assumption. For a survey including amongst others, discriminant analysis, factor analysis, canonical correlations and principal components, see Hardle and Simar (2003). Recently, there has been work on relaxing normality for many of these techniques. For instance in linear discriminant analysis, Zhu and Hastie (2002) analyze the case in which data is classified into categories based on general types of distributions by using non-parametric techniques. In reduced rank models, a prominent example of this is Yee and Hastie (2002), who extend reduced-rank ideas to vector generalized linear models (VGLM), and base their development on the example of the reduced-rank multinomial logit model of Anderson (1984a), also called the stereotype model. Yee and Hastie (2002) also show the relation of RR-VGLM with many other classes of models that have been proposed. Amongst them is canonical correspondence analysis of Ter Braak (1986), whose aim it is to model how a group of exogenous variables influences a table of counts.

These multivariate tools are mainly applied to reduce the number of variables and to detect a structure (if it exists) in the relationship among variables. In Chapter 2 we introduced the multivariate autoregressive double Poisson model (MDACP) and we presented as a special case the single factor structure. In this chapter we are interested in comparing the results obtained using this particular case of the MDACP model with the ones obtained using RRR techniques, as a way to determine the robustness of our findings. The immediate problem that we face when working with tick-by-tick data is the non-normal distribution of the series (number of trades, for example), a problem that impairs the direct use of common techniques as the principal components analysis (PCA) and canonical correlations (CC), that are special cases of the RRR model.

Our contribution is twofold. Firstly, we propose a new procedure of RRR for potentially non-Gaussian data based on the multivariate dispersion model (MDM) of Song (2000). An MDM is a multivariate distribution obtained by taking univariate dispersion models and linking them with a multivariate Gaussian copula. Dispersion models, introduced by Jorgensen (1987), are a very general class of distributions, which include, amongst others, continuous distributions like the normal and the gamma, and discrete distributions like the Poisson and the binomial. We establish that under small dispersion asymptotics of Jorgensen (1987), our procedure converges to the RR-VGLM of Yee and Hastie (2002). We show in a Monte Carlo simulation that estimates of our model are more efficient than the traditional estimates of RRR.

Secondly, we introduce a procedure analogous to canonical correlations (CC) and principal component analysis (PCA), but which takes into account departures from normality in the distribution of the data. We show how this can be viewed under a tail area approximation as a maximization that looks like RRR, but with the appropriate deviance residual instead of the Gaussian residual.

The chapter is organized as follows. Section 4.2 briefly explains Gaussian RRR, develops non-Gaussian RRR, introduces the canonical correlations and principal components analysis for non-Gaussian data and explains how the procedures can be adapted for discrete variables. Section 4.3 presents some simulation results. Section 4.4 develops an empirical application related to the one presented in Chapter 2. Section 4.5 concludes and presents some topics for future research.

4.2 Non-Gaussian reduced rank regression

In the first subsection we briefly present the Gaussian RRR, and we explain how we extend it to non-Gaussian contexts, using multivariate dispersion models of Song (2000). In the next subsection we introduce the MDM-CC, a procedure analogous to CC, but which takes into account departures from normality in the distribution of the data. Finally we show how our procedures can be amended to deal with discrete variables.

4.2.1 The Gaussian case

We are interested in the effect on an $(N \times K_1)$ matrix Y of an $(N \times K_2)$ matrix X of explanatory variables. Let $Y_i = (Y_{i,1}, Y_{i,2}, \ldots, Y_{i,K_1})'$ denotes the vector of the i-th observation of the K_1 variables in Y, $\mu_i = (\mu_{i,1}, \mu_{i,2}, \ldots, \mu_{i,K_1})'$ denotes the corresponding mean vector and similarly $X_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,K_2})'$. We assume that $Y_i \sim \mathcal{N}(\mu_i, \Omega)$ and that μ_i is linear in X. If the model is of full rank, we have:

$$\mu_i = \mu_i^{(f)} \equiv \omega + C^{(f)} X_i, \tag{4.2.1}$$

where $C^{(f)}$ has full rank, therefore $rank(C^{(f)}) = t = min(K_1, K_2)$. In some cases the full model has too many parameters to estimate and for reasons of parsimony it is preferable to estimate a lower rank model. Alternatively, one might have some theoretical reasons for imposing a factor structure, as is often the case in financial models. In those cases the new assumption is that $rank(C^{(r)}) = r < t$. We can then write $C^{(r)}$ as the product of two matrices $A^{(r)}$ and $B^{(r)}$, of rank r and the RRR model is

$$\mu_i = \mu_i^{(r)} \equiv \omega + A^{(r)} B^{(r)} X_i \ , \tag{4.2.2}$$

where $A^{(r)}$ is a $(K_1 \times r)$ matrix of regression coefficients and $B^{(r)}$ a $(r \times K_2)$ matrix of factor loadings.
Estimation of the parameters is done using least-squares or maximum likelihood (MLE), resulting in the following maximisation:

$$\hat{\theta}_{RRR}^{(r)} = \underset{\theta^{(r)}\ ,\Omega}{\arg\max} \ \ \mathcal{L}^G(Y, \mu^{(r)}, \Omega) \ , \tag{4.2.3}$$

where μ^r is conformable with Y, $\theta^{(r)} = (\omega, vec(A^{(r)}), vec(B^{(r)}))$ and $L^G(Y, \mu, \Omega)$ denotes the Gaussian log-likelihood function of the observations Y. As such, this is an underidentified system. In order to identify the model we impose some normalisation. There are several possibilities and we choose to impose $B^{(r)} = [I_r, \tilde{B}^{(r)}]$.

Izenman (1980) shows, in the Gaussian case, how principal components and canonical correlations can be obtained as special cases of reduced rank regression under different assumptions on the variance-covariance matrix.

4.2.2 The non-Gaussian case

In the non-Gaussian case a similar procedure can be defined by replacing the multivariate Gaussian distribution by a multivariate dispersion model (MDM) of Song (2000), obtained by applying a multivariate Gaussian copula to univariate dispersion models. We assume that conditionally on the explanatory variables X, $Y_{i,j}$ (the $i-th$ observation of the $j-th$ component) is distributed according to a univariate dispersion model distribution DM_j with mean $\mu_{i,j}$ and dispersion parameter γ_j:

$$Y_{i,j} \mid X_i \sim DM_j(\mu_{i,j}, \gamma_j). \tag{4.2.4}$$

These models are characterized by their density, which can be written as:

$$f(y; \mu, \gamma) = a(y; \gamma) \cdot exp\left(-\frac{1}{2\gamma} d(y; \mu)\right), \tag{4.2.5}$$

where $a(y;\gamma)$ is positive, $d(y;\mu)$ is the unit deviance and γ is the dispersion parameter. For more details we refer to Jorgensen (1997). This is a very general class of distributions, which contains as special cases continuous distributions like the normal, the gamma, the inverse Gaussian, and discrete distributions like the Poisson, binomial, negative binomial and compound Poisson.

In order to get a multivariate version of the univariate DM, we use a Gaussian copula,[1] i.e. we write the density of a matrix Y which follows an MDM as the product of its marginals DM_j times the Gaussian copula:

$$f(Y_i;\mu_i,\gamma,\Omega) = c(q_i,\Omega)\prod_{j=1}^{K_1} f_j(Y_{i,j},\mu_{i,j},\gamma_j)\ , \tag{4.2.6}$$

where μ_j is the mean of Y_j defined in equation (4.2.2), $\gamma = (\gamma_1,\ldots,\gamma_{K_1})$ is the vector of dispersion parameters, $f_j(.,\mu_{i,j},\gamma_j)$ is the p.d.f. corresponding to the marginal DM_j, and $c(q,\Omega) =\mid \Omega \mid^{-1/2} \exp\left(\frac{1}{2}(q'(I_{K1} - \Omega^{-1})q\right)$ is the multivariate Gaussian copula. By definition $c(q,I) = I_{k_1}$. The vector

$$q_i = (\Phi^{-1}(z_{i,1}),\ldots,\Phi^{-1}(z_{i,K_1}))', \tag{4.2.7}$$

collects the quantiles[2] of the $z_{i,j}$, which are the probability integral transforms (PIT) of the data under the marginal densities:

$$z_{i,j} = F_j(Y_{i,j};\mu_{i,j},\gamma_j), \tag{4.2.8}$$

where $F_j(.;\mu_{i,j},\gamma_j)$ is the c.d.f. corresponding to the marginal model DM_j. One caveat applies at this point: the method relies on the fact that the marginal distributions are correctly specified. In empirical work this assumption should be tested. We suggest to apply the tests proposed by Diebold, Gunther, and Tay (1998) in the context of the evaluation of density forecasts. The basic idea underlying these tests is to make sure that the $z_{i,j}$'s are uniform $[0,1]$ and i.i.d. The uniformity assumption can also be tested with a Kolgomorov-Smirnov test.

In the case of the MDM, we replace the linear mean equation (4.2.2) by

$$\mu_i = g\left(\omega + A^{(r)}B^{(r)}X_i\right), \tag{4.2.9}$$

where $g(\mu_i) = (g_1(\mu_{i,1}),\ldots,g_{K_1}(\mu_{i,K_1}))$ and $g(.)$ is the inverse of the link function, as defined in the literature of generalized linear models (GLM).[3] For variable j, the mean is:

$$\mu_{i,j} = g_j\left(\omega_j + A_j^{(r)}B^{(r)}X_i\right), \tag{4.2.10}$$

where $A_j^{(r)}$ is the j-th row of $(K_1 \times r)$ matrix $A^{(r)}$.

The estimation is done by maximum likelihood using the MDM distribution instead of the multivariate normal to build the likelihood function. Denote by $\hat{\theta}$ the estimator of the RRR:

[1] See Chapter 2 Section 2.2.1 for a detailed description of the main ideas underlying the use of copulas.

[2] $\Phi(.)$ is the distribution function of a standard normal random variable.

[3] The *link function* $g = g(\mu)$ relates the linear predictor $x'\beta$ to the mean μ. For example, the Poisson model with mean $\mu = exp(x'\beta)$ corresponds to the log link function $g = ln\ \mu$.

$$\hat{\theta} = \arg\max_{\theta\ ,\ \Omega}\ \mathcal{L}(\theta, \Omega)\ , \tag{4.2.11}$$

where

$$\mathcal{L}(\theta, \Omega) = \mathcal{L}(\theta) + \log(c(q, \Omega))\ , \tag{4.2.12}$$

and

$$\mathcal{L}(\theta) = \sum_{i=1}^{N}\sum_{j=1}^{K_1} \log\left(f_j(Y_{i,j}, g_j(\omega_j + A_j^{(r)}B^{(r)}X_i))\right) \tag{4.2.13}$$

is the part of the likelihood that comes from the marginal models. Maximizing $\mathcal{L}(\theta)$ alone corresponds to the joint estimation of common parameters of the mean, under the assumption of uncorrelated marginal distributions. By definition, $\mathcal{L}(\theta, I) = \mathcal{L}(\theta)$, since $c(q, I) = 1$.

It is clear that if the marginals are Gaussian, the procedure outlined above reduces to the classic RRR, considered by Izenman (1980), by virtue of the fact that using a multivariate Gaussian copula along with Gaussian marginals is equivalent to using a multivariate Gaussian distribution.

Song (2000) shows, in the context of panel data models for distributions in the exponential family, that the generalized estimating equation (GEE) approach of Zeger and Liang (1986) provides estimates that approximate the MDM estimates of the same model, under the conditions of small dispersion asymptotics of Jorgensen (1997). A similar relationship exists between RRR using MDM and the RR-VGLM procedure proposed by Yee and Hastie (2003). This means that the benefits of our procedure relative to RR-VGLM in terms of efficiency should be particularly important in data with large dispersion. This can be seen in the following development for multivariate exponential dispersion models (MED), [4] which parallels Song (2000).

Assume $Y_i \sim MED_{K1}(\mu_i, \gamma)$ and define $var(Y_i) = \gamma_i \nu(\mu_i)$ and $\mu = g(\eta)$, where $\eta = \omega + A^{(r)}B^{(r)}X_i$ is the linear reduced rank predictor. Then, under small dispersion asymptotics and Theorem 2 of Song (2000), the log-likelihood (4.2.12) can be approximated by:[5]

$$-\frac{N}{2} log \mid \Omega \mid + \sum_{i=1}^{N}\sum_{j=1}^{K_1} log(a(Y_{i,j}; \gamma_j)) - \frac{1}{2\gamma_j}\sum_{i=1}^{N}(Y_i - \mu_i)' V_i^{-\frac{1}{2}} \Omega^{-1} V_i^{-\frac{1}{2}}(Y_i - \mu_i)\ , \tag{4.2.14}$$

where $V_i = diag(\nu(\mu_{i,1}), \ldots, \nu(\mu_{i,K1}))$. The first order condition of this with respect to θ, the vector containing the reduced-rank coefficients, is given as:

$$\sum_{i=1}^{N}\left(\frac{\partial \mu_i'}{\partial \theta}\right)\Sigma^{-1}(Y_i - \mu_i) = 0\ , \tag{4.2.15}$$

[4]The exponential dispersion model is dispersion model with mean μ_i, dispersion γ_i and $d(y_i, \mu_i) = a(\mu_i) + b(y_i) + c(\mu_i)y_i$, for given functions a, b and c.

[5]The order of approximation is equal to $o(\gamma_{max})$, where $\gamma_{max} = max(\gamma_1, \ldots, \gamma_{K1})$.

where $\Sigma = \gamma\, V_i^{\frac{1}{2}}\ \Omega\ V_i^{\frac{1}{2}}$. Finally, this can be written as:

$$\sum_{i=1}^{N} \left(\frac{\partial \eta_i'}{\partial \mu_i} \right) W(Y_i - \mu_i) X_j = 0 \ , \tag{4.2.16}$$

where $W^{-1} = (\frac{\partial \eta_i'}{\partial \mu_i})' \Sigma (\frac{\partial \eta_i'}{\partial \mu_i})$ is the variance of the adjusted dependent variable, as defined in the GLM literature. Equation (4.2.16) is the first order condition of Yee and Hastie (2003).

4.2.3 Canonical correlations and principal component analysis

Canonical correlations (CC) between two sets of variables Y and X, is a very widely used technique of multivariate analysis. It can be thought of as finding successive pairs of linear combinations of X and Y, which are most correlated, in order to summarize the dependence between the two sets of variables. Principal component analysis (PCA) is a similar technique with only one set of variables, where the aim is to find linear combinations that best represent the variation in the original data set.

Even though there is no reference to any specific distribution in these techniques, and they are typically used in practice for all sorts of data, there exist strong links between CC (or PCA) and RRR under the assumption of Gaussian errors. In particular Tso (1981) shows that maximum likelihood estimation of a reduced-rank model under normal errors is equivalent to CC. Izenman (1980) works in the context of the following estimation procedure:

$$\hat{\theta}_{RRR} = \underset{\theta\, \in\, \Theta}{\text{arg min}} \quad \sum_{i=1}^{N} [(Y_i - \omega - ABX_i)' \Omega^{-1} (Y_i - \omega - ABX_i)] \ , \tag{4.2.17}$$

and shows that when $\Omega = \Sigma_{YY} \equiv Var(Y)$, this is equivalent to CC, and when $X = Y$ and $\Omega = I_{K_1}$, the procedure is equivalent to PCA of Y. Moreover the procedure in (4.2.17) is equivalent to maximum likelihood estimation under normality and with a given variance-covariance matrix Ω. We use this last equivalence as the basis for a new procedure which we call MDM-CC. It is analogous to CC but takes into account the distribution of the data. We have shown in section 4.2 how we can take into account the distribution of the data in a RRR using the MDM. In that case, the traditional variance-covariance matrix is replaced as a measure of dependence by a copula variance-covariance matrix, which is the variance of the normal score q associated with the original data. We propose to estimate MDM distributions instead of the Gaussian with a given copula variance-covariance matrix equal to the unconditional copula variance-covariance of the data. This essentially mimics the way in which CC are obtained from Gaussian RRR, but in the case of the RR-MDM.

Denote by $\hat{\theta}_\Omega$ the estimator of the RRR when the copula variance-covariance of the data is assumed to be Ω:

$$\hat{\theta}_\Omega = \underset{\theta\, \in\, \Theta}{\text{arg min}} \quad \mathcal{L}(\theta, \Omega) \ . \tag{4.2.18}$$

By analogy to the Gaussian RRR, we want Ω to be Ω_{YY}, the unconditional copula covariance matrix of the dependent variable. As we are using the multivariate Gaussian copula to model the dependence, we have to map the covariance matrix of Y into the corresponding copula covariance. In order to do this, we note that the input into the copula is the normal quantile $q^0_{i,j}$ of the probability integral transform (PIT) $z^0_{i,j} = F^0_j(Y_{i,j}, \mu^0_{i,j})$ of the raw data. As we want to consider the unconditional variance-covariance of the dependent variables (i.e. without considering the impact of the explanatory variables X), we use the unconditional distribution of Y, which consists in taking a distribution with a constant mean for every variable in Y. We denote the c.d.f. of the unconditional distribution of Y_j by F^0_j. Several possibilities arise at this stage. Firstly, we can have a distribution with no other parameter than the mean. Examples of this are the exponential in the continuous case or the Poisson in the discrete case. As mentioned before, if we believe that the data follows such a distribution, the assumption should be tested, for instance with the density forecast evaluation techniques of Diebold, Gunther, and Tay (1998) and if it is found to be satisfactory we can proceed. However, if the data is assumed to follow a distribution $F^0_j(., \gamma_j)$, which depends on some unknown parameter γ_j, then this parameter has to be estimated first. An estimate $\hat{\Omega}_{YY}$ of the unconditional Gaussian copula variance-covariance matrix of the dependent variables Y can thus be obtained as:

$$\hat{\Omega}_{YY} = Var[q^0] = \frac{1}{N}\sum_{i=1}^{N} q^0_i q^{0'}_i \ , \tag{4.2.19}$$

where the probability integral transform of the data is:

$$q^0_i = (q^0_{i,1}, \ldots, q^0_{i,K_1}), \tag{4.2.20}$$

$$q^0_{i,j} = \Phi^{-1}(z^0_{i,j}) \ , \tag{4.2.21}$$

$$z^0_{i,j} = F^0_j(Y_{i,j}). \tag{4.2.22}$$

where F^0_j is the c.d.f. of the j-th variable under a distribution characterized only by its mean and dispersion, which corresponds to the assumption that the variables Y depend on a constant only (without variables X):

$$Y_{i,j} \sim DM_j(\mu^0_j, \gamma_j) \ , \tag{4.2.23}$$

where μ^0_j is the constant mean of the variable Y_j and γ_j the corresponding dispersion parameter. Of course, as noted in section 4.2, when we consider this procedure in the Gaussian case, we get back the original CC.

Song (2000) shows that under a tail area approximation, the MDM approximates the multivariate dispersion density of Jorgensen and Lauritzen (1998). This density is based on the deviance residual $r(Y, \mu) = d^{\frac{1}{2}}(Y, \mu)$, where $d(Y, \mu)$ is the deviance of GLM models (see McCullagh and Nelder (1976)), which takes the place of the Pearson residual in a density which looks otherwise very much like the multivariate normal. The loglikelihood can be written as:

$$-\frac{1}{2}log \mid \Omega \mid + \sum_{i=1}^{N}\sum_{j=1}^{K_1} log(a(Y_{i,j};\gamma_j)) - \frac{1}{2}\sum_{i=1}^{N} r(Y_i,\mu_i)^{'}\Sigma^{-1}r(Y_i,\mu_i) \ , \tag{4.2.24}$$

where $\Sigma = diag(\gamma_j)\ \Omega\ diag(\gamma_j)$. Under that approximation, the MDM-CC can be seen to be the analogue of a traditional CC, but with the appropriate deviance residual instead of the Gaussian one. For instance in the case of the gamma distribution, the deviance residual takes the form $r = \frac{Y-\mu}{\mu}$, as opposed to the Gaussian one, which is simply $r = (Y - \mu)$. One remark needs to be made about MDM-CC: unlike CC, MDM-CC is not symmetric in the variables X and Y, and one therefore needs to choose which set of variables is a priori thought of as determining the other. This is certainly a limitation of the procedure, but it is inevitable if the distribution is taken into account. Another comment needs to be made about PCA. The above development has been made in the case of CC, but all the results are valid for PCA. In that case, considering $Y = X$ and $\Omega = I$ will yield a MDM-PCA, which takes into account the distribution of Y_j.

4.2.4 The discrete case

So far we have implicitly considered the case of a continuous distribution for $Y_{i,j}$. However, our method should be amended to deal with discrete distributions. We use the continuous extension argument proposed by Denuit and Lambert (2002), for multivariate time series of counts. See Section 2.2 of Chapter 2 for a detailed presentation of this case.

In addition to what is presented in that section and in order to avoid the noise introduced by the uniform random number, we average the estimates over a certain number of runs, as proposed in Machado and Santos Silva (2003), in the context of quantile regression for counts. Our estimator is therefore the average over M uniform draws of the continued estimates:

$$\hat{\theta}_{cont} = \frac{1}{M}\sum_{l=1}^{M}\hat{\theta}_{cont}^{(l)} \ , \tag{4.2.25}$$

where $\hat{\theta}_{cont}^{(l)}$ is the estimate of the parameters obtained with the l-th random uniform draw $U^{(l)}$.

4.3 Simulation

In order to evaluate the performance of our estimator we conduct several simulation studies. In all cases we generate 500 replications of a three-dimensional data set consisting of a (500×3) matrix of normally distributed explanatory variables X and a three-dimensional matrix Y, generated from the distribution $DM(\mu, \gamma)$, which is allowed to depend on a dispersion parameter γ, according to:

$$Y_{i,j} \mid X_i \sim DM(\mu_{i,j}, \gamma_j) \ , \tag{4.3.1}$$

The dispersion models used are the normal, the gamma (with shape parameter equal to 2) and the Poisson distributions. The conditional mean is specified as:

$$\mu_i = g(\omega + CX_i) \ , \tag{4.3.2}$$

where $C = A^{(r)}B^{(r)}$[6] is given and where $g(.)$ is the inverse of the canonical link function, which is the exponential function in the case of the Poisson distribution and the identity function in the case of the Gaussian. For the gamma case, we also use the exponential function, which avoids the possibility of having a negative mean. Depending on the particular dispersion model, we use the appropriate deviance residual as explained in Section 4.2.3. Note that the density depends on a dispersion parameter in the case of the Gaussian and the gamma distributions, but not of the Poisson. In the various simulations, we compare the estimators to the traditional Gaussian-RRR which, according to Tso (1981), reduces to simply canonical correlations analysis as soon as the errors are normally distributed. We apply this procedure on the raw data Y (RRR) and on the data to which we apply the link function $g^{-1}(Y)$ (RRR-link). We assume that applying the correct link function to the data will improve the performance of RRR.

