Empirical Models of the Intraday Process of Price Changes and Liquidity — A Transaction Level Approach

Frank Gerhard

Nuffield College, Oxford, OX1 1NF, United Kingdom

E-mail address: frank.gerhard@nuffield.oxford.ac.uk

Dissertation der Universität Konstanz

Tag der mündlichen Prüfung: 8. Februar 2001

Referent: Prof. Dr. Winfried Pohlmeier

Referent: Prof. Dr. Günter Franke

Contents

List of Figures

List of Tables

Variables

β	Coefficient vector of exogeneous variables x_t.
c	Constant.
d_t	Price change between transaction $t-1$ and t.
e_i	Unit vector for the dimension i, i.e. $e_{i[i]} = 1$ and $e_{i[j]} = 0$ for $i \neq j$.
$f_a(u)$	P.d.f. of a at u
$F_a(u)$	C.d.f. of a at u
γ	Coefficient vector of exogeneous variables w_t.
Γ_s	Impulse response function after s periods based on the difference of conditional expectations.
Γ_s^*	Impulse response function after s periods based on the partial derivative of the conditional expectation.
$\mathcal{F}_t$	Information set available for transaction t
P_t	Price level at transaction t.
Π	Transition matrix of a Markov chain.
π	Stationary distribution of a Markov chain.
π_t	A vector of conditional probabilities corresponding to transaction t.
R_t	Return of the transaction process, growth rate.
s_t	Sign of the price change d_t.
$s_t^\dagger$	Sign of the price change d_t if $z_t > 0$.
S_t	Random variable mapping the state of the market at transaction t.
$s(\delta, t, P)$	Trigonometric expansion up to order P of the time of day for observation t using coefficients δ.
σ_t^2	Conditional variance of transaction t.
t	Observation index.
T_i	Number of observations on day i.
ϑ_t	The clock time at which transaction t occurs.
τ_t	The time between transaction t and $t-1$.
v_i	Individual values the process of price changes can take on.
w_t	Weakly exogeneous regressors associated with transaction t.
x_t	Weakly exogenous regressors associated with transaction t.
ξ_t	State of the system at observation t.
y_t	Endogenous components of the transaction t.
z_t	Size of the price change d_t.

Notation

$a_{[i]}$	ith element of the vector a.
$a_{[i:j]}$	Elements i through j of the vector a.
$A_{[i,j]}$	Element of the matrix A in row i, column j.
$A_{[i,\cdot]}$	Row i of the matrix A.
a^c	If a is a vector, then each element of a is raised to the power c.
A^c	If A is a quadratic matrix and c a positive integer, then this stands for a successive multiplication of A with itself from the left.
$A^{1/2}$	Cholesky decomposition of the matrix A.
$\operatorname{diag} a$	Matrix of zeros with the vector a on its main diagonal.
$\bar{x}_t$	All past observations of x_t up to transaction t.
d_t	Observable discrete variable in a continuous time context, also: z_t, s_t.
d_t^*	Unobservable latent variable corresponding to the discrete observation d_t, also z_t^*, s_t^*.
d_t°	Observable discrete variable in a discrete time context, also: a_t°, z_t°, s_t°.

List of abbreviations

ACD	Autoregressive conditional durations
ACF	Autocorrelation function
ARMA	Autoregressive moving average
BIC	Schwarz / Bayesian information criterion
DGP	Data generating process
GLAR	Generalized linear autoregressive
GLARMA	Generalized linear autoregressive moving average
GLM	Generalized linear model
LACD	Logarithmic autoregressive conditional durations
LEF	Linear exponential family
MC	Markov chain
ML	Maximum likelihood
PACF	Partial autocorrelation function
PML	Pseudo maximum likelihood
SPDE	Stochastic partial differential equation

Part 1

The intraday transaction process

CHAPTER 1

Introduction

The development of a feasible and flexible model for the intraday transaction process is an ongoing topic in the empirical analysis of market microstructure. Here, the focus is on the two major components of the transaction process, namely the process of price changes and the process of trade intensity. This work attempts to modify and extend the models of the transaction process suggested by Rydberg and Shephard (1998) and Russell and Engle (1998). The main contribution is twofold:

First of all, a new and flexible econometric model of the intraday process of price changes is developed. Contrary to Rydberg and Shephard and Russell and Engle who use generalized linear models for model building, here, an extension to the classical ordered probit model is proposed. The ordered probit was already employed by Hausman, Lo, and MacKinlay (1992) for the analysis of discrete trade-to-trade price changes. In their study the probit turned out to be a flexible modelling tool to test economic hypotheses on the economic properties of price changes. In this work, the quantal response model is augmented by a latent ARMA type dynamic relying on the concept of generalized residuals in the spirit of Gourieroux, Monfort, Renault, and Trognon (1987) to capture the well known serial dependency of transaction price changes. This model employs an observation driven dynamic in the sense of Cox (1981). In contrast to the approaches by Rydberg and Shephard and Russell and Engle the model proposed here allows a parsimonious modelling strategy and stationarity conditions of the latent process take a well-known and simple form. It is derived from an ARMA process in the latent variable, which is a value of its own because of the substantial amount of previous research done on observable ARMA processes. In addition, this approach allows a straightforward extension to multivariate models. The extension to multivariate models is of particular importance if one attempts to assess exogeneity or simultaneity relationships among economic variables. The use of this dynamic quantal response model is not only limited to the analysis of price changes. Lately, quantal response models have become increasingly popular in the analysis of time series, e.g. on the business cycle, financial crises, interest rate changes, and credit scores. See e.g. work by Dueker (1999), Hamilton and Jorda (2000), Estrella and Mishkin (1997), Bernard and Gerlach (1996), Davutyan and Parke (1995), Eichengreen, Watson, and Wyplosz (1995), Broseta (1993), Eichengreen, Watson, and Grossman (1985).

The second major contribution of this work is an attempt to disentangle the relationship between the process of price changes and the process of trade intensity. This exploits the fact that the new dynamic quantal response model proofs to be a valuable building block for multivariate models. The simultaneity of the time between individual transactions and both, the direction of price changes and the size of price changes will be under close scrutiny. To allow for enough flexibility, the analysis of the simultaneity between the transaction intensity and direction of price changes and the size of price changes respectively is carried out in separate models. The potential simultaneity is of particular interest for the analysis of the relationship between volatility and liquidity, measured as transaction frequency. A potential simultaneity between the trade intensity and the process of price changes has far reaching consequences because volatility and liquidity will thus need to be considered jointly and are not independent risk factors. If both processes were found to be independent, however, both factors could be considered independently. In addition, the method proposed here could be easily extended to include volume as an additional dependent variable in a trivariate equation system, or the bid-ask spread might be accounted for if one is willing to estimate a bivariate quantal response model.

The structure of this work is dictated by the peculiar structure of the data under consideration. Here, disaggregated transactions in a financial market are considered. The most prominent difference to aggregated data from financial markets, as e.g. daily returns, is to be found in the sheer number of observations available over

Table 1.1. Typical sampling schemes of price changes. The rough approximations depicted in the table are based on descriptive statistics of the Bund Future traded at DTB and LIFFE. The calculations are based on 9 hours trading per day and 6 transactions per minute.

type	observations per year	scale	sampling
daily data	$\sim$ 250	continuous	regular
high-frequency data (5 min aggregates)	$\sim$ 27.000	discrete	regular
transaction data	$\sim$ 810.000	discrete	irregular

a very short sampling time span. See table 1.1 for an overview over typical sample sizes for one year of observations. Whereas on an aggregate level the analysis of volatility and liquidity typically recurs to a mixture of distributions hypothesis going back to the original work by Clark (1973), the intraday process, on the other hand, allows to disentangle the relationship of price changes, absolute price changes, and time between transactions.

The amount of observations available and the potentially clearer results we can obtain on the relationship of liquidity and volatility come however at a substantial cost. In general, the more disaggregate the chosen observation level is, the more shows the microstructure of the respective market in the data. In particular, first, the minimum price change, i.e. the tick size, which is institutionally fixed will determine the grid on which the processes of price changes and size of price changes live. Second, transaction costs in form of the bid-ask spread, generate a peculiar time dependency in returns. Both effects tend to blur results obtained by standard linear time series models. Third, at a transaction level observations occur irregularly, depending of course on the interaction of market participants on the buyer and seller side of the market.

Thus, the main task pursued in this work is to formulate an econometric model which accommodates the particular structure of the data, just outlined, and allows to assess the relationship of individual components of the transaction process, while still maintaining a computationally manageable structure.

The structure of this work is as following. The peculiarities of transaction data are the core of the second chapter. Before econometric models for the individual components are discussed in later chapters, the basic structure of models of the intraday process of volatility are discussed and it is analysed how the model suggested here fits in this framework. The modelling challenges of transaction data are briefly reviewed in the context of the relevant literature and typical hypotheses about the relationship of liquidity and volatility are illustrated in the stylized context of a Markov Chain. This grossly simplified structure allows to discuss the possible relationships between the process of price changes and trade intensity which will be carried out in later chapters in greater detail, employing a suitable empirical model.

After having discussed the necessary properties of an empirical model of the intraday transaction process, in chapter 3 different alternatives to model the process of price changes are reviewed, in particular the models of Rydberg and Shephard and Russell and Engle. The new latent dynamic for quantal response models is introduced along with a discussion of possible extensions and its dynamic properties.

The second component of the transaction process is the transaction intensity, for which econometric models are reviewed in chapter 4. The literature review concentrates on the seminal work of Engle (1996) or rather Engle (2000) and on several extensions suggested by himself and among others by Bauwens and Veredas (1999).

The models available in the literature are slightly modified to accommodate the joint structure of the transaction process outlined in chapter 5. A decomposition of the price process is suggested in chapter 5, which is designed so that the sign and the size of the price changes can be considered in separate models. Yet, this decomposition allows to recover the joint distribution of price changes and time between transactions. The simultaneous multivariate model involves the transaction intensity process and quantal response models of the direction of the price process and the size of the price changes. An appropriate method to describe the dynamics of the joint system is outlined based on the concept of a generalized impulse response functions, as suggested by Koop, Pesaran, and Potter (1996). The use of generalized impulse response functions allows to summarize the dynamic properties and the relationship between both equations in an intuitive fashion, if impulse response functions are adequately adjusted to the type of nonlinearity encountered in this modeling framework.

The empirical findings concerning the individual processes based on a sample of the BUND future trading at the DTB are given in chapter 6. The well-known intraday seasonality is analysed for this particular contract and is included in the dynamic specification of the process of the direction of price changes, the process of absolute price changes, and the process of time between transactions. Some results concerning the influence of trade frequency and trade volume on the probability to observe price changes of given size are analysed. It turns out however that a univariate analysis yields no conclusive results and that a thorough evaluation needs to be based on a multivariate model, which takes the potential simultaneity of trade frequency and volume into account. The joint transaction process and its implications for the volatility per transaction are thus under scrutiny. The empirical results are augmented by an analysis of the generalized impulse response functions implied by the model estimates. The main findings are twofold. First, and most important, evidence is provided that the size of price changes and the time between transactions are indeed simultaneous. This, implies that a meaningful study of volatility and thus risk, will need to be linked to an analysis of liquidity and thus trade frequency. Second, the sign of price changes on the other hand is found to be independent of the transaction frequency. This, however, might well be an empirical fact limited to the trading of futures.

The stability of the empirical findings is examined in chapter 7. In contrast to the standard methods to detect structural breaks in time series, here, a particular method suggested Gerhard, Hess, and Pohlmeier (1998) is employed. This method based on a minimum distance estimator is particularly well adjusted to the needs of intraday transaction analysis, i.e. although a single day contains by far enough data to estimate most nonlinear models, it is important to query whether empirical findings based on small time spans are indeed representative for the financial market under consideration. Results indicate that the main findings of this work are stable over the extended sample under consideration.

Chapter 8 concludes with a summary of the major findings of this work.

CHAPTER 2

On the microstructure of financial markets

1. The structure of empirical models of volatility

1.1. Volatility Estimation based on Intraday Data. The peculiarities of transaction data, especially the relationship between the process of price changes and the process of trade intensity, are also a relevant factor for the analysis of volatility per time. Although volatility per time will not explicitly be considered within this work, as this would necessitate and extended analysis of the time aggregation of the transaction process, still the influence of different market microstructure settings should be highlighted.

In order to clarify the implications of the chosen risk estimator in the context of a transaction based analysis, it seems worthwhile to consider the conditional volatility per time σ_t^2. The variable t, $t = 1, \ldots, T$, gives only a consecutive ordering of the observations. Observations on the price process are available at various points in time ϑ_t. Time ϑ is measured since the start of the sample, e.g. in seconds. Returns R_t are generated from these observations. For the time being we abstract from overnight and weekend returns. Also, for ease of exposition in the following growth rates R_t as opposed to log returns r_t are used, without a significant limitation of the scope. Since the same arguments put forward here, can be constructed on the basis of log returns. Especially the discreteness of price changes on a transaction basis suggests the use of growth rates and to give up the well known advantages of log returns. However, using log returns would be possible in principle, but would involve a considerable notational overhead.

Consider a first definition of volatility per time,

$$\sigma_t^2 = \mathrm{E}\left[\frac{1}{\tau_t} \left(R_t - \mathrm{E}\left[R_t \middle| \mathcal{F}_{t-1}\right]\right)^2 \middle| \mathcal{F}_{t-1} \right] \tag{2.1}$$

with $\tau_t = \vartheta_t - \vartheta_{t-1}$.

which is kept somewhat imprecise with respect to the nature of the expectation operator on purpose, as different cases will be considered in the following.

Depending on the chosen application $\mathcal{F}_{t-1}$ contains a subset of the information available at time ϑ_t. As a matter of fact it should only contain contemporaneous regressors, i.e. data available at ϑ_t, which are at least weakly exogenous. In typical time series applications $\mathcal{F}_{t-1}$ contains especially lagged values of the dependent variable, therefore the index $t-1$. Note that the time between observations τ_t needs to be bounded away from zero. So, if we consider a trading environment, where different trades are time stamped with the identical time ϑ_t we would need to consider a data processing mechanism, which accounts for this fact.

Often it is imposed that $\mathrm{E}\left[R_t \middle| \mathcal{F}_{t-1}\right] = 0$ as the estimation of this relationship is subject to a substantial estimation risk and spurious results induced by sample selection. See e.g. the discussion in Merton (1980), Bai, Russell, and Tiao (1999), and the extensive discussion in Lo and MacKinlay (1999) on the stability of deviations from various versions of the random walk hypothesis. Concentrating on transaction data, this assumption is clearly inappropriate. On the one hand for reasons of the bid-ask bounce and also because of the presence of strong intraday serial correlations due to ongoing market movements, which might induce a positive autocorrelation.

The definition in (2.1) is somewhat imprecise with respect to two details. First, it is not explicitly stated with respect to which distribution the conditional expectation is evaluated. This however will be treated subsequently in greater detail, when either τ_t or R_t or both will be assumed stochastic. The three possible cases all have virtues of their own. However, it turns out that an analysis at the transaction level usually necessitates to consider both variables stochastic, τ_t and R_t. The second issue concerns the assumed data generating process (DGP) underlying the observable bivariate vector which describes the transaction process $y_t = [\tau_t \, R_t]$. Considering (2.1) the transaction process should satisfy the condition that for $\tau_t \to 0$ the returns should also converge in some appropriate sense to their conditional expectation, i.e. $R_t \to \mathrm{E}\,[R_t | \mathcal{F}_{t-1}]$ so that the volatility per time remains finite. This however is clearly beyond the scope of this work as it would involve a clear cut stochastic concept of a stochastic partial differential equation (SPDE) driving y_t. The latter point is not of interest here, as the focus of this work is on the flexible modeling of the individual components of y_t and thereby necessarily prevents a rigorous stochastic treatment which would impose considerable bounds on the class of models involved. To avoid this theoretical problem, throughout this work it is assumed that τ_t is bounded away from zero. A condition which is usually fulfilled in practice or can be achieved by an appropriate processing of the data.

1.2. Volatility Estimation Based on Price Changes. In order to apply standard GARCH-type procedures to the estimation problem implied by eq. (2.1) a new process is derived from the original observations y_t. In the context of standard GARCH-type models the time between observations τ_t is not modeled explicitly. More concisely, it is considered deterministic and constant over all t, $\tau_t = \tau$. In this new process with a constant time between observations τ, the returns R_t° are defined over these time intervals. Hence, intraday aggregates are used, e.g. on the basis of 5 minute intervals. See e.g. Andersen and Bollerslev (1998a) and Andersen, Bollerslev, Diebold, and Labys (1999) for further references.

In the resulting volatility per time the observation intensity $\frac{1}{\tau}$ can thus be taken out of the integral, which now involves only the conditional density of the return process.

$$\sigma_{1,t}^2 = \mathrm{E}\left[R_t^{\circ 2} \middle| \mathcal{F}_{t-1} \right] \frac{1}{\tau} \tag{2.2}$$

$$\text{using } \mathrm{E}\,[R_t^\circ | \mathcal{F}_{t-1}] = 0 \tag{2.3}$$

Seasonalities over the trading day, like the time varying trade intensity, and the size of the aggregation interval raise substantial problems for this volatility specification, since both effects might actually interact to create spurious heterogeneity due to the aggregation mechanism.

To account for intraday seasonalities in the volatility some modifications are proposed in the literature. Seasonalities can be corrected for by an inclusion of particular explanatory variables in the information set $\mathcal{F}_t$ as suggested by Andersen and Bollerslev (1997), see also chapter 6, section 1.2, or by the introduction of time varying coefficients as in Bollerslev and Ghysels (1996) for the parameterization of the conditional expectation in (2.2), or by the introduction of concepts of time deformation in the spirit of Dacorogna, Gauvreau, Müller, Olsen, and Pictet (1996). This involves the transformation of raw returns R_t^r to seasonally corrected R_t°, i.e.

$$R_t^\circ = \frac{R_t^r}{\delta(\vartheta_t, x_t)}, \quad \delta(\vartheta_t, x_t) > 0, \tag{2.4}$$

where $\delta(\cdot)$ captures deterministic components of R_t^r by functions of the time of day ϑ_t or other exogenous variables x_t related to the cyclical behaviour of R_t^r.

Such a procedure, however raises the question of an optimal aggregation level as a major problem which has to be solved before estimators of this type can be used for risk assessment. Evidently, there are two main sources of errors introduced. First, it is an unresolved issue what the criteria of optimality are for the determination of the aggregation level. It is a well documented fact that the volatility estimates based on different aggregation levels τ lead to significantly different results, which are hard to consolidate with each other. See the discussion of Andersen and Bollerslev (1998a) in the context of a continuous time DGP.

Second, it is far from clear what kind of bias is introduced by a constant aggregation level in the context of a time varying trading intensity, particularly for assets having a 'liquidity life cycle'. This is clearly the case for futures contracts where trading is concentrated in the front month contract. But also for assets which have a time varying liquidity over the trading day, which is actually true for most assets. This time varying transaction intensity implies that in periods of thin trading the variation of returns is not only due to the variation of the return process but captures also a measurement error component of the return R_t° in this period.

Third, this bias might be found to be even more severe if the trade intensity process and the process of price changes show a joint dynamic, which implies e.g. that short durations come along with small absolute price changes as opposed to large price changes. This might induce a systematic bias independent of the well-known daily seasonalities, which might be assumed to aggregate out over the trading day.

1.3. Volatility estimation based on price intensities. Instead of deriving a process with a stochastic return from the original y_t as in (2.2), now a process is derived which features a conditionally deterministic return for every observation, which nevertheless takes a stochastic time τ_t° to realize. This model class rests on the assumption that a decision maker in need of a risk measure is able to express the size of a significant price change. This size is denominated by c. Using this c in the given context, the risk measure boils down to the question of how long it might take to realize this significant price change, i.e. the empirical analysis of first passage times. This volatility measure was first suggested by Cho and Frees (1988) to estimate the volatility of discrete stock price changes. A similar specification of a instantaneous volatility is put forward by Engle and Russell (1998) in the context of their highly appraised autoregressive conditional duration (ACD) model. An extension to larger price changes c which cause the price duration to extend over night or even over several trading days is suggested by Gerhard and Hautsch (1999).

The bivariate distribution of R and τ is no longer of interest since R is reduced to the ratio of a constant and a conditionally deterministic variable. So, the conditional volatility $\sigma_{2,t}^2$ per time can be formulated as

$$\sigma_{2,t}^2 = \mathrm{E}\left[\frac{1}{\tau_t^{\circ}}\middle|\,\mathcal{F}_{t-1}\right]\left(\frac{c}{P_{t-1}}\right)^2 \tag{2.5}$$

$$= \sigma_t^{*2} \cdot \frac{1}{P_{t-1}^2}, \tag{2.6}$$

assuming that $\mathrm{E}\left[r_t \middle|\, \mathcal{F}_{t-1}\right] = 0$. The variable σ_t^{*2} stands for the conditional volatility of price changes from which the conditional volatility of returns can easily be recovered.

Thus, a form of the conditional volatility per time is obtained which does not involve the conditional expectation over a function of the return process any more. The variable τ_t° does not capture the time between individual observations but stands for the time it takes to complete the price change $|P_t - P_s| \geq c$ where $t > s$. Thus τ_t° maps the price intensity in this context.

This concept might seem awkward at first sight, yet it has the significant benefit that for typical values of c a data dependent aggregation scheme is obtained. This implied aggregation scheme rids the researcher of many market microstructure effects which are a nuisance in the analysis of risk, if c is chosen large enough. And it allows to aggregate data based on a parameter c, which allows a straightforward interpretation, as opposed to the aggregation interval τ in the classical context.

Although this model class of price intensities shows quite a few advantages over models based on time aggregates, it is not quite feasible for an analysis on the transaction level. Even if c is set to the minimal value of one tick, this estimator would not discriminate between the occurrence of a trade without a price change and the absence of a trade. If the role of liquidity for the volatility process is in the focus of an analysis, the volatility estimation based on price intensities does not seem to be the appropriate model, at least not in the basic form laid out here, because the trade intensity is aggregated out in the context of this model.

1.4. Volatility estimation based on the joint process of price changes and trade intensity. A third approach is to actually use the original process y_t and assume that both R_t and τ_t are stochastic. Obviously, this necessitates the use of empirical models which are at least bivariate, some of which will be discussed more in depth in chapter 5. The random variable of interest in the context of empirical market microstructure models is the price change d_t from one transaction to the next. We have thus

$$\sigma_{3,t}^2 = \mathrm{E}\left[\frac{1}{\tau_t} \cdot \left(d_t - \mathrm{E}\left[d_t \middle| \mathcal{F}_{t-1} \right] \right)^2 \middle| \mathcal{F}_{t-1} \right] \frac{1}{P_{t-1}^2}. \tag{2.7}$$

The limitation of these models to the analysis of price changes d_t is standard and not substantial, considering that these models are usually estimated on the basis of transaction data and the price level changes only very slowly. Of course this limits the use of these models for the use in risk management considerably. In theory measures of interest in this area, e.g. volatility per day, or the probability of a large price change of a given size, could be recovered from estimation results. Practically however, these models are tuned to analyse the short run behaviour of the transaction process and reveal no particular insights into the long run dynamics. Also, this approach raises a significant problem, because the volatility implied by (2.7) shows quite an erratic behaviour from one transaction to the next. Thus, in order to gain a meaningful risk measure, the volatility of price changes needs to be aggregated over time using the stochastic model of trade intensities. Yet, this volatility estimator has the significant advantage that it explicitly accounts for the relationship of volatility and trade intensity. This advantage however is bought by the need for a model dependent aggregation scheme. As this is quite a formidable task and the major focus is on the dynamics of the transaction process and not its aggregation, no attempt in this direction is made within this work. Additionally, a transaction based approach might actually reveal other properties of the time series, e.g. the simultaneity of individual components, which are very useful, if a model for the long run behaviour is constructed.

Some approaches actually attempt to combine a GARCH specification with the analysis of trade intensity as e.g. Ghysels and Jasiak (1998), who analyse a bivariate process including returns in a GARCH specification which accounts explicitly for the stochastic nature of the price intensity. These models refrain however from taking market microstructure effects into account.

2. Properties of transaction price changes

2.1. Discrete price changes. When considering differences between the analysis of the process of price changes on an aggregate level, e.g. daily returns, and the analysis of transaction data, the most prominent distinction is the discreteness of price changes. Discreteness of price changes is mainly due to institutional regulations setting permissible prices at multiples of a smallest divisor, called a ’tick’. The tick size varies from traded asset to traded asset and depends mainly on the price level at which one unit of the asset is traded. Exchanges typically fix the minimum tick size balancing a trade-off between providing an efficient grid for price formation and allowing market participants to realize a price reasonably close to their valuation of the asset. See the discussion in Harris (1994) on the economic aspects of the tick size of a traded asset. The analysis of discreteness was from the very beginning on not only concerned with these institutional aspects but also with different reasons for discreteness, e.g. habits, see e.g. Niederhofer (1965) and Niederhofer (1966). Those studies are primarily related to the fact that traders do not always use the whole range of prices allowed by the trading rules, but use actually a coarser grid which suffices their needs. This coarser grid is not fixed by trading rules but is used by custom. The degree of rounding is even related to information flow, see e.g. Ball, Torous, and Tschoegl (1985) for a study on the gold market or Harris (1991) for a similar analysis on price clustering at the NYSE.