We present three results: the bias, the efficiency and the mean squared error (MSE) of our parameter estimates of the given C matrix. In order to have a clear idea of the comparison of the models we present the results in terms of the Euclidean norm $\|C\| = trace(C^{'}C)^{1/2}$. We are going to evaluate our models in terms of the bias, efficiency and mean squared errors. In terms of these statistics, the best model would be the one with smallest bias, variance (measured by the efficiency) and mean squared errors. We define:

$$Bias = \left\| \frac{1}{N_{sim}} \sum_{i=1}^{N_{sim}} \hat{C}^i - C \right\| \tag{4.3.3}$$

$$Efficiency = \|V\| \tag{4.3.4}$$

$$MSE = \|W\| \ , \tag{4.3.5}$$

where $V = (v_{i,j}) = \left(\frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} (\hat{C}_{i,j}^{(k)} - \bar{C}_{i,j}^{(k)})^2 \right)$, $W = (w_{i,j}) = \left(\frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} (\hat{C}_{i,j}^{(k)} - C_{i,j}^{(k)})^2 \right)$, and $\bar{C}_{i,j}^{(k)}$ is the mean of the $(i,j) - th$ component of the estimated C matrix.

Finally and in order to be sure that the results are not driven by some outliers, we present the number of times (in percentage) that our estimates are better than the competing ones.

In terms of computational time, it takes more or less 6 hours to do the reduced rank gamma regression (MDM-RRGR) and about 14 hours to do the reduced rank Poisson regression (MDM-RRPR). This happens because in the last case we perform simulations in three different cases: the simple MDM-RRPR, the MDM-RRPR continued with a simple uniform draw and the MDM-RRPR continued averaging over 20 uniform draws of the continued estimates. Using

[6] We assume a rank-2 C matrix:

$$C = \begin{pmatrix} (0.50 & 2.21 & 0.50 \\ -2.20 & -1.41 & -2.20 \\ 1.10 & 1.78 & 1.10) \end{pmatrix}$$

the normalization defined above, we have not had numerical problems in the optimization procedure.[7]

4.3.1 Reduced rank gamma regression (MDM-RRGR)

In the first simulation, we generate data according to gamma distributions, with shape parameters γ equal to $(2, 2, 2)$. We estimate the parameters of RRR, RR-link and Reduced Rank Gamma Regression (RRGR). Table 4.1 presents (in terms of the Euclidean norm) the bias, efficiency and MSE of the parameter estimates of our $MDM - RRGR$ compared with the RRR and the $RRR - link$. We can see that our procedure is the most efficient and that also has the smallest bias. Moreover, we present in the last two columns the number of times (in percentage) that our procedure is better than the other ones. With this we can see that over 70% of the time, our estimates are closer to the given value of the C matrix.

Table 4.1: **Bias, Efficiency and Mean Squared Error of MDM-RRGR.**

Model	Bias	Eff	MSE	RRR	RRR-link
RRR	0.38	0.67	0.68		
$RRR - link$	0.04	0.70	0.70	0.52	
$MDM - RRGR$	0.04	0.50	0.50	0.74	0.72

The model estimates are based on Gamma marginals and multivariate gaussian copula (MDM-RRGR). The competing models are the Gaussian-RRR estimated on the raw data (RRR) and on the data to which we apply the link function $g^{-1}(Y)$ (RRR-link). The table presents the bias, efficiency and MSE of the parameter estimates according to equations (4.3.3), (4.3.4) and (4.3.5). The last two columns contain the number of times (in percentage) that the model $MDM - RRGR$ is better than the competing models.

4.3.2 Reduced rank Poisson regression (MDM-RRPR)

As discussed in section 4.2.4, MDM models include many discrete distributions. For this kind of distributions we propose to apply the continuous extension argument in order to satisfy the conditions for use of the copula. First we estimate the simple MDM-RRPR model, which uses the Poisson distribution, but does not consider the continuous extension argument (RRPR). Then, we estimate a RRPR model with a simple continuous extension argument, i.e. we use a single uniform random variable $[0, 1]$ ($M = 1$ in (4.2.25), RRPR-cont). Finally we average in each run over $M = 20$ draws of the uniform (RRPR-cont-20). There is a trade-off between more efficiency and simulation time, and we chose $M = 20$ based on a graph of the added benefit of an additional simulation. Arguably, efficiency gains would bit somewhat higher if we moved to 100 uniform draws. In a sense our results here can be seen as being lower bounds on the gains that could be obtained from a higher number of uniform draws. Table 4.2 presents the bias, efficiency and MSE of the parameter estimates of the $RRPR$, the $RRPR - cont$, the $RRPR - cont - 20$ and compares them with Gaussian RRR (RRR) and Gaussian RRR on transformed data (RRR-link). First of all the RRPR is better than RRR, but not than RRR-link. The second and very striking result is how much better the estimates are with only

[7] All the simulations where done using a Pentium IV 2 Ghz.

one uniform draw. This procedure is more efficient than all the other procedures by a very large margin. Finally averaging over $M = 20$ draws instead of doing a simple continuous extension argument has only a small impact on the quality of our estimates, as it improves results by $5-10\%$. In order to be sure that the result is not driven by some outliers, we check that a similar picture holds in terms of the number of times that the model outperforms the benchmarks. We verify that the RRPR-cont outperforms the RRR, RRR-link and the RRPR without the continuous extension argument respectively 100%, 83% and 96% of the time, and the RRPR-cont-20 outperforms RRR, RRR-link, RRPR without the continuous extension argument and RRPR-cont respectively 100%, 91%, 99% and 82% of the time, which establishes that averaging over continued estimates does help.

Table 4.2: **Bias, Efficiency and Mean Squared Error of MDM-RRPR.**

Model	Bias	Eff	MSE	RRR	RRR-link	RRPR	RRPR-cont
RRR	9.15	3.28	34.64				
$RRR-link$	1.21	0.45	0.40	1.00			
$RRPR$	2.38	2.90	3.99	1.00	0.11		
$RRPR-cont$	0.19	0.39	0.39	1.00	0.83	0.96	
$RRPR-cont-20$	0.18	0.35	0.36	1.00	0.91	0.99	0.82

The model estimate is based on Poisson marginals and multivariate gaussian copula. We have applied our procedure without considering the discreteness of the marginal distributions (RRPR) and we have developed a simple continuous extension argument procedure (RRPR-cont) and we have developed a Monte Carlo simulation on the continued estimates (RRPR-cont-20). Our last estimator is therefore the average over $M = 20$ uniform draws of the continued estimates according to equation 4.2.25. The competing models are the Gaussian-RRR estimated on the raw data (RRR) and on the data to which we apply the link function $g^{-1}(Y)$ (RRR-link). The table presents the bias, efficiency and MSE of the parameter estimates according to equations (4.3.3), (4.3.4) and (4.3.5). The last two columns contain the number of times (in percentage) that the model $RRPR-cont$ and $RRPR-cont-20$ are better than the competing models, including the $RRPR$.

4.4 Empirical results: sector- and stock-specific news

In Section 2.3 of Chapter 2 we presented an empirical application where the single factor MDACP model was used. This application was based on the microstructure literature that studies the existence of asymmetric information and consequently the presence of two types of traders: the uninformed who trade for liquidity reasons and informed traders who possess superior information. This superior information can be macroeconomic, sector- or stock-specific information. Accordingly, the trading activity of one asset does not only convey information about that specific asset, but can also contain information about the whole sector that this asset belongs to. We consider this analysis helpful for the purpose of identifying leaders from the point of view of dissemination of sectorial information.

We analyze the same data as in Chapter 2: the five most important US department stores traded on the New York Stock Exchange during the year 1999: May Department Stores (MAY), Federated Department Stores (FD), J.C. Penney Company, Inc (JCP), Dillar's INC (DDS) and Saks Inc (SKS). We work with the number of trades in 5-minute intervals. The data we use was taken from the Trades and Quotes (TAQ) data set, produced by the New York Stock

Exchange (NYSE). The data used was from January 2nd 1999 to December 30th 1999. This means that the sample covers 252 trading days, that represent 18,900 observations, as there are 75 5-minute intervals every day between 9:45 AM and 4 PM.[8]

4.4.1 Estimation results

In the context of this chapter, we develop a MDM-PCA of rank 1 on the number of transactions of the five assets described above. We consider that the marginal distributions of each of them is negative binomial. We choose this distribution due to the fact that the number of transactions per 5-minute intervals have small means and are overdispersed (the variance is bigger than the mean) making the use of the normal distribution inappropriate. Moreover, this distribution belongs to the multivariate dispersion models that is the basis for the method we propose.

In this case we are particulary interested in the factor loadings because they represent the weights of a linear combination that explains the comovements of the whole system that we take as a proxy for the sector-specific news. Table 4.3 presents these factor loadings. According to the results presented in this table, the ranking of sectorial influence seems to be given by FD, JCP, SKS, DDS, and MAY. Even though we do not see a clear difference in the factor loadings, this result is closely related to the one founded by Spierdijk, Nijman, and van Soest (2002) and with the factor-only MDACP model of Heinen and Rengifo (2003) already discussed in Chapter 2. This ranking of the assets in terms of sectorial information is related to the average number of transactions (see Table 2.1 in Chapter 2), except for MAY. However, this finding is also supported by the other authors' findings, that also rank MAY as the one that contains less sectorial information. Thus, even though MAY has the third largest average number of transactions, its sector-specific information effect on the other assets seems to be the smallest one.

Looking at the sector-specific information conveyed by the assets, i.e. the common factor that drives the behavior of this sector, we conclude that the stocks with most sectorial information are the most frequently traded ones (FD and JCP), a result that goes in hand with the findings of Spierdijk, Nijman, and van Soest (2002) and with the factor-only MDACP model of Heinen and Rengifo (2003).

Table 4.3: Common Factor results

	DDS	FD	JCP	MAY	SKS
Factor loadings	1.000	1.028	1.025	0.970	1.005

Factor loadings of RR-MDM of rank 1 on the number of transactions of 5 US department stores traded on the New York Stock Exchange during the year 1999. Sample goes from 02/01/99 to 30/12/99.

[8] We refer the reader to Section 2.3.1 for more details about the data used here.

4.5 Conclusion and future work

In this chapter we introduce a set of new techniques designed to apply the reduced-rank ideas to potentially non-Gaussian data. We use to that effect the multivariate dispersion models (MDM), which provide a convenient statistical framework. We show that reduced-rank multivariate dispersion models (RR-MDM) include Gaussian reduced rank regression (RRR) as a special case, and that under small dispersion asymptotics they are equivalent to RR-VGLM of Yee and Hastie (2003). We introduce multivariate dispersion models canonical correlations (MDM-CC), a procedure similar to CC, but which takes into account the distribution of the data. Finally, we describe how our methods can be amended in the case of discrete data. We show in a Monte Carlo study that our RR-MDM yields significant gains in efficiency compared to RRR.

Finally, we present an empirical application on the number of trades of five US department stores traded on the New York Stock Exchange during the year 1999. The results show that there exist a common factor among them. We present a ranking of the sector leaders based on this sampled assets. The leadership in terms of sectorial informational content is related to the average number of transactions of the assets.

Future work related with the model presented here include the non-parametric estimation of the link function that can allow better results in terms of efficiency and mean squared error. The empirical applications of our method are not limited to microstructure analysis. These applications can be from a large range of research linked to the reduced rank regression framework, including the well-known principal components analysis and canonical correlations.

Part II

Dynamic optimal portfolio models

Chapter 5

Portfolio selection

In this chapter I present an overview of theoretical and empirical approaches to optimal portfolio models.[1] In particular, I present the well-known mean-variance framework and point out its main weaknesses. My dynamic model strategy is going to be different from this framework in that it uses a different risk measure: the Value-at-Risk. In order to understand the relation among these two approaches I summarize the main results obtained by Alexander and Baptista (2002).[2] Next, I present the asymmetric power autoregressive conditional heteroscedastic (APARCH) model of Ding, Granger, and Engle (1993). This is an econometric model that encompasses many other GARCH models and that I use in the empirical model of the next chapter. Finally, I briefly present the skewed-t distribution and describe its main features.

The remainder of the chapter is organized as follows. Section 5.1 describes the basic mean-variance model. Section 5.2 presents the relation between the mean-variance and mean-VaR frameworks based on the paper of Alexander and Baptista (2002). Section 5.3 presents the APARCH model of Ding, Granger, and Engle (1993) and Section 5.4 describes the skewed-t distribution.

The basic references for Sections 5.1 and 5.2 are given by Alexander and Baptista (2002) and Campbell (2000). For Section 5.3, Ding, Granger, and Engle (1993) for the asymmetric power autoregressive conditional heteroscedastic (APARCH) model and Lambert and Laurent (2001a) for the skewed-t distribution. I refer the reader to those references for a more detailed exposition of the concepts presented in this chapter.

5.1 Mean-variance portfolio analysis

The basic finance theory underlying modern portfolio analysis is due to the pioneering work of Markowitz (1952) who suggested the mean-variance optimization. Since then many attempts have been done in order to provide a theoretical structure for pricing assets with uncertain returns. The capital-asset pricing model (CAPM), originally proposed by Sharpe (1964) and Lintner (1965), has provided a simple and powerful theory of asset pricing for more than 20

[1]For a more detailed survey see Campbell (2000).
[2]The definitions, propositions, corollaries and proofs are taken from their paper.

years. In its simplest form the CAPM predicts that the expected return on an asset above the risk-free rate is proportional to the non diversifiable risk, which is measured by the covariance between the asset return and the return of a portfolio composed of all the available assets in the market.

The mean-variance model of asset choice is based on the second degree stochastic dominance, i.e. when risky asset A second order stochastically dominates risky asset B, risky asset A must have the at least the same expected rate of return as risky asset B and lower variance. The same idea is used when we can form portfolios: if portfolio A second order stochastically dominates all other portfolios which have the same expected rate of return as it has, then this dominant portfolio must have the minimum variance among all the portfolios.

This theory is based on the assumption that the investors would like to increase the expected rate of return on their portfolios and to reduce the standard deviation of the return. A preference for expected return and an aversion to variance is implied by monotonicity and strict concavity of an individual's utility function.[3] However, for arbitrary distributions and utility functions, expected utility cannot be defined just from the expected return and variance.[4] To see this, an individual's utility function may be expanded as a Taylor series around his expected end of period wealth,

$$U(W) = U(E(W)) + U'(E(W))(W - E(W)) + \frac{1}{2}(U"(E(W))(W - E(W))^2 + R_3 \tag{5.1.1}$$

where

$$R_3 = \sum_{n=3}^{\infty} \frac{1}{n!} U^{(n)}(E(W))(W - E(W))^n \tag{5.1.2}$$

and where $U^{(n)}$ denotes the n-th derivative of U.

Assuming that the Taylor series converges and that all moments exist, the individual's expected utility could be expressed as

$$E[U(W)] = U(E(W)) + \frac{1}{2}(U"(E(W))\sigma^2(W) + E[R_3] \tag{5.1.3}$$

where

$$E[R_3] = \sum_{n=3}^{\infty} \frac{1}{n!} U^{(n)}(E(W))\mu^n(W) \tag{5.1.4}$$

and where $\mu^n(W)$ denotes the n-th central moment of W.

The last relation illustrates that the expected utility cannot be defined solely by the expected value and variance of wealth for arbitrary distributions and preferences, as indicated by the remainder term which involves higher order moments.

Taking this fact into account, the mean-variance model assumes quadratic utility functions under which the third and higher order derivatives are assumed to be zero ($E(R_3) = 0$) for arbitrary distributions.

[3] A detailed discussion of individual's preferences and utility functions is found in Leroy and Werner (2001).
[4] See Huang and Litzenberger (1988).

Unfortunately, quadratic utility functions display undesirable properties of satiation and increasing absolute risk aversion. The satiation property implies that an increase in wealth beyond the satiation point decreases utility. Increasing absolute risk aversion implies that risky assets are inferior goods.

Furthermore, the mean-variance model assumes that the rates of return of the risky assets are jointly normally distributed. The normal distribution is completely described by its mean and variance and the third and higher order moments can be expressed as functions of the first two moments. Normal distributions are also stable under linear combinations, i.e. the rate of return on a portfolio made up of assets having returns that are normally distributed is also normally distributed. Unfortunately, the normal distribution is unbounded from below, which is inconsistent with the limited liability and with economic theory, which attributes no meaning to negative consumption. Fortunately, multivariate normality is only a sufficient distributional condition for all individuals to choose mean-variance efficient portfolios, not a necessary condition.

Based on the above, the mean-variance model is not a general model for asset choice. Its central role in financial theory can be attributed to its analytical tractability and the richness of its empirical predictions.

In the next section I follow the presentation of Alexander and Baptista (2002) in order to explain how the mean-variance and the mean-VaR approaches can be reconciled. The definitions, lemmas and propositions are taken literarily from these authors.

5.2 Mean-variance and mean-VaR approaches

Nowadays banks and financial institutions have adopted Value-at-Risk (VaR) as the measure for market risk. For example, the Basle Capital Adequacy Accord establishes the minimum capital requirement to cover the market risk exposure in terms of a given VaR.[5] The VaR is defined as the worst expected loss on an investment over a specified horizon given some confidence level. For example a 95% VaR for a 30-day holding period, implies that the worst loss incurred over the next 30 days should only exceed the VaR limit in five cases out of 100. It therefore reflects the potential downside risk faced on investments in terms of nominal losses. Furthermore, the VaR at the $100\alpha\%$ confidence level of a risky portfolio for a specific time horizon can be defined as the rate of return v such that the probability of that portfolio having a rate of return of $-v$ or less is $1-\alpha$.

Let Assume that there is not a riskfree security in the market and that there are $n \geq 2$ risky assets traded in a frictionless economy,[6] where unlimited short selling is allowed and where the rates of return on the assets have finite variances and unequal expectations. Also assume that the random rate of return of any asset cannot be expressed as a linear combination of the

[5] Actually, the Basle committee defines a 99% confidence level VaR over 10 days period. This resulting VaR is multiplied by a safety factor of 3 to provide the minimum capital requirement for regulatory purposes.

[6] A frictionless economy is one in which the market is fully efficient in that: (i) there exist no transaction costs on risky assets (buying and selling a stock at the same price is possible) and, (ii) there exist no transaction costs on riskfree assets (borrowing and lending is possible at the same rate) (iii) there exists no taxes.

rates of return on other assets.[7] Under these assumptions, asset returns are said to be linearly independent and their variance-covariance matrix Σ is nonsingular. Furthermore, it is assumed that the rates of return have a multivariate normal distribution whit mean $\mu \in R^n$, the vector of expected rates of return, and variance-covariance matrix Σ.[8] Let $W \equiv [w \in R^n : \sum_{j=1}^{n} w_j = 1]$ be the set of portfolios with well-defined expected rates of return (w_j is the proportion of wealth invested in security j).[9] For any $w \in W$, define r_w as the random rate of return of portfolio w, and let $E[r_w]$ and $\sigma[r_w]$ denote the expected rate of return and the standard deviation of the rate of return of portfolio w. Formally, the VaR is defined as follows:

Definition 5.2.1. *The VaR for a given time period, at the* $100\alpha\%$ *confidence interval of a risky portfolio is the rate of return* v *such that* $F(-v) = 1 - \alpha$*, where* $\alpha \in (1/2, 1)$ *and* $F(.)$ *is the cumulative distribution function of the portfolio rate of return at the end of the period.*

In particular the VaR at $100\alpha\%$ confidence interval for a portfolio with a normally distributed rate of return r is given by:

$$V[\alpha, r] = \eta\sigma[r] - E[r] \tag{5.2.1}$$

for any $\alpha \in (1/2, 1)$ with $\eta \in (0, \infty)$ such that $\Phi(-\eta) = 1 - \alpha$. The function $\Phi(.)$ is the standard normal cumulative distribution.

The mean-VaR and the mean-variance portfolio frontier

Definition 5.2.2. *A portfolio* $w \in W$ *belongs to the mean-VaR portfolio frontier at the* $100\alpha\%$ *confidence level if and only if for some* $\bar{r} \in R$*,* w *solves the following problem:*

$$\min_{w \in W} \quad \eta\sigma[r_w] - E[r_w] \tag{5.2.2}$$

s.t.

$$E[r_w] = \bar{r} \tag{5.2.3}$$

Definition 5.2.3. *A portfolio is a mean-variance frontier portfolio if it has the minimum variance among portfolios that have the same expected rate of return. A portfolio p is a frontier portfolio if and only if* w_p*, the n-vector portfolio weights of p, is the solution to the following problem:*

$$\min_{w \in W} \quad \frac{1}{2} w' \Sigma w \tag{5.2.4}$$

[7]This corresponds to the idea of "Effectively complete market" of Arrow. For a further discussion see Mas-Colell, Whinston, and Green (1995).

[8]Embrechts, McNeil, and Straumann (2002) showed that the VaR and variance constraints lead to the same efficient frontiers not only when rates of return have a multivariate normal distribution but in general when the distribution used belongs to the family of the elliptical distributions.

[9]Note that in this formulation we do not rule out the possibility of short sales in that w_j is allowed to be negative.

s.t.

$$w'\mu = \bar{r} \tag{5.2.5}$$

and

$$w'\iota = 1 \tag{5.2.6}$$

where μ is the n-vector of expected rates of return on the n risky assets, ι is a n-vector of ones.