Rounding is recognized as a serious cause of estimation error for the variance of a price process. This can be easily demonstrated on the basis of a simple example taken from Gottlieb and Kalay (1985). Assume that $b(t)$ is a Brownian motion defined on $t \in [0, 1]$, with zero drift and unknown variance σ^2. The rounded process is defined by

$$\bar{b}(t) = \begin{cases} \vdots & \\ -1 & \text{if } b(t) \in (-1.5, -0.5] \\ 0 & \text{if } b(t) \in (-0.5, 0.5] \\ 1 & \text{if } b(t) \in (0.5, 1.5] \\ \vdots & \end{cases} \tag{2.8}$$

For the variance of the continuous process there is a consistent estimator S_n of σ^2 available as

$$S_n = \sum_{t=0}^{2n-1} \left(b\left(\frac{t+1}{2n}\right) - b\left(\frac{t}{2n}\right)\right)^2$$

$$\text{plim}_{n\to\infty} S_n = \sigma^2$$

The same estimator applied to the discrete valued process however diverges

$$\bar{S}_n = \sum_{t=0}^{2n-1} \left(\bar{b}\left(\frac{t+1}{2n}\right) - \bar{b}\left(\frac{t}{2n}\right)\right)^2$$

$$\text{plim}_{n\to\infty} \bar{S}_n = \infty$$

The work of Gottlieb and Kalay (1985) and Ball (1988) concentrates on the derivation of volatility estimators corrected for observation rules like 2.8. They find that the effects of discreteness are higher for low priced assets with a low variance compared to high priced assets or assets with a high variance.

Hausman, Lo, and MacKinlay (1992) extend this work as they recognize that the observation rule in 2.8 is very restrictive. They employ a classical ordered probit approach which will be discussed and extended in chapter 3 and find that a linear regression approach suffers indeed from the well-known deficiencies of standard inference methods applied to discrete valued data. See e.g. the extended discussion in Judge, Griffiths, Hill, Lütkepohl, and Lee (1985, chap. 18.2.1). The ordered probit is less restrictive concerning the observation rule in (2.8) as it does not impose fixed intervals $\ldots, (-1.5, -0.5], (-0.5, -0.5], \ldots$ but allows for an additional flexibility by parameterizing these intervals as $(-\infty, \mu_1], (\mu_1, \mu_2], \ldots$ using model parameters μ_i. In particular they find that the probability distribution over the price changes implied by a conventional linear approach shows often serious deviations from the more appropriate conditional probabilities implied by a quantal response model.

In addition Hausman, Lo, and MacKinlay claim that on the basis of their empirical results the observation rule in (2.8) cannot be justified. The latter point needs however some careful interpretation, as Hausman, Lo, and MacKinlay work with a fixed volatility level of the unobservable continuous process and estimate the thresholds which were identified by Gottlieb and Kalay (1985) and Ball (1988) as $k + 1/2$ for $k = \ldots, -1, 0, 1, \ldots$. Prima facie it is not possible to estimate both, the thresholds which map the latent in the observable process and the level of the volatility of the latent process. This is the well-known identification problem of the scale of the latent variable, which necessitates the introduction of additional identifying assumptions. If indeed both sets of parameters are to be estimated another set of identifying restrictions on the scale of the latent variable needs to be introduced. This last idea will however not be pursued in this work. Thus, one can conclude that Hausman, Lo, and MacKinlay as opposed to Gottlieb and Kalay and Ball are just using a different set of parameters to accommodate for the proportions of observed price changes. However, the specification of Hausman, Lo, and MacKinlay involves more degrees of freedom, as they include an individual parameter μ_i for each threshold instead of just one parameter of the latent process, i.e. σ^2.

Nevertheless, the findings of Hausman, Lo, and MacKinlay (1992) indicate that discreteness matters for the analysis of the transaction process, if one is interested in the probability distribution of price changes and even more so, if the joint distribution with another economic variable like time between transactions is under consideration. Thus, this work will concentrate on models which are adapted to the price process which lives on a discrete grid.

2.2. Bid-ask spread. There is a long tradition in market microstructure theory to analyse trading costs and thereby in some sense the cost of liquidity in market maker markets. The theoretical background of trading costs is amply surveyed e.g. in O'Hara (1995), Goodhart and O'Hara (1997), or more recently in Madhavan (2000).

The focus of this subsection rests on the implications of the bid-ask bounce for empirical work. The movement of the transaction price is termed bid-ask bounce which takes place between the price offered to a buyer and the price offered to a seller. A further characteristic of the bid-ask bounce is that it may take place without a change in the level of the mid-point between the bid and the ask price.

Roll (1984) develops a simple approximate relationship between the autocorrelation function (ACF) of returns R_t and the percentage bid-ask spread s_r, which is defined as[1]

$$s_r = \frac{P_{a,t} - P_{b,t}}{\sqrt{P_{a,t} P_{b,t}}}, \tag{2.9}$$

using $P_{a,t}$ and $P_{b,t}$ as the untransformed ask and bid price. If we define R_t on the basis of transaction prices P_t as a growth rate $R_t := \frac{P_t - P_{t-1}}{P_{t-1}}$, then we can write the result of Roll (1984) as

$$s_r = 2\sqrt{-\text{Cov}\left[R_{t-1}, R_t\right]}. \tag{2.10}$$

Note that this implies a MA(1) structure of the return series. Campbell, Lo, and MacKinlay (1997, chap. 3.4.2) discuss the obvious deficiencies of the above relationship and provide some extensions suggested in the literature. The most prominent problem lies in the obvious fact that relationship (2.10) is only defined for a negative covariance.

The relationship between the bid-ask spread and the autocorrelation of returns yields inconsistent estimators of the processes' variance if they do not take into account the time series properties. French and Roll (1986) suggest in the context of a study of variance components the use of an adapted estimator which accounts explicitly for the first order autocorrelation in price changes. Alternatively, one might as well use a variance estimator in the form suggested by Newey and West (1987), which is consistent under a wide range of forms of serial dependency in first and second moments.

Harris (1990) extends the work of Roll (1984) to include the discreteness of price changes as in the model of Gottlieb and Kalay (1985). He shows that the discreteness of price changes induces a negative ACF and that the relationship (2.10) consequently overstates the bid-ask spread as it does not take into account discreteness. He shows that the variance estimator suggested by French and Roll (1986) captures the effect of the bid-ask bounce as well as the serial dependency due to

[1] Using the notation of Campbell, Lo, and MacKinlay (1997).

the discreteness of price changes. This correction compensates for the MA(1) effect already discussed.

Empirical studies on the bid-ask bounce include the work of Hausman, Lo, and MacKinlay (1992) allowing to calculate the explicit probability to observe a bounce. A similar methodology will be employed in the empirical part of this work. Other empirical studies relating to the time varying evolution of the spread are surveyed by Goodhart and O'Hara (1997, chap. 4.2).

2.3. Irregular time between transactions. The second striking feature of transaction data as opposed to aggregates either on an intra daily or daily level is the irregular occurrence of observations. This feature of the data has two important aspects. First, the arrival of transactions depends of the interaction of individuals buying or selling assets. Therefore, it is usually depicted as a stochastic process. Second, the occurrence of trades is subject to a strong intra daily seasonality, which is usually seen as a deterministic function of the time of day. For different assets there will be additional effects governing the arrival of transactions, e.g. the time to maturity in the life cycle of a futures contract. The intra daily seasonality of the volatility and other components of the transaction process such as trade intensity or trading costs is a well known problem in empirical research of market microstructure.

The problem of those patterns is that in contrast to the bid-ask spread, they are not easily explained by theory or as Goodhart and O'Hara (1997, p. 86) put it "The intriguing feature of this temporal intra daily pattern is that it has not proven easy to explain theoretically, at least using the basic model that splits agents in the market into informed, uninformed and market maker, ...". Although the economic background of these seasonalities is yet to be determined, they have quite stable and distinct patterns, even if one leaves markets apart, which trade 24 hours, like the foreign exchange market. Those seem to show the strongest variation due to seasonalities. The seasonalities of the U.S. markets as they are described e.g. by Wood, McInish, and Ord (1985) and other asset markets as they are amply surveyed by Goodhart and O'Hara (1997). Usually, they show a clear U-shaped pattern with peaks around the opening and the closing of markets. This is often attributed to a price formation phase in the morning and the need of traders to close positions they do not want to hold overnight. European markets on the other hand show quite often a peak in the morning and an even more pronounced high around 14.30 CET right after a trading low in the lunch hour. See e.g. the study by Gourieroux, Jasiak, and Fol (1999) based on trading in France. These seasonalities are mainly driven by the fact that traders from different regions trade in this market during their local business hours and trade with a lower intensity during their lunch hours. The most convincing hypotheses concerning this fact is that traders based in the U.S. become active at this point in time, as 14.30 CET corresponds to 8.30 ET.

The main contributions to the theoretical understanding of the informational content of time between transactions are Admati and Pfleiderer (1988), Foster and Viswanathan (1990), Foster and Viswanathan (1993), Diamond and Verrecchia (1987), and Easley and O'Hara (1992), Easley, Kiefer, O'Hara, and Paperman (1996), and Easley, Kiefer, and O'Hara (1997).

The models of Admati and Pfleiderer and Foster and Viswanathan rest mainly on the motivation of liquidity traders who have some discretion concerning the timing of their trades to trade when transaction costs are low and on the motivation of informed traders to pool with uninformed traders to hide their information from the market maker. Thereby it implies phases where informed and discretionary liquidity traders pool. In these phases a high liquidity, measured by trading volume, and a high volatility are observed. Trading frequency is not used as an information channel in this model but is only a consequence of the optimizing behaviour of market participants.

In the model of Diamond and Verrecchia (1987) time between transactions itself carries information via the introduction of short sales constraints. This model relies on a steady information flow and no trading would actually indicate that in the absence of constraints traders would sell, i.e. the informed market participant has a negative price signal.

The model of Easley and O'Hara (1992) extends this notion without the use of a short sales restriction by the simple assumption that informed traders trade only if they have private information available which is not already contained in the market price. In this model the market maker is able to identify periods in which no information is available by a lack of trades, as the trade intensity of noise traders is constant.

An entirely different approach was initiated by the work of Clark (1973). He puts forward the notion that the irregular occurrence of trades has a significant influence on the properties of aggregates. He recognizes that the liquidity of markets interpreted as the number of transactions per day plays an important role for the observed price changes over a given time span. These can be interpreted as the sum over a random number of random variables. In his empirical study he argues that the number of transactions can be approximated by trading volume, which in turn is related to the information flow. This notion is extended into explicit economic models by Epps and Epps (1976) and Tauchen and Pitts (1983). For a general survey on the volume-volatility relationship see Karpoff (1987) or more recently also Goodhart and O'Hara (1997).

Clark originally imposes that aggregate log returns r_t^a might be characterized by[2]

$$r_t^a = \sum_{t=1}^{\tilde{T}} r_t \tag{2.11}$$

assuming that r_t is i.i.d. with mean zero and variance σ^2. Furthermore, he assumes that $\tilde{T}$ is an integer valued random variable with mean α, independent of r_t. $\tilde{T}$ is the random variable which maps the number of transactions per time interval. The time interval is assumed to be given, e.g. five minutes, one trading day. Clark develops the distribution of r_t^a and finds that the unconditional variance of r_t^a is a function of σ^2 and α,

$$\mathrm{Var}\left[r_t^a\right] = \alpha\sigma^2, \tag{2.12}$$

[2]For expositional purposes the outlined model differs slightly from the original one, but makes use of the theorems developed by Clark on the asymptotic distribution of the random variables.

but more interestingly, he develops that the kurtosis of r_t^a is an increasing function of the variance of $\tilde{T}$. Thereby, he clearly relates the properties of the intraday process of transaction intensity to well known properties of aggregates. Clark's theoretical arguments are developed under quite restrictive assumptions. Particularly, the independence of the return process and the directing process is a crucial assumption. In the empirical section of this work it will be shown that his assumptions do not hold in practice and that a more elaborate theory is called for in this type of analysis.

3. Alternative hypotheses on the relationship of volatility and liquidity

3.1. A simple model of the market. To demonstrate the far reaching consequences of alternative relationships of liquidity and price changes and their volatility, a simple example in the form of a Markov chain is used to discuss alternative hypotheses in a more concise manner. Assume that a financial market can be described by four states at each point in time $t = 1, \ldots, T$:

s_1	no trade	no price change
s_2	trade	no price change
s_3	trade	price increases by one tick
s_4	trade	price decreases by one tick

The random variable which maps the state of the market S_t can thus take on four different values $S_t \in \{s_1, s_2, s_3, s_4\}$. If subsequent S_t were independent, the model would be fully described by the unconditional probabilities π to observe a state,

$$\pi_{[i]} = \text{Prob}\left[S_t = s_i\right], \; i = 1, \ldots, 4. \tag{2.13}$$

Using the vector d of price changes corresponding to a certain state the expectation and variance of a price change d_t is given by[3]

$$\begin{aligned} \mathrm{E}\left[d_t\right] &= \pi' d, & \mathrm{E}\left[d_t^2\right] &= \pi'(d^2), \\ \mathrm{Var}\left[d_t\right] &= \pi'(d^2) - d'\pi\pi' d, \\ \text{using } d' &= \begin{bmatrix} 0 & 0 & +1 & -1 \end{bmatrix}. \end{aligned} \tag{2.14}$$

This model shows no serial dependence and would imply that past price changes and trade frequency do not carry any information for future transactions, i.e.

$$\mathrm{Prob}\left[S_t = s_i \,|\mathcal{F}_{t-1}\right] = \mathrm{Prob}\left[S_t = s_i\right],$$

if $\mathcal{F}_{t-1}$ captures all the information up to and including $t-1$. This leaves however no room in the model to capture e.g. effects of information diffusion or other effects leading to a serial dependency in trade frequency or price changes. Nor allows this unconditional model for a relationship between liquidity and the volatility of the price process.

The serial dependency of trade frequency and price changes at the transaction level are well established facts, which have to be taken into account and which will be in the centre of the empirical analysis. For ease of exposition it is assumed that the conditional distribution of S_t given the information up to $t-1$, $\mathcal{F}_{t-1}$ is completely described by a Markov chain. This imposes the assumption that the conditional distribution of market states in t depends only on the state of the model in $t-1$, i.e. $\mathrm{Prob}\left[S_t = s_i \,|\mathcal{F}_{t-1}\right] = \mathrm{Prob}\left[S_t = s_i \,|S_{t-1}\right]$. The system dynamics are fully described by the transition matrix

$$\Pi_{[j,i]} = \mathrm{Prob}\left[S_t = s_i \,|S_{t-1} = s_j\right],\; i, j = 1, \ldots, 4. \tag{2.15}$$

[3] A vector raised to a power means in this context that each element is raised to the given power.

Given the state of the market at any point in time the probability distribution after l periods is simply given by[4]

$$\text{Prob}\left[S_{t+l}\,|S_t = s_i\right] = \pi(l,i) \tag{2.16}$$

$$= \left(\prod_{j=1}^{l} \Pi'\right) e_i \tag{2.17}$$

by construction, this conditional probability does not depend on t but only on the initial state s_i and the number of forecast periods l. The vector e_i is a unit vector for the dimension i, i.e. $e_{i[i]} = 1$ and $e_{i[j]} = 0$ for $i \neq j$. The unique stationary distribution π is defined if the Markov chain is positive recurrent and irreducible, properties which are given in this context by assumption as transient or absorbing states make no sense in this context and should thus be ruled out. See e.g. Resnick (1992, Prop. 2.14.1). The unique stationary distribution π can be obtained from the definition

$$\pi = \Pi'\pi$$

by solving for π, under given conditions, where $\imath$ is an appropriately defined vector of ones

$$\pi = \left(I - \Pi + \imath\imath'\right)^{-1}{}'\imath. \tag{2.18}$$

The conditional moments of the price changes are derived by replacing π in (2.14) by $\pi(l,i)$ from (2.17)

$$\begin{aligned} \text{E}\left[d_{t+l}\right| S_t = s_i] &= \pi(l,i)'d, \\ \text{E}\left[d_{t+l}^2\right| S_t = s_i\Big] &= \pi(l,i)'(d^2), \\ \text{Var}\left[d_{t+l}\right| S_t = s_i] &= \pi(l,i)'(d^2) - d'\pi(l,i)\pi(l,i)'d. \end{aligned} \tag{2.19}$$

This model allows for serial dependency in the conditional mean and conditional volatility and in the process of time between transactions.

[4]The product operator for matrices stands for a repeated multiplication from the left side.

The expected number of periods between transactions is calculated using the distribution after n periods. The calculation is eased by defining the probability to hit state j starting from state k after n periods, f_{jk}^{n}. See e.g. Resnick (1992, chap. 2.6). Collect those values in the vector f_k^n to obtain

$$f_k^n = \begin{cases} \Pi_{[\cdot,k]}, & \text{if } n = 1 \\ \Pi_{(k)}^{n-1} f_k^{n-1}, & \text{if } n > 1 \end{cases} . \tag{2.20}$$

For the unconditional expected time between transactions the derivation is straightforward from the distribution of hitting times and from the transition probabilities.[5]

$$\begin{aligned} \mathrm{E}\left[\tau_t\right] &= \pi_{[2:4]}' \imath + \sum_{t=2}^{\infty} \pi_{[1]}^{t-1} \pi_{[2:4]}' \imath t \\ &= \pi_{[2:4]}' \imath \sum_{t=1}^{\infty} \pi_{[1]}^{t-1} t \\ &= \pi_{[2:4]}' \imath \frac{1}{(1-\pi_{[1]})^2} \\ &= \frac{1}{\pi_{[2:4]}' \imath} \end{aligned} \tag{2.21}$$

Here, τ_t denotes the time between two transactions. Note that $\pi_{[2:4]}' \imath$ is the unconditional probability to observe a transaction.

After having laid out a simple model of the market and having briefly sketched the relationship between the dynamics of the model and conditional moments of price changes and the expected time between transactions, this simple model can be used to outline raw models of financial markets which are characterized by different relationships between volatility and liquidity.

3.2. The potential simultaneity between liquidity and volatility. The informativeness of past trade occurrences for the current price changes and the information content of past price changes for the present probability to observe a transaction can be analysed on the basis of individual components of the transition matrix Π.

[5] This uses the fact that $\sum_{t=1}^{n} x^{t-1} t = \frac{1-(n+1)x^n + nx^{n+1}}{(1-x)^2}$, $x \neq 1$.

The notation is considerably lightened by introducing the following partition

$$\Pi = \left[\begin{array}{c|c} \pi^{(1)} & \pi^{(3)\prime} \\ \hline \pi^{(2)} & \pi^{(4)} \end{array}\right]. \tag{2.22}$$

Furthermore, two matrices Π^a and Π^d are defined. Π^a is a (2×2) matrix which contains the conditional probability to observe a transaction given the past state of the market. The (3×3) matrix Π^d collects the transition probabilities for the case that two transactions are observed consecutively.

If the type of price change observed is not informative for the probability to observe a transaction in the current period then the following equality should hold

$$\text{Prob}\left[S_{t+1} = s_1 \left| S_t = s_i \right.\right] = \text{Prob}\left[S_{t+1} = s_1 \left| S_t \in \{s_2, s_3, s_4\} \right.\right], \quad i = 2, \ldots, 4,$$

i.e. $\pi^{(2)}$ should contain three identical elements and is simply given by

$$\pi^{(2)} = \Pi^a_{[2,1]}\imath. \tag{2.23}$$

If the occurrence of a trade in the last period has no implication for the present distribution, then

$$\text{Prob}\left[S_{t+1} = s_i \left| S_t = s_1 \right.\right] = \text{Prob}\left[S_{t+1} = s_i \left| S_t \in \{s_1, s_2\} \right.\right] \quad i = 2, \ldots, 4,$$

i.e. the distribution over the potential price changes depends only on the fact that no price changes were observed and it is straightforward to derive

$$\pi^{(2)} = \Pi^a_{[1,2]}\Pi^d_{[1,\cdot]}. \tag{2.24}$$

For $\pi^{(1)}$ and $\pi^{(4)}$ we obtain consequently

$$\begin{aligned} \pi^{(1)} &= \Pi^a_{[1,1]} \\ \pi^{(4)} &= \Pi^a_{[2,2]}\Pi^d. \end{aligned} \tag{2.25}$$

A dynamic is obviously not ruled out under this particular specification of Π. All these restrictions amount to is that the type of trade should not alter the conditional distribution of the trade intensity et vice versa.

Under this specification it is quite easy to show that the liquidity of the market can be assessed independently from the conditional moments of the process of price changes, in particular independently from the volatility of the process. To do so S_t is decomposed into a variable $a_t \in \{1, 0\}$ which indicates whether a trade occurs, $a_t = 1$, at time t or not, $a_t = 0$. The other variable $d_t \in \{-1, 0, 1\}$ maps the price changes observed at t. The observation $d_t = 0$ might arise if either there was no trade observed, $S_t = s_1$, or there was no price change, $S_t = s_2$, at time t.

The relationship of both variables is determined by properties of the conditional joint distribution

$$\begin{aligned}\text{Prob}\left[S_t \,|S_{t-1}\right] &= \text{Prob}\left[a_t, d_t \,|a_{t-1}, d_{t-1}\right] \\ &= \text{Prob}\left[d_t \,|a_t, a_{t-1}, d_{t-1}\right] \text{Prob}\left[a_t \,|a_{t-1}, d_{t-1}\right]. \qquad (2.26)\end{aligned}$$

If the type of price change observed is not informative for the probability to observe a transaction in the current period as in (2.23) then we have

$$\text{Prob}\left[a_t \,|a_{t-1}, d_{t-1}\right] = \text{Prob}\left[a_t \,|a_{t-1}\right].$$

Although the random variables a_t and d_t cannot be independent which can be easily seen by noting that

$$\text{Prob}\left[d_t \,|a_t, a_{t-1}, d_{t-1}\right] \neq \text{Prob}\left[d_t \,|d_{t-1}\right]$$

in (2.26). It is still possible that the liquidity of the market contains no information for individual moments of the process which is described using the following scenarios of practical interest. Liquidity is defined in this context as the conditional probability to observe a trade given a particular information set, $\text{Prob}\left[a_t = 1 \,|\cdot\right]$. The direction and the size of price changes refer to the conditional expectations $\text{E}\left[d_t | \cdot\right]$ and $\text{E}\left[d_t^2 | \cdot\right]$.

Independence: Liquidity and the size of price changes are two independent factors.

Direction: Liquidity is informative for the direction of price changes but not vice versa and it is not informative for the size of price changes.

Size: Liquidity is informative for the size of price changes but not vice versa and it is not informative for the mean of price changes.

The aforementioned hypotheses on various degrees of dependence between the process of price changes and the liquidity process translate into several hypotheses on the elements of $\pi^{(2)}$ and $\pi^{(3)}$.

$$\pi^{(3)} = \begin{bmatrix} p_{1,2} & p_{1,3} & p_{1,4} \end{bmatrix} \tag{2.27}$$

The hypothesis *Independence* was already discussed in the context of the decomposition of S_t into a_t and d_t. From the arguments given there it is clear that a_t and d_t cannot be independent by definition.

The hypothesis *Direction* and *Size* translate to a structure, where the conditional probabilities of $\pi^{(3)}$ are not necessarily given by the product of the marginal probabilities as in (2.24) and (2.23) but restrictions on the conditional expectations are imposed. If $\pi^{(2)}$ also deviates from the product of marginals restriction then the type of trade would also carry information for the trade occurrence.

As a special case *Size* arises if $\pi^{(3)}$ is of the form (2.23) and $\pi^{(2)}$ deviates from (2.24) but the conditional expectation of the price change given that there was no trade is identical with the conditional expectation given that there was no price change, if

$$\mathrm{E}\left[d_t \middle| S_{t-1} = s_1 \right] = \mathrm{E}\left[d_t \middle| S_{t-1} \in \{s_1, s_2\} \right]. \tag{2.28}$$

This means that probability mass between $p_{1,2}$ and $\pi^{(1)}$ can be shifted without changing the conditional expectation compared to the one implied by (2.24).

The scenario *Direction* corresponds to a deviation of $\pi^{(2)}$ from (2.24) while maintaining the restriction that the conditional expectation of the squared price change given that there was no trade is identical with the conditional expectation given that there was no price change, i.e. if one is not able to distinguish between the no trade and the no price change states s_1 and s_2.

$$\mathrm{E}\left[d_t^2 \middle| S_{t-1} = s_1 \right] = \mathrm{E}\left[d_t^2 \middle| S_{t-1} \in \{s_1, s_2\} \right]. \tag{2.29}$$

This amounts to shifting probability mass between $p_{1,3}$ and $p_{1,4}$ compared to the reference in (2.24). It is only necessary to maintain the sum of $\pi^{(1)}$ and $p_{1,2}$ constant.