Forming the Lagrangian we get:

$$\min_{[\,w,\,\lambda,\,\gamma]} \qquad L = \frac{1}{2}w'\Sigma w + \lambda(\bar{r} - w'\mu) + \gamma(1 - w'\iota). \tag{5.2.7}$$

By solving the first order conditions [10] with respect to w, λ and γ the unique set of portfolio weights for the frontier portfolio having the expected rate of return $\bar{r}$ is found to be:

$$\hat{w}_p = g + h\bar{r} \tag{5.2.8}$$

where

$$g = \frac{1}{D}[B(\Sigma^{-1}\iota) - A(\Sigma^{-1}\mu)] \tag{5.2.9}$$

and

$$h = \frac{1}{D}[C(\Sigma^{-1}\mu) - A(\Sigma^{-1}\iota)] \tag{5.2.10}$$

with

$$A = \iota'\Sigma^{-1}\mu \tag{5.2.11}$$

$$B = \mu'\Sigma^{-1}\mu \tag{5.2.12}$$

$$C = \iota'\Sigma^{-1}\iota \tag{5.2.13}$$

$$D = BC - A^2 \tag{5.2.14}$$

Notice that $D > 0$. To see this note that $(A\mu - B\iota)\Sigma^{-1}(A\mu - B\iota) = B(BC - A^2)$, moreover, the left-hand side of this relation is strictly positive, since Σ^{-1} is positive definite. Hence the right-hand side is strictly positive, wherefrom $D > 0$.

Note that all these conditions are necessary and sufficient for $\hat{w}_p$ to be a frontier portfolio having an expected rate of return equal to $\bar{r}$. Therefore, any frontier portfolio can be represented by (5.2.8). The set of all frontier portfolios is called the portfolio frontier.

The covariance between the rates of return of any two frontier portfolios p and q is

$$Cov(r_p, r_q) = w_p'\Sigma w_q = \frac{C}{D}(\bar{r}_p - A/C)(\bar{r}_q - A/C) + \frac{1}{C} \tag{5.2.15}$$

where the definition of covariance and the portfolio weights for a frontier portfolio given in equation (5.2.8) were used.

[10] In this case, since Σ is a positive definite matrix, it follows that the first order conditions are necessary and sufficient for a global optimum.

Using the definition of the variance of the rate of return of a portfolio and (5.2.15) we can define:

$$\frac{\sigma^2(r_p)}{1/C} - \frac{(\bar{r} - A/C)^2}{D/C^2} = 1 \qquad (5.2.16)$$

which is a hyperbola in the standard deviation-expected rate of return space, with center $(0, A/C)$ and asymptotes $\bar{r} = A/C \pm \sqrt{D/C}\sigma(r_p)$. Furthermore, it can be noted that the "minimum variance portfolio" is at $(\sqrt{1/C}, A/C)$.[11]

Then, as Merton (1972) showed, a portfolio w belongs to the mean-variance boundary if and only if it satisfies (5.2.16).

For any portfolio ($w \in W$), the mean-VaR boundary converges to a line with slope of minus one that intersects the origin as $\alpha \to 1/2$. To see this, observe that $lim_{\alpha \to 1/2} V[\alpha, r_w] = -E[r_w]$ since $\eta \to 0$ as $\alpha \to 1/2$.

The mean-VaR efficient frontier

Definition 5.2.4. *A portfolio $w \in W$ belongs to the mean-VaR efficient frontier at the $100\alpha\%$ confidence level if and only if no portfolio $v \in W$ exists such that $E[r_v] \geq E[r_w]$ and $V[\alpha; r_v] \leq V[\alpha; r_w]$, where at least one of the inequalities is strict.*

Definition 5.2.5. *A portfolio $w \in W$ belongs to the mean-variance efficient frontier if and only if no portfolio $v \in W$ exists such that $E[r_v] \geq E[r_w]$ and $\sigma[r_v] \leq \sigma[r_w]$, where at least one of the inequalities is strict. In other words, those frontier portfolios which have expected rates of return strictly higher than that of the minimum variance portfolio, A/C, are called "efficient portfolios".*

Figure 5.1 presents the main ideas given by the definitions stated above.

The minimum VaR portfolio

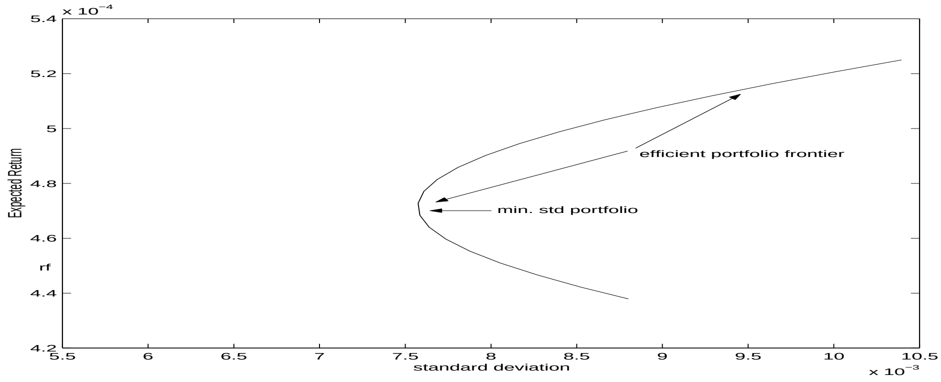
The frontier portfolio that has the smallest VaR among all the other frontier portfolios is the minimum VaR portfolio. The next lemma allows us to determine the existence of such a portfolio:

Lemma 5.2.6. *If the minimum VaR portfolio at the $100\alpha\%$ confidence level exists, then it is mean-variance efficient.*

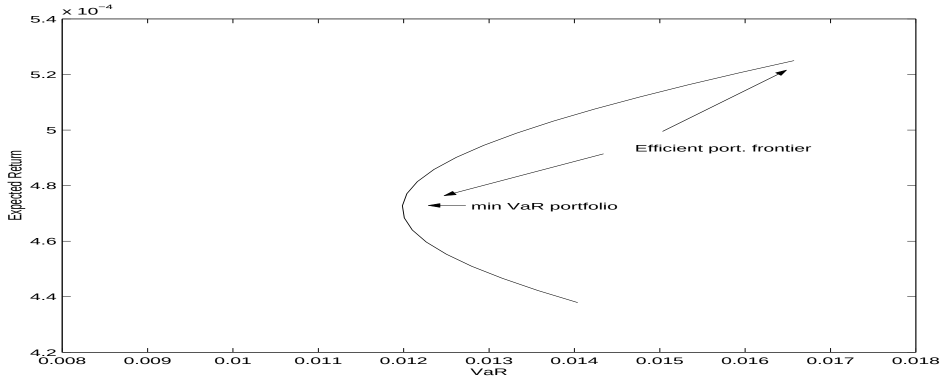
Looking at equation (5.2.1) one can observe that if the confidence level at which the VaR is determined is not large enough so that the standard deviation effect ($\eta\sigma[r]$) outweighs the mean effect ($E[r]$), then the problem of minimizing VaR has no solution, i.e. one must be careful in choosing α to calculate the VaR in a way that minimizing VaR is an obtainable objective. If α is sufficiently small, then the VaR minimization problem does not have a solution.

The following result shows the existence of the minimum VaR portfolio and a closed-form solution to the VaR minimization problem. Let $m_\alpha \in W$ denote the minimum VaR portfolio at the $100\alpha\%$ confidence level, then:

[11] This follows directly from (5.2.16).



(a) Minimum mean-standard deviation efficient frontier



(b) Minimum mean-VaR efficient frontier

Figure 5.1: **Mean-variance and mean-VaR efficient frontiers.**

Proposition 5.2.7. *The minimum VaR portfolio at the* $100\alpha\%$ *confidence level exists if and only if* $\alpha > \Phi(\sqrt{D/C})$. *Furthermore, if* $\alpha > \Phi(\sqrt{D/C})$, *then* $m_\alpha \in W$ *is given by*

$$m_\alpha = g + h\left[\frac{A}{C} + \sqrt{\frac{D}{C}\left(\frac{(\alpha^*)^2}{C(\alpha^*)^2 - D} - \frac{1}{C}\right)}\,\right] \tag{5.2.17}$$

where g and h are n-dimensional vectors defined by (5.2.9) and (5.2.10).

Using Proposition 5.2.7 and Eq.(5.2.1), if the minimum VaR portfolio exists, then its VaR is given by

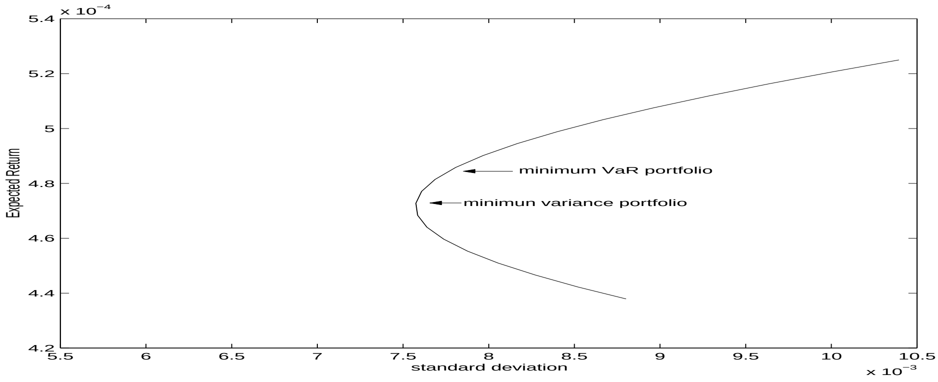
$$V[\alpha, r_{m_\alpha}] = \alpha^*\sqrt{\frac{(\alpha^*)^2}{C(\alpha^*)^2 - D}} - \left[\frac{A}{C} + \sqrt{\frac{D}{C}\left(\frac{(\alpha^*)^2}{C(\alpha^*)^2 - D} - \frac{1}{C}\right)}\,\right] \tag{5.2.18}$$

Corollary 5.2.8. *If the minimum VaR portfolio exists at a given confidence level* $\alpha < 1$, *then it lies above the minimum variance portfolio on the mean-variance efficient frontier.*

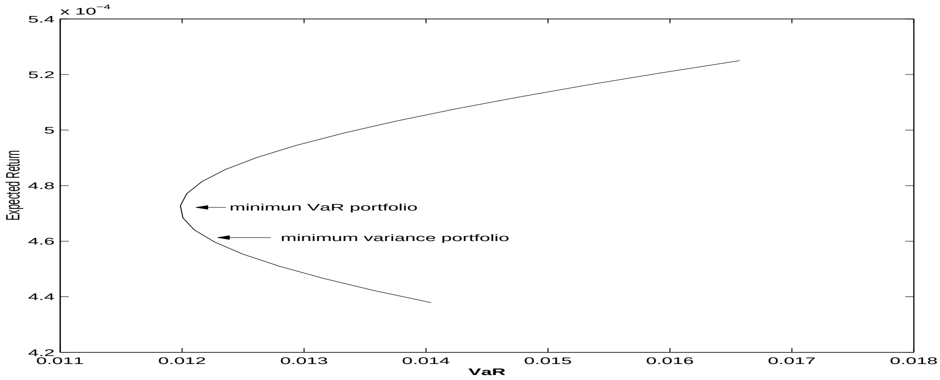
Corollary 5.2.8 indicates that the solution to the variance minimization problem is always different from the VaR minimization problem. Figure 5.2 presents the graphical idea of this corollary.

The characterization of mean-VaR efficiency

Proposition 5.2.9. *(i) If* $\alpha > \Phi(\sqrt{D/C})$, *then a portfolio* $w \in W$ *is mean-VaR efficient at that confidence level if and only if it belongs to the mean-VaR boundary and* $E[r_w] \geq E[r_{m_\alpha}]$.



(a) Minimum mean-standard deviation portfolio



(b) Minimum mean-VaR portfolio

The minimum VaR portfolio at α% confidence level lies above the minimum variance portfolio, thus, any mean-variance efficient portfolio that lies below the minimum VaR portfolio at α% confidence level is inefficient.

Figure 5.2: **Minimum variance and minimum VaR portfolios.**

(ii) If $\alpha \leq \Phi(\sqrt{D/C})$*, then no mean-VaR efficient portfolio exists at that confidence level.*

Corollary 5.2.10. *The minimum variance portfolio is mean-VaR inefficient at any confidence level* $\alpha < 1$.

Convergence results

The next corollary states that as the confidence level (α) increases, the standard deviation effect ($\eta\sigma[r]$) also increases due to the associated increase in η, and the mean effect ($E[r]$) becomes small and virtually disappears in the limit.

Corollary 5.2.11. *The minimum VaR portfolio converges to the minimum variance portfolio as* $\alpha \to 1$.

Corollary 5.2.12 shows that the set of efficient portfolios in the economy is reduced (respect to the mean-variance efficiency criterion) when the VaR is used as a measure of risk, and that the mean-VaR efficient set converges to the mean-variance efficient set as $\alpha \to 1$. Figure 5.3 gives us a clear idea of this corollary.

Corollary 5.2.12. *The set of mean-VaR efficient portfolios is a proper subset of the set of mean-variance efficient portfolios when* $\alpha < 1$*, but the former converges to the latter as* $\alpha \to 1$.

Corollary 5.2.13. *The expected rate of return of the minimum VaR portfolio converges to infinity and the set of mean-VaR efficient portfolios converges to the empty set as* $\alpha \downarrow \Phi(\sqrt{D/C})$.

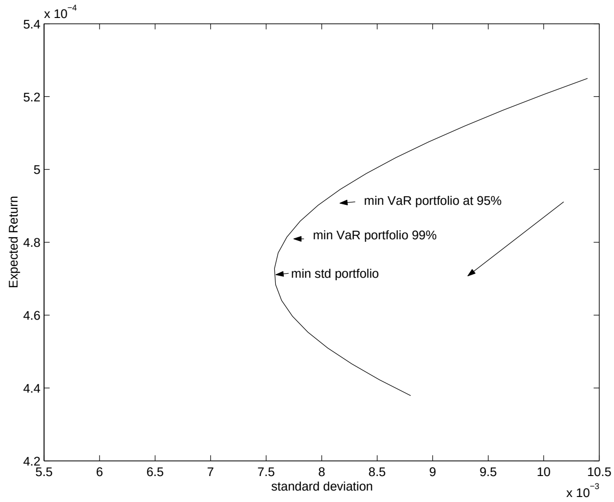


Figure 5.3: Convergence results.

VaR and non-normality

The VaR is very sensitive to the parametric distribution used for computing the quantiles. Stylized facts observed in the empirical literature on high frequency financial time series have shown that the normality assumption is in general unrealistic and that it is important to consider higher moments such as the skewness (asymmetry) and kurtosis (fat tails). The most obvious way to relax the normality assumption in order to capture the excess of kurtosis present in the returns' distribution is to instead assume that the rates of return of n risky securities have a multivariate Student-t distribution, denoted by $t_n(\mu, \Omega, \upsilon)$, where $\mu \in R^n$ is the vector of expected rates of return, Ω is the $(n \times n)$ scale matrix, and $\upsilon > 2$ is the number of degrees of freedom.[12] Then, the variance-covariance matrix of the rates of return is $\Sigma \equiv \frac{\upsilon}{\upsilon-2} \times \Omega$.

For any $\alpha \in (1/2, 1)$, let $\eta(\upsilon)$ be such that $F_\upsilon(-\eta(\upsilon)) = 1-\alpha$, where $F_\upsilon(.)$ is the cumulative distribution function of a Student-t distribution with $\upsilon > 2$ degrees of freedom. Note that if $Z \sim t_n(\mu, \Omega, \upsilon)$, then $a'Z \sim t_1(a'\mu, a'\Omega a, \upsilon)$ for every $a \in R^n$. Hence, for any portfolio $w \in W$, we have $r_w \sim t_1(w'\mu, w'\Omega w, \upsilon)$ and $V[\alpha, r_w, \upsilon]$, portfolio $w's$ VaR, is:

$$V[\alpha, r_w, \upsilon] = \eta(\upsilon)[(\upsilon-2)/\upsilon]^{1/2}\sigma[r_w] - E[r_w] \tag{5.2.19}$$

That is, for a fixed υ, portfolio $w's$ VaR is a linear function of $\sigma[r_w]$and $E[r_w]$. Therefore, the previous results still hold if rates of return have a multivariate Student-t distribution, and $V[\alpha, r_w, \upsilon]$ and $\eta(\upsilon)[(\upsilon-2)/\upsilon]^{1/2}$ are used instead of $V[\alpha, r_w]$ and η, respectively.

[12]In Section 5.4, we present the skewed-t distribution. This distribution is able to capture both the skewness and the kurtosis of the portfolio returns.

5.3 The asymmetric power ARCH model

Modern portfolio theories aim to allocate assets by maximizing expected return per unit of risk. In a mean-variance framework risk is defined in terms of possible variation of portfolio returns. The focus on standard deviation as the appropriate measure of market risk implies that investors weigh the probability of negative returns equally against positive returns. However, there is ample evidence that agents often treat losses and gains asymmetrically. Furthermore, it is a stylized fact that the distribution of many financial returns is non normal, with skewness and excess kurtosis. Therefore, the choice of a mean-variance efficient portfolio is likely to give rise to an inefficient strategy for optimizing expected returns for financial assets whilst minimizing risk. Another important fact is that those models work in an unconditional setup.

Therefore, it seems relevant to include in the optimal portfolio selection problem additional moments of the return distribution, which captures additional risk factors (along with the use of standard deviation), and to use a model that allows to perform conditional analysis.

5.3.1 A brief literature review

During the last years, the statistical analysis of financial time series has been focused on the conditional second moment due to the volatility clustering found in most financial asset returns. Engle (1982) proposed the ARCH-type model to describe the conditional variance while an ARMA structure is used to describe the conditional mean. These models often rely on the simplistic assumption of normality for the estimation of the conditional mean and for the conditional variance. But, accumulated empirical evidence shows that the distribution of financial returns on a weekly, daily or intradaily basis, has fat tails and may even be skewed. Then, it is interesting to note that modelling skewness and kurtosis has an impact on all conditional quantiles. However, if one is only interested in the first two conditional moments, the normality assumption may be justified by the fact that the quasi-maximum likelihood estimator is consistent assuming that the conditional mean and the conditional variance are correctly specified. This estimator is, however, inefficient with the degree of inefficiency increasing with the degree of departure from normality (Engle and Gonzalez-Rivera (1991)).

To take into account the skewness and excess of kurtosis, GARCH models have been used with Student-distributed errors. However, this density cannot capture the skewness and leptokurtosis. Liu and Brorsen (1995) and Lambert and Laurent (2001b) consider the asymmetric stable density in combination with a GARCH model. This new model has an undesirable characteristic (except when the tail parameter $\alpha = 2$, i.e. normality), i.e. the inexistence of the variance, something that is hard to support empirically and to consider for practical purposes. Lee and Tse (1991), Knight, Satchell, and Tran (1995) and Harvey and Siddique (1999) propose alternative skewed fat-tailed densities with respectively the Gram-Charlier expansion, the double-gamma distribution and the non-central t. But, as pointed out by Bond (2000) estimation of these densities in a GARCH framework often generates computational problems due to the high sensitivity to initial values. McDonald (1991) introduced the exponential generalized beta distribution of the second kind (EGB2), a flexible distribution that is able to accommodate not only thick tails but also asymmetry. However, based on the works of Wang, Fawson, Barret, and McDonald (2000) in a GARCH framework, goodness-of-fit test reject the EGB2 for all the series that they consider, even if it seems that it does a better job than normal and the Student-t. Hansen (1994) proposed a skewed Student-t distribution in which the conditional higher moments may vary over time. This density nests the symmetric Student-t when the asymmetry coefficient (λ) equals 0, with $-1 < \lambda < 1$. This density is easy to implement and does not imply serious problems of convergence in estimation. Recently Fernandez and Steel (1998) developed a more general tool (based on the method of inverse scaling of the probability density function on the left and the right of the mode) in which all the parameters have a clear interpretation. Finally and in order to keep in the ARCH tradition, Lambert and Laurent (2001a) reexpressed the skewed Student-t density in terms of the mean and the variance, i.e., parameterize this density in such a way that the innovation process has zero mean and unit variance. Otherwise, it is difficult to separate the fluctuations in the mean and variance from the fluctuations in the shape of the conditional density.

5.3.2 VaR models

From an empirical point of view, the computation of VaR for a sample of returns requires the computation of the empirical quantile at level α of the distribution of the returns of the portfolio, because quantiles are direct functions of the variance in parametric models.

ARCH class models immediately translate into conditional VaR models. However, it is important to note that volatility forecastability deteriorates quickly with the time horizon of the forecasts, as a consequence volatility forecasting is more relevant for short time horizons than for long time horizons.

To define a model, let us consider a sample of daily returns, r_t, with t=1....T. The class of AR(p)-APARCH(1,1) model is defined by:

$$\phi(L)(r_t - \mu) = e_t \tag{5.3.1}$$

$$e_t = \epsilon_t h_t \tag{5.3.2}$$

$$h_t^\delta = \omega + \alpha_1(|e_{t-1}| - \alpha_n e_{t-1})^\delta + \beta_1 h_{t-1}^\delta \tag{5.3.3}$$

In equation (5.3.1) $\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$ is an AR lag polynomial of order p and e_t is an innovation with $E(e_t) = 0$. Accordingly, the conditional mean of r_t is equal to $\mu + \sum_{j=1}^p \phi_j(r_{t-j} - \mu)$ and the unconditional mean is $\mu/(1 - \sum_{j=1}^p \phi_j)$ if $|\sum_1^p \phi_j| < 1$. The innovation e_t is specified by equation (5.3.2) as conditionally heteroskedastic, with conditional variance given by h_t^2. In equation (5.3.2), ϵ_t is IID with mean zero and variance 1. Finally, equation (5.3.3) defines the dynamics of h_t with ω, α_1, α_n, β_1 and δ parameters to be estimated: δ ($\delta > 0$) is the parameter of a Box-Cox transformation of h_t, while α_n ($-1 < \alpha_n < 1$) reflects the leverage effect. A positive (negative) value of α_n means that a past negative (positive) shock has a deeper impact on current conditional volatility than a past positive shocks of the same magnitude.

The inclusion of the power coefficient δ has been motivated by a stylized fact detected by Taylor (1986), who observed that the absolute returns ($|r_t|$) in financial time series are positively autocorrelated, even at long lags. Ding, Granger, and Engle (1993) found that the closer δ to 1, the larger the memory of the process. It is interesting to note that the APARCH(1,1) model includes seven other ARCH extensions as special cases:

- The ARCH of Engle (1982) when $\delta = 2$, $\alpha_1 = 0$ and $\beta_1 = 0$.
- The GARCH of Bollerslev (1986) when $\delta = 2$ and $\alpha_1 = 0$.
- Taylor (1986) and Schwert (1990)'s GARCH when $\delta = 1$ and $\alpha_1 = 0$.
- The GJR of Glosten, Jagannathan, and Runkle (1993) when $\delta = 2$.
- The Threshold-ARCH (TARCH) of Zakoian (1994) when $\delta = 1$.
- The non-linear-ARCH (NARCH) of Higgins and Bera (1992) when $\alpha_1 = 0$ and $\beta_1 = 0$.
- The Log-ARCH of Geweke (1986) and Pentula (1986), when $\delta \to 0$.

5.4 The skewed-t distribution

The empirical literature on high frequency financial time series shows that the normality assumption of the assets' returns is in general unrealistic and that it is important to consider distributions that allow for skewness (asymmetry) and kurtosis (fat tails). In this section I present the skewed-t distribution to deal with these features. This distribution is going to be extensively used in the next chapter.

Using the parametrization developed by Lambert and Laurent (2001a) for the skewed Student distribution in such a way that the innovation process has zero mean and unit variance, an innovation ϵ is said to be standarized skewed-t with parameters $\upsilon > 2$ (number of degrees of freedom) and $\xi > 0$ (a parameter related to the skewness) if its density is given by:

$$f(\epsilon \mid \xi, \upsilon) = \begin{cases} \frac{2}{\xi + \frac{1}{\xi}} s g[\xi(s\epsilon + m) \mid \upsilon] & \text{if } \epsilon < -\frac{m}{s} \\ \frac{2}{\xi + \frac{1}{\xi}} s g[(s\epsilon + m)/\xi \mid \upsilon] & \text{if } \epsilon \geq -\frac{m}{s}. \end{cases} \tag{5.4.1}$$

where $g(. \mid \upsilon)$ is a symmetric (zero mean and unit variance) Student density with υ (> 2) degrees of freedom,[13] denoted $x \sim t(0, 1, \upsilon)$, and defined by

$$g(x \mid \upsilon) = \frac{\Gamma(\frac{\upsilon+1}{2})}{\sqrt{\pi(\upsilon - 2)}\Gamma(\frac{\upsilon}{2})}[1 + \frac{x^2}{\upsilon - 2}]^{-(\upsilon+1)/2} \tag{5.4.2}$$

where $\Gamma(.)$ is Euler's gamma function.

[13] The number of degrees of freedom is restricted to be larger than 2, since it is wanted to construct a distribution with finite variance.

In (5.4.1) the constants $m = m(\xi, \upsilon)$ and $s = \sqrt{s^2(\xi, \upsilon)}$ are respectively the mean and the standard deviation of the non-standarized skewed-t density skewed-t(m, s^2, ξ, υ), and are defined as follows:

$$m(\xi, \upsilon) = \frac{\Gamma(\frac{\upsilon-1}{2})(\sqrt{\upsilon-2})}{\sqrt{\pi}\Gamma(\frac{\upsilon}{2})}(\xi - \frac{1}{\xi}) \tag{5.4.3}$$

and

$$s^2(\xi, \upsilon) = (\xi^2 + \frac{1}{\xi^2} - 1) - m^2. \tag{5.4.4}$$

It can be shown that in (5.4.1), ξ^2 is equal to the ratio of the masses above and below the mode:

$$\xi^2 = \frac{Pr(\epsilon \geq 0|\xi)}{Pr(\epsilon < 0|\xi)} \tag{5.4.5}$$

This makes the use of this density very attractive because ξ^2 can be interpreted as a skewness measure. Note that the density $f(\epsilon \mid 1/\xi, \upsilon)$ is the symmetric of $f(\epsilon \mid \xi, \upsilon)$ respect to the zero mean, i.e. $f(\epsilon \mid 1/\xi, \upsilon) = f(-\epsilon \mid \xi, \upsilon)$. Therefore, the sign of log ξ indicates the direction of the skewness: the third moment is positive (negative), and the density is skewed to the right (left), if log $\xi > 0$ (< 0).

Lambert and Laurent (2001b) showed that the quantile function skewed-t(α, υ, ξ) of a non standarized skewed-t density is:

$$skewed - t^*(\alpha, \upsilon, \xi) = \begin{cases} \frac{1}{\xi}t(\alpha, \upsilon)[\frac{\alpha}{2}(1+\xi^2)] & \text{if } \alpha < \frac{1}{1+\xi^2} \\ -\xi t(\alpha, \upsilon)[\frac{1-\alpha}{2}(1+\xi^{-2})] & \text{if } \alpha \geq \frac{1}{1+\xi^2}. \end{cases} \tag{5.4.6}$$

where $t(\alpha, \upsilon)$ is the quantile function of the (zero mean, unit variance) Student-t density. To obtain the quantile function of the standarized skewed-t we use:

$$skewed - t(\alpha, \upsilon, \xi) = \frac{skewed - t^*(\alpha, \upsilon, \xi) - m}{s}. \tag{5.4.7}$$

Efficient estimation of the model defined by equations (5.3.1)-(5.3.3) under the assumption that ϵ_t is IID skewed-t$(0, 1, \xi, \upsilon)$ is performed by maximizing the log-likelihood function $L_T(\theta) = \sum_{t=1}^{T} l_t(\theta)$ where $\theta = (\mu', \eta', \xi, \upsilon)'$ denotes the vector of parameters, with $\mu = (\phi_1, \ldots, \phi_p)$ and $\eta = (\omega, \alpha_1, \alpha_n, \beta_1, \delta)$ in the case of the APARCH model of equation (5.3.3) and,

$$\begin{aligned} l_t(\theta) \quad = \quad & log(\frac{2}{\xi + \frac{1}{\xi}}) + log\Gamma(\frac{\upsilon+1}{2}) - 0.5log[\pi(\upsilon-2)] - log\Gamma(\frac{\upsilon}{2}) + log\frac{s}{h_t} \\ & -0.5(1+\upsilon)log[1 + \frac{(s\epsilon_t + m)^2\xi^{-2I_t}}{\upsilon - 2}] \end{aligned} \tag{5.4.8}$$

with $\epsilon_t = (r_t - \mu_t)/h_t$ where $\mu_t = \phi_0 + \phi_1 r_{t-1} + ... + \phi_p r_{t-p}$ and,

$$I_t = \begin{cases} 1 & \text{if } \epsilon_t \geq -\frac{m}{s} \\ 0 & \text{if } \epsilon_t < -\frac{m}{s}. \end{cases}$$

In equation (5.4.8), μ_t, h_t, m and s are functions of the parameters defined by equations (5.3.1), (5.3.3), (5.4.3) and (5.4.4), respectively.

Following Ding, Granger, and Engle (1993) and Paolella (1997), if it exists, a stationary solution of (5.3.3) is given by:

$$E(h^\delta) = \frac{\omega}{1 - \alpha_1 E(|\epsilon| - \alpha_n \epsilon)^\delta - \beta_1} \tag{5.4.9}$$

which depends on the density of ϵ. Such solution exists if $\alpha_1 E(|\epsilon| - \alpha_n \epsilon)^\delta + \beta_1 < 1$ and where

$$E(|\epsilon| - \gamma\epsilon)^\delta = [\xi^{-(1+\delta)}(1+\gamma)^\delta + \xi^{(1+\delta)}(1-\gamma)^\delta] \frac{\Gamma(\frac{\delta+1}{2})\Gamma(\frac{\upsilon-\delta}{2})(\upsilon-2)^{\frac{1+\delta}{2}}}{(\xi + \frac{1}{\xi})\sqrt{(\upsilon-2)\pi}\Gamma(\frac{\upsilon}{2})}. \tag{5.4.10}$$

Chapter 6

Dynamic optimal portfolio selection in a VaR framework

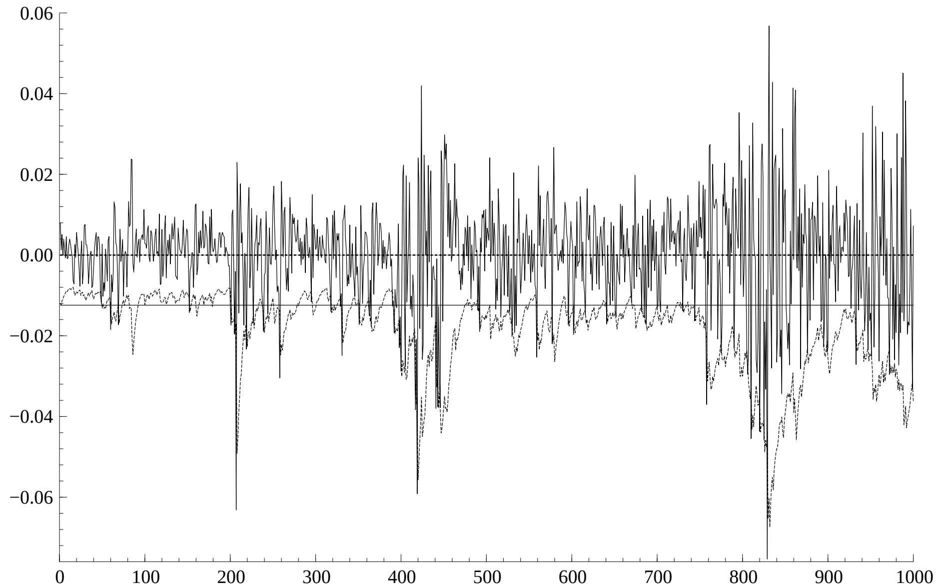
This chapter reports on research done jointly with Jeroen Rombuts.

6.1 Introduction

One important venue of portfolio allocation research started with Markowitz (1952). According to the mean-variance model, investors maximize the expected return for a given risk level, where risk is measured by the variance. In this framework Fleming, Kirby, and Ostdiek (2001) study the economic value of volatility timing and de Roon, Nijman, and Werker (2003) show its usefulness in currency hedging for international stock portfolios. Recently, models have been proposed where the variance is replaced by another risk measure, the Value-at-Risk (VaR) being one of them. The VaR is defined as the worst expected financial loss on an investment over a specified horizon given some confidence level, see Jorion (1997) for more information. Campbell, Huisman, and Koedijk (2001) propose a model which allocates financial assets by maximizing the expected return subject to the constraint that the expected maximum loss should meet the VaR limits. Their model is applied in a static context to find optimal weights between stocks and bonds for a past period. In this context the VaR is estimated by computing the quantiles from parametric distributions or by non parametric procedures such as empirical quantiles or smoothing techniques. See Gouriéroux, Laurent, and Scaillet (2000) for an example of the latter techniques.

Contrary to many papers that evaluate statistically the accuracy of the VaR estimation for individual assets (see for example Mittnik and Paolella (2000), Giot and Laurent (2003) and Giot and Laurent (2004)), this chapter proposes to generalize the work of Campbell, Huisman, and Koedijk (2001), CHK hereafter, to a flexible forward looking dynamic portfolio selection model. This model combines assets in order to maximize the portfolio expected return subject to a VaR risk constraint, allowing to give future investment recommendations. We determine, from both a statistical and economic point of view, the best daily investment recommendations in terms of percentage to borrow or lend and the optimal weights of the assets in the risky portfolio. For the estimation of the VaR we use ARCH-type models and we investigate the importance of several parametric innovation distributions.

Figure 6.1 shows the importance of estimating the 95% level-VaR dynamically in an out-of-sample, or forward looking, context using Russell2000 index return data (see Section 6.4.3 for more details), a GARCH(1,1) model and a skewed-t innovation distribution. The static case assumes that it is necessary to estimate the VaR only at the beginning of the investment period. The dynamic approach assumes that the VaR should be re-estimated every day in order to capture the changes in the moments of the returns' distributions. In the dynamic case, the failure rate is 6.7% and in the static case it is 13.5%, i.e. in the last case the realized confidence level is more than twice the desired one. Thus, from a risk management point of view it could pay off to shift from the static to the dynamic framework.



Returns, constant and dynamic VaR of Russell2000, during the out-of-sample period. The dynamic VaRs are estimated using the GARCH(1,1) model with the skewed-t innovation distribution. The confidence level is 95%. The constant VaR is equal to −0.012. Out-of-sample period from 02/01/1997 until 20/12/2000 (1000 days).

Figure 6.1: **Returns, constant and dynamic VaR**

Guermat and Harris (2002) working with three equity return series find more accurate VaR forecasts using a model that allows for time variation not only in the variance but also in the kurtosis of the return distribution. Jondeau and Rockinger (2003), investigating the time-series behavior of five stock indices and of six foreign exchange rates, find time dependence of the asymmetry parameter and generally a constant degrees of freedom parameter. Patton (2004) in the context of asset allocation studies the skewness in the distribution of individual stocks and the asymmetry in the dependence between stocks. Our approach, apart from time variation in the variance, also allows for an evolution of the skewness and kurtosis of the portfolio distributions. This is done by estimating the model parameters by Weighted Maximum Likelihood (WML) in a increasing window setup.

For two datasets, one consisting of indices and another of stocks, we perform out-of-sample forecasts applying our dynamic portfolio selection model to determine the daily optimal portfolio allocations. We work with two stock indices and two individual stocks and not with bonds indices in order to capture the asymmetric dependence documented only for stock returns, see Patton (2004). The dynamic model we propose outperforms the CHK model in terms of failure rates, defined as the number of times the desired confidence level used for the estimation of the VaR is violated and in terms of the dynamic quantile test of Engle and Manganelli (2004), used to determine the quality of our results.[1] Based on these statistical criteria, the APARCH model gives as good results as the GARCH model. However, if we consider not only the failure rates and the dynamic quantile test but also economic criteria like the achieved wealth and the risk-adjusted returns, we find that the APARCH model outperforms the GARCH model. A sensitivity analysis with respect to the distribution innovation shows that the skewed-t is preferred to the normal and Student-t.

The plan of the paper is as follows: In Section 6.2 we present the optimal portfolio selection in a VaR framework. In Section 6.3, we describe different model specifications for the estimation of the VaR. Section 6.4 presents two empirical applications using out-of-sample forecasts to determine the optimal investment strategies. We use portfolios formed by either two US indices (SP500-RUSELL2000) or by two stocks (Colgate-IBM). We compare the performance of the different models using the failure rates, the dynamic quantile test, the wealth achieved and the risk-adjusted returns as instruments to determine the best model. Section 6.5 evaluates several related aspects of the models and Section 6.6 concludes and provides an outlook for future research.

6.2 Optimal portfolio selection

This section follows Campbell, Huisman, and Koedijk (2001). The portfolio model allocates financial assets by maximizing the expected return subject to a risk constraint, where risk is measured by a Value-at-Risk (VaR). The optimal portfolio is such that the maximum expected loss should not exceed the VaR for a chosen investment horizon at a given confidence level α. We consider the possibility of borrowing and lending at the market interest rate, considered as given.

Denote by W_t the investor's wealth at time t, b_t the amount of money that can be borrowed ($b_t > 0$) or lent ($b_t < 0$) at the risk free rate r_f. Consider n financial assets with prices at time t given by $p_{i,t}$, with $i = 1, \ldots, n$. Define $\Omega_t \equiv [w_t \in R^n : \sum_{i=1}^n w_{i,t} = 1]$ as the set of portfolios weights at time t, with well-defined expected rates of return, such that $x_{i,t} = w_{i,t}(W_t + b_t)/p_{i,t}$ is the number of shares of asset i at time t. The budget constraint of the investor is given by:

$$W_t + b_t = \sum_{i=1}^{n} x_{i,t} p_{i,t} = x_t' p_t. \qquad (6.2.1)$$

The value of the portfolio at $t+1$ is:

[1] Recently, Christoffersen and Pellettier (2004) have proposed a duration based approach to back test the VaR.

$$W_{t+1}(w_t) = (W_t + b_t)(1 + R_{t+1}(w_t)) - b_t(1 + r_f), \tag{6.2.2}$$

where $R_{t+1}(w_t)$ is the portfolio return at maturity. The VaR of the portfolio is defined as the maximum expected loss over a given investment horizon and for a given confidence level α:

$$P_t[W_{t+1}(w_t) \leq W_t - VaR^*] \leq 1 - \alpha, \tag{6.2.3}$$

where P_t is the probability conditioned on the available information at time t and VaR^* is the cutoff return or the investor's desired VaR level. Note that $(1 - \alpha)$ is the probability of occurrence. Equation (6.2.3) represents the second constraint that the investor has to take into account. The portfolio optimization problem can be expressed in terms of the maximization of the expected returns $E_t W_{t+1}(w_t)$, subject to the budget restriction and the VaR-constraint:

$$w_t^* \equiv \arg\max_{w_t} \, (W_t + b_t)(1 + E_t R_{t+1}(w_t)) - b_t(1 + r_f), \tag{6.2.4}$$

s.t. (6.2.1) and (6.2.3). $E_t R_{t+1}(w_t)$ represents the expected return of the portfolio given the information at time t. The optimization problem may be rewritten in an unconstrained way. To do so, replacing (6.2.1) in (6.2.2) and taking expectations yields:

$$E_t W_{t+1}(w_t) = x_t' p_t (E_t R_{t+1}(w_t) - r_f) + W_t(1 + r_f). \tag{6.2.5}$$

Equation (6.2.5) shows that a risk-averse investor wants to invest a fraction of his wealth in risky assets if the expected return of the portfolio is bigger than the risk free rate. Substituting (6.2.5) (before taking the E_t) in (6.2.3) gives:

$$P_t[x_t' p_t (R_{t+1}(w_t) - r_f) + W_t(1 + r_f) \leq W_t - VaR^*] \leq 1 - \alpha, \tag{6.2.6}$$

so that,

$$P_t \left[R_{t+1}(w_t) \leq r_f - \frac{VaR^* + W_t r_f}{x_t' p_t} \right] \leq 1 - \alpha, \tag{6.2.7}$$

defines the quantile $q(w_t, \alpha)$ of the distribution of the return of the portfolio at a given confidence level α or probability of occurrence of $(1 - \alpha)$. Then, the budget constraint can be expressed as:

$$x_t' p_t = \frac{VaR^* + W_t r_f}{r_f - q(w_t, \alpha)}. \tag{6.2.8}$$

Finally, substituting (6.2.8) in (6.2.5) and dividing by the initial wealth W_t we obtain:

$$\frac{E_t(W_{t+1}(w_t))}{W_t} = \frac{VaR^* + W_t r_f}{W_t r_f - W_t q(w_t, \alpha)} (E_t R_{t+1}(w_t) - r_f) + (1 + r_f), \tag{6.2.9}$$

and therefore,

$$w_t^* \equiv \arg\max_{w_t} \frac{E_t R_{t+1}(w_t) - r_f}{W_t r_f - W_t q(w_t, \alpha)}. \quad (6.2.10)$$

The two fund separation theorem applies, i.e. the investor's initial wealth and desired $VaR = W_t q(w_t, \alpha)$ do not affect the maximization procedure. As in traditional portfolio theory, investors first allocate the risky assets and second the amount of borrowing and lending. The latter reflects by how much the VaR of the portfolio differs according to the investors' degree of risk aversion measured by the selected VaR level. The amount of money that the investor wants to borrow or lend is found by replacing (6.2.1) in (6.2.8):

$$b_t = \frac{VaR^* + W_t q(w_t^*, \alpha)}{r_f - q(w_t^*, \alpha)}. \quad (6.2.11)$$

In order to solve the optimization problem (6.2.10) over a large investment horizon T, we partition this in one-period optimizations, i.e. if T equals 30 days, we optimize 30 times one-day periods to achieve the desired final horizon.

We illustrate the method by a simple hypothetical example with $n = 2$, an initial investor's wealth of US$ 10000 and an annual risk free rate equal to 1.24%. Let us assume that the optimal portfolio weights that maximize Equation (6.2.10) are those presented in Table 6.1. We also assume that the quantile $q(w_t, \alpha)$ corresponds to a non-normal innovation distribution in order to show the fact noted by Campbell, Huisman, and Koedijk (2001): the portfolio VaR in absolute value increases when the confidence level increases. However, the portfolio weights are non-monotonic functions of the confidence level, unless the normal distribution is used. Table 6.1 presents these hypothetical results.

Table 6.1: **Optimal portfolio selection under VaR.**

α(%)	Asset1(%)	Asset2(%)	$Portfolio\ VaR$($)
90	30	70	-5.0
94	35	65	-5.6
95	40	60	-6.5
97	30	70	-7.5
99	25	75	-8.5

Next, using Equation (6.2.11) we determine the amount of money to borrow or lend. First, assume that the desired VaR^* is equal to 6.5 (that corresponds to the 95% confidence level) and that we have two kinds of investors. One who is less risk averse (Investor 1) and chooses a confidence level of 90% and the other who is more risk averse (Investor 2) and chooses a confidence level of 99%. Table 6.2 presents the decisions based on their particular types.

Table 6.2: **Investment decision of different type of investors.**

Type of Investor	b(%)	Asset1(%)	Asset2(%)	Tot-portfolio
Less risk averse	28.08	38.42	89.66	128.08
More risk averse	-22.62	19.35	58.03	77.38

We observe that Investor 1 borrows ($b > 0$) at the risk-free rate an amount equivalent to 28.08% of his initial wealth investing everything (128.08%) in the portfolio made of the two assets. Investor 2 prefers to lend ($b < 0$) 22.62% of his wealth at the risk-free rate, investing the difference (77.38%) in the risky portfolio.

6.3 Methodology

We observe the following steps in the estimation of the optimal portfolio allocation and its evaluation:

1. Portfolio construction:

   We construct portfolios for different weights and form univariate time series on which we base the rest of the methodology.

2. Estimation of portfolio returns:

   A typical model of the portfolio return R_t may be written as follows:

   $$R_t = \mu_t + \epsilon_t, \tag{6.3.1}$$

   where μ_t is the conditional mean and ϵ_t an error term. As mentioned for example by Merton (1980) and Fleming, Kirby, and Ostdiek (2001), forecasting returns is more difficult than forecasting of variances and covariances.