Neither *Direction* nor *Size* rule out that past price changes have an influence on the expected price change. This would lead to a model where the past state is not informative for the mean at all, then

$$\mathrm{E}\left[d_t \middle| S_{t-1} = s_i \right] = \mathrm{E}\left[d_t \right] \qquad \text{for } i = 1, \ldots, 4$$

which implies that

$$\Pi_{[i,3]} - \Pi_{[i,4]} = \pi_{[3]} - \pi_{[4]}. \tag{2.30}$$

This restriction is much stronger than the ones involved in *Direction* and *Size*. Note that (2.30) imposes only restrictions on the differences of probabilities. Shifting probability mass between s_1 and s_2 will not affect the mean. The same is true for a shift between s_1, s_2 and s_3 and s_4 as long as the difference is held constant.

The latter shift in probability however affects the conditional volatility of the process as is obvious from the restrictions implied by the restriction where the conditional expectations of the squared price changes are supposed to equal the unconditional ones, i.e.

$$\mathrm{E}\left[d_t^2 \middle| S_{t-1} = s_i \right] = \mathrm{E}\left[d_t^2 \right],$$

$$\Pi_{[i,3]} + \Pi_{[i,4]} = \pi_{[3]} + \pi_{[4]}.$$

The two special cases just discussed should make clear that it is indeed possible to construct models in which trade frequency is only informative for either the conditional mean or the conditional variance of the process, and of course the absence of this relationship does not rule out a dynamic in the mean or in the variance.

Part 2

Models of the price process

CHAPTER 3

The dynamics of transaction price changes

1. Count models for discrete transaction price changes

1.1. Modeling approaches. Most of the early literature taking the existence of market microstructure into account attempted to develop estimators, e.g. volatility estimators, corrected for the adverse effects of the bid-ask bounce or minimum tick size. This impeded however a thorough analysis of financial markets at the transaction level as it did away with effects which might actually help to explain the workings of financial markets. In particular this is true for the analysis of liquidity, since most of the early work does not distinguish between the effects of trade frequency and the incremental information contained in traded volume. See e.g. the extensive literature reviews in Karpoff (1987) and Goodhart and O'Hara (1997). The economic virtues of a disaggregate analysis were already discussed in the preceding chapter. This chapter and the following outline a modeling framework for price changes and absolute price changes and for the time between transaction on a disaggregate level. No attempt is made in this chapter to develop methods to reduce the effects of discreteness but it is attempted to propose a suitable model for the process of transaction price changes. It has already been recognized that the discreteness of observable price changes d_t is a significant feature of the data so that some type of count data model has to be used in the analysis of d_t. If an analysis of the trading process is carried out at the transaction level three approaches to model a discrete valued processes can be distinguished in general[1]:

(1) Discrete valued distributions
(2) Generalized linear models

[1]See e.g. Lee (1992) for a similar categorization.

(3) Discrete observations of an unobservable continuous variable

A thorough exposition of applications using discrete valued distributions and count data models can be found e.g. in Cameron and Trivedi (1996) and Cameron and Trivedi (1998). The typical reference for generalized linear models is the monograph by McCullagh and Nelder (1989). Cameron and Trivedi (1998) contains also a survey of generalized linear models in the context of count data applications with an extended treatment of dynamic models. Quantal response models are discussed by Maddala (1983) in the context of economic applications.

Before embarking on a concise discussion of alternative models we characterize briefly the three types of models considered above:

(1) When discrete distributions are applied to the analysis of price changes, it is assumed that the observable price changes d_t are generated from a discrete distribution with a continuous parameter λ linked to some conditional moments of d_t. For example in the Poisson model we have

$$\lambda_t = \mathrm{E}\left[d_t | \, x_t \right] = \mathrm{Var}\left[d_t | \, x_t \right] \tag{3.1}$$

using e.g. a linear parameterization with regressors x_t and coefficients β

$$\lambda_t = \exp(x_t'\beta). \tag{3.2}$$

The Poisson model is obviously not a serious candidate for an appropriate model of price changes, as it is far too restrictive on the necessary properties of the DGP. First of all, it is not straightforward to modify d_t in a way to take on only positive values. Second, the strict relationship between the conditional expectation and the conditional variance is almost surely violated. Cameron and Trivedi (1998, chap. 4) give a broad survey over sensible alternative count data models. The extension of the stationary framework to incorporate some form of dynamics is provided e.g. by Harvey and Fernandes (1989). They employ an iterative scheme to incorporate the information contained in observations into the conditional moments of the random variable using conjugate distributions.

(2) Generalized linear models are based on the specification of an arbitrary link function G, which is a monotonic mapping from the support of the observed random variable, e.g. $[0, 1]$, to $(-\infty, +\infty)$. This circumvents the problem of tailoring a distribution to the particular dependent variable and allows to use standard regression methods. In contrast to the explicit specification of a distribution for the dependent variable, these models rely on a pseudo-maximum likelihood (PML) estimator to obtain parameter estimates. It will be shown that in some special cases the link function G can be related to a distributional assumption yielding a proper maximum likelihood interpretation in the context of the third model class considered here. For a binary response model with regressors x_t a typical model specification leaving the distribution of d_t unspecified would be

$$G(\mathrm{E}\left[\left. d_t \right| x_t\right]) = x_t'\beta \tag{3.3}$$

Russell and Engle (1998) propose a generalized quantal response model based on the work of Shephard (1995) and Zeger and Qaqish (1988) on generalized linear autoregressive (GLAR) models. The main target of this research is to propose an ARMA-type framework for the analysis of discrete valued time series. Rydberg and Shephard (1998) propose to decompose the transaction process into parts which can in turn be modeled by distinct generalized linear models.

(3) Closely related to GLMs are models based on the discrete observation, d_t, of a latent continuous and unobservable variable, d_t^*. Both random variables are linked through a deterministic function. In the case of e.g. an ordered probit with three categories the mapping works through a threshold

function

$$d_t = \begin{cases} -1 & \text{if} \quad d_t^* < \mu_1 \\ 0 & \text{if} \quad \mu_1 \leq d_t^* < \mu_2 \\ 1 & \text{if} \quad d_t^* \geq \mu_2 \end{cases}, \tag{3.4}$$

where the distribution of the observable random variable d_t follows from the distributional assumptions about the latent model, e.g.

$$d_t^* = x_t'\beta + u_t$$

$$\text{with } u_t \sim N(0, 1).$$

A general description of ordered response models is given later on in the chapter.

This type of model was first applied to the analysis of transaction price changes by Hausman, Lo, and MacKinlay (1992). Cameron and Trivedi (1998, chap. 3.6) point out that the ordered discrete choice model is particularly attractive if the random variable is similar to a count, but also takes on negative values. This certainly stretches the notion of a count variable. Although it seems awkward to speak of negative count variables, this is well justified on the grounds of the properties of the models involved. Similar to a count variable means in this context that the dependent variable takes on only few distinct values which can be ordered from smallest to largest and which are measured on a metric scale. Most of the advantages of this class of models are related to the fact that the latent model takes the form of a standard linear regression model.

This last point is made more concise in the work of Gourieroux, Monfort, Renault, and Trognon (1987) who describe the relationship of the likelihood and conditional moments of the latent and the observable model. In the course of this chapter we will exploit this relationship and construct a dynamic for the latent model. This is one of the major contributions of this work. We will demonstrate that particularly the last type of modeling strategy is easy to estimate and is a useful basis for various extensions. In subsequent chapters the ordered probit model will be used as a solid building block for nonlinear multivariate systems.

Before turning to the development of a dynamic model in this context, we will first summarize other attempts to formulate a dynamic model for limited dependent variables in the context of generalised linear models.

1.2. Generalized linear models.

1.2.1. *Static models.* The main problem of modeling a count variable, i.e. a variable which takes on only a few distinct values on a metric scale, boils down to the crucial point that such a random variable has a limited support as opposed to the usual parameterization chosen in econometrics which features a mean function, e.g. $x'\beta$ or $\exp(x'\beta)$, taking on values in $(-\infty, \infty)$ or $(0, \infty)$. Thus the standard approach allows for forecasts which cannot possibly be observed in realisations of the random variable. This is a feature which is clearly undesirable. The solution proposed by Nelder and Wedderburn (1972) is the use of a link function $G(\cdot)$, which maps the domain of the random variable, e.g. $[0, 1]$ to the domain of the systematic component, typically $(-\infty, \infty)$. This function should be differentiable and monotone.

Generalized linear models (GLM) are usually described to consist of three components:

- A random component, here d_t.
- A systematic component, here m_t.
- A link function between random and systematic components, $G(\cdot)$.

Let us first consider the systematic component, i.e. the mean function. We denote the conditional expectation of the observable dependent variable as $\mu_t = \mathrm{E}\left[d_t | \, x_t\right]$, limiting our attention for the time being to a static context with weakly exogenous regressors x_t. The expectation μ_t of d_t is not modeled as constant but is to be conditioned on regressors x_t, so the systematic component m_t is in general specified as a linear function

$$m_t = x_t'\beta. \tag{3.5}$$

See Cameron and Trivedi (1998, p. 34). Alternatively, the systematic component can be specified to mimic an ARMA model, as will be outlined in the following subsection.

The link function G relates the conditional expectation of the random component μ_t to the systematic component m_t

$$G(\mu_t) = m_t$$

Considering e.g. Probit or Logit models in the generalized linear context, the corresponding link functions are

$$\begin{aligned} &\text{logit} \quad m_t = \log(\mu_t/(1-\mu_t)) \\ &\text{probit} \quad m_t = \Phi^{-1}(\mu_t) \end{aligned}$$

which imply the standard Probit or Logit. This can be easily seen if the link function is inverted to obtain

$$\mathrm{E}\left[d_t | \, x_t\right] = \mu_t = \Phi(x_t'\beta) \tag{3.6}$$

$$= \mathrm{Prob}\left[d_t = 1 \, | x_t\right]. \tag{3.7}$$

After these introductory remarks on static models, we focus on dynamic models which have been proposed in this context.

1.2.2. *Generalized Linear Autoregressive Models.* [2] In order to account for dynamics in the dependent variable d_t it has regularly been suggested in the literature to augment the systematic component by functions g_i of past observations of the dependent variable[3] and parameters ϕ_i to obtain

$$m_t = x_t'\beta + \sum_{i=1}^{p} \phi_i g_i(d_{t-i}). \tag{3.8}$$

[2]Henceforth abbreviated GLAR.

[3]The functions g_i should not be mistaken with the link function G() of the GLM.

In contrast to the model by Zeger and Qaqish (1988) who use heuristic arguments to motivate the form of g_i, Shephard (1995) uses a first order Taylor series expansion of the link function around the systematic component m_t to decompose the link-transformed observed value into the systematic component m_t and an error term by

$$G(d_t) \approx m_t + \frac{\partial G}{\partial \mu}(d_t - \mu_t) =: z(d_t). \tag{3.9}$$

Using $g_i(d_{t-i}) = z(d_{t-i})$ the systematic component can be written as

$$m_t := x_t'\beta + \sum_{j=1}^{p} \phi_j z(d_{t-j}) \tag{3.10}$$

$$z(d_t) := x_t'\beta + \sum_{j=1}^{p} \phi_j z(d_{t-j}) + c_t \tag{3.11}$$

$$c_t := G'(\mu_t)(d_t - \mu_t) \tag{3.12}$$

Shephard (1995) argues that term c_t takes on the role of an error term, given that it has the properties of a martingale difference sequence. In the context of a GLAR(1) with $m_t = \beta + \phi z(d_{t-1})$ where β is a constant, the lag term has the form

$$z(d_t) = m_t + c_t \tag{3.13}$$

$$= \beta + \phi z(d_{t-1}) + c_t \tag{3.14}$$

$$= \frac{\beta}{1-\phi} + \sum_{i=0}^{\infty} \phi^i c_{t-i} \tag{3.15}$$

We will see that the term $\frac{\partial m_t}{\partial \beta}$ is needed for estimation. In this context, this term needs to be calculated recursively. Shephard (1995) suggests to iterate on

$$D_t' = \frac{\partial m_t}{\partial \beta}\frac{\partial \mu_t}{\partial m_t} \qquad \frac{\partial m_t}{\partial \beta} = \left(\begin{bmatrix} x_t \\ z(d_{t-1}) \end{bmatrix} + \frac{\partial z(d_{t-1})}{\partial \beta}\phi \right)$$

$$\frac{\partial z(d_{t-1})}{\partial \beta} = \frac{\partial m_{t-1}}{\partial \beta} + \frac{\partial c_{t-1}}{\partial \mu_{t-1}} D_{t-1}'$$

1.2.3. *Extension to generalized linear autoregressive models.* Shephard (1995) as well as Russell and Engle (1998) suggest the inclusion of MA-like terms by extending the systematic component to obtain

$$m_t := x_t'\beta + \sum_{j=1}^{p} \phi_j z(d_{t-j}) + \sum_{j=0}^{q} \theta_j c_{t-j} \tag{3.16}$$

where $\theta_0 = 1$. The formulation of Russell and Engle (1998) looks at first sight somewhat broader since they specify c_{t-j} not necessarily by (3.12) but as

$$c_t = \Psi_{t-j} \left(d_{t-j} - \mu_{t-j} \right). \tag{3.17}$$

Ψ_{t-j} is a function which supposedly corrects for the dispersion of $(d_{t-j} - \mu_{t-j})$. Similar to the earlier dynamic GLM a linear heuristic is employed to include the history of the dependent variable. Furthermore, Russell and Engle (1998) suggest to include a set of terms $\sum_{j=1}^{q} \zeta_j d_{t-j}$ into the systematic component which seems inappropriate, if one recalls that $z(d_t)$ is a Taylor expansion of d_t. If one feels that the approximation is insufficient one should rather increase the order of the Taylor expansion. It is also discussed by Russell and Engle (1998) that weakly exogenous variables might be included in the specification to obtain something like a distributed lag model. The interpretation of such a model seems however complicated. In the linear case, such a model would be described as a distributed lag model and lagged values of the endogenous variable are included to account parsimoniously for a possibly infinite lag structure of the exogenous variables. Another similarity to the plain linear case is also noteworthy, namely the fact that the log-likelihood may not have a unique maximum, as is also noted by Shephard (1995). The virtue of an inclusion of MA-terms is to be seen in the reduction of model parameters as opposed to a model which consists only of AR terms. As it is well known from the linear model AR and MA terms may cancel out, see e.g. Hamilton (1994). Similar effects might also be experienced in this context.

To sum up the extensions, the full model has the following form, using coefficients ζ and ξ

$$m_t := x_t'\beta + \sum_{j=1}^{p} \theta_j \Psi_{t-j}(d_{t-j} - \mu_{t-j}) + \sum_{j=1}^{q} \zeta_j d_{t-j} + \sum_{j=1}^{r} \xi_j m_{t-j}. \tag{3.18}$$

Russell and Engle (1998) suggest to specify $\Psi_{t-j} = V_t^{-1/2}$ and V_t is the conditional variance $V_t := \mathrm{Var}\left[d_t | \, x_t, d_{t-1}\right]$. In the model of Russell and Engle the dependent variable is a dummy vector, as described in the next subsection, nevertheless their proposed methodology translates directly to the univariate case considered by Shephard.

1.2.4. *GLM as extensions to Markov chains.* The Markov chain models fit in the scheme of models given in the previous chapter as a direct model for the probabilities to experience a certain price change, given the type of price change observed last. If one identifies the K states of the model with the different price changes v_k, $k = 1, \ldots, K$ which are observable. A discrete time Markov chain is completely described by the transition probabilities $\Pi_{[ij],t}$ and the probability distribution of the first element of the chain $\pi_{[i],0}$. To be concise, the evolution of the probability to observe a certain price change $d_t = v_k$ at a certain point in time t only depends on the probability distribution at $t-1$ and not on the point in time, i.e. t itself so that the chain is stationary[4]

$$\Pi_{[ij]} = \Pi_{[ij],t} \tag{3.19}$$

$$= \mathrm{Prob}\left[d_t = v_i \, | d_{t-1} = v_j\right] \tag{3.20}$$

[4]If one considered a cross section of observations this property would rather be termed 'homogeneous'.

If one collects the indicator functions $1_{(d_t=v_i)}$ in the dummy vector $D_t := \sum_{i=1}^{K} e_i 1_{(d_t=v_i)}$ the log likelihood function of this model is[5]

$$\log\,\mathfrak{L} = \sum_t D'_{t-1}\,(\log\Pi)\,D_t + D_0\log\pi \tag{3.21}$$

See e.g. Amemiya (1985, chap. 11).

The model just outlined has some obvious shortcomings which will be discussed in turn. As a benchmark however it will remain quite important. Given the Markov property that the transition probabilities just depend on the present state, there is a need to incorporate a more flexible dynamic in this model. There are two paths which have been pursued in previous research. The first possible extension is to employ an extended state space. A straightforward method to do this is to employ an l dimensional cross-product to describe the present state on the Markov chain , i.e. the transition probabilities are modified to

$$\text{Prob}\left[d_t = v_i \,\middle|\, [d_{t-1},\ldots,d_{t-l}] = v_{(l)j}\right]. \tag{3.22}$$

Now the number of states and the number of observable price changes is no longer identical. If J is the number of different possible price changes, the Markov chain has J^K states and the $(l\times 1)$ vector $v_{(l)j}$, with $j = 1,\ldots,J$, describes each individual state. This approach has the virtue that the dynamics are completely described by the $(J \times J^K)$ transition probabilities Π and the necessary time series probabilities are easily verified using standard results for Markov chains. Yet, this approach entails the considerable problem that the number of parameters increases very fast. Russell and Engle (1998) also start out from the Markov chain model but give up the methodology of using a constant transition matrix for the sake of introducing a richer dynamic. If one abandons the stationarity assumption, the transition probabilities are serially dependent and in order to achieve an identifiable version of eq. 3.21 one needs to find a feasible expression for $\Pi_{[ij],t}$. In analogy to Amemiya (1985, chap. 11.1.3) one might use the probabilities implied by an ordered probit to introduce a serial dependence in the transition probabilities.

[5] Note that $\log A := [\log a_{ij}]_{ij}$.

Russell and Engle (1998) define their model via a $(K-1\times 1)$ dummy vector

$$D_t^* := \sum_{i=1}^{K-1} e_i 1_{(d_t=v_i)},$$

for which they model the conditional probabilities to observe the individual categories given past observations on the price changes D_{t-i}^* and weakly exogenous variables x_t collected in $\mathcal{F}_{t-1}$ as

$$\mu_t := \mathrm{E}\left[D_t^* \middle| \mathcal{F}_{t-1}\right],$$

$$\mathcal{F}_{t-1} = [x_t, D_{t-1}^*, x_{t-1}, D_{t-2}^*, x_{t-2}, \ldots].$$

The conditional variance of D_t^* implied by this specification is

$$\mathrm{Var}\left[D_t^* \middle| \mathcal{F}_{t-1}\right] = \mathrm{diag}\, \mu_t - \mu_t \mu_t'.$$

Abstracting from exogenous regressors and concentrating on the first order lag, we have a Markov chain. We use $\bar{\pi}$ as the unconditional distribution over $K-1$ non-redundant states and constant vector μ, then the model takes on the form

$$\mu_t = B^* D_{t-1}^* + \mu \tag{3.23}$$

$$= \bar{\pi} + B^*(D_{t-1}^* - \mu) \tag{3.24}$$

$$\text{with } \mu = (I - B^*)^{-1}\bar{\pi} \tag{3.25}$$

using a coefficient matrix B^*. A more flexible type of dynamic based on past state probabilities using A^* and C^* as coefficient matrices can be formulated as

$$\mu_t = A^*(D_{t-1}^* - \mu_{t-1}) + C^*\mu_{t-1} + \bar{\pi} \tag{3.26}$$

$$= A^* D_{t-1}^* + (C^* - A^*)\mu_{t-1} + \bar{\pi}$$

which is interpreted by the authors as an ARMA-type process. Russell and Engle (1998) give bounds on A^*, $C^* - A^*$, and $\bar{\pi}$ so that the probabilities μ_t take on only permissible values. The k step ahead forecast apparently takes on in both cases the same form as

$$\mathrm{E}\left[\mu_{t+k} \middle| \mu_t\right] = \bar{\pi} + (C^*)^k (\mu_t - \mu), \tag{3.27}$$

where C^* is replaced by B^* in the standard Markovian case.

Nevertheless, the use of μ_t, i.e. a probability, as a dependent variable raises substantial problems, as the dynamic of μ_t needs to account for the limited support of μ_t. In simple cases this can be achieved by straightforward parameter restrictions. In a more general setting however they propose to parameterize the transform of μ_t based on the link function G. This bounds the values implied by the model to the appropriate interval $[0, 1]$. The analogue specification to the modified Markov chain in (3.26) would be

$$\begin{aligned} G(\mu_t) = A(D^*_{t-1} - \mu_{t-1}) + CG(\mu_{t-1}) + \bar{\pi} \\ = m_t \end{aligned}$$

The k step ahead forecast of the stochastic component takes on a form known from the standard Markov case

$$\mathrm{E}\left[\mu_{t+k} | \, \mu_t\right] = \bar{G} + C^k(\mu_t - \bar{G}) \tag{3.28}$$

$$\text{with } \bar{G} = (I - C)^{-1}\bar{\pi} \tag{3.29}$$

The model can be augmented by higher order lags to obtain

$$m_t = \sum_{j=1}^{p} A_j(D^*_{t-j} - \mu_{t-j}) + \sum_{j=1}^{q} C_j m_{t-j} + \bar{\pi}. \tag{3.30}$$

This model has a structure similar to an ARMA(p,q) model. Other possible extensions, i.e. exogenous regressors and lags of D^*_t itself and the scaling of the difference term $(D^*_{t-j} - \mu_{t-j})$ were already discussed in the preceding subsection, see (3.18).

1.2.5. *PML estimators for generalised linear models.* Up to this point different specifications of the mean function m_t have been outlined which extend standard generalised linear models to include dynamics. The estimation of parameters in the mean function m_t relies typically on quasi/pseudo maximum likelihood estimators (PMLEs). See e.g. White (1982) or Gourieroux, Monfort, and Trognon (1984). The distribution f_d used in the model and indexed by the parameters of interest θ and nuisance parameters σ is not necessarily assumed to be identical with the population distribution, as in classical maximum likelihood estimation (MLE). The key issue is the selection of f_d from the linear exponential family (LEF), which allows parameters of interest θ to be consistently estimated, if they are parameters of the mean function. Apart from the limitation to the LEF, the criterion to choose a pseudo-true distribution f_d is primarily the ease of estimation. This allows us prima facie to focus on the specification of the mean function, without a need to explore further distributional details. The simplification and increased robustness of PMLE compared to MLE, i.e. consistency of the parameters estimates of interest $\hat{\theta}$ even in the case of a severe misspecification of the distribution of random variables in the models, comes at a cost. First, the derivation of consistency and normality of the estimators relies on asymptotic arguments, which demands some care in the application of estimators to small samples. Second, the efficiency of the estimator could be greatly increased if the true population distribution was used.

To keep the exposition as simple as possible we consider first the static context and illustrate then how dynamic models can be accommodated and outline some additional considerations necessary in the time series context. Following the exposition in Cameron and Trivedi (1998, p. 34) we concentrate first on one observation of the random component d_t, which is assumed to have a constant mean μ. The pseudo-true distribution of the random variable is again denoted by $f_d(u; m(\theta), \sigma)$, θ and σ being parameters of the distribution. We should however keep in mind that the true distribution of d_t is not necessarily from this family, but f_d is merely chosen as a vehicle for the estimation of θ.

The notation in the context of GLMs is usually somewhat different from the standard notation chosen for the LEF, see Cameron and Trivedi (1998, p. 34). Here we use

$$f_d(u, \theta, \sigma) = \exp\left((um(\theta) - b(m(\theta)))\frac{1}{a(\sigma)} + c(u, \sigma)\right), \tag{3.31}$$

where $a(\sigma)$ is usually assumed to be equal to σ, and σ is a nuisance parameter capturing dispersion.

The component $b(m)$ is a function of the mean m of the process and $c(u, \sigma)$ is a scaling factor.