   In the empirical application we forecast the expected return by the unconditional mean using observations until day $t-1$. We also modelled the expected return by autoregressive processes, but the results were not satisfactory, neither in terms of failure rates nor in terms of wealth evolution.

3. Estimation of the conditional variance:

   The error term ϵ_t in equation (6.3.1) can be decomposed as $\sigma_t z_t$ where z_t is an IID innovation with mean zero and variance 1. We distinguish three different specifications for the conditional variance σ_t^2:

   - The CHK model, similar to the model presented in Section 6.2, where σ_t^2 is estimated as the empirical variance using data until $t - 1$. In fact, this can be interpreted as a straightforward dynamic extension of the application presented in Campbell, Huisman, and Koedijk (2001).
   - The GARCH(1,1) model of Bollerslev (1986), where

     $$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2.$$

   - The APARCH(1,1) model of Ding, Granger, and Engle (1993), where

$$\sigma_t^\delta = \omega + \alpha_1(|\epsilon_{t-1}| - \alpha_n \epsilon_{t-1})^\delta + \beta_1 \sigma_{t-1}^\delta.$$

with ω, α_1, α_n, β_1 and δ parameters to be estimated. The parameter δ ($\delta > 0$) is the Box-Cox transformation of σ_t. The parameter α_n ($-1 < \alpha_n < 1$), reflects the leverage effect such that a positive (negative) value means that the past negative (positive) shocks have a deeper impact on current conditional volatility than the past positive shocks of the same magnitude. Note that if $\delta = 2$ and $\alpha_n = 0$ we get the GARCH(1,1) model.

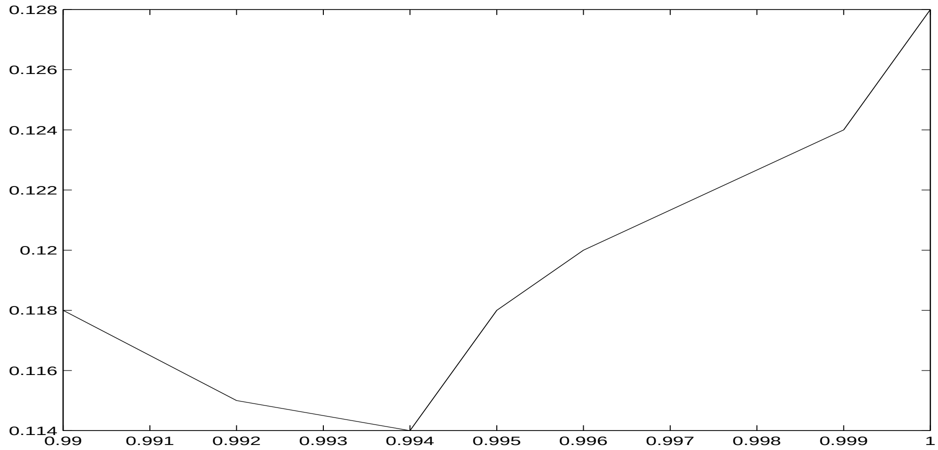
With respect to the innovation distribution, several parametric alternatives are available in the literature. In the empirical application, see Section 6.4, we consider the normal, Student-t and skewed-t distributions. The skewed-t distribution was proposed by Hansen (1994) and reparameterized in terms of the mean and the variance by Lambert and Laurent (2001a) in such a way that the innovation process has zero mean and unit variance. The skewed-t distribution depends on two parameters, one for the thickness of tails (degrees of freedom) and the other for to the skewness.

Following Mittnik and Paolella (2000) the parameters of the models are estimated by Weighted Maximum Likelihood (WML). We use weights which multiply the log-likelihood contributions of the returns in period t, $t = 1, \ldots, T$. This allows to give more weight to recent data in order to obtain parameter estimates that reflect the "current" value of the "true" parameter. The weights are defined by:

$$\omega_t = \rho^{T-t}. \tag{6.3.2}$$

If $\rho < 1$ more weight is given to recent observations than those far in the past. The case $\rho = 1$ corresponds to usual maximum likelihood estimation. The decay factor ρ is obtained by minimizing the failure rate (defined later in this section) for a given confidence level. Figure 6.2 illustrates the failure rate-ρ relationship for portfolios made of Russell2000 and SP500 indices for an investor using VaR at 90% level. The model used is the GARCH(1,1) with normal innovation distribution. The optimal ρ that minimizes the failure rate is equal to 0.994. We find similar results for other cases. Moreover, the value of the optimal ρ is robust to different innovation distributions. We use WML in an increasing window setup, i.e. the number of observations of the sample increases through time in order to consider the new information available. The improvement, in terms of better approximation to the desired confidence levels, using WML in an increasing window setup instead of ML is of the order of 10%. See also Section 6.5 for more details.

By using WML in a increasing window setup, $q_{1-\alpha}$ in (6.3.3) takes into account the time evolution of the degrees of freedom and asymmetry parameters when we use the skewed-t distribution. We do not specify an autoregressive structure for the degrees of freedom and for the asymmetry parameter like Jondeau and Rockinger (2003). They find that



Failure rates (vertical axis) obtained with different ρ values (horizontal axis) using the geometric weighting scheme for a 1000 out-of-sample period. Portfolios made of Russell2000 and SP500 indices for an investor with VaR-90. The model used is the GARCH with normal innovation distribution Out-of-sample period from 02/01/1997 until 20/12/2000 (1000 days).

Figure 6.2: **Failure rates-ρ relationship**

this approach is subject to numerical instabilities.

4. Estimation of the VaR:

   The VaR is a quantile of the distribution of the return of a portfolio, see Equations (6.2.3) and (6.2.7). In an unconditional setup the VaR of a portfolio may be estimated by the quantile of the empirical distribution at a given confidence level α. In parametric models, such as the ones we are using, the quantiles are functions of the variance of the portfolio return R_t. The $VaR_{t,\alpha}$ (VaR for time t at the confidence level α) is calculated as:

$$VaR_{t,\alpha} = \hat{\mu}_t + \hat{h}_t q_{1-\alpha}, \tag{6.3.3}$$

   where $\hat{\mu}_t$ and $\hat{h}_t$ are the forecasted conditional mean and variance using data until $t-1$ and, $q_{1-\alpha}$ is the $(1-\alpha)$-th quantile of the innovation distribution.

5. Determine the optimal risky portfolio allocation:

   Once we have determined the VaR for each of the risky portfolios, we solve equation (6.2.10) to find the optimal portfolio weights. These weights correspond to the portfolio that maximizes the expected returns subject to the VaR constraint.

6. Determine the optimal amount to borrow or lend:

   As shown in section 6.2, the two fund separation theorem applies. Then, in order to determine the amount of money to borrow or lend, we simply use equation (6.2.11).

7. Evaluate the models:

   A criterion to evaluate the models is the failure rate:

$$f = \frac{1}{n} \sum_{t=T-n+1}^{T} \mathbf{1}_{[R_t < -VaR_{t-1,\alpha}]}, \tag{6.3.4}$$

where, n is the number of out-of-sample days, T is the total number of observations, R_t is the observed return at day t, $VaR_{t-1,\alpha}$ is the threshold value determined at time $t-1$ and $\mathbf{1}$ is the indicator function. A model is correctly specified if, the observed return is bigger than the threshold values in 100α percent of the forecasts. One can perform a likelihood ratio test to compare the failure rate with the desired VaR level, as proposed by Kupiec (1995). We will call this test the Kupiec-LR test in the rest of the paper.

Another statistical test that we use is the dynamic quantile test proposed by Engle and Manganelli (2004). According to this test and in order to determine the quality of the results, a property that a VaR measure should have besides respecting the level is that the VaR violations (hits) should not be serially correlated. This can be tested by defining the following sequence:

$$h_t = \mathbf{1}[R_t < -VaR_{t,\alpha}] - \alpha, \tag{6.3.5}$$

such that the expected value of this sequence is zero. The dynamic quantile test is computed as an F-test under the null that all coefficients, including the intercept, are zero in a regression of the variable h_t on its own past, on current VaR and on any other regressors. In our case, we perform the test using the current VaR and 5 lags of the VaR violations as explanatory variables.

We also evaluate the models by analyzing the wealth evolution generated by the application of the portfolio recommendations of the different models. With this economic criterion, the best model will be the one that reports the highest wealth for similar risk levels. Finally, we evaluate the models by comparing the risk-adjusted returns using equation (6.2.10), where the expected return is changed by the realized return. With this test we can compare the risk premium adjusted by the risk, measured by the VaR.

6.4 Empirical results

We develop two applications of the model presented in the previous sections. We construct 1000 daily out-of-sample portfolio allocations based on conditional variance forecasts of GARCH and APARCH models and compare the results with the ones obtained with the CHK model. The parameters are estimated using WML in a rolling window setup. Moreover, we use the normal, Student-t and skewed-t distributions to investigate the importance of the choice of several innovation densities for different confidence levels. Each of the three models can be combined with the three innovation distributions resulting in nine different specifications. In the applications we consider an agent's problem of allocating his wealth (set to 1000 US dollars) among two different American indices and two stocks, Russell2000-SP500 and Colgate-IBM respectively. For the riskfree rate we use the one-year Treasury bill rate in January 1998 (approximately 4.47% annual). We have considered only the trading days in which both indices or stocks where traded. We define the daily returns as log price differences from the adjusted closing price series .

With the information until time t, the models forecast one-day ahead the percentage of the cumulated wealth that should be borrowed ($b_t > 0$) or lent ($b_t < 0$) according to the agent's risk aversion expressed by his confidence level α, and the percentage that should be invested in the portfolio made of the two indices or the two stocks. The models give the optimal weights of each of the indices or stocks in the optimal risky portfolio. Then, with the investment recommendations of the previous day, we use the real returns and determine the agent's wealth evolution according to each model suggestions. Since the parameters of the GARCH and APARCH models change slowly from one day to another, these parameters are re-estimated every 10 days to take into account the expanding information and to keep the computation time low. We also re-estimate the parameters daily, every 5, 15 and 20 days (results not shown). We find similar results in terms of the parameter estimates. However, in the case of daily and 5-day re-estimation, the computational time was about 10-times bigger.

For the estimation of the programs we use a Pentium Xeon 2.6 Ghz. The time required for the GARCH and APARCH models is 90 and 120 minutes on average, respectively. Estimating the models with a fixed window requires 60 and 90 minutes on average to run the GARCH and APARCH models respectively.

In the next section we present the statistical characteristics of the data. Then, we present generally how the models work only for two specific examples due to space limitations. Finally, we present the key results for all the models in terms of failure rates, the dynamic quantile test, the total achieved wealth and the risk-adjusted returns and stress the models' differences.

6.4.1 Data

SP500 - Russell2000

We use daily data of the SP500 composite index (large stocks) and the Russell2000 index (small stocks). The sample period goes from 02/01/1990 to 20/12/2000 (2770 observations). Descriptive statistics are given in the left panel of Table 6.3. We see that for all indices skewness and excess kurtosis is present and that the means and standard deviations are similar. Figure 6.3 presents the daily returns during the out-of-sample period for both indices.

Note that our one-day ahead forecast horizon is four years (more or less 1000 days). During this period we observe mainly a bull market, except for the last days, when the indices start a sharp fall. The lower panel of Table 6.3 presents the descriptive statistics corresponding to the out-of-sample period. Note that the volatility in this period is higher than the previous one.

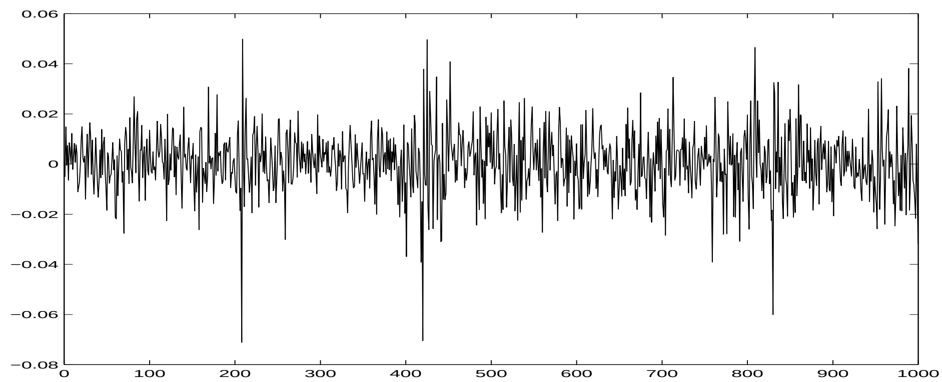
Colgate - IBM

The daily sample period for these two stocks goes from 10/01/1990 to 31/12/2000 (2870 observations). Descriptive statistics are given in the right panel of Table 6.3. Both series present skewness and excess kurtosis. However, Colgate is only slightly positively skewed while IBM is negatively skewed. The excess of kurtosis is higher than in the indices case due to the presence

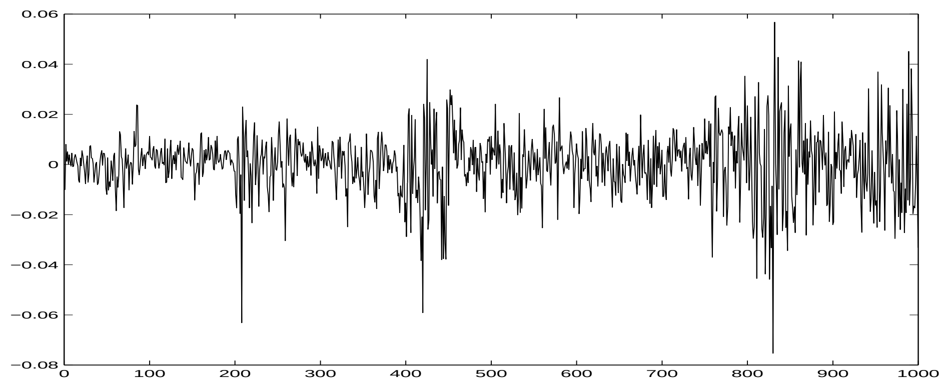
Table 6.3: Descriptive statistics

	02/01/1990 - 20/12/2000 N=2770		10/01/1990 - 31/12/2000 N=2870	
	SP500	**Russell2000**	**Colgate**	**IBM**
Mean	0.045	0.035	0.073	0.045
Standard deviation	0.946	0.937	1.730	2.012
Skewness	-0.293	-0.642	0.012	-0.101
Kurtosis	7.741	9.084	13.108	10.203
Minimum	-7.114	-7.533	-17.329	-16.889
Maximum	4.990	5.678	18.499	12.364
	02/01/1997 - 20/12/2000 N=1000		10/01/1997 - 31/12/2000 N=1000	
	SP500	**Russell2000**	**Colgate**	**IBM**
Mean	0.053	0.020	0.090	0.085
Standard deviation	1.247	1.279	2.311	2.481
Skewness	-0.306	-0.454	0.035	-0.317
Kurtosis	6.059	6.308	10.915	8.648
Minimum	-7.114	-7.533	-17.329	-16.889
Maximum	4.990	5.678	18.499	12.364

Descriptive statistics for the daily returns of the corresponding indices (left panel) and stocks (right panel). The mean, the standard deviation, the minimum and maximum values are expressed in %.



(a) SP500 daily returns. Out-of-sample period from 02/01/1997 until 20/12/2000 (1000 days)

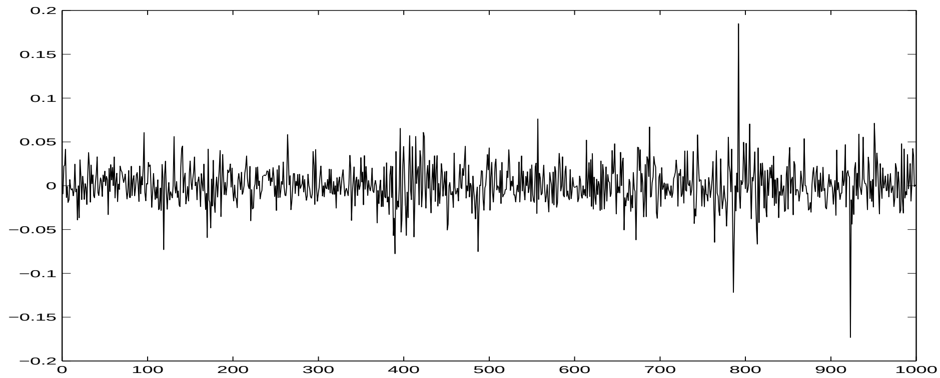


(b) Russell2000 daily returns. Out-of-sample period from 02/01/1997 until 20/12/2000 (1000 days)

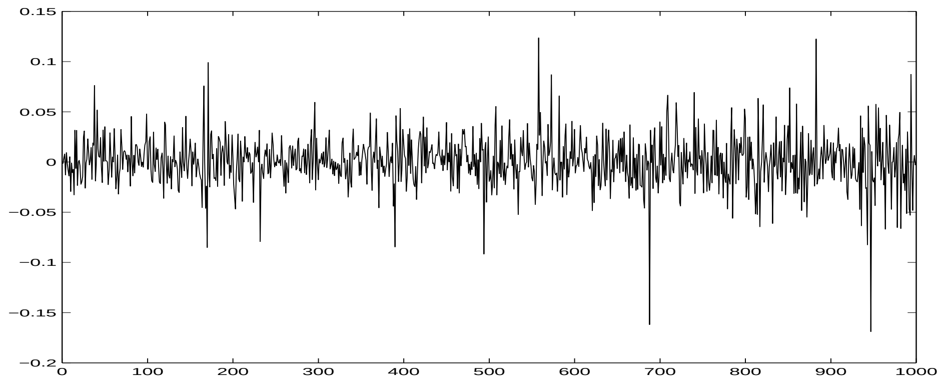
Figure 6.3: **SP500 and Russell2000 out-of-sample returns**

of more extreme returns (either positive or negative), which is a common finding when stocks are used instead of indices. The mean of the Colgate returns is higher than the mean of the IBM returns and interestingly, the standard deviation of Colgate is also smaller. In Figure 6.4 we present the daily returns during the out-of-sample period for both assets.

As observed in the case of the indices, during the forecast period we observe mainly a bull market, except for the last days, where the stock prices start to fall. The right panel of Table 6.3 also presents the descriptive statistics of the out-of-sample period. As noted in the previous case, the volatility in this out-of-sample period is higher than the previous period.



(a) Colgate daily returns. Out-of-sample period from 10/01/1997 until 31/12/2000 (1000 days)



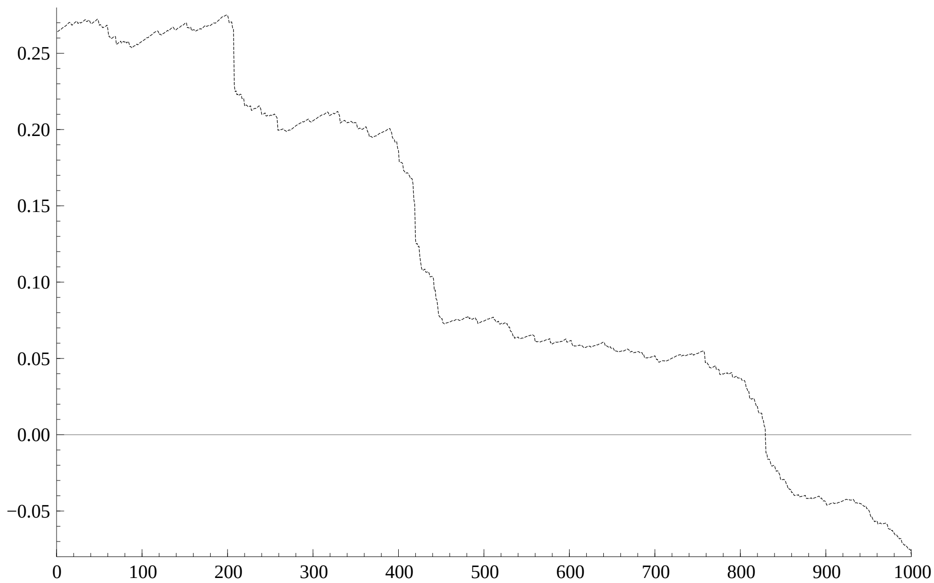
(b) IBM daily returns. Out-of-sample period from 10/01/1997 until 31/12/2000 (1000 days)

Figure 6.4: **Colgate and IBM out-of-sample returns**

6.4.2 A general view of the daily recommendations

We present two examples of model configurations to illustrate the main results. For all the cases the investor's desired VaR_t^* is set to 1% of his cumulated wealth at time $t-1$. First, we explain the investment decisions based on the CHK model using the normal distribution for portfolios made of Russell2000-SP500. The agent desired VaR confidence level is $\alpha = 90\%$, i.e. a less risk-averse investor. Figure 6.5 shows the evolution of the percentage of the total wealth to be borrowed ($b_t > 0$) or lent ($b_t < 0$). In this case the model suggests until day 829 to borrow at the risk-free rate and to invest everything in the risky portfolio. However, after that day the model recommendation is to change from borrowing to lending. This is a natural response to the negative change in the trend of the indices and to the higher volatility observed in the stock market during the last days of the out-of-sample period (Figure 6.3). Figure 6.6 presents the evolution of the share of the risky portfolio to be invested in the Russell2000 index. The model suggests for 807 days to invest 70% of the wealth (on average) in Russell2000 index and the difference in SP500 index. After that day, the model recommendations change drastically favoring the investment in SP500, which increases its portfolio weights to 66%, i.e. going from 30% to almost 50% at the end of the out-of-sample period. Again, this responds to the higher volatility of the Russell2000 compared with the SP500 during the last days. Thus, the model recommend to shift from the more risky index to the less risky one and from the risky portfolio to the risk free investment.

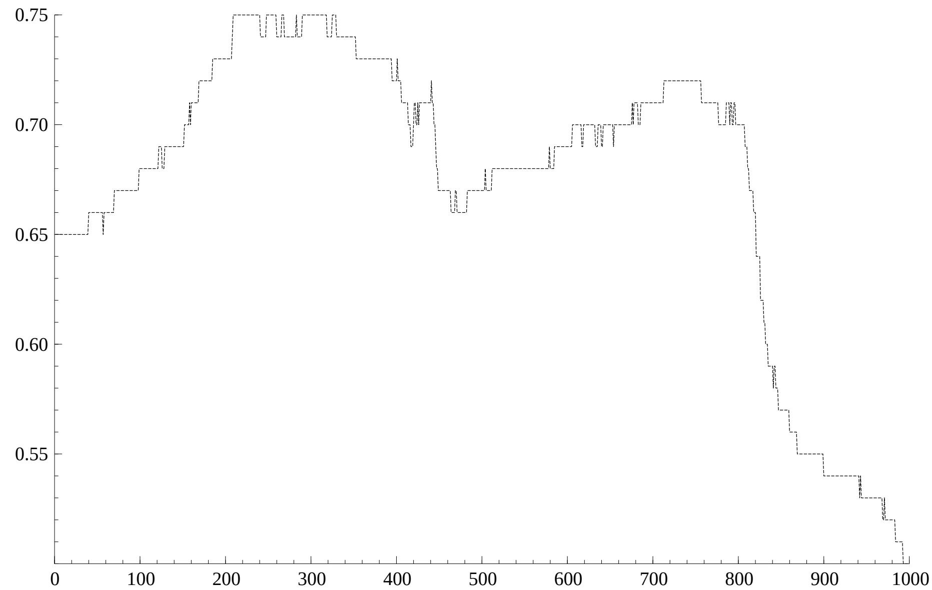
Figure 6.7 compares the wealth evolution obtained by applying the CHK model suggestions with investments made in either one or the other index. The wealth evolution is higher than the one that could be obtained by investing only in Russell2000 but lower if investing only in SP500 during the out-of-sample forecast period. We also include the wealth evolution that an agent can realize when investing everything at the risk-free rate (assumed constant during the whole forecasted period). More details can be found in Section 6.4.3.



Riskfree weights for portfolios made of Russell2000 and SP500 indices for an investor with VaR-90, based on the CHK model using the normal distribution. Out-of-sample period from 02/01/1997 until 20/12/2000 (1000 days).

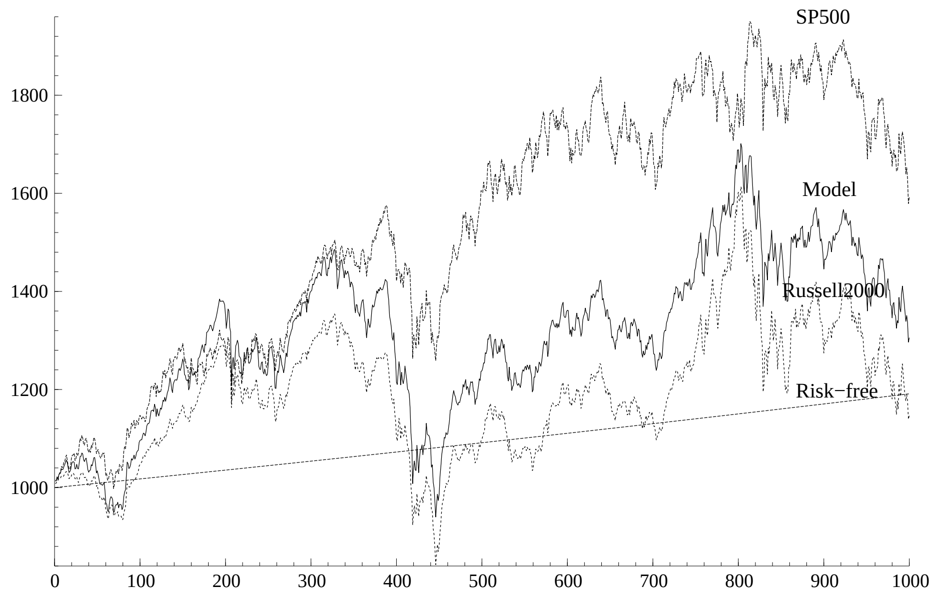
Figure 6.5: **Riskfree weights using CHK model with normal distribution**

As a second example we present the results of applying our dynamic optimal portfolio selection model to the Colgate-IBM data for which the conditional variance is estimated using the APARCH model. The agent's desired VaR confidence level is $\alpha = 99\%$, i.e. a risk-averse investor and the distribution is the skewed-t distribution. In Figure 6.8 we observe how the model accommodates its recommendations to higher risk aversion. The model suggests during the whole forecasted period to lend a big proportion of the wealth at the risk free rate (70% on average) which comes as no surprise given the desired confidence level. Figure 6.9 shows the model recommendations with respect to the weight invested in Colgate. It varies considerably,



Risky weights of Russell2000 for an investor with VaR-90, based on the CHK model using the normal distribution. Out-of-sample period from 02/01/1997 until 20/12/2000 (1000 days).

Figure 6.6: **Risky weights on Russell2000 using CHK model with normal distribution**

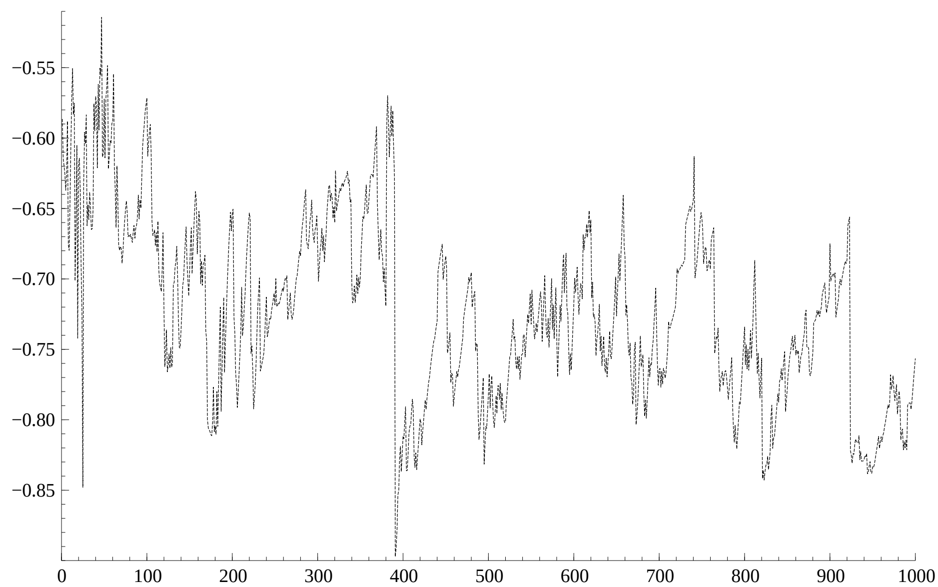


Portfolios made of Russell2000 and SP500 indices for an investor with VaR-90. Wealth evolution for 1000 out-of-sample forecast using the model recommendations (Model) compared with the wealth evolution obtained by investments made on Russell2000 or SP500 alone and with investments done at the risk-free rate. Out-of-sample period from 02/01/1997 until 20/12/2000.

Figure 6.7: **Wealth evolution using CHK model**

showing how the model adjusts its suggestions in order to maximize the expected return subject to the VaR constraint.

Figure 6.10 presents the wealth evolution obtained by applying the model suggestions and compares it with investments in either one or the other stock. An agent that desires a 99% VaR confidence level is a highly risk-averse investor. As a result, the investment decisions are very conservative, since his risk constraint is tight. Even though the returns are lower than the ones obtained by investing in either one of the two stocks, it is higher (during the whole period) than the investment at the risk-free rate.[2]



Riskfree weights for portfolios made of Colgate and IBM for an investor with VaR-99, based on the APARCH model using the skewed-t distribution. Out-of-sample period from 10/01/1997 until 31/12/2000 (1000 days).

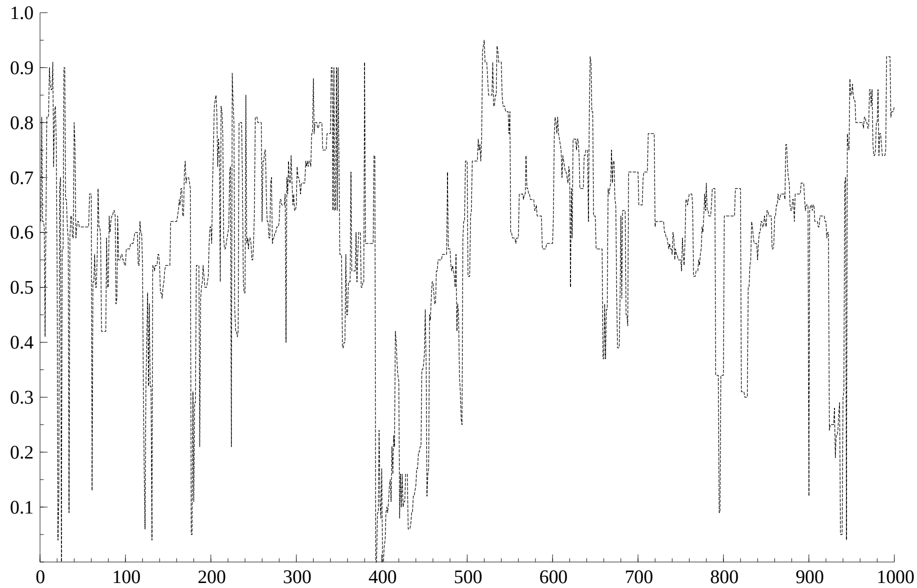
Figure 6.8: **Riskfree weights using APARCH model with skewed-t distribution**

The two previous illustrations show how the model recommendations change according to new information coming to the market, allowing the agent to maximize expected return subject to budget and risk constraints in a dynamic way. The next section presents more synthetically the comparison of all models for different distributional assumptions and different confidence levels.

6.4.3 Results

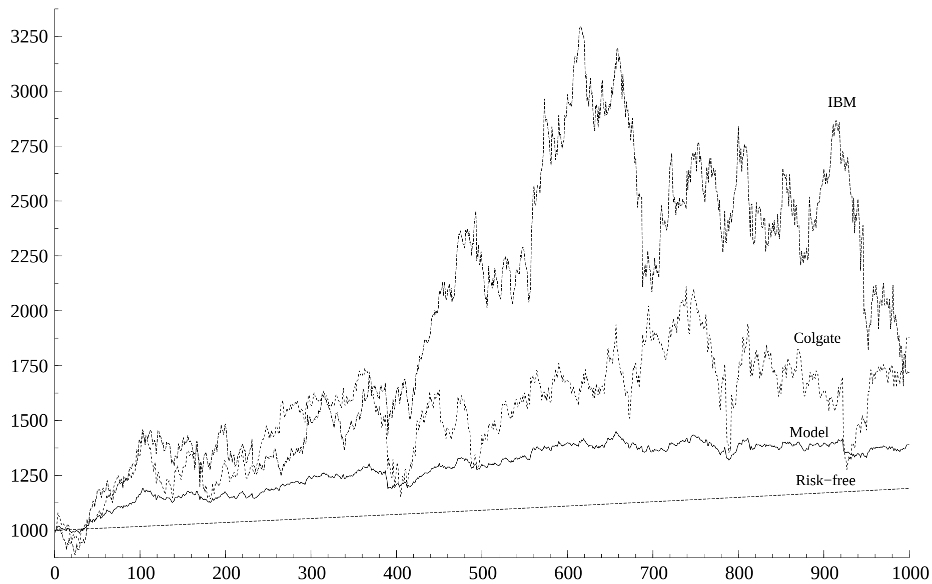
This section presents concisely the results of all the model configurations we used. We compare the three different models explained in Section 3: the CHK model in which the variance is estimated simply from the observed past returns and the parametric dynamic model in which the conditional variance is estimated using either the GARCH or the APARCH model. Moreover, we investigate three different distributional assumptions: the normal, the Student-t and the skewed-t. We consider three VaR confidence levels: 90%, 95% and 99%, corresponding

[2]The same graph for a more risky investor, i.e. with a desired VaR confidence level of 90% for example, shows that the wealth evolution is always higher than the one resulting of investing only in Colgate and sometimes higher than only investing in IBM. Moreover, the final wealth attained with the model recommendations is higher than the final wealth achieved by investing only in IBM.



Risky weights on Colgate for an investor with VaR-99, based on the APARCH model using the skewed-t distribution. Out-of-sample period from 10/01/1997 until 31/12/2000 (1000 days).

Figure 6.9: **Risky weights on Colgate using APARCH model with skewed-t distribution**



Portfolios made of Colgate and IBM for an investor with VaR-99. Wealth evolution for 1000 out-of-sample forecast using the model recommendations (Model) compared with the wealth evolution obtained by investments made on Colgate or IBM alone and with investments made at the risk-free rate. Out-of-sample period from 10/01/1997 until 31/12/2000.

Figure 6.10: **Wealth evolution using APARCH model**

to increasing risk aversion and show how these levels affect the results. The parameters are estimated using WML in a rolling window setup.

From the optimization procedure presented in Section 2, see Equation (6.2.10), we determine the weights of the risky portfolio and, considering the agent's desired risk expressed by the desired VaR (VaR^*), the amount to borrow or lend, see Equation (6.2.11). With this information at time t the investment strategy for day $t+1$ is set: percentage of wealth to borrow or lend and percentage to be invested in the risky portfolio. In order to evaluate the models we consider the wealth evolution of the initial invested amount and the failure rate of the returns obtained by applying the strategies with respect to the desired VaR level. A model is good when the wealth is high and when the failure rate is respected.

We expect that the forecasted VaR's by the different models be less or equal than the threshold values. To test this we perform a likelihood ratio test comparing the failure rate with the desired VaR level, as proposed by Kupiec (1995). We present the Kupiec-LR test for the portfolios made of Russell2000-SP500 (Table 6.4) and of Colgate-IBM (Table 6.6), for the probabilities of occurrence of $1-\alpha = 10\%$ (upper panel), 5% (middle panel) and 1% (lower panel). Several failure rates are significantly different from their nominal levels when we do out-of-sample forecasts. For in-sample forecast (results not presented) we found p-values as high as those presented by Giot and Laurent (2004) for example. This is understandable since the information set, on which we condition, contains only past observations so that the failure rates tend to be significantly different from their nominal levels. However, these failures rates are not completely out of scope of the desired confidence level, see for example Mittnik and Paolella (2000) for similar results.

Moreover, we estimate the dynamic quantile test proposed by Engle and Manganelli (2004) in order to determine the quality of our results. With this test we evaluate the property that any VaR measure should have besides respecting the level: the VaR violations should not be serially correlated. We perform the test using the current VaR and 5 lags of the VaR violations as explanatory variables. The results (not presented here) suggest that for all the portfolios and for all the dynamic models this test is satisfied, meaning that the VaR violations are not serially correlated.

Table 6.4 presents the failure rates and p-values for the Kupiec LR ratio test for portfolios made of Russell2000 and SP500. In general we observe that the dynamic model performs considerably better than its CHK counterpart for any VaR confidence level α. In terms of the distributional assumption we see that in the case of the probability of occurrence of $1-\alpha = 10\%$ the normal distribution performs better than the Student-t even for low degrees of freedom (7 on average). This happens because the two densities cross each other at more or less that confidence level. See Guermat and Harris (2002) for similar results. Looking at lower probabilities of occurrence (higher confidence levels), one remarks that the skewed-t distribution performs better than the other two distributions. This is due to the fact that the skewed-t distribution allows not only for fatter tails but it can also capture the asymmetry present in the long and short sides of the market. This result is consistent with the findings of Mittnik and Paolella (2000), Giot and Laurent (2003) and Giot and Laurent (2004) who used single indices, stocks, exchange rates or a portfolio with unique weights.

With respect to the conditional variance models, we observe that for all the confidence levels, the APARCH model performs almost as good as the GARCH model. However, considering that an agent wants to maximize his expected return subject to a risk constraint, we look after good results for the portfolio optimization (in terms of the final wealth achieved, presented in Table 6.4 in terms of the annualized returns), respecting the desired VaR confidence level (measured by the failure rate). We can appreciate the following facts: first, it happens that the annualized rate of return obtained by the static model not only is lower than those attained by the dynamic models but also, as pointed out before, has a higher risk. Second, even though we cannot select a best model between the APARCH and GARCH models in terms of failure rates, we can see that for almost the same level of risk the APARCH model investment recommendations allow the agent to get the highest annualized rate of return. Therefore, we infer that the APARCH model outperforms the GARCH model. Thus, if an investor is less risk averse ($1 - \alpha = 10\%$) he could have earned an annualized rate return of 9.5%, two times bigger than simple investing at the risk-free rate.

We compute finally the evolution of the risk-adjusted returns. On average, 90% of the time the risk-adjusted returns obtained by the APARCH model outperforms the ones obtained using the GARCH model. With this test, we can confirm that indeed the APARCH model outperforms the GARCH model.

Table 6.4: Failure rates for portfolios made of Russell2000 - SP500

$1-\alpha$	Model	Normal	p	Student-t	p	Skewed-t	p
0,10	**CHK**	0,177	0,000	0,200	0,000	0,188	0,000
	GARCH	0,114	0,148	0,130	0,002	0,117	0,080
	APARCH	0,126	0,008	0,129	0,003	0,118	0,064
0,05	**CHK**	0,127	0,000	0,135	0,000	0,120	0,000
	GARCH	0,071	0,004	0,074	0,001	0,060	0,159
	APARCH	0,083	0,000	0,081	0,000	0,062	0,093
0,01	**CHK**	0,068	0,000	0,048	0,000	0,032	0,000
	GARCH	0,029	0,000	0,021	0,002	0,011	0,754
	APARCH	0,030	0,000	0,027	0,000	0,012	0,538

Empirical tail probabilities for the out-of-sample forecast for portfolios made of linear combinations of Russell2000 and SP500 indices. The Kupiec-LR test is used to determine the specification of the models. The null hypothesis is that the model is correctly specified, i.e. that the failure rate equal to the probability of occurrence $1 - \alpha$. Results obtained using WML with $\rho = 0.994$.

Tables 6.6 and 6.7 present the results for the Colgate-IBM dataset. Like for the previous dataset, the dynamic models outperform the CHK model in terms of the failure rate. The normal distribution behaves better than the Student-t when the VaR confidence level is set to 90% ($1 - \alpha = 10\%$). In general, we see that the skewed-t distribution outperforms the other distributions. In terms of the failure rate, the APARCH is slightly better than the GARCH but this difference is not striking enough to conclude that the APARCH model outperforms the GARCH model. If we also consider the wealth achieved by the application of the model recommendations (Table 6.7) and the times that the APARCH risk-adjusted returns are bigger

Table 6.5: Final wealth achieved by investing in portfolios made of Russell2000-SP500

$1-\alpha$	Model	Normal	r	Student-t	r	Skewed-t	r
0,10	**CHK**	1306	6,9	1303	6,8	1303	6,8
	GARCH	1355	7,9	1351	7,8	1346	7,7
	APARCH	1586	12,2	1630	13,0	1439	9,5
0,05	**CHK**	1297	6,7	1300	6,8	1296	6,7
	GARCH	1324	7,3	1328	7,3	1317	7,1
	APARCH	1497	10,6	1517	11,0	1368	8,2
0,01	**CHK**	1277	6,3	1270	6,2	1263	6,0
	GARCH	1290	6,6	1296	6,7	1281	6,4
	APARCH	1409	8,9	1388	8,5	1310	7,0

Final wealth achieved by investing in portfolios made of Russell2000-SP500. r is the annual rate of return in (%). The risk-free interest rate is 4.47% annual.

than the GARCH ones (93% in average), we conclude that the APARCH model outperforms the GARCH model.

Table 6.6: Failure rates for portfolios made of Colgate - IBM

$1-\alpha$	Model	Normal	p	Student-t	p	Skewed-t	p
0,10	**CHK**	0,145	0,000	0,175	0,000	0,166	0,000
	GARCH	0,100	1,000	0,112	0,214	0,122	0,024
	APARCH	0,097	0,751	0,115	0,122	0,114	0,148
0,05	**CHK**	0,092	0,000	0,102	0,000	0,085	0,000
	GARCH	0,060	0,159	0,065	0,037	0,066	0,027
	APARCH	0,058	0,257	0,063	0,069	0,064	0,051
0,01	**CHK**	0,037	0,000	0,028	0,000	0,020	0,005
	GARCH	0,024	0,000	0,022	0,001	0,016	0,079
	APARCH	0,025	0,000	0,018	0,022	0,015	0,139

Empirical tail probabilities for the out-of-sample forecast for portfolios made of linear combinations of Colgate and IBM. The Kupiec-LR test is used to determine the specification of the models. The null hypothesis is that the model is correctly specified, i.e. that the failure rate equal to the desired probability of occurrence $1-\alpha$.

Finally,in order to be sure that the final wealth is not just caused by an outlier, we present as examples, the wealth evolution of the portfolios made of Russell2000 - SP500 (Figure 6.11) and Colgate - IBM (Figure 6.12). The distributional assumption used was the skewed-t. The VaR confidence level used in the first case was 90% and in the second case 99%. Figures 6.11 and 6.12 show that the final wealth achieved by the recommendations of the APARCH model is consistently larger than the wealth achieved by the GARCH model suggestions. Thus, even though the estimated wealth is also uncertain, it seems to be consistently better in the APARCH case than in the GARCH case.