We consider the log likelihood implied by the pseudo-true density

$$\log \mathfrak{L} = \sum_{t=1}^{T} \log f_d(d_t, m, \sigma) = \sum_{t=1}^{T} \left((d_t m - b(m)) \frac{1}{a(\sigma)} + c(d_t, \sigma) \right). \tag{3.32}$$

The first order condition implied by the maximisation of the log likelihood with respect to m is of course the score

$$\frac{\partial \log \mathfrak{L}}{\partial m} = \sum_{t=1}^{T} \frac{\partial \log f_d(d_t, m, \sigma)}{\partial m}, \tag{3.33}$$

if we limit our attention to the mean parameter m. One representative element of the score reveals the key idea behind the PML estimation of the generalised linear models

$$\frac{\partial \log f_d(d_t, m, \sigma)}{\partial m} = \frac{d_t - \partial b(m)/\partial m}{a(\sigma)}. \tag{3.34}$$

Here, we reveal $\partial b(m)/\partial m$ as the unconditional mean of d_t. In order to derive the relationship between $b(m)$ and $a(\sigma)$ and the moments of the observable variable two properties of the log-likelihood function derived from the information equality are employed.

$$\mathrm{E}\left[\frac{\partial f_d}{\partial m}\right] = 0 \qquad \mathrm{E}\left[\frac{\partial^2 f_d}{\partial m^2}\right] + \mathrm{E}\left[\frac{\partial f_d}{\partial m}\right]^2 = 0 \tag{3.35}$$

In conjunction with the corresponding derivatives

$$\frac{\partial f_d}{\partial m} = \frac{d - b'(m)}{a(\sigma)} \qquad \frac{\partial^2 f_d}{\partial m^2} = \frac{-b''(m)}{a(\sigma)}. \tag{3.36}$$

the variance of d_t can be found as

$$\mathrm{Var}\left[d_t\right] = b''(m)a(\sigma). \tag{3.37}$$

After this, the form of the static and dynamic GLMs with a time varying mean function is now obvious, once we have identified b' as the inverse of the link function. We replace the constant mean m by a suitable mean function as e.g. (3.5) or (3.8) and obtain a likelihood from the prediction error decomposition. See e.g. Harvey (1990, ch. 3.5).

An extension which allows to model heteroskedasticity conditional on observable variables or functions thereof could possibly devised in the context of the quadratic exponential family, along the lines of the discussion in Gourieroux and Monfort (1995).

Gourieroux and Monfort (1995, Property 8.17) show that under certain regularity conditions a consistent pseudo maximum likelihood estimator associated with a linear exponential family is asymptotically normally distributed with

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\left(J^{-1} I J^{-1}\right) \tag{3.38}$$

where

$$I = \mathrm{E}\left[\frac{\partial m'}{\partial \theta} \Sigma^{-1} \Omega \Sigma^{-1} \frac{\partial m}{\partial \theta'}\right] \tag{3.39}$$

$$J = \mathrm{E}\left[\frac{\partial m'}{\partial \theta} \Sigma^{-1} \frac{\partial m}{\partial \theta'}\right] \tag{3.40}$$

Where Ω is the conditional variance of d given x based on the DGP, and Σ is the corresponding conditional covariance based on the pseudo true distribution.

To clarify the relationship between GLMs and PML estimation the first order condition of a GLM can be easily derived from (3.31) as

$$\begin{aligned}
\frac{\partial \log f}{\partial \theta} &= 0 \\
\frac{\partial \mu_t}{\partial \theta} \frac{\partial m_t}{\partial \mu_t} \cdot \frac{\partial \log f}{\partial m_t} &= \\
\frac{\partial \mu_t}{\partial \theta} \left(\frac{\partial^2 \mu_t}{\partial m_t^2}\right)^{-1} \cdot \frac{1}{\sigma} \left(d_t - \frac{\partial b(m_t)}{\partial m_t}\right) &= \\
&= \mathrm{Var}\left[d_t\right]^{-1} (d_t - \mu_t) \frac{\partial \mu_t}{\partial \theta}.
\end{aligned} \tag{3.41}$$

This shows that the first order conditions implied by the GLMs and the corresponding PML estimator are indeed identical. It is also quite obvious, that the validity of the first order condition does not depend on the dispersion parameter σ. This also shows the close relationship to the weighted least squares estimator, see also the discussion in Cameron and Trivedi (1998), and the nonlinear least squares estimator, see also Gourieroux and Monfort (1995, Chap. 8.4.2).

Note the restriction which is introduced in the GLM context in contrast to PML models, namely that the conditional variance of d_t is multiplicative in σ, see (3.37). This allows an estimation of θ without the nuisance parameter σ. See also Cameron and Trivedi (1998, p. 35). The nuisance parameter σ describing the dispersion can be estimated in a second step as

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial \mu_t}{\partial m_t}\right)^{-1}(d_t - \mu_t)^2. \tag{3.42}$$

2. A latent model for discrete price changes

2.1. The ordered probit as a model for price changes.

2.1.1. *The ordered probit.* The ordered probit has proven to be a reliable work horse for the analysis of ordered quantal response data. It has been used before by Hausman, Lo, and MacKinlay (1992) for the analysis of discrete price changes to reveal the influence of other economic variables assumed to be at least weakly exogenous. In this section we will suggest a dynamic for limited dependent variables which has on the one hand well-defined statistical properties and on the other hand maintains a computational ease of estimation, very much like the generalised linear models outlined previously.

We have seen in the discussion of generalised linear models (GLM) that simple probits are nested in this framework, see (3.7). Yet, the standard view of probits is quite different from the typical GLM perspective. The starting point for this class of models is a latent linear specification

$$\begin{aligned} d_t^* &= m_t + \epsilon_t^* \qquad (3.43)\\ \epsilon_t^* &\sim iid \quad N(0, \sigma^2). \end{aligned}$$

The latent variable d_t^* is unobserved and assumed to be continuous. If the conditional expectation m_t of the latent variable d_t^* is dependent only on weakly exogenous variables x_t, it is usually parameterized linearly as

$$m_t = \mathrm{E}\left[d_t^*\middle|\, x_t\right] \qquad (3.44)$$
$$= x_t'\beta.$$

Yielding of course the normal regression model, with the added complication that d_t^* is not observed directly, but through an observation rule $g()$. The observation rule used for the analysis of discrete price changes is a threshold function. This many-to-one function maps the unobservable, latent d_t^* into the observable, discrete d_t,

$$d_t = g_{OP}(d_t^*) = \sum_{j=1}^{J} v_j 1_{(d_t^* \in A_j)} \qquad (3.45)$$
$$A_j = (\mu_{j-1}; \mu_j], \text{for } j = 2, \ldots, J-1$$
$$A_1 = (-\infty; \mu_1],\ A_J = (\mu_{J-1}; \infty)$$
$$-\infty < \mu_1 < \mu_2 < \ldots < \mu_{J-1} < \infty$$

where the variable v_j contains the distinct values d_t can take on. For the analysis of price changes one would set e.g. $v_1 = -1$, $v_2 = 0$, and $v_3 = 1$. As in the original paper by McKelvey and Zavoina (1975) the thresholds μ are assumed to be unknown and thus parameters of the model. For an exposition of alternative settings see e.g. Lee (1992).

Note that in this setting neither the level of the latent variable nor the scale of the latent variable are identified. To focus our attention on the dynamic we restrict the unconditional variance of the latent error to one, $\sigma^2 = 1$. The identification issue is discussed at length in Pohlmeier and Gerhard (2000), who propose an identifying restriction adapted to the given problem. Here however, the standard interpretation is maintained where scale and level are not identified. This will ease the exposition considerably and furthermore will allow to decompose the process of price changes into a component of the sign and the size of price changes.

The exposition relies on the fact that the v_i contain values on a metric scale. This is not necessarily so as the estimator would also work with values on an ordinal scale.

Thus the dynamic specified for the latent model does not rely on the fact that prices are observed and the application range includes also time series of categorial data e.g. credit scores or the like.

2.1.2. *Score and likelihood of the static model.* To derive the likelihood, the probability to observe a certain category given an information set $\mathcal{F}_{t-1}$ is evaluated as

$$\begin{aligned} p_{j,t} &:= \text{Prob}\left[d_t = v_j \,|\mathcal{F}_{t-1}\right] & j = 1, \ldots, J \qquad (3.46)\\ &= \Phi(\nu_{j,t}) - \Phi(\nu_{j-1,t}) & \\ \nu_{j,t} &= \mu_j - m_t, \qquad \mu_0 = -\infty,\ \mu_J = +\infty \end{aligned}$$

Usually the information set $\mathcal{F}_{t-1}$ is equal to a set of exogenous regressors x_t, whereas the inclusion of these variables raises no particular problems, in the dynamic extension $\mathcal{F}_{t-1}$ consists of the past observable discrete price changes $\bar{d}_{t-1} = \begin{bmatrix} d_{t-1} & d_{t-2} & \ldots & d_1 \end{bmatrix}$. These are used to gain information on the dynamics of the latent variable d_t^* and exactly this difference between an observable information set and a latent variable will necessitate the particular dynamic model proposed in this section. The log likelihood and the score take on the well known form

$$\log\, \mathfrak{L} = \sum_{t=1}^{T} \sum_{j=1}^{J} 1_{(d_t = v_j)} \log p_{j,t} \qquad (3.47)$$

$$\frac{\partial \log\, \mathfrak{L}}{\partial \theta} = \sum_{t=1}^{T} \sum_{j=1}^{J} 1_{(d_t = v_j)} \frac{1}{p_{j,t}} \frac{\partial p_{j,t}}{\partial \theta} \qquad (3.48)$$

$$\theta' := \begin{bmatrix} \mu_1 & \ldots & \mu_{J-1} & \beta' \end{bmatrix}$$

$$\frac{\partial p_{j,t}}{\partial \beta} = -\left(\phi(\nu_{j-1,t}) - \phi(\nu_{j,t})\right) \frac{\partial m_t}{\partial \beta}$$

$$\frac{\partial p_{j,t}}{\partial \mu_i} = \delta_{i,j}\phi(\nu_{j,t}) - \delta_{i,j-1}\phi(\nu_{j-1,t})$$

where $\delta_{i,j}$ takes on the value of one if $i = j$ and is zero in all other cases. In the static linear parameterization of (3.44) we have of course $\partial m_t / \partial \beta = x_t$. We will simplify the exposition in this section by concentrating on this case.

From the score of the model with respect to the coefficients of the mean function the relationship of the ordered probit to PML estimators is evident. If the latent variable d_t^* was observable, then the score of the model would be given by the score of the latent model, which is

$$\frac{\partial \log \mathfrak{L}^*}{\partial \beta} = \sum_{t=1}^{T} \frac{1}{\sigma^2} x_t \epsilon_t^*. \tag{3.49}$$

Gourieroux, Monfort, Renault, and Trognon (1987) have shown that the score of the observable model involves the generalized error $\mathrm{E}\left[\epsilon_t^* \middle| d_t, x_t\right]$ in place of the unobservable error ϵ_t^* which is used in the latent linear model and that the score of the observable model is the conditional expectation of the score of the latent model given the contemporaneous observation. This is of course only a special version of the more general EM algorithm approach to missing observations problems proposed by Dempster, Laird, and Rubin (1977).

$$\begin{aligned} \frac{\partial \log \mathfrak{L}}{\partial \beta} &= \mathrm{E}\left[\frac{\partial \log \mathfrak{L}^*}{\partial \beta} \middle| d_t, x_t\right] \\ &= \sum_{t=1}^{T} \frac{1}{\sigma^2} x_t \mathrm{E}\left[\epsilon_t^* \middle| d_t, x_t\right] \end{aligned} \tag{3.50}$$

Whereas the consistency and asymptotic normality of the latent model (3.49) does not depend on the distributional assumption on ϵ_t^* as long as the pseudo true distribution used for estimation is from the LEF, see e.g. White (1994) or Gourieroux and Monfort (1995) and further references given there. For the ordered probit this result needs to be modified. In this model, the calculation of the generalized errors which appear in the first order condition relies on the distributional assumptions and thus the consistency is not independent from the distributional assumptions. See e.g. the extended discussion in Gourieroux, Monfort, Renault, and Trognon (1987). Assuming Normality of the error term, Poirier and Ruud (1988) have shown consistency and asymptotic Normality of the estimated parameters in the mean function even under severe dynamic misspecification. The relationship to conditional moment restrictions implied by (3.49) is also evident. See Newey (1990) and Newey (1993).

The conditional expectations of the latent variable given the contemporaneous observation of the dependent variable and the exogenous regressors is the sum of the mean function, here $x_t'\beta$, and the generalized residuals.

$$\mathrm{E}\left[d_t^* \middle| d_t = v_i, x_t\right] = \mathrm{E}\left[x_t'\beta + \epsilon_t \middle| d_t = v_i, x_t\right] \tag{3.51}$$

$$= x_t'\beta + \mathrm{E}\left[\epsilon_t \middle| d_t = v_i, x_t\right] \tag{3.52}$$

$$= \begin{cases} x_t'\beta + \frac{\phi(\nu_{t,j-1})-\phi(\nu_{t,j})}{\Phi(\nu_{t,j})-\Phi(\nu_{t,j-1})} & \text{if } j = 2, \ldots, J-1 \\ x_t'\beta + \frac{-\phi(\nu_{t,1})}{\Phi(\nu_{t,1})} & \text{if } j = 1 \\ x_t'\beta + \frac{\phi(\nu_{t,J-1})}{1-\Phi(\nu_{t,J-1})} & \text{if } j = J \end{cases} \tag{3.53}$$

Note however that the stochastic nature of the error even given the full sample information. This prevents the straightforward application of the prediction error decomposition as it is the usual practice for the estimation of dynamic models.

2.2. A dynamic process of limited dependent variables.

2.2.1. *Parameter and observation driven models.* This just outlined class includes Probits, as well as ordered Probits or Tobit type models, yet, it is not limited to these classical limited dependent variable models but could be extended to a wider class of models defined through different choices of the observation function $g(\cdot)$. For a broad range of applications in economics and statistics along with different observation rules, see e.g. Maddala (1983), Cox and Snell (1989), McCullagh and Nelder (1989). The main feature of these models is the mapping function $g(\cdot)$, which of course needs to be a Borel measurable function, but will in general not be a one-to-one mapping but many-to-one. This gives rise to two important information sets which will dominate the description of the model and derivation of its properties.

First, there is the information set generated by the latent innovations ϵ_t^* up to t denoted by $\mathcal{F}_t^* = \sigma(\epsilon_t^*, \epsilon_{t-1}^*, \ldots, \epsilon_1^*)$. This information set might be augmented by weakly exogenous variables x_t in an obvious way. The extension is understood implicitly and no additional notation needs to be introduced. This full information set reveals the latent variable perfectly, i.e.

$$\mathrm{E}\left[\left(\mathrm{E}\left[d_t^* \middle| \mathcal{F}_t^*\right] - d_t^*\right)^2\right] = 0, \tag{3.54}$$

but is not available.

The observation rule $g()$ maps the latent variable d_t^* into the observable d_t and thus the information set $\mathcal{F}_t^*$ into an available information set $\mathcal{F}_t$. Denote the information set generated by the observable limited dependent variables d_t up to time t by $\mathcal{F}_t = \sigma(d_t, d_{t-1}, \ldots, d_1)$, again with an obvious extension if weakly exogenous variables x_t are available.

The two information sets will only coincide if the observation rule $g(\cdot)$ is a one-to-one function, or more formally a Borel measurable isomorphism See e.g. Davidson (1994, theorem 10.3). In typical applications however, especially in the limited dependent case, we have that $\mathcal{F}_t \subset \mathcal{F}_t^*$. This implies of course that $\mathcal{F}_t$ does only partially reveal the latent variable

$$\mathrm{E}\left[(\mathrm{E}\left[d_t^* \vert\, \mathcal{F}_t^*\right] - d_t^*)^2\right] \geq 0, \tag{3.55}$$

where, again, equality occurs only if $g()$ is a one-to-one function.

To illustrate the complications for time series models implied by this difference, we consider the simplest illustration possible: an AR(1) process in the latent variable, LA-AR(1) in short, without exogenous variables:

$$d_t^{**} = \phi d_{t-1}^{**} + \epsilon_t^*. \tag{3.56}$$

We use here d^{**} to denote the latent endogenous variable since we want to emphasize the difference to d^*, which usually denotes the latent variable and is used later on to introduce an alternative process. To relate to the first paragraph outlining this type of models and defining a threshold observation rule, we have $\mu_t = \phi d_{t-1}^{**}$, and assume $J = 2$, with $v_1 = 0$ and $v_2 = 1$ and fix $\mu_1 = 0$, which yields a simple probit observation rule. Since we limit ourselves to the stationary interval $\phi \in (-1, 1)$. For ease of exposition we assume a fixed initial d_0^{**} and omit explicit reference, where we can do so without loss of clarity.

The log likelihood of the latent model is follows from a straightforward prediction error decomposition of the latent variable d_t^{**}. We slightly abuse notation to put a limit on the number of variables introduced and use ν_t and d_t for either of the two processes d_t^* and d_t^{**} since the context resolves the overload. We define thus

$$\nu_t = d_t^{**} - \mu_t, \tag{3.57}$$

where m_t is measurable with respect to $\mathcal{F}_{t-1}^*$ for the given LA-AR(1) model and obtain

$$\log \mathfrak{L}^* = \text{const} - \frac{1}{2}\sum_{t=1}^{T}(\nu_t)^2. \tag{3.58}$$

This is only of theoretical interest, as $\mathcal{F}_{t-1}^*$ is not available.

The available information $\mathcal{F}_{t-1}$ implies a somewhat more involved expression. We note that under given assumptions on the errors ϵ_t^* and the LA-AR(1) dynamic the latent variables d_t^{**} follow a joint normal distribution. For the T vector d^{**} collecting the time series of the dependent variable we have

$$d^{**} \sim N(0,\Sigma), \qquad \Sigma_{[t,t+s]} = \phi^{|s|}, \qquad t = 1,\ldots,T, \quad s = -t+1,\ldots,T-t. \tag{3.59}$$

We denote the set implied for the latent variable d_t^{**} by the observation of d_t by the mapping $G(d_t)$. In the case of the threshold observation rule g_{OP} given in (3.45) we have the corresponding

$$G_{OP}(v_j) = A_j, \qquad j = 1,\ldots,J. \tag{3.60}$$

Equipped with this notation we can write the likelihood of the observable model as[6]

$$\log \mathfrak{L} = \log \int_{G_{OP}(d_T)} \cdots \int_{G_{OP}(d_1)} \varphi_T(u;\Sigma)du_{[1]}\ldots du_{[T]}. \tag{3.61}$$

We recognize this as a T-fold integral which poses prima facie a considerable problem. There are two main roads to attack this problem, in chronological order they are: Either avoid maximum likelihood estimation altogether and use measurable

[6] $\phi_T(u;\Sigma)$ denotes the density of the T-variate standard normal distribution with correlation matrix Σ.

sampling moments of d_t implied the dynamic model of d_t^{**} to estimate the parameters of Σ. Or use simulation techniques to evaluate the above T fold integral or to solve similar equivalent problems.

The use of sampling moments has a long tradition in statistics and goes back at least to Lomnicki and Zaremba (1955), although they refer to earlier work by Kendall only published in working papers. This line of research was continued by Kedem (1980) and Keenan (1982). Gourieroux, Monfort, Renault, and Trognon (1987) re-discover the use of sampling moments to solve this problem in econometrics, however without making use of earlier useful results. Finally Poirier and Ruud (1988) explicitly approximate the above integral by an approach which relates very closely to the direct use of sampling moments.

The wide availability of fast computing resources favored however the use of simulation methods, especially Markov Chain Monte Carlo, to overcome this inherent inferential hurdle of parameter driven dynamic models, see the general framework proposed in Chib and Greenberg (1998) and Manrique and Shephard (1998) for an emphasis on time series applications and further literature given there. All of the simulation approaches involve however a considerable computational overhead.

Cox (1981) introduced the useful notion of parameter driven and observation driven models in this context. The distinction stems from the information set, which is used to condition the distribution of the dependent variable. An observation driven model uses the observable information set $\mathcal{F}_t$, e.g. GLMs or GARCH models. A parameter driven model uses the latent information set $\mathcal{F}_t^*$, e.g. LA-AR(p) or SV models. Both approaches have in general different virtues. The observation driven class is characterised by a straightforward estimation strategy, yet the derivation of statistical properties becomes often burdensome. Parameter driven models on the other hand allow often a straightforward derivation of statistical properties, yet parameter estimation might pose a considerable computational burden.

We propose in this section a variation to the latent ARMA model, which is observation driven. It retains the ease of estimation experienced in the GLMs by using a

$\mathcal{F}_{t-1}$ measurable mean function, thus avoiding unwieldy expressions like (3.61). Yet, it allows to derive statistical properties of the process, e.g. conditions for stationarity.

2.2.2. *An observation driven dynamic.* The parameter driven LA-AR(1) model in (3.56) can be given a different form, to see where the updating of the latent mean function would rely on observations which are not available,

$$d_t^{**} = \mu_t + \epsilon_t^*, \qquad \mu_t = \phi\left(\mu_{t-1} + \epsilon_{t-1}^*\right). \tag{3.62}$$

The latent error ϵ_{t-1}^* is not available and does not become available through the observation of d_t. Only the observation of d_t^{**} would allow a prediction-error decomposition and thus a simple MLE.

Adopting a missing observations perspective to solve this problem would mean to use the likelihood of the fully observed model in (3.58) and taking the conditional expectation of the likelihood with respect to the available information $\mathcal{F}_T$. This general procedure has been widely used with success in missing observations problems, particularly after the systematic treatment in Dempster, Laird, and Rubin (1977) in form of the EM algorithm. For an iid. data-generating process an EM type analysis is suggested for limited dependent variables by Gourieroux, Monfort, Renault, and Trognon (1987), mainly for the purpose of tests. In the given context however, this would involve the evaluation of conditional expectations like $\mathrm{E}\left[\epsilon_{t-1}^* \middle| \mathcal{F}_T\right]$, which again need the solution of integrals of order T

$$\mathrm{E}\left[\epsilon_{t-1}^* \middle| \mathcal{F}_T\right] = \int_{G_{OP}(d_T)} \cdots \int_{G_{OP}(d_1)} u_{[t-1]} \varphi_T(u; \Sigma) du_{[1]} \dots du_{[T]}. \tag{3.63}$$

If we reduce the amount of information used to update the mean function, from using the set $\mathcal{F}_t^*$ in the LA-AR(1) model to the information set $\mathcal{F}_t$ we obtain a perfectly feasible model. Instead of (3.62) we have an alternative data-generating process of the form

$$d_t^* = m_t + \epsilon_t^*, \qquad m_t = \phi\left(m_{t-1} + \mathrm{E}\left[\epsilon_{t-1}^* \middle| \mathcal{F}_{t-1}\right]\right), \tag{3.64}$$

which has a very useful connection to the LA-AR(1) model if we remember that $\epsilon_{t-1}^* = \mathrm{E}\left[\epsilon_{t-1}^* \middle| \mathcal{F}_{t-1}^*\right]$.

In the LD-AR(1) model, we suggest to use for the mean function m_t most of the information available $\mathcal{F}_{t-1}$, but not all of it, i.e. $\mathcal{F}_T$, and we also avoid using unobservable information, i.e. $\mathcal{F}^*_T$. The important difference between the mean functions is therefore the measurability of m_t with respect to $\mathcal{F}_{t-1}$ as opposed to μ_t, which not measurable with respect to $\mathcal{F}_{t-1}$ but to $\mathcal{F}^*_{t-1}$. Although the latter two approaches are clearly favourable with respect to the amount of information they use, i.e.

$$\mathcal{F}_{t-1} \subseteq \mathcal{F}_T \subseteq \mathcal{F}^*_T, \tag{3.65}$$

the latter two involve by construction a considerable computational overhead.

The new LD-AR(1) model, although not making full use of the available information set, has three considerable advantages:

(1) a computationally simple maximum likelihood estimator based on the prediction error decomposition, and
(2) favourable time series properties, i.e. it shares the stationary parameter space with the LA-AR(1) models,
(3) it nests the standard AR(1) model, if the observations rule $g()$ is chosen to be the identity mapping.

2.2.3. *A maximum likelihood estimator.* The great advantage of the LD-AR(1) specification is that the conditional expectation m_t of the latent variable d^*_t is measurable with respect to the information available up to time $t-1$, $\mathcal{F}_{t-1}$, and thus allows to rely on a prediction error decomposition of the likelihood.

The evaluation of the likelihood follows a standard recursive scheme:

(1) The conditional expectation of the latent variable given no available past information is assumed to equal the unconditional expectation

$$m_0 := \mathrm{E}\left[m_t\right] = 0. \tag{3.66}$$

(2) The likelihood contribution of observation t given the probit observation rule, the Gaussian assumption on the error term and most important, the

measurable mean function is

$$\operatorname{Prob}\left[d_t \,|\mathcal{F}_{t-1}\right] = \begin{cases} \Phi(-m_t), & \text{if } d_t = 0, \\ 1 - \Phi(-m_t), & \text{if } d_t = 1. \end{cases} \tag{3.67}$$

(3) The generalized error c_t, which makes up the mean function is a (conditionally) deterministic function of the observations, concisely, of the $\mathcal{F}_{t-1}$ measurable mean function m_t and the current observation d_t

$$c_t = \begin{cases} \frac{-\phi(-m_t)}{\Phi(-m_t)}, & \text{if } d_t = 0, \\ \frac{\phi(-m_t)}{1-\Phi(-m_t)}, & \text{if } d_t = 1. \end{cases} \tag{3.68}$$

See the original paper by Gourieroux, Monfort, Renault, and Trognon (1987) for an extended discussion of generalized errors in the context of non-dynamic models.

(4) Calculation of the conditional expectation of the future latent variable given the present information,

$$m_{t+1} = \phi(m_t + c_t). \tag{3.69}$$

(5) Steps 2 through 4 are repeated for all d_t, $t = 1, \ldots, T$.

(6) The likelihood $\mathfrak{L}_d$ of the observable model can be directly evaluated, using $\bar{d}_t$ which contains all observations of d_t up to t, as

$$\begin{aligned} \mathfrak{L}_d(\bar{d}_T|\phi) &= \int_{G_P(d_1)} \int_{G_P(d_2)} \cdots \int_{G_P(d_T)} f(u_1, u_2, \ldots, u_T) du_1\, du_2 \ldots du_T \\ &= \int_{G_P(d_1)} \int_{G_P(d_2)} \cdots \int_{G_P(d_T)} f(u_1) f(u_2|\mathcal{F}_1) \ldots f(u_T|\mathcal{F}_{T-1}) du_1\, du_2 \ldots du_T \\ &= \prod_{t=1}^{T} \operatorname{Prob}\left[d_t = 1 \,|\mathcal{F}_{t-1}\right]^{d_t} \operatorname{Prob}\left[d_t = 0 \,\middle|\mathcal{F}^d_{t-1}\right]^{(1-d_t)}. \end{aligned} \tag{3.70}$$

Thus by the use of the LD-ARMA process (3.64) and the implied likelihood $\mathfrak{L}_d$, the quite cumbersome likelihood implied by the parameter driven model (3.56) can be circumvented.

2.2.4. *A bound on variances and autocorrelations.* To appreciate the LD-AR(1) model as a sensible process of LD variables in its own right, we will establish two important properties:

(1) Variance stationarity of the latent process,
(2) stationary autocovariance of the latent process, and

We will proceed to show these properties by establishing a bound on variances and autocovariances by a corresponding process with well-known properties. Variances and autocovariances of the LD-AR(1) process are shown to be bounded from above by the corresponding LA-AR(1) process. The analogue is true for the latent and the observable LD-AR(1) process.

In order to show these properties we will exploit that the observable information $\mathcal{F}_t$ is a subset of the latent information $\mathcal{F}_t^*$,

$$\mathcal{F}_t \subseteq \mathcal{F}_t^*. \tag{3.71}$$

The following proposition from Davidson (1994, theorem 10.27) will be extremely valuable in this pursuit:

Proposition 2.1. *For a random variable x which is measurable with respect to $\mathcal{F}$ and if $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}$ we have that*

$$\mathrm{E}\left[\mathrm{E}\left[x|\,\mathcal{F}_1\right]^2\right] \leq \mathrm{E}\left[\mathrm{E}\left[x|\,\mathcal{F}_2\right]^2\right]. \tag{3.72}$$

Proof: *See Davidson (1994, theorem 10.27).*

Notably this proposition is fairly general and this generality can actually be maintained for most of this section. The boundedness of the variance is formulated in

Proposition 2.2. *For a LD-AR(1) process d_t^* with dynamic parameter ϕ and a LA-AR(1) process d_t^{**} with identical dynamic parameter we have for all $\phi \in (-1, 1)$ that*

$$\mathrm{Var}\left[d_t^*\right] \leq \mathrm{Var}\left[d_t^{**}\right]. \tag{3.73}$$

Proof: *Note that we have* $\mathrm{E}[d_t^*] = \mathrm{E}[c_t] = 0$ *for the given specification. For the variance of the latent process we proceed in the usual way after having defined an error* $v_t = c_t - \epsilon_t$

$$\mathrm{E}\left[(d_t^*)^2\right] = \phi^2 \mathrm{E}\left[\left(d_{t-1}^* + v_{t-1}\right)^2\right] + \sigma^2 \tag{3.74}$$

$$= \phi^2 \mathrm{E}\left[\left(d_{t-1}^*\right)^2\right] + \phi^2\left(\mathrm{E}\left[v_{t-1}^2\right] + 2\mathrm{E}\left[d_{t-1}^* v_{t-1}\right]\right) + \sigma^2. \tag{3.75}$$

See e.g. Lütkepohl (1991, chap. 2.1). Using $\mathrm{E}\left[d_{t-1}^* v_{t-1}\right] = \mathrm{E}[\epsilon_t c_t] + \sigma^2$ *and* $\mathrm{E}\left[v_{t-1}^2\right] = \mathrm{E}[c_t^2] - 2\mathrm{E}[\epsilon_t c_t] + \sigma^2$ *we obtain the following*

$$\mathrm{E}\left[(d_t^*)^2\right] = \frac{\sigma^2 + \phi^2\left(\mathrm{E}[c_t^2] - \sigma^2\right)}{1 - \phi^2}. \tag{3.76}$$

The term in parentheses, $(\mathrm{E}[c_t^2] - \sigma^2)$*, is always less than zero as a consequence of proposition 2.1 and the proposition* (2.2) *follows immediately.*

This result helps our intuition considerably, as we can observe that the larger the amount of information lost by imposing the observation rule $g()$ on the process d_t^* the larger will be the deviation in the variances. Here loss in information is quantified by the difference $(\mathrm{E}[c_t^2] - \sigma^2)$.

The proposition and proof for the autocorrelation follow similar lines.

Proposition 2.3. *For a LD-AR(1) process* d_t^* *with dynamic parameter* ϕ *and a LA-AR(1) process* d_t^{**} *with identical dynamic parameter we have for all* $\phi \in (-1, 1)$ *that*

$$\left|\mathrm{Cor}\left[d_t^*, d_{t-l}^*\right]\right| \leq \left|\mathrm{Cor}\left[d_t^{**}, d_{t-l}^{**}\right]\right|, \ \textit{for all } l > 0. \tag{3.77}$$

Proof: *For the autocovariance of the process we have analogue to the variance*

$$\mathrm{E}\left[d_t^* d_{t-l}^*\right] = \frac{\phi^l}{1 - \phi^2} \mathrm{E}[\epsilon_t c_t]. \tag{3.78}$$

Note that by definition we have $\mathrm{E}[v_t] = 0$ *and* $\mathrm{E}[\epsilon_t c_t] = \mathrm{E}[c_t^2]$*. For the autocorrelation follows directly*

$$\mathrm{Cor}\left[d_t^*, d_{t-l}^*\right] = \frac{\phi^l \mathrm{E}[c_t^2]}{\sigma^2 + \phi^2\left(\mathrm{E}[c_t^2] - \sigma^2\right)} \tag{3.79}$$

If q *denotes the ratio* $q = \mathrm{Cor}\left[d_t^*, d_{t-l}^*\right] / \mathrm{Cor}\left[d_t^{**}, d_{t-l}^{**}\right]$*, then*

$$q = \frac{\mathrm{E}\left[c_t^2\right]}{\sigma^2 + \phi^2 \left(\mathrm{E}\left[c_t^2\right] - \sigma^2\right)}, \tag{3.80}$$

and the condition for $q \in (0;1)$ *is of course*

$$\mathrm{E}\left[c_t^2\right] \leq \sigma^2 + \phi^2 \left(\mathrm{E}\left[c_t^2\right] - \sigma^2\right) \quad \Leftrightarrow \quad (1-\phi^2)\mathrm{E}\left[c_t^2\right] \leq (1-\phi^2)\sigma^2. \tag{3.81}$$

Making again use of proposition 2.1 we obtain proposition 2.3.

2.3. Extensions of the simple limited dependent dynamic.

2.3.1. *An ARMA(p,q) process in the latent variable.* Before detailing the ARMA(p,q) extension to the limited dependent model, we will first recall a well-known state space form of ARMA(p,q) models.

Consider first the AR(p) model

$$d_t^{**} = \sum_{i=1}^{p} \phi_i d_{t-i}^{**} + \epsilon_t^*,$$

The state space representation of the AR(p) model is formulated using the state vector ξ_t as

$$d_t^{**} = e_1' \xi_t \tag{3.82}$$

$$\xi_t = F\xi_{t-1} + e_1 \epsilon_t^* \tag{3.83}$$

$$F = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & & 0 \\ 0 & 1 & 0 & \dots & \vdots \\ \vdots & & \ddots & & 0 \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix}$$

See e.g. the exposition in Hamilton (1994, chap. 13).

Given this definition of the state space we note that the conditional expectation of the state is evaluated recursively as

$$\mathrm{E}\left[\xi_t \middle| \bar{d}_{t-1}\right] = F\left(\epsilon_{t-1}^* e_1 + \mathrm{E}\left[\xi_{t-1} \middle| \bar{d}_{t-2}\right]\right), \tag{3.84}$$

and the conditional expectation of the latent variable is just

$$\mathrm{E}\left[d_t^{**}\middle|\,\bar{d}_{t-1}\right] = e_1'\mathrm{E}\left[\xi_t\middle|\,\bar{d}_{t-1}\right]. \qquad (3.85)$$

Recall that $\bar{d}_{t-1}$ denotes all d_t up to and including d_{t-1}. It remains to include the MA(q) components in the latent dynamics to obtain

$$d_t^{**} = \sum_{i=1}^{p} \phi_i d_{t-i}^{**} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}^{*} + \epsilon_t^{*}.$$

The observation equation of the state space model is slightly modified to accommodate the MA terms as

$$d_t^{**} = \begin{bmatrix} 1 & \theta_1 & \dots & \theta_q \end{bmatrix} \xi_t. \qquad (3.86)$$

See e.g. Hamilton (1994, 13.1). The dimension of the state space is equal to $\max(p, q+1)$. If $q+1$ is larger than p the matrix F is $(q+1 \times q+1)$ and padded with zeros to obtain the following shape

$$F = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p & 0 \\ 1 & 0 & \dots & & 0 & 0 \\ 0 & 1 & 0 & \dots & \vdots & \vdots \\ \vdots & & \ddots & & 0 & 0 \\ 0 & \dots & 0 & 1 & 0 & 0 \end{bmatrix}.$$

We note the slightly augmented updating equation of the latent mean (3.84), compared to the simple case (3.62) considered earlier.

2.3.2. *An ARMA(p,q) process for limited dependent variables.* The algorithm proposed for the AR(1) model relies on the simple forecast function used to iterate the conditional expectation of the latent variable forward. This simple forecast function does not translate directly to the ARMA(p,q) case. We use therefore the state-space notation introduced above. It is however important to note that the estimation algorithm employed is not a Kalman filter. The latent variable is defined as in the AR(1) case as the sum of a mean function and an error term

$$d_t^{*} = \begin{bmatrix} 1 & \theta_1 & \dots & \theta_q \end{bmatrix} \xi_t, \qquad \xi_t = m_t + \epsilon_t^{*} e_1. \qquad (3.87)$$

Denote the coefficient matrix of ξ_t by H. The mean function

$$m_t = H m_t^{\xi} \tag{3.88}$$

is again a modification of the ARMA(p,q) mean function, where we replace the error in the updating step by its conditional expectation, given the observable information set.

$$m_t^{\xi} = F\left(c_{t-1}e_1 + m_{t-1}^{\xi}\right), \tag{3.89}$$

To sum up the procedure the individual steps necessary to evaluate the likelihood are briefly described below:

(1) Initialization of the state variables

$$m_0^{\xi} = \mathrm{E}\left[\xi_t\right]$$

(2) Calculation of the generalized residuals, given the state variables and the contemporaneous observation $\mathrm{E}\left[\epsilon_t^* \middle| \bar{d}_t\right]$.
(3) Updating of the state variables, by the forward iteration in (3.89).
(4) Repeat steps 2 and 3 for $t = 2, \ldots, T$
(5) Calculate the Likelihood from $\{p_{j,t}\}_{t=1}^{T}$.

This completes the derivation of the ARMA(p,q) dynamic for the latent variable d_t^* of a discrete valued process d_t. It should be noted that at no point of the derivation the fact was employed that the dependent variable is a modified count variable in this context. Thus, the proposed dynamic is applicable to standard ordered probits with a conventional categorial variable as observable or even to a probit, where only a binary outcome is observed.

2.3.3. *Exogenous variables in the mean function of a discretely observed ARMA(p,q) model.* There are two ways to include exogenous regressors in the dynamic model. The standard procedure is to include the regressors in the dynamic specification and obtaining thereby an infinite distributed lags model, see e.g. Judge, Griffiths, Hill, Lütkepohl, and Lee (1985, chap.10). The flexibility of this approach is directly demonstrated by introducing m lags of a regressor x_t with coefficients β in the context of the latent ARMA(p,q) model:

$$\begin{aligned} m_t^\xi &= F(m_{t-1}^\xi + e_1 c_{t-1}) + e_1 \beta(L) x_t \qquad (3.90) \\ &= \sum_{j=0}^{\infty} \Psi_j L^j e_1 (\beta(L) x_t + F c_{t-1}) \\ &= \sum_{j=0}^{\infty} \sum_{i=1}^{m} \Psi_j e_1 \beta_j x_{t-i-j} + \sum_{j=0}^{\infty} \Psi_{j+1} e_1 c_{t-1-j} \\ m_t &= \sum_{j=0}^{\infty} \sum_{i=1}^{m} H \Psi_j e_1 \beta_j x_{t-i-j} + \sum_{j=0}^{\infty} H \Psi_{j+1} e_1 c_{t-1-j} \qquad (3.91) \end{aligned}$$

Here we use Ψ_j as a repeated product defined as $\Psi_j = \prod_{k=1}^{j} F$, where the product operator is understood as a matrix multiplication from the right. The third equality for m_t^ξ makes clear that the latent state is decomposed into a weighted sum of all the past x_t and an already known MA(∞) term. This flexibility is sometimes needed. A typical candidate for the inclusion as a regressor with an infinite lag structure is the observed volume per transaction.

Other variables however are rendered virtually uninterpretable by a dynamic inclusion, e.g. regressors capturing the seasonality. A viable alternative in the context of state-space models is the inclusion of explanatory variables in the "observation" equation of the latent model like

$$m_t = H m_t^\xi + w_t' \gamma. \qquad (3.92)$$

This is equivalent to a linear regression of d_t^* on w_t with ARMA(p,q) errors.

Of further interest is the calculation of descriptive statistics of the infinite lag structure to gain some insights into the distribution of weights over the history of the regressors. See e.g. Hendry (1995, chap. 6.5) for an extensive treatment and the discussion of alternatives. For ease of exposition $m = 1$ is used in (3.91) without loss of generality. Considering the infinite lag structure, the coefficient of the regressor at lag j is directly given by

$$k_j = H\Psi_j e_1 \beta$$

The sum over k_j should be finite for reasons of stationarity, i.e.

$$\sum_{j=0}^{\infty} k_j = \kappa, \text{ with } |\kappa| < \infty. \tag{3.93}$$

If the k_j are either positive for all j or negative for all j, then the weight of an individual lag k_j/κ can be calculated and the median lag can be evaluated as the first integer m for which the following inequality holds

$$\sum_{i=0}^{m-1} \frac{k_j}{\kappa} \leq \frac{1}{2} \leq \sum_{i=0}^{m} \frac{k_j}{\kappa}. \tag{3.94}$$

See Hendry (1995, p. 215). By the specification proposed in this subsection a parsimonious way to include lags of weakly exogenous variables is proposed.

2.4. The treatment of rare events and moments of the observable process. The application of the general dynamic limited dependent variables model proposed in the preceding section to the analysis of the transaction price process is complicated by the rare occurrence of larger price changes. As will be demonstrated in the empirical analysis of price changes, the values this process takes on are often limited to a set of three values, i.e. +1, 0, −1. Rarely, in less that 1% of the cases, larger price changes are observed. The rare event would pose considerable problems, since they do not suffice for the estimation of individual threshold parameters in an ordered probit model. Therefore, these values will be collected in the +1 and −1 categories. Here, we analyse the effects of this procedure on the implied conditional moments.

As the price change from one transaction to the next is limited to a multiple of the tick size, the conditional moments of the process of price changes take on a very peculiar form, which is based on an unspecified model for quantal response variables. The individual values the process of price changes can take on are given by the variable v_i for which we have $v_i = i \cdot c$ if c is the tick size and i an integer including zero, i.e. $i = \ldots, -2, -1, 0, 1, 2, \ldots$.

The conditional variance of the process of price changes is decomposed into conditional first and second moments as

$$\operatorname{Var}\left[d_t \middle| \mathcal{F}_{t-1}\right] = \operatorname{E}\left[d_t^2 \middle| \mathcal{F}_{t-1}\right] - \operatorname{E}\left[d_t \middle| \mathcal{F}_{t-1}\right]^2 . \tag{3.95}$$

Both estimators can in turn be specified independently of each other for the sake of modeling flexibility. The conditional mean of the process is given by[7]

$$\operatorname{E}\left[d_t \middle| \mathcal{F}_t\right] = \sum_{i=-\infty}^{\infty} v_i \operatorname{Prob}\left[d_t = v_i \left|\mathcal{F}_t\right.\right] \tag{3.96}$$

It is rather a formal necessity to let v_i take on values from $(-\infty, \infty)$ rather than a model component necessitated by properties of observed price changes. As a matter of fact absolute values of the dependent variable larger than two, i.e. $v_i \notin [-2, 2]$, are hardly ever observed in the liquid market considered in this context. By introducing the integer valued variable $v_i \in \mathbb{J} = [v_l, v_u]$ and by limiting the model to this range, one can explicitly account for this fact.

A straightforward approximation is used which treats the values outside the relevant support of the dependent variable, $d_t \notin \mathbb{J}$, either as the lower or upper boundary, $d_t = v_l$ or as $d_t = v_u$. This allows to build flexible models of the conditional probabilities to observe individual price changes, $\operatorname{Prob}\left[d_t = v_i \left|\mathcal{F}_{t-1}\right.\right]$, because there are indeed enough observations on the different values d_t is allowed to take on in this reduced setting. The alternative is to specify a functional relationship, e.g. in the form of a parameterized probability function, over the theoretically possible support of d_t, which allows to extrapolate from the categories with a substantial amount of observations to the categories with hardly any observations. Here, the former method of reducing the support of d_t is chosen for the sake of an increased modeling flexibility.

The approximation error incurred can be analysed using the decomposition of the conditional expectation into three parts defined by the intervals $\mathbb{J}$, $\mathbb{J}_l = (-\infty, v_l - 1]$,

[7] Regularity conditions on the employed conditional probabilities so that the respective conditional moments exist are obvious and thus not discussed for the sake of a brief exposition.

and $\mathbb{J}_u = [v_u + 1, \infty)$:

$$\begin{aligned}
\mathrm{E}\left[d_t \middle| \mathcal{F}_{t-1}\right] = \mathrm{E}\left[d_t \middle| d_t \in \mathbb{J}, \mathcal{F}_t\right] \mathrm{Prob}\left[d_t \in \mathbb{J} \left| \mathcal{F}_t\right.\right] \\
+ \mathrm{E}\left[d_t \middle| d_t \in \mathbb{J}_l, \mathcal{F}_t\right] \mathrm{Prob}\left[d_t \in \mathbb{J}_l \left| \mathcal{F}_t\right.\right] \\
+ \mathrm{E}\left[d_t \middle| d_t \in \mathbb{J}_u, \mathcal{F}_t\right] \mathrm{Prob}\left[d_t \in \mathbb{J}_u \left| \mathcal{F}_t\right.\right] \quad (3.97)
\end{aligned}$$

For the second and third term of the sum in this equation the following approximation is used

$$\begin{aligned}
&\mathrm{E}\left[d_t \middle| d_t \in \mathbb{J}_l, \mathcal{F}_t\right] \mathrm{Prob}\left[d_t \in \mathbb{J}_l \left| \mathcal{F}_t\right.\right] \\
&= \sum_{i=-\infty}^{v_{l-1}} v_i \, \mathrm{Prob}\left[d_t = v_i \left| v_i \in \mathbb{J}_l, \mathcal{F}_t\right.\right] \cdot \mathrm{Prob}\left[d_t \in \mathbb{J}_l \left| \mathcal{F}_t\right.\right] \\
&= v_l \cdot \left(\sum_{i=-\infty}^{v_{l-1}} \mathrm{Prob}\left[d_t = v_i \left| v_i \in \mathbb{J}_l, \mathcal{F}_t\right.\right] \cdot \mathrm{Prob}\left[d_t \in \mathbb{J}_l \left| \mathcal{F}_t\right.\right] \right) + r_l \\
&= v_l \cdot \mathrm{Prob}\left[d_t \in \mathbb{J}_l \left| \mathcal{F}_t\right.\right] + r_l \qquad (3.98)
\end{aligned}$$

The approximation error for this component of the conditional expectation is evaluated as

$$r_l = \sum_{i=-\infty}^{v_{l-1}} (v_i - v_l) \, \mathrm{Prob}\left[d_t = v_i \left| d_t \in \mathbb{J}_l, \mathcal{F}_t\right.\right] \cdot \mathrm{Prob}\left[d_t \in \mathbb{J}_l \left| \mathcal{F}_t\right.\right]$$

As the events $d_t \in \mathbb{J}_l$ and $d_t \in \mathbb{J}_u$ are not only unconditionally very unlikely but proof to be virtually unforecastable on the basis of common information sets, the probabilities $\mathrm{Prob}\left[d_t \in \mathbb{J}_l \left| \mathcal{F}_t\right.\right]$ and $\mathrm{Prob}\left[d_t \in \mathbb{J}_u \left| \mathcal{F}_t\right.\right]$ are virtually zero and the approximation error is negligible, in particular because price changes larger than four ticks are not observed at all. If evidence is found that this approximation error is too large, the solution is an extension of the explicit modeling range J. As a matter of fact if the approximation error is large due to the size of the conditional probability to observe a price change out of the reduced support, $\mathrm{Prob}\left[d_t \in \mathbb{J}_u \left| \mathcal{F}_t\right.\right]$, then it is guaranteed that there are enough observations to extend the support of the dependent variable in the model and to obtain meaningful results.

The model of $\mathrm{E}\left[d_t^2 \middle| \mathcal{F}_{t-1}\right]$ follows the same line of arguments. In the empirical application it turns out that for $\mathrm{E}\left[d_t \middle| \mathcal{F}_t\right]$ only the sign of the price change is needed in the approximation, so that $\mathbb{J} = \{-1, 0, 1\}$. For $\mathrm{E}\left[d_t^2 \middle| \mathcal{F}_t\right]$ the set of admissible values of the dependent variable can be extended to $\mathbb{J} = \{0, 1, 2\}$. An explicit derivation of the errors incurred by the truncation beyond $\mathbb{J}$ for the conditional expectation of the squared price change is omitted for sake of brevity. It raises however no additional problems as v_i is merely replaced by v_i^2 in the calculations.

An alternative to the modeling scheme outlined here would be an adapted parameterization of $\mathrm{Prob}\left[d_t = v_i \middle| \mathcal{F}_t\right]$ which pays tribute to the fact that $d_t \notin \mathbb{J}$ are rare events so that the probability of price changes for which $d_t \notin \mathbb{J}$ are linked by some functional form to price changes $d_t \in \mathbb{J}$. This could be accomplished by the use of an explicit distribution for count variables. This is complicated by the fact that d_t can take on negative values so that some modifications compared to standard count data model would be needed anyway.

CHAPTER 4

The time component of the transaction process

1. Standard duration models

1.1. The transaction event in discrete and continuous time. Trade frequency at the transaction level can be described recurring either to a concept of discrete or continuous time. Either way, the model needs to account for the obvious time series properties of the sequence of trade intensities. Different solutions have been proposed in the literature.

Most of the contributions adopt a continuous time framework based on the seminal work of Engle (1996). In a sequence of papers the estimator proposed originally was analysed more in depth, see Engle and Russell (1997), Engle and Russell (1998), and Engle (2000). Engle and Russell employ the well known methodology of GARCH processes to capture the dynamics of the process of price intensities. Various extensions have been proposed in the literature so far. These include a relaxation of the distributional assumptions as in Grammig and Maurer (2000). Or the inclusion of a more flexible dynamic. Bauwens and Veredas (1999) employ a model structure similar to a stochastic volatility (SV) model yielding a stochastic conditional duration model. Jasiak (1999) proposes a fractionally integrated ACD model which allows to account for possible long memory effects in observed durations. Ghysels, Gourieroux, and Jasiak (1998) even allow for a dynamic in the conditional second moments of the series of time between transactions. This sums up only the major extensions to the original ACD model. These extensions are all very closely related to ARMA and GARCH models and the well documented extensions thereof. As these models are well covered in the literature it does not seem worthwhile to repeat this discussion in the given context. This section rather puts a focus on the peculiarities of duration models compared to well known GARCH models.

As an alternative to the continuous time framework of the ACD type models a discrete time structure might be imposed. Rydberg and Shephard (1998) employ an indicator variable a_t° which indicates whether a transaction occurred at time t or not, i.e. $a_t^\circ = 1$ or $a_t^\circ = 0$. Since the Rydberg and Shephard methodology is a decomposition of the entire transaction process, their modeling framework will be outlined in the context of other joint models of the transaction process in chapter 5. A comparison of Rydberg and Shephard's work with the evidence presented in this context might indicate that the use of a continuous time setting yields a more parsimonious model structure.

A mixture of both concepts is pursued by Hautsch (1999) and Gerhard and Hautsch (2000a) who discretise the dependent variable and assume that there exists a continuous latent variable building on the work of Han and Hausman (1990) and Meyer (1990) in order to achieve a semiparametric estimation of the baseline hazard. As this approach of price intensities as opposed to transaction intensities is tailored to the optimal aggregation of transaction data as an input for risk measurement, it will not be covered in this context.