6.5 Evaluation

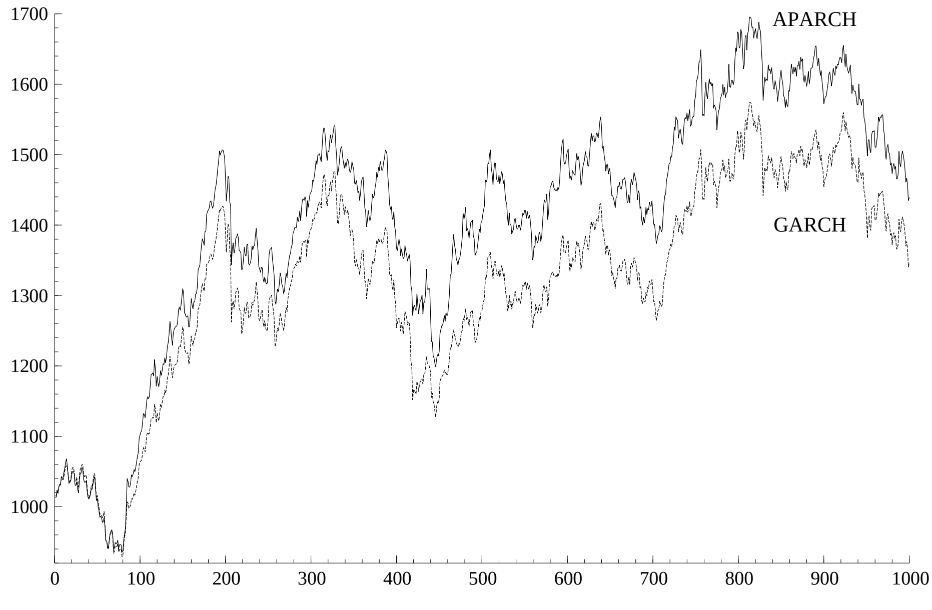
6.5.1 Risk-free interest rate sensitivity

We have used as a risk-free interest rate the one-year Treasury bill rate in January 1998 (approximately 4.47% annual) as an approximation for the average risk-free rate during the whole

Table 6.7: Final wealth achieved by investing in portfolios made of Colgate-IBM

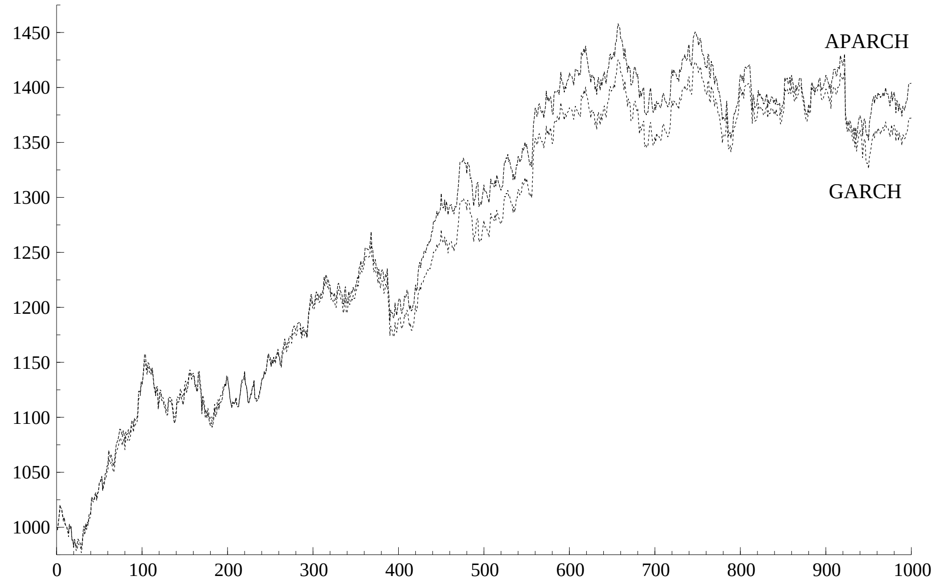
$1-\alpha$	Model	Normal	r	Student-t	r	Skewed-t	r
0,10	**CHK**	1758	15,2	1830	16,3	1799	15,8
	GARCH	1559	11,7	1602	12,5	1624	12,9
	APARCH	1658	13,5	1641	13,2	1691	14,0
0,05	**CHK**	1638	13,1	1661	13,5	1622	12,8
	GARCH	1491	10,5	1476	10,2	1491	10,5
	APARCH	1577	12,1	1521	11,1	1506	10,8
0,01	**CHK**	1506	10,8	1470	10,1	1432	9,4
	GARCH	1415	9,1	1353	7,8	1392	8,6
	APARCH	1496	10,6	1446	9,7	1400	8,8

Final wealth achieved by investing in portfolios made of Colgate-IBM. r is the annual rate of return in (%). The risk-free interest rate is 4.47% annual.



Wealth evolution of portfolios made of Russell2000 - SP500. The distribution used is the skewed-t and the confidence level for the VaR is 90%. Out-of-sample period goes from 02/01/1997 until 20/12/2000.

Figure 6.11: **Compared Wealth evolution using GARCH and APARCH models**



Wealth evolution of portfolios made of Colgate - IBM. The distribution used is the skewed-t and the confidence level for the VaR is 99%. Out-of-sample period goes from 10/01/1997 until 31/12/2000.

Figure 6.12: **Compared Wealth evolution using GARCH and APARCH models**

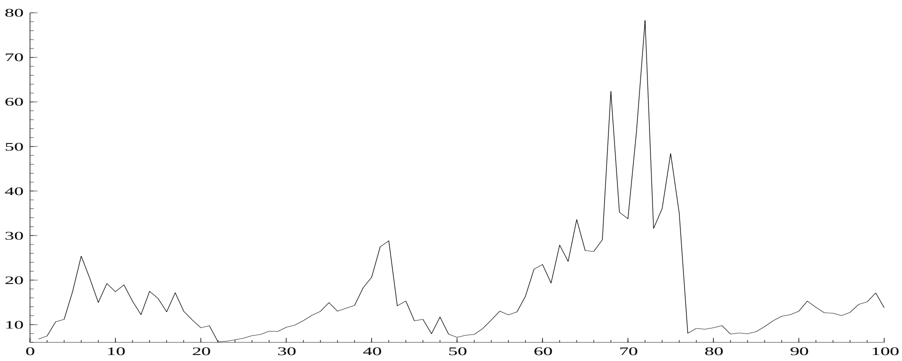
out-of-sample period (January 1997 to December 2000). In order to analyze the sensitivity of our our results to changes of the risk-free rate, we develop four scenarios based on increments (+1% and +4%) or decrements (−1% and −4%) with respect to the benchmark.

The results show that neither the borrowing/lending nor the risky portfolios weights are strongly affected by either of these scenarios. This is due to the fact that we are working with daily optimizations, and that those interest rates at a daily frequency are low. For example 4.47% annual equals 0.01749% daily (based on 250 days).

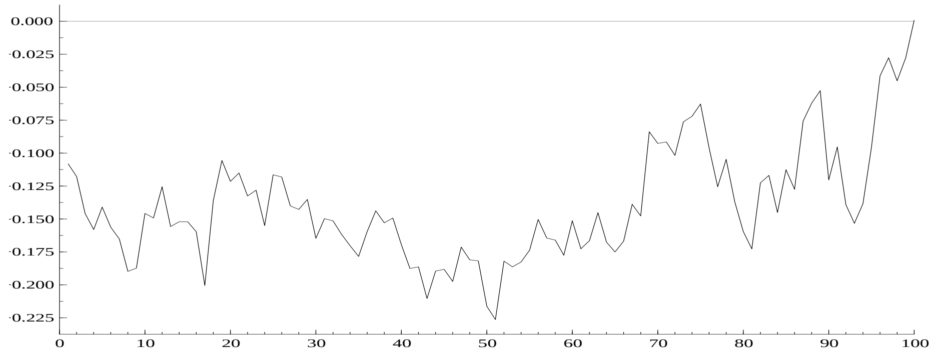
6.5.2 Time varying kurtosis and asymmetry

As in Guermat and Harris (2002), our framework allows for time varying degrees of freedom parameters, related to the kurtosis, when working with either the Student-t or the skewed-t distributions. Moreover, when the skewed-t distribution is used we allow for time varying asymmetry parameters. Figure 6.13 presents the pattern of the degrees of freedom and asymmetry parameter of the skewed-t distribution estimated using the APARCH model. Similarly to Jondeau and Rockinger (2003), we find time dependence of the asymmetry parameter but we also remark that the degrees of freedom parameter is time varying.

We also test the significance of the asymmetry parameter and of the asymmetry and degrees of freedom parameters, for the Student-t and skewed-t respectively. We find that they are highly significant. As an example, Table 6.8 presents the results for the first out-of-sample day for portfolios made of linear combinations of Russell2000 and SP500 using the WML procedure with $\rho = 0.994$. The skewed-t distribution was estimated using the APARCH model. Similar results are observed in the other procedures.



(a) Degrees of freedom



(b) Asymmetry parameter

Time varying degrees of freedom and Asymmetry for the skewed-t innovation distribution estimated using the APARCH model. The parameters are estimated every 10 days during the out-of-sample forecast. The figure corresponds to a portfolio made only of RUSSELL2000.

Figure 6.13: **Time varying degrees of freedom and asymmetry parameters**

6.5.3 Weighted maximum likelihood vs maximum likelihood

We study the effect of using Weighted Maximum Likelihood (WML) instead of Maximum Likelihood (ML). Note that when $\rho = 1$ WML is equal to ML. Table 6.9 presents a comparison of failure rates for portfolios made of Russell2000-SP500. Both dynamic models improve their failure rates by using WML in a rolling window setup instead of ML. In terms of the p-values (not presented) it turns out that when ML is used almost none of the failure rates were significant at any level. Thus, using WML helps to satisfy the investor's desired level of risk.

6.5.4 Rolling window of fixed size

We analyze the effect of using a rolling window of fixed size. The idea behind this procedure is that we assume that information until n days in the past convey some useful information for the prices, meanwhile the rest does not. We use a rolling window of fixed size of $n = 1000$ days for performing the out-of-sample forecasts. The results presented in Tables 6.10 and 6.11 show that the gains in better model specification are nil: the failure rates are worse and the final wealth achieved are almost the same. The computational time decreases (about 30% less).

Table 6.8: **Significance of the asymmetry and degrees of freedom parameters.** This table presents the parameter estimates and the statistical significance of the asymmetry and degrees of freedom parameters of the skewed-t distribution estimated using the APARCH model. The results correspond to the first day of the out-of-sample forecast for portfolios made of linear combinations of Russell2000 and SP500 using the WML procedure with $\rho = 0.994$.

Parameter	Estimates	Std-errors	T-value	p-value
asymmetry	-0.064	0,025	-2,582	0.000
degrees of freedom	6,918	0,897	7,712	0.000

Table 6.9: Comparison of failure rates

		Normal		**Student-t**		**Skewed-t**	
α	**Model**	**ML**	**WML**	**ML**	**WML**	**ML**	**WML**
0,90	**CHK**	0,177	0,177	0,200	0,200	0,188	0,188
	GARCH	0,128	0,114	0,153	0,130	0,139	0,117
	APARCH	0,132	0,126	0,149	0,129	0,126	0,118
0,95	**CHK**	0,127	0,127	0,135	0,135	0,120	0,120
	GARCH	0,085	0,071	0,094	0,074	0,069	0,060
	APARCH	0,085	0,083	0,086	0,081	0,068	0,062
0,99	**CHK**	0,068	0,068	0,048	0,048	0,032	0,032
	GARCH	0,037	0,029	0,026	0,021	0,011	0,011
	APARCH	0,040	0,030	0,030	0,027	0,014	0,012

Comparison of empirical tail probabilities for the out-of-sample forecast for portfolios made of linear combinations of Russell2000 and SP500 using the ML procedure ($\rho = 1$) with WML with $\rho = 0.994$.

6.5.5 VaR subadditivity problem

According to Artzner, Delbaen, Eber, and Heath (1999), Frey and McNeil (2002) and Szegö (2002), a coherent risk measure satisfies the following axioms: translation invariance, subadditivity, positive homogeneity and monotonicity. They show that VaR satisfies all but one of the requirements to be considered as a coherent risk measure: the subadditivity property. Subadditivity means that "a merger does not create extra risk", i.e. that diversification must reduce risk. Moreover, the VaR is not necessarily convex with respect to portfolio rebalancing no matter what is the assumption made on the return distribution. This non-convexity problem generates the problem of non-unique solutions. It is for this reason that beside the global properties of the VaR, it is also important to study their local second-order behavior. To accomplish this, the knowledge of partial derivatives of the VaR with respect to the portfolio allocation is very useful. Moreover, partial derivatives are required to check the convexity of the VaR. However, such derivatives are difficult to derive when the multivariate normal distribution is not used. Gouriéroux, Laurent, and Scaillet (2000) propose analytical forms for these derivatives in a general framework. This derivatives are used to easy the statistical inference and to perform local risk analysis.

In this chapter we following Consigli (2002) and we do not discuss the limits of the VaR and instead we try to generate more accurate VaR estimates considering the asymmetry and kurtosis of the financial data.

Table 6.10: Failure rates for portfolios made of Russell2000-SP500, using ML with rolling window of fixed size

$1-\alpha$	**Model**	**Normal**	**p**	**Student-t**	**p**	**Skewed-t**	**p**
0,10	**CHK**	0,177	0,000	0,200	0,000	0,188	0,000
	GARCH	0,133	0,001	0,148	0,000	0,129	0,003
	APARCH	0,141	0,000	0,145	0,000	0,130	0,002
0,05	**CHK**	0,127	0,000	0,135	0,000	0,120	0,000
	GARCH	0,081	0,000	0,088	0,000	0,069	0,009
	APARCH	0,085	0,000	0,087	0,000	0,067	0,019
0,01	**CHK**	0,068	0,000	0,048	0,000	0,032	0,000
	GARCH	0,039	0,000	0,028	0,000	0,012	0,538
	APARCH	0,045	0,000	0,031	0,000	0,016	0,079

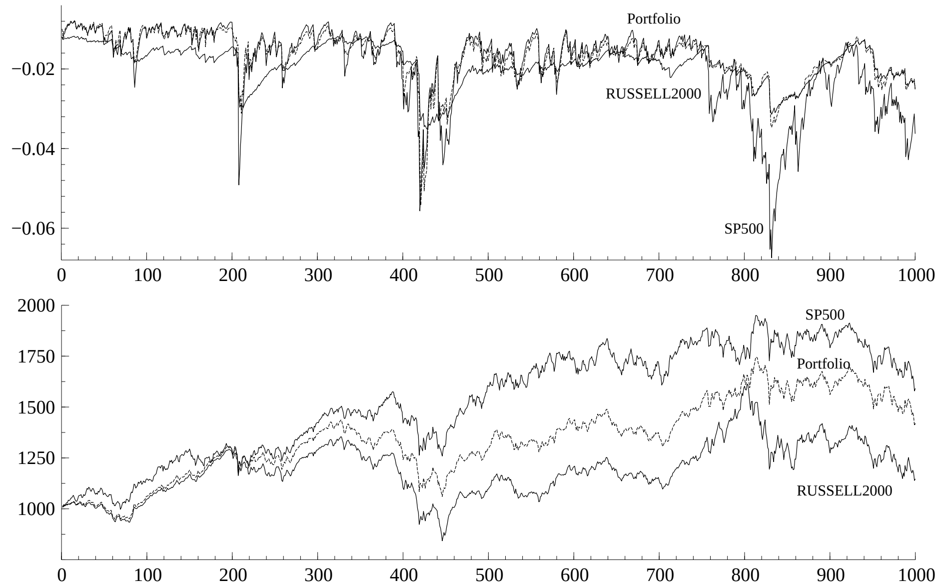
Empirical tail probabilities for the out-of-sample forecast for portfolios made of linear combinations of Russell2000 and SP500 using a rolling window of fixed size of 1000 days. The Kupiec-LR test is used to determine the specification of the models. The null hypothesis is that the model is correctly specified, i.e. that the failure rate equal to the desired probability of occurrence $1-\alpha$.

Table 6.11: Final wealth achieved by investing in portfolios made of Russell2000-SP500, using ML with rolling window of fixed size

$1-\alpha$	**Model**	**Normal**	**r**	**Student-t**	**r**	**Skewed-t**	**r**
0,10	Static	1306	6,9	1303	6,8	1303	6,8
	GARCH	1311	7,0	1283	6,4	1300	6,8
	APARCH	1663	13,6	1704	14,3	1461	9,9
0,05	Static	1297	6,7	1300	6,8	1296	6,7
	GARCH	1292	6,6	1284	6,5	1284	6,5
	APARCH	1579	12,1	1564	11,8	1390	8,6
0,01	Static	1277	6,3	1270	6,2	1263	6,0
	GARCH	1271	6,2	1258	5,9	1248	5,7
	APARCH	1465	10,0	1436	9,5	1336	7,5

Final wealth achieved by investing in portfolios made of Russell2000-SP500. r is the annual rate of return in (%). The risk-free interest rate is 4.47% annual.

Figure 6.14 presents the VaR and wealth evolution for an investor whose desired confidence level is 5%, the model used is GARCH and the innovation distribution is the skewed-t. The optimal portfolio VaR's are consistently smaller that the VaR's of the individual series. This is the case for all the models in our empirical application implying that by combining the two indices or stocks optimally we are reducing the risk. Moreover, the portfolio model not only allows to decrease risk but also to obtain portfolio returns between the returns of the individual indices.



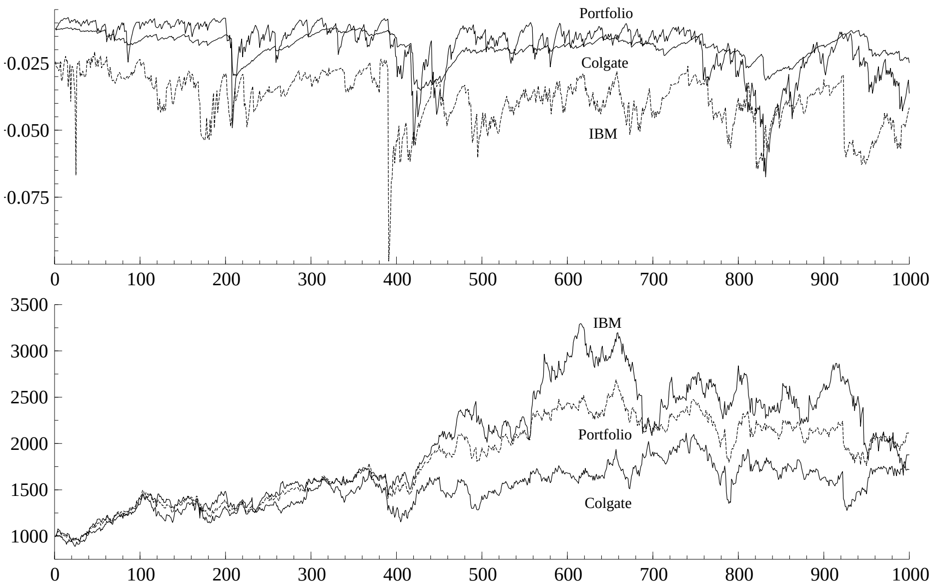
VaR-95 evolution (above) and Wealth evolution (below) for SP500, Russell2000 and for the optimal portfolios using GARCH model with skewed-t innovation distribution. Out-of-sample period goes from 02/01/1997 until 20/12/2000.

Figure 6.14: **Compared VaR and Wealth evolution: Russell2000-Sp500**

Figure 6.15 presents the same graph for portfolios made of Colgate-IBM. The VaR of the optimal portfolios are always smaller than the individual VaRs. Moreover, we can appreciate the advantages of diversification by looking at the wealth evolution at the end of the out-of-sample period (lower panel). The wealth evolution by only investing on IBM reduces rapidly while the portfolio wealth does not. At the end of the out-of-sample period the final wealth is almost the same.

6.6 Conclusions and future work

The dynamic portfolio selection model we propose performs well out-of-sample statiscally in terms of failure rates, defined as the number of times the desired confidence level used for the estimation of the VaR is violated and in terms of the dynamic quantile test, used to determine the quality of our results. Based on this criteria, the APARCH model gives as good results as the GARCH model. However, if we consider not only the failure rate but also the wealth achieved and the risk-adjusted returns, we find that for similar levels of risk, the APARCH model outperforms the GARCH model. A sensitivity analysis with respect to the distributional



VaR-95 evolution (above) and Wealth evolution (below) for Colgate, IBM and for the optimal portfolios using APARCH model with skewed-t innovation distribution. Out-of-sample period goes from 02/01/1997 until 20/12/2000.

Figure 6.15: **Compared VaR and Wealth evolution: Colgate-IBM**

innovation hypothesis shows that in general the skewed-t is preferred to the normal and Student-t. Estimating the model parameters by Weighted Maximum Likelihood in an increasing window setup allows us to account for a changing time pattern of the degrees of freedom and asymmetry parameters of the innovation distribution and to improve the forecasting results in the statistical and economical sense: smaller failure rates and larger final wealth.

There are a number of directions for further research along the lines presented here. A potential extension could use the dynamic model to study the optimal time of portfolio rebalancing, as day-to-day portfolio rebalancing may be neither practicable nor economically viable. A more ambitious extension is to work in a multivariate setting, where a group of different financial instruments are used to maximize the expected return subject to a risk constraint. Another interesting extension of the model is to investigate its intra-daily properties. This extension could be of special interest for traders who face the market second by second during the trading hours in the financial markets.

Finally, other interesting topic for future research can be to estimate optimal portfolios with large number of financial instruments. To achieve this goal, some econometrical methods and tools should be used. One possibility is to build a two step procedure. In the first step, develop a cluster analysis among the whole set of financial instruments in order to reduce the dimensionality of the portfolio components. In the next step, one can use this "reduced" financial instruments to estimate the optimal portfolios.

Bibliography

Admati, A., and P. Pfleiderer, 1988, A theory of intraday patterns: Volume and price variability, *Review of Financial Studies* 1, 3–40.

Al-Osh, M., and A. Alzaid, 1987, First order integer valued autoregressive (inar(1)) process, *Journal of Time Series Analysis* 8, 261–275.

Alexander, B., and M. Baptista, 2002, Economic implications of using a mean-var model for portfolio selection: a comparison with mean-variance analysis, *Journal of Economic Dynamics and Control* 26, 1159–1193.

Amihud, Y., and H. Mendelson, 1980, Dealership markets: Market making with inventory, *Journal of Financial Economics* 8, 31–53.

Andersen, T. G., and T. Bollerslev, 1997, Intraday periodicity and volatility persistence in financial markets, *Journal of Empirical Finance* 4, 115–158.

Anderson, J. A., 1984a, Regression and ordered categorical variables, *Journal of the Royal Statistical Society, B* 46, 1–30.

Anderson, Thomas W., 1984b, *An Introduction to Multivariate Statistical Analysis* (John Wiley and Sons: New York).

Angel, J., 1994, Limit versus market orders, Unpublished working paper. School of Business Administration, Georgetown University.