There are no really obvious reasons why one should prefer either a continuous or a discrete time framework. A continuous time model can be interpreted as the limiting case of a suitably defined discrete time model. If one considers a Markov chain as the raw model to analyse transaction data, like in the stylized model proposed in chapter 2, then the relationship between both frameworks is quickly outlined. It is possible to interpret a continuous-time Markov chain as the limit of a discrete-time Markov chain model where the clock time distance between two adjacent periods approaches 0. Thus the selection of a continuous or a discrete time framework is rather a question of practicability and of the context the duration model is needed in.

Apart from the question of what support is used for the process of trade intensity, the modeling framework needs to be selected. Generally speaking, two types of models can be distinguished which are characterized by the dependent variable they employ, see e.g. Lancaster (1994). Either the conditional expectation of the event duration is modeled based on explanatory variables or the conditional probability that an event is terminated the next instant given that it lasted until the present time is parameterized as a function of explanatory variables. Both strategies are briefly outlined in order to introduce the modeling strategy.

1.2. Hazard rate models. The hazard rate is a particular model of the intensity of event occurrence given the time since the last event was observed. This is sometimes preferable compared to a specification based on conditional expectations of inter event durations if economic theory suggests a particular specification.

In order to introduce the hazard rate and to describe duration models, the notation is briefly outlined first. The index $t = 1, \ldots, T$ gives as usually only the consecutive order of observations. The clock time in seconds since the trading start at which a certain transaction was observed is given by ϑ_t and the time between transactions can thus be defined as $\tau_t = \vartheta_t - \vartheta_{t-1}$. Note that in the discrete time context the transaction time is identical with the observation index, if a one second grid is used, i.e. $\vartheta_t = t$. The number of transactions observed until the time of the tth transaction, ϑ_t, is given by $N(\vartheta_t)$. This can be simplified in the discrete time context to N_t.

The concept of hazard rates is most easily understood in the discrete time context where it is defined as the conditional probability to observe a transaction in t given the state of the counting process in $t-1$, N_{t-1}, i.e.

$$\lambda_t := \text{Prob}\left[N_t > N_{t-1} \,|N_{t-1}\right]. \tag{4.1}$$

See e.g. Lancaster (1994, p. 12). The extension of this definition to the continuous time framework is achieved by the use of the random duration τ in the marginal form

$$\lambda_t(s) := \lim_{c \to 0} \frac{\text{Prob}\left[s \leq \tau_t < s + c\,|\tau_t \geq s\right]}{c}. \tag{4.2}$$

Several econometric models are defined through a parameterization of the hazard function $\lambda(s)$. See e.g. Lancaster (1994, chap. 3.3) or Cox (1972), who originally proposed to modify a baseline hazard by the inclusion of exogenous variables. These types of models usually decompose the hazard rate into some exclusively time dependent component, the baseline hazard $\lambda_0(s)$, and the hazard function depending exclusively on a set of exogenous regressors x_t in $\lambda_1(x_t)$, yielding proportional hazards models

$$\lambda_t(s, x_t) := \lambda_0(s)\lambda_1(x_t) \tag{4.3}$$

under the assumption that the exogenous regressor x_t is constant over the time interval between ϑ_t and ϑ_{t-1}.

In spite of the great value of these models in econometrics, especially if the dependent variable is observed only through a censoring scheme, in this context those specifications are not favourable compared to a direct specification using $\log \tau_t$ as dependent variable. The latter builds on the conditional moments of the time between transactions $\mathrm{E}\left[\log \tau_t | \, x_t\right]$. There are two aspects which need to be compared. On the one hand there is the question whether anything can be learned from the hazard function in the context of a given application. If that is the case, then this would necessitate the use of either a specific parametric assumption for the baseline hazard λ_0, which makes the model of course vulnerable to a parametric misspecification of the hazard rate. Or a flexible semi- or nonparametric procedure is employed as suggested in Meyer (1990), Horowitz (1999), or Gerhard and Hautsch (2000b). On the other hand if the baseline hazard is not of importance in a certain application, then a straightforward model of the conditional moments, usually the mean, suffices and has the additional benefit of being robust against distributional misspecification. This point will be made more explicit in the next subsection.

1.3. Models of expected time between transactions. If one is willing to assume explicit distributions for the durations, there is a close link between models of the conditional expectation of the time between transactions and hazard rate models. On the one hand conditional moments of the time between transactions can be evaluated on the basis of the hazard rate $\lambda_t(s, x_t)$.

On the other hand, if one specifies directly a model in the general class of accelerated failure time models, and is willing to work with explicit distributional assumptions, then the implied hazard rate can be easily evaluated.

Accelerated failure time models are defined using a function of the exogenous variables, $g(x_t)$, which scales the random variable ϵ_t. Where ϵ_t has a distribution which does not depend on x_t so that it makes sense to interpret ϵ_t as an error term

$$\tau_t = \frac{\epsilon_t}{g(x_t)}. \tag{4.4}$$

The density function of ϵ_t is denoted by f_ϵ, the distribution function by F_ϵ. Since the time between transactions τ_t is positive, is quite sensible to restrict ϵ_t to a positive support and to choose functions $g()$ which are strictly positive. Thus we can use a log linear model structure in the error term,

$$\log \tau_t = -\log g(x_t) + \log \epsilon_t. \tag{4.5}$$

The most simple model which leads directly to a linear specification uses $g_t := \exp(-x_t'\beta)$ to obtain

$$\log \tau_t = x_t'\beta + \log \epsilon_t. \tag{4.6}$$

If one chooses ϵ_t to be from the unit exponential distribution, then the model is not only an accelerated failure time model but also from the class of proportional hazard rate models with $\lambda_1(x_t) = \exp x_t'\beta$ and constant $\lambda_0(s)$. Extensions involve the Weibull or Gamma distribution and allow for more flexible shapes of the hazard function. See e.g. Lancaster (1994, chap. 3.2.1).

The relationship between the assumed distribution of the time between transactions and the hazard rate implied by the distributional assumption conditional on the regressors is

$$\lambda(\tau_t|x_t) = \frac{f_\epsilon(\epsilon_t)}{1 - F_\epsilon(\epsilon_t)}. \tag{4.7}$$

The latter expression makes clear that once an explicit assumption about the trade intensities' distribution is chosen, the corresponding hazard rate is given by (4.7).

The models outlined so far all rely on an explicit distributional assumption which in turn implies a particular hazard rate function. This is unnecessarily cumbersome, if one is only interested in the conditional moments of the time between transactions given certain information sets which are of economic interest. A direct parameterization in the continuous case taking into account that τ_t cannot be negative is e.g.

$$\log \tau_t = x_t'\beta + \epsilon_t \tag{4.8}$$

where ϵ_t in this case is either left unspecified apart from the usual regularity conditions for OLS in large samples or it is assumed that ϵ_t has a pseudo-true distribution from the linear exponential family. This route is followed in most of the models employed in this work. Note that this implies a model from the accelerated failure time class, as outlined above.

The estimation of these models can be greatly simplified since the main interest in the applications considered here lies on the influence of the specified regressors on expected durations and not primarily on the hazard rate which would explicitly involve the distribution f_ϵ. Thus it is straightforward to use pseudo maximum likelihood with the conditional mean restriction

$$\mathrm{E}\left[\log \tau_t |\, x_t\right] = -\log g(x_t, \beta) - \beta_0, \tag{4.9}$$

while assuming that $\log g$ contains no constant, to obtain an estimator from the accelerated failure time class. This straightforward linear estimator for durations is only appropriate if there is no censoring and no time-varying regressors. The term β_0 is a necessary correction for the unspecified and thus unknown mean of the true distribution $\mathrm{E}\left[\log \epsilon_0\right]$, which presumably differs from the pseudo-true $\mathrm{E}\left[\epsilon\right] = 0$ in the specification

$$\log \tau_t = x_t'\beta - \beta_0 + \epsilon_t, \qquad \epsilon_t \sim N(0, 1). \tag{4.10}$$

The pseudo-true model assumed above is identical to the log normal duration model with $\sigma = 1$

$$\log \tau_t \sim N(x_t'\beta - \beta_0, \sigma) \tag{4.11}$$

See e.g. Lancaster (1994, chap. 3.4.1). Remember that we denote the standard normal density and distribution function by φ and Φ. To complete the properties of the pseudo-true model (4.11), it is worthwhile to note that its hazard function has a unique maximum and takes on the form

$$\lambda(\tau_t|x_t) = \frac{\varphi(\epsilon_t)}{\sigma\,\tau_t\,(1-\Phi(\epsilon_t))}. \tag{4.12}$$

To provide some intuition about the hazard rate implied by the log normal duration model the hazard rate is provided in figure 4.1 for different values of σ.

The above mentioned restrictions with respect to the dependence of the hazard rate on the explicit p.d.f. employed apply directly so that these properties should be used with care since most likely they are not valid for the true density and the $\lambda(\tau_t|x_t)$ is to be considered a pseudo true hazard rate.

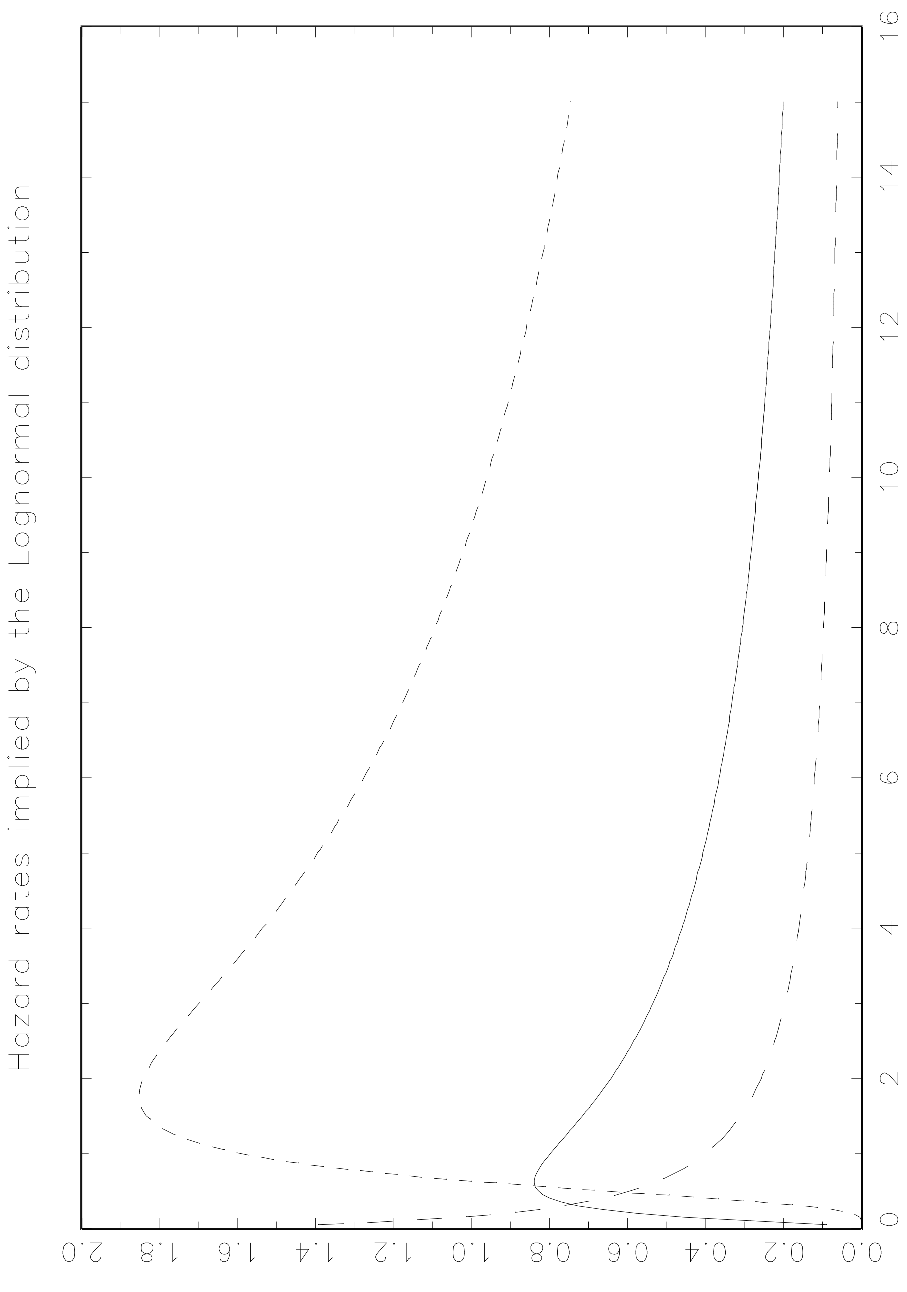


Figure 4.1. Hazard rate implied by the log normal duration model for different variances σ^2. Solid line $\sigma = 1$, dashed line $\sigma = 0.5$, small dashes $\sigma = 2$.

2. Autoregressive conditional durations

2.1. The parametric interpretation of ACD models. The models described so far all rely on the implicit assumption that there are exogenous variables available which are suitable as regressors for the trade intensity. However, the evolution of the dependent variable might not be explained by regressors in a satisfactory manner. Especially, the widely documented serial dependence of durations is a crucial issue. It might be considered a value of its own to have a model which describes the stochastic of the random variable but yields no information which opens itself for a direct economic interpretation.

In a sequence of papers a particular model for the transaction intensity of financial markets was put forward by Engle (1996), Engle and Russell (1997), and Engle and Russell (1998). These models are based on the conditional expectation of the time between transactions given past observations on the time between transactions $\bar{\tau}_{t-1} = \begin{bmatrix} \tau_{t-1} & \tau_{t-2} & \dots \end{bmatrix}$

$$\psi_t = \mathrm{E}\left[\tau_t | \, \bar{\tau}_{t-1}\right] \tag{4.13}$$

In the context of accelerated failure time models, see (4.5), one would replace $g(x_t)$ by $g(\bar{\tau}_{t-1}) = \frac{1}{\psi_t}$. Engle and Russell suggest parametric forms of the conditional expectation to be used for estimation. A rather general ACD(p,q) model suggested by Engle and Russell (1998) is

$$\psi_t = \phi_0 + \sum_{i=1}^{p} \phi_i \tau_{t-i} + \sum_{i=1}^{q} \theta_i \psi_{t-i}. \tag{4.14}$$

The specification is not only related to the accelerated failure time models but also to the standard GARCH type model introduced by Engle and Bollerslev, see Engle (1982), Bollerslev (1986). The relationship is apparent from the characterization of the observable duration as

$$\tau_t = \psi_t \epsilon_t, \tag{4.15}$$

where ψ_t is the conditional expected duration parameterized according to (4.14). The distributional assumptions will be discussed subsequently. The analogue GARCH model in the absence of an explicit mean function would model the return process r_t by

$$r_t = h_t \cdot \epsilon_t^{\dagger}, \tag{4.16}$$

where h_t is the square root of the conditional variance of r_t and is parameterized like ψ_t. The models diverge however concerning the parametric assumptions on the error terms because $\epsilon_t^{\dagger}$ is usually chosen to be from a distribution which is symmetric around zero and ϵ_t from a distribution with positive support.

Engle and Russell (1998) point out that the simplest specification in this context is a conditional exponential distribution of ϵ_t which implies a constant baseline hazard and a conditional intensity of

$$\lambda(\tau_t) = \psi_t^{-1}. \tag{4.17}$$

Alternative specifications for the density of ϵ_t based on a Weibull distribution allow more flexibility in terms of the implied hazard rate. The constant hazard rate of the exponential density is nested by the Weibull distribution which in turn allows for an upward or downward sloping behaviour in the hazard function. Even greater flexibility for the hazard functions is implied by introducing a Burr ACD model as it was suggested by Grammig, Hujer, Kokot, and Maurer (1998). This model allows even for specific non-monotonic shapes of the hazard function and nests the Weibull as well as the exponential ACD. These models allow a parametric interpretation of the baseline hazard. Nevertheless, the results of Grammig and Maurer (2000) should be acknowledged. They show in a simulation study that these parametric models are highly sensitive to their distributional assumptions.

2.2. The pseudo maximum likelihood interpretation of ACD models. If the hazard rate is not in the focus of the analysis not much can be gained by the parametric interpretation of the ACD models. If the mean function of durations is the target of empirical research the PML estimators developed for GARCH models can be readily employed.

Engle and Russell (1998) build consequently on the work of Lee and Hansen (1994) and Lumsdaine (1996) and others in order to show consistency and asymptotic normality of the ACD(1,1) estimator under very mild assumptions in the context of pseudo ML estimators.

The asymptotic properties developed by Lee and Hansen and Lumsdaine and invoked by Engle and Russell (1998) rely on standard assumptions on ϵ_t like strict stationarity, ergodicity, bounded conditional second moments. Apart from the usual compactness condition on the parameter space, the pseudo-true distribution should be selected from the linear exponential family. Quite a spectrum of strictly parametric specifications of the ACD model have been suggested so far. The full parameterization as opposed to the pseudo maximum likelihood interpretation has the already mentioned advantage that the baseline hazard obtained is indeed meaningful. Yet, it rests on the crucial assumption that the chosen error distribution is identical with the true distribution and that the model of the conditional expectation is correctly specified.

If one recurs to the PML interpretation of ACD models it is quite important to keep in mind that the hazard function implied by the pseudo true distribution is not really meaningful, as it can only be considered to be pseudo-true.

As a solution Engle and Russell suggest the use of a non parametric estimator based on the estimated residuals of an ACD model. Given however that usually the raw durations are corrected in a first step for intraday seasonals and the parameters of the conditional expectation of the time between transactions are estimated in a second step, which does not explicitly account for the seasonal adjustments made, it is questionable whether a nonparametric third estimation step which does not account for the first two steps yields really meaningful results.

The original specification by Engle and Russel which has just been outlined has however some significant drawbacks. First, restrictions need to be imposed on ϕ and θ in order to ensure positive expected durations. Second, estimation can be still simplified by a slight change of specification. It is possible to formulate the problem in an ARMA-like framework instead of a GARCH-type one, which in turn greatly simplifies the specification of a simultaneous model, as will be shown in chapter 5.

3. Logarithmic ACD models

3.1. A parametric model specification. A slight change of the model specification as suggested by Bauwens and Giot (1997) yields the well-known log-linear specification of accelerated failure time models

$$\log \tau_t = \xi_t + \log \epsilon_t, \tag{4.18}$$

where ξ_t is the conditional expectation of the log duration, i.e. $\xi_t^* := \mathrm{E}\left[\log \tau_t | \cdot\right]$. This model relieves the user of any restrictions on the model parameters to ensure that expected durations remain positive. The main advantage can however be seen in the flexibility of this approach to incorporate a wide range of models as will be shown subsequently. This specification is easily cast into a state space form which eases the model handling, in particular its estimation. Harvey, Ruiz, and Shephard (1994), Ruiz (1994) and others have shown an approximate state-space form can also be derived for GARCH- and SV-type models. These come however at the price of a more elaborate derivation and more restrictive assumptions on the data generating process. Particularly, assumptions on the existence of higher-order moments and the true dynamics are more severe. Second, the model allows for a conditionally stochastic ξ_t, which resembles SV-models, in the context of a standard Kalman filter. Third, seasonalities and corrections for measurement errors are easily incorporated. Finally, a simultaneous specification of the price process decomposed into a process of price changes and durations is readily available.

Bauwens and Veredas (1999) propose a stochastic conditional duration model in the classical state space form as

$$\log \tau_t = \xi_t + \log \epsilon_t \tag{4.19}$$

$$\xi_t = \phi_0 + \phi \xi_{t-1} + \eta_t \tag{4.20}$$

with ξ_t being the scalar state of the system, ϕ the corresponding coefficient of the dynamic and ϕ_0 a constant. The stationarity condition $|\phi| < 1$ and distributional assumptions

$$\eta_t \sim N(0, \sigma^2)$$

$$\epsilon_t \sim W(\gamma, 1) \qquad \text{or } \epsilon_t \sim G(\nu, 1)$$

η_t independent of ϵ_s for all s, t

complete the model. The key difference between this model and the standard ACD model proposed by Engle and Russell (1998) is the presence of the additional error term η_t which renders the conditional duration ψ_t stochastic even if all observable contemporaneous information is available. This random variable however leads to the problem that for ML estimation purposes this random factor needs to be integrated out, as ψ_t is not observed directly. The straight integration would involve a N-dimensional integral, where N is the length of the time series.

If however a PML interpretation is applied and a Normal distribution is chosen as the pseudo-true distribution of the error term in the observation equation (4.19) then the Kalman filter still yields minimum mean square linear estimates of ξ_t, even in the PML setting. See Watson (1989) or the extensive discussions in Jazwinski (1970) and Anderson and Moore (1979) on different versions of the Kalman filter under different sets of assumptions. Although they do discuss a number of alternatives to estimate the model, this approach is chosen by Bauwens and Veredas (1999). Nevertheless, this approach seems to invalidate the interpretation of the hazard rate. The Gaussian pseudo-true distribution imposes stricter assumptions on the hazard function than the Weibull or Gamma distribution.

3.2. A pseudo maximum likelihood specification. As the shape of the hazard rate is not the focus of this work but an estimator as robust as possible within a framework which allows the extension to a multivariate estimator, a direct specification of a PML estimator is proposed here. Just as Bauwens and Veredas (1999) the state space of the dynamic of $\log \tau_t$ is employed.

The extensions are briefly illustrated on the basis of a logarithmic autocorrelated conditional duration, LACD(1,1), model. This is done for ease of exposition. As demonstrated in chap. 3 the extension to a LACD(p,q) model raises no particular problems and is achieved by the expansion of the state space.

Assume that the expected conditional duration given past observations is described by

$$\xi_t = \phi \log \tau_{t-1} + \theta \log \epsilon_{t-1} \tag{4.21}$$

so that the log duration is given by

$$\begin{aligned} \log \tau_t &= \phi \log \tau_{t-1} + \log \epsilon_t + \theta \log \epsilon_{t-1} &(4.22)\\ &= \sum_{j=0}^{\infty} \phi^j L^j \log \epsilon_t + \theta \sum_{j=0}^{\infty} \phi^j L^j \log \epsilon_{t-1}. &(4.23) \end{aligned}$$

If we define a two dimensional linear state space model, with the first dimension mapping the weighted sum of past innovations

$$s_{[1t]} = \sum_{j=0}^{\infty} \phi^j L^j \log \epsilon_t, \tag{4.24}$$

and the second dimension equal to the lag of the first dimension $s_{[2t]} = Ls_{[1t]}$, we obtain the measurement equation of a state space model with no measurement error,

$$\log \tau_t = H s_t, \tag{4.25}$$

with $H = \begin{bmatrix} 1 & \theta \end{bmatrix}$. It remains to resolve the dynamics of the latent equation yielding a model equivalent to (4.22)

$$\begin{aligned} s_t &= \begin{bmatrix} \phi \sum_{j=1}^{\infty} \phi^{j-1} L^j \log \epsilon_t & + \log \epsilon_t \\ \sum_{j=1}^{\infty} \phi^{j-1} L^j \log \epsilon_t & \end{bmatrix} \\ &= \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \sum_{j=1}^{\infty} \phi^{j-1} L^j \log \epsilon_t \\ L \sum_{j=1}^{\infty} \phi^{j-1} L^j \log \epsilon_t \end{bmatrix} + e_1 \log \epsilon_t \\ &= F s_{t-1} + \eta_t, &(4.26) \end{aligned}$$

with $\eta_t = e_1 \log \epsilon_t$. The equations (4.25) and (4.26) together with the distributional assumption

$$\log \epsilon_t \sim N(0, 1) \tag{4.27}$$

along with the usual assumption on the uncorrelatedness of regressors and residuals make up a complete description of the state space model. See e.g. Hamilton (1994, chap. 13) for a general treatment of ARMA(p,q) models in the state space context.

The inclusion of intra daily seasonal effects proceeds straightforward through the additional set of regressors $w_t'\gamma$ which corresponds to the usual specification where $\tau_t = \exp(\xi_t + A'x_t)\epsilon_t$. The augmented model measurement equation is

$$\log \tau_t = Hs_t + w_t'\gamma. \tag{4.28}$$

Alternatively, the set of regressors x_t might also be included in the latent model as has already been discussed in the context of the dynamic quantal response model in chapter 3.