Artzner, P., F. Delbaen, J. Eber, and D. Heath, 1999, Coherent measures of risk, *Mathematical Finance* 9, 203–228.

Back, K., 1992, Insider trading in continuous time, *Review of Financial Studies* 5, 435–472.

Bae, K.-H., H. Jang, and K.S. Park, 2003, Trader's choice between limit and market orders: Evidence from nyse stocks, *Journal of Financial Markets* 6, 517–538.

Bagehot, W. [pseud.], 1971, The only game in town, *Financial Analists Journal* 27, 12–14.

Bauwens, L., and P. Giot, 2000, The logarithmic acd model: an application to the bid-ask quote process of three nyse stocks, *Annales d'Economie et Statistique* 60, 117–149.

——— , 2001, *Econometric Modelling of Stock Market Intradaily Activity* (Kluwer Academic Publishers: Dordrecht, the Netherlands).

——— , J. Grammig, and D. Veredas, 2004, A comparison of financial duration models via density forecasts, CORE mimeo.

Bauwens, L., and N. Hautsch, 2003, Dynamic latent factor models for intensity processes, Working paper 2003/103, Center of Operactions Research and Econometrics, Catholic University of Louvain.

Bauwens, L., and D. Veredas, 2004, The stochastic conditional duration model: a latent factor model for the analysis of financial durations, *Journal of Econometrics*.

Beltran-Lopez, H., P. Giot, and J. Grammig, 2004, How to summarize the information in the order book?, Mimeo. Center for Operations Research and Econometrics-CORE.

Bertsimas, Dimitris, and Andrew W. Lo, 1998, Optimal control of execution costs, *Journal of Financial Markets* 16, 1–50.

Biais, B., P. Hillion, and C. Spatt, 1995, An empirical analysis of the limit order book and the order flow in the paris bourse, *Journal of Finance* 52, 655–681.

——— , 1999, Price discovery and learning during the preopening period in the paris bourse, *Journal of Political Economy* 107, 1218–1248.

Bisière, C., and T. Kamionka, 2000, Timing of orders, orders aggressiveneess and the order book at the paris bourse, *Annales d'Economie et de Statistique* 56, 43–72.

Bollerslev, Tim, 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 52, 5–59.

Bond, S., 2000, A review of asymmetric conditional density functions in autoregressive conditional heteroskedasticity models, Mimeo, Duke University.

Bowsher, C.G., 2002, Modelling security markets in continuous time: Intensity based, multivariate point process models, Discussion paper 2002-W22, Nuffield College.

Brännäs, K., 1995, Estimation and testing in integer valued AR(1) models, Umea Economic Studies, number 355, University of Umeå.

Brännäs, Kurt, and Per Johansson, 1994, Time series count data regression, *Communications in Statistics: Theory and Methods* 23, 2907–2925.

Brémaud, P., 1981, *Point Processes and Queues, Martingale Dynamics* (Springer: New York).

Cameron, A. Colin, and Pravin K. Trivedi, 1998, *Regression Analysis of Count Data* (Cambridge University Press: Cambridge).

Campbell, J., 2000, Asset pricing at the millennium, *Journal of Finance* 55, 1515–1567.

Campbell, M.J., 1994, Time series regression for counts: an investigation into the relationship between sudden infant death syndrome and environmental temperature, *Journal of the Royal Statistical Society A* 157, 191–208.

Campbell, R., R. Huisman, and K. Koedijk, 2001, Optimal portfolio selection in a value-at-risk framework, *Journal of Banking and Finance* 25, 1789–1804.

Cebrián, A., M. Denuit, and O. Scaillet, 2004, Testing for concordance ordering, *Astin Bulletin* 34, 151–173.

Christoffersen, P., and D. Pellettier, 2004, Backtesting value-at-risk: A duration-based approach, *Journal of Financial Econometrics* 2, 84–108.

Clark, Peter K., 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* 41.

Cohen, J. Kalman, S. Maier, R. Schwartz, and D. Whitcomb, 1981, Transaction costs, order placement strategy, and existence of the bid-ask spread, *Journal of Political Economy* 89, 287–305.

Consigli, G., 2002, Tail estimation and mean-var portfolio selection in markets subject to financial inestability, *Journal of Banking and Finance* 26, 1355–1382.

Copeland, T;, and D. Galai, 1983, Information effects and the bid-ask spread, *Journal of Finance* 38, 1457–1469.

Cox, D.R., 1981, Statistical analysis of time series, some recent developments, *Scandinavian Journal of Statistics* 8, 93–115.

Cox, D., and V. Isham, 1980, *Point Processes* (Chapman and Hall: London).

Danielson, J., and R. Payne, 2001, Measuring and explaining liquidity on an electronic limit order book: evidence from reuters d200-2, London School of Economics, Working paper.

de Roon, F., T. Nijman, and B. Werker, 2003, Currency hedging for international stock portfolios: the usefulness of mean-variance analysis, *Journal of Banking and Finance* 27, 327–349.

Degryse, H., F. de Jong, M. Ravenswaaij, and G. Wuyts, 2003, Aggressive orders and the resiliency of a limit order market, Working paper, K.U. Leuven, Department of Economics.

Demsetz, H., 1968, The cost of transacting, *Quarterly Journal of Economics* 82, 33–53.

Denuit, M., and P. Lambert, 2002, Constraints on concordance measures in bivariate discrete data, Discussion Paper 02-12, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

——— , 2004, Constraints on concordance measures in bivariate discrete data, *Journal of Multivariate Analysis.* Article in press.

Diebold, Francis X., Todd A. Gunther, and Anthony S. Tay, 1998, Evaluating density forecasts with applications to financial risk management, *International Economic Review* 39, 863–883.

Ding, Z., C. Granger, and R. Engle, 1993, A long memory property of stock market returns and a new model, *Journal of Empirical Finance* 1, 83–106.

Easley, D., and M. O'Hara, 1987, Price, trade size and information in security markets, *Journal of Financial Economics* 19, 69–90.

——— , 1992, Time and the process of security price adjustment, *Journal of Finance* 47, 577–605.

Efron, Bradley, 1986, Double exponential families and their use in generalized linear regression, *Journal of the American Statistical Association* 81, 709–721.

Embrechts, P., A. McNeil, and D. Straumann, 2002, Correlation and dependence in risk management: properties and pitfalls, in M.A.H. Dempster, ed.: *Risk Management: Value at Risk and Beyond* . pp. 176–223 (Cambridge University Press: Cambridge).

Engle, Robert, 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica* 50, 987–1008.

Engle, R., and G. Gonzalez-Rivera, 1991, Semiparametric arch model, *Journal of Business and Economic Statistics* 9, 345–360.

Engle, R., and A. Lunde, 2003, Trades and quotes: a bivariate point process, *Journal of Financial Econometrics* 1, 159–188.

Engle, R., and S. Manganelli, 2004, Caviar: Conditional autoregressive value at risk by regression quantile, *Journal of Business and Economic Statistics* 22, 367–381.

Engle, Robert F., and Jeffrey Russell, 1998, Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica* 66, 1127,1162.

Fernandes, M., and J. Grammig, 2003, A family of assymetric autoregressive conditional duration models, Graduate School of Economics, Getulio Vargas Foundation.

Fernandez, C., and M. Steel, 1998, On bayesian modelling of fat tails and skewness, *Journal fo the American Statistical Association* 93, 359–371.

Fisher, R. A., 1932, *Statistical Methods for Research Workers.*

Fleming, J., C. Kirby, and B. Ostdiek, 2001, The economic value of volatility timing, *Journal of Finance* pp. 329–352.

Foster, F., and S. Viswanathan, 1990, A theory of the interday variations in volume, variance, and trading costs in securities markets, *Review of Financial Studies* 3, 593–624.

——— , 1993, Variations in trading volume, return, volatility, and trading costs: Evidence on recent price formation models, *Journal of Finance* 48, 187–211.

——— , 1995, Can speculative trading explain the volume-volatility relation?, *Journal of Business and Economic Statistics* 13, 379–396.

Foucault, T., 1999, Order flow composition and trading costs in a dynamic limit order market, *Journal of Financial Markets* 2, 99–134.

Frey, R., and A. McNeil, 2002, Var and expected shorfall in portfolios of dependent credit risks: Conceptual and practical insights, *Journal of Banking and Finance* 26, 1317–1334.

Garman, M., 1976, Market microstructure, *Journal of Financial Economics* 3, 257–275.

Genest, C., K. Ghoudi, and L.-P. Rivest, 1995, A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika* 82, 543–52.

Geweke, J., 1986, Modelling the persistence of conditional variances: A comment, *Econometric Review* 5, 57–61.

Ghysels, E., C. Gouriéroux, and J. Jasiak, 2004, Stochastic volatility duration models, *Journal of Econometrics* 119, 413–433.

Giot, P., and S. Laurent, 2003, Value-at-risk for long and short trading positions, *Journal of Applied Econometrics* 18, 641–663.

——— , 2004, Modelling daily value-at-risk using realized volatility and arch type models, *Journal of Empirical Finance* 11, 379–398.

Glosten, L., R. Jagannathan, and D. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 48, 1779–1801.

Glosten, Lawrence R., 1994, Is the electronic open limit order book inevitable?, *The Journal of Finance* 49, 1127–1161.

——— , and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71–100.

Gouriéroux, C., J. Laurent, and O. Scaillet, 2000, Sensitivity analysis of values at risk, *Journal of Empirical Finance* 7, 225–245.

Grammig, J., and K. Maurer, 2000, Non-monotonic hazard functions and the autoregressive conditional duration model, *Econometrics Journal* 3, 16–38.

Grandell, J., 1976, *Doubly Stochastic Poisson Process* (Springer: Berlin).

Griffiths, M., Smith-B., A. Turnbull, and R. White, 2000, The cost and determinants of order aggressiveness, *Journal of Financial Economics* 56, 65–88.

Grossman, S., 1976, On the efficiency of competitive stock markets where traders have diverse information, *Journal of Finance* 31, 573–585.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393–408.

Guermat, C., and D.F. Harris, 2002, Forecasting value at risk allowing for time variation in the variance and kurtosis of portfolio returns, *International Journal of Forecasting* 18, 409–419.

Hall, A., N. Hautsch, and J. Mcculloch, 2003, Estimating the intensity of buy and sell arrivals in a limit order book market, Mimeo, University of Technology, Sidney and University of Konstanz.

Handa, P., R. Schwartz, and A. Tiwari, 2003, Quote setting and price formation in an order driven market, *Journal of Financial Markets* 6, 461–489.

Handa, Puneet, and Robert A. Schwartz, 1996, Limit order trading, *The Journal of Finance* 51, 1835–1861.

Hansen, B., 1994, Autoregressive conditional density estimation, *International Economic Review* 35, 705–730.

Hardle, W., and L. Simar, 2003, *Applied Multivariate Statistical Analysis* (Springer: UK).

Harris, L., 1998, Optimal dynamic order submission strategies in some stylized trading problems, *Financial Markets* 7, 1–76.

Harvey, A.C., and C. Fernandes, 1989, Time series models for count or qualitative observations, *Journal of Business and Economic Statistics* 7, 407–417.

Harvey, C., and A. Siddique, 1999, Autoregressive conditional skewness, *Journal of Finance and Quantitative Analysis* 34, 465–487.

Hasbrouck, Joel, 1999, Trading fast and slow: Security markets events in real time, mimeo.

Hasbrouck, J., 2004, Empirical market microstructure: Economic and statistical perspectives on the dynamics of trade in securities markets, Teaching notes, Department of Finance, Stern School of Business,New York University.

Hasbrouck, Joel, and Duane J. Seppi, 2001, Common factors in prices, order flows and liquidity, *Journal of Financial Economics* 59, 383–411.

Hasbrouck, J., and G. Sofianos, 1993, The trades of market makers: An empirical examination of nyse specialists, *Journal of Finance* 48, 1565–1593.

Hausman, J., A. Lo, and C. Mackinlay, 1992, An ordered probit analysis of transaction stock prices, *Journal of Financial Economics* 31, 319–379.

Hawkes, A.G., 1971, Spectra of some self-exciting and mutually exciting point processes, *Biometrika* 58, 83–90.

Heinen, A., 2003, Modeling time series count data: An autoregressive conditional poisson model, CORE Discussion Paper 2003/62.

———, and E. Rengifo, 2003, Multivariate modeling of time series count data: An autoregressive conditional poisson model, CORE Discussion Paper 2003/25.

Higgins, M., and A. Bera, 1992, A class of nonlinear arch models, *International Economic Review* 33, 137–158.

Ho, T., and R. Macris, 1984, Dealer bid-ask quotes and transaction prices: An empirical study of some amex options, *Journal of Finance* 39, 23–45.

Ho, T., and H. Stoll, 1981, Optimal dealer pricing under transactions and return uncertainty, *Journal of Financial Economics* 9, 47–73.

Holden, C., and A. Subrahmanyam, 1992, Long-lived private information and imperfect competition, *Journal of Finance* pp. 247–270.

——— , 1994, Imperfect competition, and long-lived information, *Economics Letters.*

Huang, C., and R. Litzenberger, 1988, *Foundations for Financial Economics* (North-Holland: Amsterdam).

Izenman, Alan Julian, 1980, Assessing dimensionality in multivariate regression, in P.R. Krishnaiah, ed.: *Handbook of Statistics*, vol. 1 . pp. 571–591 (North-Holland Publishing Company).

Jacobs, P., and P. Lewis, 1978a, Discrete time series generated by mixtures i: Correlational and runs properties, *Journal of the Royal Statistical Society B* 40, 94–105.

——— , 1978b, Discrete time series generated by mixtures ii: Asymptotic properties, *Journal of the Royal Statistical Society B* 40, 222–228.

——— , 1983, Stationary discrete autoregressive moving average time series generated by mixtures, *Journal of Tima Series Analysis* 4, 19–36.

Joe, Harry, 1997, *Multivariate Models and Dependence Concepts* (Chapman and Hall: London).

Jondeau, E., and M. Rockinger, 2003, Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements, *Journal of Economic Dynamics and Control* 27, 1699–1737.

Jorgensen, Bent, 1987, Exponential dispersion model, *Journal of the Royal Statistical Society* 49, 127–162.

——— , 1997, *The Theory of Dispersion Models* (Chapman and Hall: London).

Jorgensen, B., and S.L. Lauritzen, 1998, Multivariate dispersion models, Research Report 2, Department of Statistics and Demography, Odense University, Denmark.

Jorion, P., 1997, *Value-at-Risk: the New Benchmark for Controlling Market Risk* (McGraw-Hill: New York).

Kawasaki, Y., 2002, Modelling periodicity in high frequency financial data, Discussion paper, Institute of Statistical Mathematics, Tokio.

Knight, J., S. Satchell, and K. Tran, 1995, Statistical modelling of asymmetric risk in asset returns, *Applied Mathematical Finance* 2, 155–172.

Kupiec, P., 1995, Techniques for verifying the accuracy of risk measurement models, *The Journal of Derivatives.*

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1335.

Lambert, P., and S. Laurent, 2001a, Modelling financial time series using garch-type models with a skewed student distribution for the innovations, Discussion Paper 0125, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

——— , 2001b, Modelling skewness dynamics in series of financial data using skewed location-scale distributions, Discussion Paper 0119, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Lee, Y., and T. Tse, 1991, Term structure of interest rates in the singapore asian dollar market, *Journal of Applied Econometrics* 6, 143–152.

Leroy, F., and J. Werner, 2001, *Principles of Financial Economics* (Cambridge University Press: Cambridge).

Lintner, John, 1965, The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13–37.

Liu, S., and B. Brorsen, 1995, Maximum likelihood estimation of a garch-stable model, *Journal of Applied Econometrics* 2, 273–285.

Lunde, A., 1999, A generalized gamma autoregressive conditional duration model, University of Aarthus.

MacDonald, Iain L., and Walter Zucchini, 1997, *Hidden Markov and Other Models for Discrete-valued Time Series* (Chapman and Hall: London).

Machado, J., and J.M.C. Santos Silva, 2003, Quantiles for counts, Working Paper, ISEG, Universidade Técnica de Lisboa.

Madhavan, A., 2000, Market microstructure: A survey, *Journal of Financial Markets* 3, 205–258.

——— , and S. Smidt, 1991, A bayesian model of intraday specialist pricing, *Journal of Financial Economics* 30, 99–134.

——— , 1993, An analysis of changes in specialist inventories and quotations, *Journal of Finance* 48, 1595–1628.

Markowitz, Harry M., 1952, Portfolio selection, *Journal of Finance* 7, 77–91.

Mas-Colell, A., M. Whinston, and J. Green, 1995, *Microeconomic Theory* (Oxford University Press).

McCullagh, P., and J.A. Nelder, 1976, *Generalized Linear Models* (Chapman and Hall: London).

McDonald, J., 1991, Parametric models for partially adaptive estimation with skewed and leptokurtic residuals, *Economic Letters* 37, 272–278.

McKensie, E., 1986, Autorregressive moving-average processes with negative binomial and geometric marginal distributions, *Advances in Applied Probability* 18, 679–705.

Merton, R., 1972, An analytic derivation of the efficient portfolio frontier, *Journal of Financial and Quantitative Analysis* pp. 1851–1872.

——— , 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* pp. 323–361.

Mittnik, S., and M. Paolella, 2000, Conditional density and value-at-risk prediction of asian currency exchange rates, *Journal of Forecasting* 19, 313–333.

Nelsen, Roger B, 1999, *An introduction to Copulas* (Springer: New York).

Newey, Whitney K, and Daniel McFadden, 1994, Large sample estimation and hypothesis testing, in Robert F. Engle, and Daniel McFadden, ed.: *Handbook of Econometrics* . , vol. 4 (North-Holland: Amsterdam).

O'Hara, M., 1995, *Market microstructure* (Blackwell business).

——— , and G. Oldfield, 1986, The microeconomics of market making, *Journal of Financial and Quantitative Analysis* 21, 361–376.

Paolella, M.., 1997, Using flexible garch models with asymmetric distributions, Working paper, Institute of Statistic and Econometrics Christian Albrechts University at Kiel.

Parlour, C., 1998, Price dynamics in limit order markets, *Review of Financial Studies* 11, 789–816.

Pascual, R., and D. Veredas, 2004, What pieces of limit order book information do are informative? an empirical analysis of a pure order-driven market, Mimeo. Dept. of Business. Universidad de las Islas Baleares.

Patton, Andrew, 2002, Skewness, asymmetric dependence, and portfolios, mimeo.

Patton, A., 2004, On the out-of-sample importance of skewness and asymmetric dependence for asset allocation, *Journal of Financial Econometrics* 2, 130–168.

Pentula, S., 1986, Modelling the persistence of conditional variances: A comment, *Econometric Review* 5, 71–74.

Ranaldo, A, 2003, Order aggressiveness in limit order book markets, *Journal of Financial Markets* 7, 53–74.

Reiss, P., and I. Werner, 1998, Does risk-sharing motivate interdealer trading?, *Journal of Finance* 53, 1657–1703.

Richard, J. F., 1999, Efficient high-dimensional monte carlo importance sampling, Discussion paper, University of Pittsburgh.

Russell, J., 1999, The econometric modeling of multivariate irregularly-spaced high-frequency data, Working Paper Tilburg University and CentER.

Sandas, P., 2001, Adverse selection and competitive market making: Empirical evidence from a limit order market, *Review of Financial Studies* 14, 705–734.

Schwert, W., 1990, Stock volatility and the crash of '87, *Review of Financial Studies* 3, 77–102.

Seppi, D., 1997, Liquidity provision with limit orders and a strategic specialist, *Review of Financial Studies* 10, 103–150.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.

Sklar, A, 1959, Fonctions de répartitions à n dimensions et leurs marges, *Public Institute of Statistics of the Univeristy of Paris* 8, 229–231.

Song, Peter Xue-Kun, 2000, Multivariate dispersion models generated from gaussian copulas, *Scandinavian Journal of Statistics* 27, 305–320.

Spierdijk, L., T. Nijman, and A. van Soest, 2002, Modelling comovements in trading intensities to distinguish sector and specific news, Working paper, Econometric and Finance Group, CentER, Tilburg University.

Stoll, H., 1976, Dealer inventory behavior: An empirical examination of nasdaq stocks, *Journal of Financial and Quantitative Analysis* 11, 359–380.

——— , 1978, The supply of dealer services in security markets, *Journal of Finance* 33, 1133–1151.

Subrahmanyam, A., 1991a, Risk aversion, market liquidity, and price efficiency, *Review of Financial Studies* 4, 416–441.

——— , 1991b, A theory of trading in stock index futures, *Review of Financial Studies* 4, 17–51.

Szegö, G., 2002, Measures of risk, *Journal of Banking and Finance* 26, 1253–1272.

Taylor, S.J., 1982, Financial returns modelled by the product of two stochastic processes - a study of daily sugar prices, in O.D. Anderson, ed.: *Time Series Analysis: Theory and Practice* (North-Holland, Amsterdam).

Taylor, S., 1986, *Modelling Financial Time Series* (J.Wiley & Sons: New York).

Ter Braak, Cajo J. F., 1986, Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis, *Ecology* 67, 1167–1179.

Tso, M. K-S., 1981, Reduced-rank regression and canonical analysis, *Journal of the Royal Statistical Society, B* 43, 183–189.

Wang, K., C. Fawson, C. Barret, and J. McDonald, 2000, A flexible parametric garch model with an application to exchange rates, *Journal of Applied Econometrics* 16, 521–536.

Yee, Thomas, and Trevor Hastie, 2002, Reduced-rank vector generalized linear models, Working Paper, University of Auckland.

Yee, T.W., and T. J Hastie, 2003, Reduced-rank vector generalized linear models, *Statistical Modelling* 3, 15–41.

Zakoian, J., 1994, Threshold heteroskedastic models, *Journal of Economic Dynamics and Control* 18, 931–955.

Zeger, Scott L., 1988, A regression model for time series of counts, *Biometrika* 75, 621–629.

——— , and Kung-Yee Liang, 1986, Longitudinal data analysis for discrete and continuous outcomes, *Biometrics* 42, 121–130.

Zhu, Mu, and Trevor Hastie, 2002, Feature extraction for non-parametric discriminant analysis, Working Paper, University of Waterloo.