An additional error term can be included in the measurement equation to obtain the general state space model

$$\log \tau_t = Hs_t + w_t'\gamma + R\zeta_t \tag{4.29}$$

$$s_t = Fs_{t-1} + x_t'\beta + Q\eta_t \tag{4.30}$$

$$\text{with Cov}\left[\zeta_t, \eta_t\right] = 0 \tag{4.31}$$

$$\text{and } \zeta_t \sim N(0,1)\text{and} \qquad \eta_t \sim N(0,I) \tag{4.32}$$

β and γ are appropriately defined coefficient vectors of the exogenous variables x_t and w_t. As ζ_t and η_t are assumed as standard normal distributed random variables, the corresponding variances and covariances are parameterized by R and Q. The ML estimation of the identified model under the assumption of normality proceeds using the prediction error decomposition on the basis of the Kalman filter algorithm. A rigorous derivation of this well-known algorithm along with a variety of extensions can be found in the excellent monographs by Jazwinski (1970), Anderson and Moore (1979), and Harvey (1989), or in chapter 13 of Hamilton (1994). This algorithm relies on the step-wise projection of observations into the state space, as they become available. The conditional expectation of the latent variable and its mean-squared error can be calculated iteratively starting from

$$s_{1|0} = 0 \tag{4.33}$$

$$\text{vec}P_{1|0} = (I - (F \otimes F))^{-1} \text{vec}Q. \tag{4.34}$$

Where $s_{t|t-1}$ is the conditional expectation of the state vector given all observations up to $t-1$ and $P_{t|t-1}$ is the corresponding mean squared error, MSE. Once the next observation $\log \tau_t$ becomes available, the prediction error v_t is used to update the conditional expectation of the latent state. The updating is accomplished via a projection using the Kalman gain k_t

$$v_t = \log \tau_t - Hs_{t|t-1} - w_t'\gamma \tag{4.35}$$

$$k_t = P_{t|t-1}H'(HP_{t|t-1}H' + R)^{-1} \tag{4.36}$$

$$s_{t|t} = k_t v_t. \tag{4.37}$$

The prediction of the observable $\log \tau_{t+1}$ will be computed on the base of the expected state, iterated forward one period

$$s_{t+1|t} = Fs_{t|t} + x_{t+1}'\beta. \tag{4.38}$$

The forecast MSE is iterated forward as well

$$P_{t+1|t} = F(P_{t|t-1} - k_t HP_{t|t-1})F' + Q. \tag{4.39}$$

In this context it should be emphasized that the true distribution of the error terms need not be Gaussian for the Kalman filter to work. If the true distribution is Gaussian it is well-known that the filter yields the minimum mean square error estimate of the state. However, if the true distribution is not Gaussian, the estimate $s_{t|t}$ of the state produced by the algorithm is the minimum mean square error linear estimate of the true state.

A similar argument is valid for the estimation of the model parameters. The prediction error decomposition form of the Gaussian likelihood with error v_t and associated variance σ_t is given as usually by

$$\log \mathfrak{L} = -T/2 \log 2\pi - 1/2 \sum_{t=1}^{T} \left(\log \sigma_t + \frac{v_t' v_t}{\sigma_t} \right) \tag{4.40}$$

$$\sigma_t = H P_{t|t-1} H' + R. \tag{4.41}$$

If the true distribution of the errors is not Gaussian this would be the pseudo log likelihood function and it is merely assumed that the pseudo-true distribution is normal.

Consistency and asymptotic normality in the PML case are shown by Watson (1989) under rather mild assumptions. The true data generating process (DGP) should be of the form (4.27). Apart from the usual compactness assumptions on the parameters strict stationarity of the errors and ergodicity are needed. Furthermore, the MA representation of $\log \tau_t$ needs to be identifiable. The eigenvalues of F should lie inside the unit circle and the vector ARMA representation of $\log \tau_t$ should be invertible. The first two conditional moments of v_t should be well-defined and v_t should have finite fourth moments. The proof of consistency can be found e.g. in Dunsmuir and Hannan (1976). The proof of asymptotic normality can be adapted from Watson (1989).

This completes the econometric model of the time between transactions employed in this work. It is a dynamic duration model of the ACD type which allows to include regressors either in a standard fashion or via an infinite lag structure. It is formulated to match the latent process of price changes as close as possible to ease the construction of a joint model of price changes and time between transactions.

CHAPTER 5

Joint models of the transaction process

1. On the components of the transaction process

1.1. Models of the transaction process. This chapter concentrates on the analysis of the relationship between the two components of the process of price changes and the time between transactions. Other variables such as the volume associated with a certain transaction or the transaction costs, i.e. the bid-ask spread prevailing at this point in time are left for a later treatment as these can be assessed in a straightforward extension of the econometric model suggested in this work.

As this work concentrates on the analysis of transaction data with some well-known properties which have been outlined in chapter 2, approaches which are not adapted to the properties of the data will not be discussed in detail. Hasbrouck (1991) e.g. in his work on price changes and quote revisions plainly applies standard VAR techniques, which can be justified in some sense, as he concentrates on the effects of explanatory variables for the conditional mean function. This work is extended by Dufour and Engle (2000). They build on the setup by Hasbrouck in a way which allows the parameters of the process of price changes to depend on the trade intensity modelled by an ACD specification. Ghysels and Jasiak (1998) and Grammig and Wellner (1999) modify their estimators of volatility to account for the irregular spacing of observations. However, they apply a standard GARCH approach going back to the work of Drost and Nijman (1993) which allows the scaling of a modified GARCH model to account for a change in the aggregation interval.

In the literature there are mainly two approaches available which take the microstructure properties of transaction data into account and which allow for a flexible time series structure to capture the salient features of the data. First, there is the already mentioned approach by Rydberg and Shephard (1998) which decomposes the price change d_t observed at time t into three components. An indicator for the occurrence of a price change, an indicator for the sign of the price change and the size of the price change itself. This decomposition allows to apply dynamic extensions to count data models in the context of GLMs as they were already outlined in chapter 3. A feature of the modeling strategy pursued by Rydberg and Shephard worth emphasizing is the discrete nature of time. They consider each second of the trading day and characterize it by the above mentioned random variables.

Second, Russell and Engle (1998) pursue a somewhat different approach by specifying a dynamic model for the evolution of price changes in the form of a generalized Markov Chain which was also outlined in chapter 3. Their approach is to combine this model with an ACD model, which is described in chapter 4, to obtain a joint model of price changes and time between transactions in a continuous time framework.

The model proposed in this work combines both approaches of Rydberg and Shephard and Russell and Engle. The decomposition framework of Rydberg and Shephard is modified to allow for a joint model of the size of price changes and the time between transactions within a continuous time framework. This approach yields a parsimonious model compared to both alternatives, as will be pointed out in the course of this chapter.

1.2. On the relationship of components of the transaction process. All of these procedures share the property that the joint distribution of the potentially vector valued price process is decomposed into marginal and conditional models which permit the estimation of the involved parameters and in turn an economic interpretation of the observed relationships. Assume for the moment that the transaction process y_t consists of two components, the time between transactions τ_t and the price change d_t, i.e. $y_t := \begin{bmatrix}\tau_t & d_t\end{bmatrix}$, the conditional joint density is given by $f(\tau_t, d_t|\bar{\tau}_{t-1}, \bar{d}_{t-1}, \bar{x}_{t-1}, \theta)$ using θ as the parameter vector of the joint distribution. Engle, Hendry, and Richard (1983) have defined the notions of weak, strong, and super exogeneity to describe the properties of the joint distribution concisely. The concepts of exogeneity are always related to a set of parameters which are of interest to the researcher. The different concepts indicate which conclusions might be drawn from particular distributional assumptions or rather how results of inference are to be interpreted economically.

If one defines $f_{d|\tau}$ as the conditional density with parameters θ_1 of the price changes d_t given the standard information set and the contemporaneous time between transactions τ_t and f_τ as the marginal density of the time between transactions with parameters θ_2, then weak exogeneity as given in Engle, Hendry, and Richard (1983, def. 2.5) expresses whether inference on θ_1 is independent from inference on θ_2. Consider the factorization of the joint density

$$f(\tau_t, d_t|\bar{\tau}_{t-1}, \bar{d}_{t-1}, \bar{x}_{t-1}, \theta) = f_{d|\tau}(d_t|\tau_t, \bar{\tau}_{t-1}, \bar{d}_{t-1}, \bar{x}_{t-1}, \theta_1) \cdot f_\tau(\tau_t|\bar{\tau}_{t-1}, \bar{d}_{t-1}, \bar{x}_{t-1}, \theta_2). \quad (5.1)$$

Two prerequisites have to be fulfilled for weak exogeneity. The factorization as given in (5.1) has to be possible under the condition that the parameters θ_1 and θ_2 are variation free. This means that there should be no parameter restrictions involving both θ_1 *and* θ_2. If this is the case, then the estimation is greatly simplified as the likelihood based on f_τ can be maximized independently with respect to θ_2 without loss of information.

The notion of strong exogeneity relies on the additional condition that d_t does not Granger cause τ_t, i.e. the conditional distribution of τ_t is independent of past d_t given past values of τ_t. Then, the joint distribution can be reduced to

$$f(\tau_t, d_t|\bar{\tau}_{t-1}, \bar{d}_{t-1}, \bar{x}_{t-1}, \theta) = f_{d|\tau}(d_t|\tau_t, \bar{\tau}_{t-1}, \bar{d}_{t-1}, \bar{x}_{t-1}, \theta_1) \cdot f_\tau(\tau_t|\bar{\tau}_{t-1}, \bar{x}_{t-1}, \theta_2).$$

This means that not only estimation can be carried out independently of the second distribution, but also forecasts of τ_t depend exclusively on the history of τ_t and other exogenous variables and not on the history of d_t. The same line of arguments can obviously be drawn for the weak and strong exogeneity of d_t with respect to τ_t. If strong exogeneity would hold in both directions one could consider both processes to be independent.

Super exogeneity is strictly related to the question whether it is possible to control the process by manipulating a particular variable. What is needed to ensure this property is that in addition to weak exogeneity the parameter of interest is either invariant for any change in the distribution of conditioning variables or least invariant with respect to a particular class of interventions. See definitions 2.7 and 2.8 of Engle, Hendry, and Richard (1983). However, this concept will not play a significant role in the empirical analysis pursued in this work.

2. A GLM model in discrete time

2.1. A decomposition in terms of conditional probabilities. Due to the fact that the approach suggested by Rydberg and Shephard (1998) employs a discrete time framework, the index t of observations is here identical with the clock time measured in seconds since the trading start. Thus an observation is available for each second of the trading day. To discriminate between the price change d_t in a continuous time framework, where d_t is only observed if a trade occurs and the price change in a discrete time model, random variables are marked by a circle in the latter case. So d_t° stands for the price change in the discrete time context.

The decomposition of Rydberg and Shephard heavily utilizes the dynamic framework for GLMs outlined in chapter 3, to capture the dynamics of the three components, price activity a_t°, direction of price changes s_t° and size of price changes z_t°. The price activity, a_t°, is a variable taking on a value of one, if a transaction was observed at time t and a price change has taken place and is zero otherwise. The price change itself is decomposed into its direction s_t° taking on values $+1$ and -1 and its size z_t°, which is integer valued, i.e. $z_t^\circ \in \{0, 1, 2, \ldots\}$. As already mentioned the discrete time framework leads to an observation in each second having the form of a product

$$d_t^\circ = a_t^\circ s_t^\circ z_t^\circ. \tag{5.2}$$

Using appropriate definitions of the three components, the distribution of d_t° can be formulated without loss of generality based on the conditional distribution of the components, where v_i is one of J possible distinct price changes.

$$\text{Prob}\left[d_t^\circ = v_i \,|\mathcal{F}_{t-1}\right] =$$
$$\begin{cases} \text{Prob}\left[a_t^\circ = 0 \,|\mathcal{F}_{t-1}\right] & \text{if} \quad v_i = 0 \\ \text{Prob}\left[z_t^\circ = v_i \,|\mathcal{F}_{t-1}, a_t^\circ = 1, s_t^\circ = 1\right] \\ \cdot \text{Prob}\left[s_t^\circ = 1 \,|\mathcal{F}_{t-1}, a_t^\circ = 1\right] \cdot \text{Prob}\left[a_t^\circ = 1 \,|\mathcal{F}_{t-1}\right] & \text{if} \quad v_i > 0 \\ \text{Prob}\left[z_t^\circ = -v_i \,|\mathcal{F}_{t-1}, a_t^\circ = 1, s_t^\circ = -1\right] \\ \cdot \text{Prob}\left[s_t^\circ = -1 \,|\mathcal{F}_{t-1}, a_t^\circ = 1\right] \cdot \text{Prob}\left[a_t^\circ = 1 \,|\mathcal{F}_{t-1}\right] & \text{if} \quad v_i < 0 \end{cases} \tag{5.3}$$

Note that this specification does not discriminate between the occurrence of a trade without a price change and the event that no trade occurs at time t. Both events are captured by $a_t^\circ = 0$ and thereby $d_t^\circ = 0$. Note that $s_t^\circ = 0$ and $z_t^\circ = 0$ are defined to equal zero if $a_t^\circ = 0$.

2.2. GLMs for individual components. Price activity is modeled as a GLM based on a logistic distribution including ARMA(p,q) type dynamics as outlined in chapter 3. The dynamic model for a_t° is thus specified as

$$\mu_{a,t} = \text{Prob}\left[a_t^\circ = 1 \,|\mathcal{F}_{t-1}\right] \tag{5.4}$$

$$G\left(\mu_{a,t}\right) = x_t'\beta + g_{a,t}. \tag{5.5}$$

The variable $\mu_{a,t}$ is the conditional expectation of the random variable a_t° given the relevant information $\mathcal{F}_{t-1}$, i.e. $\text{Prob}\left[a_t^\circ = 1 \,|\mathcal{F}_{t-1}\right] = \text{E}\left[a_t^\circ |\, \mathcal{F}_{t-1}\right]$. The link function G is motivated from the logistic distribution, thus defined as

$$G(\mu_{a,t}) = \frac{\exp(\mu_{a,t})}{1 + \exp(\mu_{a,t})}. \tag{5.6}$$

The systematic component capturing the dynamics of a_t° is defined as in chapter 3 as

$$g_{a,t} = \sum_{j=1}^{p} \gamma_j g_{a,t-j} + \sum_{j=1}^{q} \delta_j u_{a,t-j}, \tag{5.7}$$

$$\text{with } u_{a,t} = \frac{a_t^\circ - \mu_{a,t}}{\sqrt{\mu_{a,t}(1-\mu_{a,t})}} \tag{5.8}$$

where $u_{a,t}$ is a martingale difference sequence, mimicking the error term in a conventional ARMA model. The model of the trade direction is quite similar to this specification. Actually, it is just an auto logistic model as the empirical results show that the direction of price changes seems to be driven mainly by the bid-ask bounce.

The size of price changes z_t° is at least potentially a count variable. However, price changes larger than one are observed only rarely in liquid markets. To accommodate a count model Rydberg and Shephard employ again a GLM based on a negative binomial (NegBin) distribution with an ARMA dynamic to model the size of price changes minus one, $z_t^\circ - 1$. The NegBin generalizes the Poisson distribution as it allows for overdispersion, i.e. conditional mean and variance need not necessarily be restricted to a constant ratio. Using Γ as the Gamma function and $v_{z,i} \in \{1, 2, 3, \ldots\}$ as the possible values z_t° can take on, the model is specified as

$$\text{Prob}\left[z_t^\circ = v_{z,i} \,|\mathcal{F}_{t-1}, a_t^\circ = 1, s_t^\circ\right] = \frac{\Gamma(\alpha + v_{z,i} - 1)}{\Gamma(\alpha)(v_{z,i}-1)!}\left(\frac{\alpha}{\mu_{z,t}+\alpha}\right)^{\alpha}\left(\frac{\mu_{z,t}}{\mu_{z,t}+\alpha}\right)^{v_{z,i}-1}.$$

The additional parameter α allows for some flexibility in the relationship of the conditional mean and the conditional variance as can be seen from the implied moments

$$\text{E}\left[z_t^\circ|\,\mathcal{F}_{t-1}, a_t^\circ = 1, s_t^\circ\right] = 1 + \mu_{z,t}$$

$$\text{Var}\left[z_t^\circ|\,\mathcal{F}_{t-1}, a_t^\circ = 1, s_t^\circ\right] = \mu_{z,t} + \frac{1}{\alpha}\mu_{z,t}^2.$$

Note that the mean is shifted by one, due to the fact that the model describes the size of price changes minus one. Note that this shift cancels out in the conditional variance. If the parameter α goes to infinity the model specializes to the Poisson model and the conditional variance is equal to the conditional mean minus the deterministic shift. The GLM specification is quite straightforward as $\mu_{z,t}$ is restricted to be positive we have

$$\mu_{z,t} = \exp(x_t'\beta + g_{z,t}), \tag{5.9}$$

where $g_{z,t}$ is specified similar to $g_{a,t}$ in (5.7) with $u_{z,t}$ appropriately defined as

$$u_{z,t} = \frac{z_t^{\circ} - 1 - \mu_{z,t}}{\sqrt{\mu_{z,t} + \frac{1}{\alpha}\mu_{z,t}^2}}. \tag{5.10}$$

The modeling framework proposed by Rydberg and Shephard has the considerable advantage that the decomposition of the price process allows to build flexible models for each component and allows to consider and interpret each component in turn. On the other hand, the individual processes can be combined via (5.2) or a similar definition of a composed random variable of interest, to obtain empirical evidence on the trading process as a whole.

Yet, the main problem associated with the discrete time framework chosen by Rydberg and Shephard is the identity of the observation index t with the clock time. The problem becomes apparent if one considers a very active trading phase, e.g. trading after the opening of an exchange, and a very slow trading phase, e.g. during lunch time. Assume further, to have an easy example, yet without invalidating the argument in a more complex setting, that the true DGP is such that the current time between transactions depends via an ACD(1,0) model only on the preceding time between transactions and the time of day, governing the intraday seasonals. In the log ACD context outlined in chapter 4, the difference between the slow and fast trading could be captured by an additive component, e.g. a trigonometric seasonal, in the mean function. For the model in discrete time however, a time varying coefficients model would be necessary to capture the seasonality effects. This is easy to understand: in a fast trading phase lags of the dependent variable with a lower order should have the largest weight. In a slow trading phase, on the other hand, lags with a higher order should have a greater weight. A model for this kind of seasonally varying coefficients was suggested in the GARCH context by Bollerslev and Ghysels

(1996) in the form of a periodic GARCH model. Their work was also motivated by an intraday analysis on the basis of equally spaced observations. They, however, used time aggregates which of course suffer from the same effect, if the aggregation interval is chosen small enough. A modification of the given GLM models in the sense of Bollerslev and Ghysels would however complicate matters considerably, if it is possible at all. These effects might be the reason for the large number of lags Rydberg and Shephard (1998) need to model the price activity a_t° in a satisfactory fashion. This is also the reason why a continuous time framework employing an ACD type dynamics is used in this work.

3. A GLM model in continuous time

3.1. A decomposition in terms of the conditional hazard rate. Engle and Russell (1998) in their work on ACD models propose to decompose the price process into price changes and transaction intensity, i.e. time between transactions

$$y_t' := \begin{bmatrix} d_t & \tau_t \end{bmatrix}. \tag{5.11}$$

The joint density of y_t can be written as the product of the conditional density of the price changes $f_{d|\tau}$ and the marginal density of the time between transactions f_τ without any further implications as

$$f_y(y_t|\bar{d}_{t-1}, \bar{\tau}_{t-1}) = f_{d|\tau}(d_t|\bar{d}_{t-1}, \bar{\tau}_t) \cdot f_\tau(\tau_t|\bar{\tau}_{t-1}, \bar{d}_{t-1}). \tag{5.12}$$

The autoregressive conditional multinomial model (ACM) which is used to model $f_{d|\tau}$ has already been outlined in chapter 3 as a generalization of a Markov chain. Compared to the latter it allows for a more complex dynamic. The marginal density of the duration f_τ is modeled using the ACD model which was already described in chapter 4.

Russell and Engle (1998) specify their joint model as a special competing risks model in which the hazard rate is split into intensities of price changes of size v_j or more concisely, the hazard rate $\lambda^{ACM}(\tau|\mathcal{F}_{t-1})$ is a J-dimensional vector, assuming that d_t

can take on J different values. It has the form

$$\lambda_{t[j]}^{ACM}(s|\mathcal{F}_{t-1}) = \lim_{c\to 0} \frac{\text{Prob}\left[s \leq \tau_t < s + c, d_t = v_j \,|\tau_t \geq s, \mathcal{F}_{t-1}\right]}{c}, \qquad j = 1, \ldots, J. \tag{5.13}$$

Note that here a joint probability is used as a basis to formulate the hazard rate as opposed to the standard definition of the hazard rate in (4.2). The joint probability in (5.13) is then factored by Russell and Engle into a product of the conditional density to observe a certain price change and the standard ACD hazard rate λ^{ACD}, as it is implied by a regular ACD model, e.g. in(4.17),

$$\lambda_t^{ACM}(s|\mathcal{F}_{t-1}) = \lambda_t^{ACD}(s|\bar{d}_{t-1}, \bar{\tau}_{t-1})\mu_t^{ACM}(\bar{d}_{t-1}, \bar{\tau}_t). \tag{5.14}$$

Thus, it remains to specify a process for the conditional probabilities μ_t^{ACM}.

3.2. A GLM the process of price changes. The conditional probability μ_t^{ACM} is derived from a modified GLM based on a multinomial Logit with an ARMA dynamic and some additional terms included to account for the time between transactions.

$$G(\mu_t^{ACM}) = m_t^{ACM} \tag{5.15}$$

$$\text{with } G(\mu_t^{ACM}) = \frac{\exp(\mu_t^{ACM})}{1 + \exp(\mu_t^{ACM})}$$

$$\text{and } m_t^{ACM} = \alpha_1 \log \tau_t + \alpha_2 \tau_t + \alpha_3 \frac{\tau_t}{\psi_t} + \alpha_4 \psi_t + g_t \tag{5.16}$$

$$g_t = \sum_{j=1}^{p} \gamma_j m_{t-j}^{ACM} + \sum_{j=1}^{q} \delta_j u_{t-j} + \sum_{j=1}^{r} \rho_j D_{t-j}$$

$$u_t = V_t^{-1/2}\left(D_t - \mu_t^{ACM}\right),$$

where ψ_t is the conditional expectation of the time between transactions τ_t implied by the ACD model. The matrix V_t is a scaling factor which possibly compensates the variation of the "residual" $D_t - \mu_t^{ACM}$. In the modeling framework of Rydberg and Shephard this factor is defined as the conditional variance of this difference term. Russell and Engle leave the definition of the factor more or less open.

Russell and Engle claim that τ_t could not be considered weakly exogenous in this model because τ_t enters the conditional likelihood of the transaction price changes' model. Yet, simultaneity of the of the time between transactions τ_t and the price changes d_t would involve a dependence between the martingale difference sequence u_t and the error term of the ACD model ϵ_t in (4.15). If however both error terms were correlated the estimates of the coefficients of the functions of the durations τ_t in (5.16) would be rendered inconsistent by a simultaneous equations bias. Because of this reason a slightly different approach to build a multivariate model is chosen in this work. Here, the hazard rate is not specified state dependent as in (5.13) but an assumption is made on the joint distribution of the error terms of the dynamic for the durations τ_t and price changes d_t.

A comparison of the solutions of Russell and Engle and Rydberg and Shephard yields not a really clear cut picture. The advantage of Russell and Engle (1998) is the use of an ACD model to obtain a parsimonious duration model in continuous time, which allows e.g. a straightforward inclusion of seasonalities. The advantage of Rydberg and Shephard's approach is the stringent and parsimonious inclusion of dynamics in the specifications of the components of the process of price changes and the decomposition of the transaction process into components of interest, e.g. the size of price changes. Both models unites however the drawback that the simultaneity of individual components of the transaction process cannot be assessed from the estimation results but needs to be imposed by assumption. A model which allows to assess simultaneity issues between the individual variables and which is based on a decomposition of the price process while using a continuous time framework is proposed in the next section.

4. A latent linear model of price changes in continuous time

4.1. The decomposition of the price process in sign and size. As the work of Rydberg and Shephard has shown, some flexibility in modeling can be gained by decomposing the price process. A significant drawback of their approach lies however in the discrete nature of the price frequency model employed. Therefore a modified approach is pursued in this context, where the time between transactions is taken as a continuous variable which can be modeled by an ACD type model. However, before the bivariate model of trade frequency and price changes is proposed, the decomposition of the price process in different components is discussed, leaving open the duration for later treatment. By conditioning all probabilities on the time between transactions τ_t, these components can then be easily employed in the next section to obtain a model of the joint distribution of the transaction process y_t consisting in this context of durations τ_t and individual components of the price changes d_t.

The price change d_t which is in this context only observed if a trade occurs, is decomposed into the size of price changes z_t and the sign of price changes s_t. This deviates substantially from the price change d_t° employed by Rydberg and Shephard which is observed at every period (second) of the trading day. Here again, the sign takes on values $s_t \in \{-1, 0, 1\}$ and the size $z_t \in \{0, 1, 2, \ldots\}$. The definition of the decomposition is thus

$$d_t = z_t \cdot s_t. \tag{5.17}$$

Contrary to Rydberg and Shephard there is no particular component which accounts for the occurrence of a price change, like a_t°, this is captured in this model by the random variable z_t. Finally, the occurrence of a transaction is explicitly modelled by the time between transactions random variable τ_t. Trade occurrence is a component of this model and not only price changes.

The probability to observe a certain price change d_t of size v_i at time t

$$\text{Prob}\left[d_t = v_i \,|\tau_t, \mathcal{F}_{t-1}\right] = \text{Prob}\left[z_t = \text{abs}\left(v_i\right), s_t = \text{sgn}\left(v_i\right) |\tau_t, \mathcal{F}_{t-1}\right],$$

can thus be decomposed into

$$\text{Prob}\left[z_t = \text{abs}\left(v_i\right), s_t = \text{sgn}\left(v_i\right) |\tau_t, \mathcal{F}_{t-1}\right] =$$
$$\text{Prob}\left[s_t = \text{sgn}\left(v_i\right) |z_t = abs(v_i), \tau_t, \mathcal{F}_{t-1}\right] \cdot \text{Prob}\left[z_t = abs(v_i) \,|\tau_t, \mathcal{F}_{t-1}\right].$$

The two processes of size z_t and sign s_t as defined above can not be independent, as it is quite obvious that the size of the price change is informative for the direction of the price change with respect to the zero price change event, i.e.

$$\text{Prob}\left[s_t = 0 \,|z_t = abs(v_i) = 0, \tau_t, \mathcal{F}_{t-1}\right] = 1 \tag{5.18}$$

$$\text{and } \text{Prob}\left[s_t = 0 \,|z_t = abs(v_i) > 0, \tau_t, \mathcal{F}_{t-1}\right] = 0. \tag{5.19}$$

However, it can be assumed that the size of price changes beyond the zero event carries no additional information for the sign of price changes, or more concisely

$$\text{Prob}\left[s_t = \text{sgn}\,(v_i)\,|z_t = abs(v_i) > 0, \tau_t, \mathcal{F}_{t-1}\right] = \text{Prob}\left[s_t = \text{sgn}\,(v_i)\,|z_t > 0, \tau_t, \mathcal{F}_{t-1}\right], \tag{5.20}$$

which reflects the fact that the explicit size $abs(v_i)$ of the price changes, in addition to the fact that the size is larger than zero, $z_t > 0$, carries no additional information with respect to the sign of price changes. The nature of this assumption and its consequences is discussed in the introductory Markov chain example given in section 3 of chapter 2.

To model the conditional probabilities (5.20) in practice, there are two ways to proceed, either one models a random variable $s^\dagger_{t_k} \in \{-1, 1\}$, $k = 1, \ldots, K$, which is only observed, if there actually is a price change, i.e. $z_{t_k} > 0$. The latter defines the indices t_k which perform a thinning of the original process indexed by $t = 1, \ldots, T$, i.e. the indices are defined such that $t_k \in \{t = 1, \ldots, T : z_t > 0\}$. This in turn implies that the random variable $s^\dagger_t$ can be assumed to be independent of the size of price changes z_t, i.e.

$$\text{Prob}\left[s^\dagger_{t_k} = \text{sgn}\,(v_i)\,|z_{t_k} > 0, \tau_{t_k}, \mathcal{F}_{t_k-1}\right] = \text{Prob}\left[s^\dagger_{t_k} = \text{sgn}\,(v_i)\,|\tau_{t_k}, \mathcal{F}_{t_k-1}\right]. \tag{5.21}$$

Then again, the price change is decomposed as in (5.17) but the sign of the price change s_t is defined as

$$s_t = \begin{cases} 0 & \text{if } z_t = 0 \\ s^\dagger_{t_k} & \text{if } z_t > 0. \end{cases} \tag{5.22}$$

A dynamic model for the thinned process $s^\dagger_{t_k}$ can thus be employed to evaluate the conditional probabilities required to obtain the distribution of the price change d_t, this dynamic model could e.g. be a GLM or dynamic ordered probit model.

The thinning of the available observations yielding only those $K \leq T$ observations with a price change unequal zero raises however the problem that in a dynamic specification, where the information set $\mathcal{F}_{t_k}$ is made up mainly by past observations of the endogenous variable, the information available is rather limited. Lags of the dependent variable are easily included in the model, i.e. $\bar{s}^{\dagger}_{t_{k-1}} \in \mathcal{F}_{t_k-1}$. Yet, the inclusion of zero price change observations raises substantial problems, i.e. some observation of d_t are not in the information set $\bar{d}_s \notin \mathcal{F}_{t-1}$ if $d_s = 0$, $s \leq t-1$. If one is concerned by this loss of information, one might well choose another alternative to model the process s_t to obtain the necessary conditional probabilities $\text{Prob}\left[s_t = \text{sgn}\left(v_i\right) | z_t > 0, \tau_t, \mathcal{F}_{t-1}\right]$. This involves the specification of the conditional probabilities via a model of the sign of price changes s_t unconditional on the size of price changes. The conditional probability of the sign of the price change is thus defined for the observations showing a price change as

$$\text{Prob}\left[s^{\dagger}_{t_k} = \text{sgn}\left(v_i\right) | z_t > 0, \tau_t, \mathcal{F}_{t-1}\right] := \frac{\text{Prob}\left[s_t = \text{sgn}\left(v_i\right) | \tau_t, \mathcal{F}_{t-1}\right]}{1 - \text{Prob}\left[s_t = 0 | \tau_t, \mathcal{F}_{t-1}\right]} \quad \text{if} \quad t = t_k. \quad (5.23)$$

By the description of the decomposition of the price process in (5.17) it is shown that the distribution of the process of price changes d_t can indeed be recovered from models of the conditional probabilities of the size of price changes $\text{Prob}\left[z_t = abs(v_i) | \tau_t, \mathcal{F}_{t-1}\right]$ and the sign of price changes $\text{Prob}\left[s_t = \text{sgn}\left(v_i\right) | \tau_t, \mathcal{F}_{t-1}\right]$.

The advantage of this decomposition is that it allows to examine the size and the sign process independently of each other. Supposedly, the sign process should be heavily influenced by the well-documented bid-ask bounce. The size of price changes or rather the absolute price change on the other hand is of particular interest for the analysis of volatility. For an extended exposition of the role of absolute price changes in the economic analysis of risk, see e.g. Granger and Ding (1995). They discuss the use of the conditional distribution of absolute price changes, particularly the use of various functions thereof, where z_t^2 as the variance under the mean zero assumption is just the major example.

The assumption needed to establish this decomposition is that the sign process is independent from the size of price changes beyond the fact whether a price change occurs or not, see (5.21). If these processes were not independent, it would imply that either positive or negative price changes would come along with particularly large or small price changes. If one would reject the assumption made and strive to assess this as a hypothesis, it would well be possible. In order to model the bivariate distribution of size and sign of price changes a bivariate ordered probit would be needed. In this context however, both processes are modelled independently of each other but as joint models with the duration process. This combination is achieved quite easily given the latent dynamic developed in chapter 3 and the log ACD specification in the Kalman filter context described in section 4.

4.2. A joint model of time between transactions and the price process. In the preceding subsection a decomposition of the price change process d_t into a process of the size and the sign of price changes is proposed. This decomposition is used in this chapter and thus two joint models with dependent variables $y_{s,t}$ and $y_{z,t}$ of are defined as

$$y_{s,t} := \begin{bmatrix} s_t & \log \tau_t \end{bmatrix} \text{ and} \tag{5.24}$$

$$y_{z,t} := \begin{bmatrix} z_t & \log \tau_t \end{bmatrix}. \tag{5.25}$$

The model structure of the joint model involving the size of price changes $y_{z,t}$ and the model involving the sign of price changes $y_{s,t}$ is almost identical. Thus, for the sake of brevity, only the model for $y_{z,t}$ is described in depth, the model for $y_{s,t}$ follows directly. A few minor differences will be pointed out along the model description.

The joint density $f_{z,\tau}$ of the bivariate process $y_{z,t}$ can be factored into the conditional distribution of the size of price changes $f_{z|\tau}$ and the marginal distribution of time between transactions f_τ to obtain

$$f_{z,\tau}(z_t, \tau_t|\mathcal{F}_{t-1}) = f_{z|\tau}(z_t|\tau_t, \mathcal{F}_{t-1}) f_\tau(\tau_t|\mathcal{F}_{t-1}). \tag{5.26}$$

The conditional probability function of the discrete random variable z_t and its relationship to the process of price changes d_t was already discussed in the preceding subsection. See e.g. the discussion of (5.23). Thus, it can be used here as a building block for the joint model, i.e.

$$f_{z|\tau} = \text{Prob}\left[z_t = \text{abs}\left(v_i\right)|\tau_t, \mathcal{F}_{t-1}\right]. \tag{5.27}$$

To model $f_{z|\tau}$, and the analogue $f_{s|\tau}$, a latent dynamic model in the form of an ordered probit as it is described in chapter 3 will be employed. The marginal distribution f_τ will be derived from a log ACD model which has a very similar structure, as it also builds on the state space representation of an ARMA model possibly including exogenous regressors. The combination of both components, the duration and the price change, will allow to asses the potential simultaneity of the variables.

Concisely, the bivariate process $y_{z,t}$ of absolute price changes and the log time between transaction is given by a combination of the models in chapter 3 and chapter 4. The latent process z_t^* and the mean function $m_{z,t}^{\xi}$ of the latent model of the size of price changes take the following form,

$$z_t^* = H_z m_{z,t}^{\xi} + w_{z,t}'\gamma_z + \kappa_z \log\tau_t + \epsilon_{z,t}^* \tag{5.28}$$

$$m_{z,t}^{\xi} = F_z\left(m_{z,t-1}^{\xi} + e_1 c_{z,t-1}\right) + e_1 x_{z,t}'\beta_z. \tag{5.29}$$

In addition to the model outlined in chapter 3 the contemporary observation of the time between transactions is included. See e.g. eqs. (3.87) and (3.89). This variable is of particular interest, as it allows to assess the consequences a slow or a fast order execution has on the size of price changes. Contrary to lags of the duration variable which capture only the general state of the market, this variable characterizes the particular transaction under consideration. The contemporaneous time between transactions τ_t is however potentially simultaneous with the size of price changes z_t. The latter would render a standard maximum likelihood estimation of (5.28) and (5.29) inconsistent due to a simultaneous equation bias. Thus a nonlinear simultaneous equation system is estimated with the duration as a second component.

The model of time between transactions is based on the modified log ACD model which has already been outlined. However, we have adjusted the exposition to match the process in (5.29) and (5.28).

$$\log \tau_t = H_\tau \xi_{\tau,t} + w_{\tau,t}\gamma_\tau + \epsilon_{\tau,t} \tag{5.30}$$

$$\xi_{\tau,t} = F_\tau \left(\xi_{\tau,t-1} + e_1\epsilon_{\tau,t-1}\right) + e_1 x'_{\tau,t}\beta_\tau. \tag{5.31}$$

Given that durations are directly observable, we omit a asterisk from the error term $\epsilon_{\tau,t}$. This equation can be seen as a reduced form equation, as it does not include the contemporaneous dependent variable of the size equation. The regressors $w_{\tau,t}$ and $x_{\tau,t}$ might include lags of the size of price changes since however the latent size of price changes is not observable, the same strategy as in the latent dynamic of the ordered probit is pursued and the variable z_t^* is substituted by its conditional expectation given the past observations of the dependent variable $\mathrm{E}\left[z_t^* \middle| \mathcal{F}_{t-1}\right]$.

Both error terms are collected in a redefined $\epsilon_t := \begin{bmatrix}\epsilon_{z,t}^* & \epsilon_{\tau,t}\end{bmatrix}$. For the error terms the assumption of joint normality is introduced:

$$\epsilon_t \sim N(0, \Sigma) \tag{5.32}$$

$$\text{with } \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \tag{5.33}$$

Given that $y_t^{*\prime} := \begin{bmatrix} z_t^* & \log \tau_t \end{bmatrix}$ is a function of ϵ_t and the mean function, the bivariate Gaussian distribution translates into the needed distribution of the dependent variables given the Jacobian $J_{\epsilon_t}^*(z_t^*, \log \tau_t)$ using ϕ_Σ as the density of the bivariate Gaussian distribution with covariance Σ

$$f(z_t^*, \log \tau_t | \bar{x}_t, \bar{w}_t) = \phi_\Sigma(\epsilon_t) \cdot |J_\epsilon^*(z_t^*, \log \tau_t)|. \tag{5.34}$$

The Jacobian takes on a particularly simple form given that the derivative of $\epsilon_{\tau,t}$ with respect to z_t^* is zero because the dependent variable itself is not included in the duration equation (5.30) but only its lags. Thus the Jacobian is equal to one. If the dependent variable z_t^* had been included, then it would have equalled $(1 - \kappa_d \kappa_\tau)$, if κ_d is the corresponding coefficient.

As it has already been argued in chapter 3, the given form of the latent dynamic allows to formulate the likelihood as a product over the conditional densities of T

observations. Thus, the dimension of integration to reflect the fact that not z_t^* but only its discrete transformation z_t is observable is only one and the log Likelihood takes the form:

$$\log \mathfrak{L} = \sum_{t=1}^{T} \sum_{j=1}^{J+1} 1_{(d_t = v_j)} \cdot \log \int_{A_j} f(d_t^*, \log \tau_t) d(d_t^*). \tag{5.35}$$

The dynamic in the latent variable z_t^* raises no particular problems, if the generalized error on which the filtering algorithm proposed in chapter 3 is appropriately redefined. Remember that originally the generalized error is defined[1] as

$$c_{z,t}^{\dagger} = \mathrm{E}\left[\epsilon_{z,t}^{*} \middle| \bar{z}_t, \bar{x}_t, \bar{w}_t\right]. \tag{5.36}$$

The variables $\bar{x}_t$ and $\bar{w}_t$ collect as usually the regressors of the duration and the size equation and their history. In the bivariate context this should be redefined to include all available contemporary information to obtain

$$c_{z,t} = \mathrm{E}\left[\epsilon_{z,t} \middle| \bar{z}_t, \bar{\tau}_t, \bar{x}_t, \bar{w}_t\right]. \tag{5.37}$$

The derivation of the modified filtering algorithm proceeds in a straightforward fashion by decomposing the original error term in a component dependent on the duration innovation $\epsilon_{\tau,t}$ and an error term independent of the second equation $\nu_{z,t}$ using the assumption of a bivariate normal distribution with correlation ρ and unit variances

$$\epsilon_{z,t} = \rho \epsilon_{\tau,t} + \nu_{z,t}. \tag{5.38}$$

The conditional expectation of the independent error $\nu_{z,t}$ is simply denoted by $c_{z,t}^{\nu}$. Thus the state equation (5.31) of the price change model can be modified to obtain

$$\xi_{z,t} = F_z \left(\xi_{z,t-1} + e_1 (\rho \epsilon_{\tau,t-1} + c_{z,t-1}^{\nu}) \right) + e_1 x_{z,t}' \beta_z. \tag{5.39}$$

This modified updating equation is quite intuitive, since the conditional expectation of the latent state ξ_t is updated using not only the available information on the observations of the discrete dependent variable but also the information on the exogenous regressors and the error term of the duration equation.

[1] In the chapter 3 on the dynamic ordered probit the generalized error was denoted by c_t. The bivariate nature of the model necessitates the extension of the notation.

An alternative to the joint maximum likelihood estimation of (5.28) and (5.30) is a sequential estimation procedure involving a conditional maximum likelihood estimator of the size of price changes equation (5.28) given the estimation of the duration equation (5.30).

The joint likelihood takes on a particularly convenient form which allows to estimate the time between transactions equation in a first step and the price change model in a second step

$$\begin{aligned}\log\,\mathfrak{L} &= \log\,\mathfrak{L}_{d|\tau} + \log\,\mathfrak{L}_{\tau}\\ &= \sum_{t=1}^{T}\sum_{j=1}^{J} 1_{(d_t=v_j)} \log p^{\star}_{t,j}\\ &+ \text{const.} - \sum_{t=1}^{T} \log\sigma^2_{\tau} - \frac{1}{2}\left(\frac{\epsilon_{\tau,t}}{\sigma_{\tau}}\right)^2 .\end{aligned}$$

Note that $\epsilon_{\tau,t}$ and σ_{τ} are given from the prediction error decomposition in the Kalman filter. The conditional probability $p^{\star}_{t,j}$ is derived easily from the modified dynamic model as

$$\begin{aligned}p^{\star}_{j,t} &:= \text{Prob}\left[z_t = v_j\,|\bar{z}_{t-1}, \bar{\tau}_t, \bar{x}_t, \bar{w}_t\right] \qquad j = 1,\ldots,J \qquad (5.40)\\ &= \Phi(\nu_{j,t}) - \Phi(\nu_{j-1,t})\\ \nu_{j,t} &= \mu_j - \text{E}\left[z^{*}_t|\,\bar{z}_{t-1}, \bar{\tau}_t, x_t, w_t\right]. \qquad (5.41)\end{aligned}$$

Note that in the second step the covariance matrix of the estimator is only consistently estimated under the null hypothesis of independence. Otherwise a correction term would be needed to be introduced which accounts for the first step estimation. See the original paper by Smith and Blundell (1986) on a simultaneous Tobit, the work by Pohlmeier (1989) on simultaneous Probit or the survey by Blundell and Smith (1993).

This model can be directly extended to the case where τ_t is replaced by a multivariate process including not only the time between transactions but also variables like the volume associated with a particular trade. The only limitation for the type of included variables is that they should be observed on a scale which allows to use a multivariate state space model along the lines of the duration model described in chapter 4. Then the marginal model is also multivariate and the error term of the size equation will be decomposed into the effects due to the innovations on the multivariate marginal model and an orthogonal term analogue to (5.38).

5. The analysis of impulse responses

5.1. The standard impulse response function. The estimation results of a complex dynamic system as it is described by the specification of $y_{z,t}$ in (5.28) - (5.31) are hard to interpret, since it is a formidable task to resolve the dynamic properties by a mere inspection of the individual coefficients. The standard tool to describe systems of dynamic models is the impulse response function Γ_s. The standard definition of the impulse response function is based on conditional s-step ahead forecasts, i.e. $\mathrm{E}\left[y_{z,t+s} \middle| \bar{\epsilon}_t \right]$, given a certain state of the system in t, typically described by past innovations to the system $\bar{\epsilon}_{t-1}$ and the present innovation ϵ_t. The standard way to quantify the effects of an innovation on the system is the analysis of partial derivatives of the forecasts of the individual dependent variables, here the size of price changes z_t and the log time between transactions $\log \tau_t$, with respect to the innovation is under consideration. For a general discussion see e.g. Hamilton (1994, chap. 11.4). To demonstrate the problems of this procedure in the given context and to describe the advantages of the alternative, the standard impulse response based on partial derivatives is derived for the size component of the model assuming a latent dynamic model instead of a limited dependent model. The starting point is the definition of the impulse response function Γ^*_s after s periods to a shock $\epsilon_{z,*}$ in the reference period t

$$\Gamma^*_s := \left. \frac{\partial \mathrm{E}\left[z_{t+s} \middle| \epsilon_{z,t}, \bar{\epsilon}_{z,t-1} \right]}{\partial \epsilon_{z,t}} \right|_{\epsilon_{z,t}=\epsilon_{z,*}} . \tag{5.42}$$

Note that we drop here the asterisk of the error term, which indicated a latent error before. The standard assumption on the conditioning information of the forecast is that past innovations are jointly zero, i.e. $\epsilon_{z,s} = 0$ for $s \leq t-1$. The shock imposed on the system $\epsilon_{z,*}$ is a deterministic value, usually chosen to be proportional to the standard deviation of the error term. Note also that (5.42) also includes the implicit assumption that the innovation terms between $t+1$ and $t+s$ are jointly zero. The definition (5.42) is based on marginal impulse responses under the tacit assumption that

$$\mathrm{E}\left[z_{t+s}\right|\epsilon_{z,t}=\epsilon_{z,*},\bar{\epsilon}_{z,t-1}\right]-\mathrm{E}\left[z_{t+s}\right|\epsilon_{z,t}=0,\bar{\epsilon}_{z,t-1}\right]=\Gamma_s^*\cdot\epsilon_{z,*} \qquad (5.43)$$

i.e. that the effect of the shock is linear and there is no loss of information incurred by reducing the analysis to a scrutiny of the marginals. If one is willing to accept these assumptions for the moment, then the calculation proceeds straightforward.

The key component to the calculation of the conditional expectation is the conditional probability to observe a price change of particular size at $t+s$, $\mathrm{Prob}\left[z_{t+s}=\mathrm{abs}\left(v_j\right)|\epsilon_{z,t},\bar{\epsilon}_{z,t-1}\right]$. The assumptions on the innovations other than the impulse under consideration yield a particularly simple form of the conditional expectation of the dependent latent variable. To ease notation denote $\epsilon_{z,s}=0$ for $s\leq t-1$ and $t+1\leq s\leq t+s$, by $\bar{\epsilon}_z=0$. From this follows directly that the conditional probability at the observation $t+s$ under given assumptions is equal to

$$\begin{aligned}
\mathrm{Prob}\left[z_{t+s}=v_{z,j}\left|\epsilon_{z,t},\bar{\epsilon}_z=0\right.\right] &= \Phi(\nu_{j,t+s})-\Phi(\nu_{j-1,t+s}),\\
\nu_{j,t+s} &= \mu_j - H_z\mathrm{E}\left[\xi_{t+s}\right|\epsilon_{z,t},\bar{\epsilon}_z=0\right]\\
&= \mu_j - H_z\sum_{i=0}^{\infty}\Psi_i L^i e_1\mathrm{E}\left[\epsilon_{z,t+s}\right|\epsilon_{z,t},\bar{\epsilon}_z=0\right]\\
&= \mu_j - H_z\Psi_s e_1\epsilon_{z,t},\\
\text{with } v_{z,j} &\in \{0,1,2,\ldots\},\\
\text{using } \xi_{t+s} &= \sum_{i=0}^{\infty}\Psi_i L^i e_1\epsilon_{z,t+s}.
\end{aligned}$$

Given the above conditional probabilities, the calculation proceeds straightforward by a repeated application of the chain rule

$$\Gamma_s = \sum_{j=1}^{J} v_{z,j} \frac{\partial \operatorname{Prob}\left[z_{t+s} = v_{z,j} \,|\epsilon_{z,t}, \bar{\epsilon}_z = 0\right]}{\partial \epsilon_{z,t}}$$

$$\frac{\partial \operatorname{Prob}\left[z_{t+s} = v_{z,j} \,|\epsilon_{z,t}, \bar{\epsilon}_z = 0\right]}{\partial \epsilon_{z,t}} = \frac{\partial \Phi(\nu_{j,t+s})}{\partial \epsilon_{z,t}} - \frac{\partial \Phi(\nu_{j-1,t+s})}{\partial \epsilon_{z,t}}$$

$$= \left(\phi(\nu_{j,t+s}) - \phi(\nu_{j-1,t+s})\right) \frac{\partial z^*_{t+s}}{\partial \epsilon_{z,t}} \tag{5.44}$$

$$\frac{\partial z^*_{t+s}}{\partial \epsilon_{z,t}} = H'_z \frac{\partial \xi_{z,t+s}}{\partial \epsilon_{z,t}} \tag{5.45}$$

$$\frac{\partial \xi_{z,t+s}}{\partial \epsilon_{z,t}} = \Psi_s = \begin{cases} e_1 & \text{if } s = 0 \\ \prod_{i=1}^{s} F e_1 & \text{if } s > 0 \end{cases} \tag{5.46}$$

This completes the derivation of the impulse response of the size of price changes z_t with respect to an innovation in the same process. Note that the nonlinearity is induced by the necessity to integrate over the normal density to obtain the probability to observe a certain category $v_{z,i}$ given the available information set. For the latent process the linearity assumption in (5.43) would hold, i.e. for the eq. (5.45) and following. This is clearly not the case for (5.44).

This derivation is easily extended to the case where an impulse on the duration process τ_t is examined by the impulse response function.

$$\Gamma^{**}_s := \frac{\partial \operatorname{E}\left[\, z_{t+s} |\, \epsilon_{\tau,t}, \bar{\epsilon}_z = 0\right]}{\partial \epsilon_{\tau,t}}. \tag{5.47}$$

The derivation employs the decomposition of the size error term $\epsilon_{z,t}$ into a function of the duration error term $\epsilon_{\tau,t}$ and an orthogonal component in (5.38).

The derivative (5.47) is obtained by replacing $\partial \epsilon_{z,t}$ by $\partial \epsilon_{\tau,t}$ and it remains to note that

$$\frac{\partial \xi_{t+s}}{\partial \epsilon_{z,t}} = \frac{\partial \xi_{t+s}}{\partial \epsilon_{z,t}} \frac{\partial \epsilon_{z,t}}{\partial \epsilon_{\tau,t}} \text{ and}$$

$$\frac{\partial \epsilon_{z,t}}{\partial \epsilon_{\tau,t}} = \rho.$$

This completes the derivation of the impulse function of the size of price changes with respect to innovation in the duration equation. The impulse response function for the time between transactions model is well documented in the literature and identical to the derivation in the case described above, once only the latent model is considered. Thus, it is not reviewed here explicitly. See e.g. Hamilton (1994). By these arguments it is also clear that it is not subject to the problems due to a nonlinear mapping from the latent to the observable variable, which render this type of impulse responses uninformative.

As has already been emphasized along the lines, the use of impulse response functions based on partial derivatives are misleading in the context of quantal response models as (5.43) does not hold. A more appropriate concept of impulse response functions will be described in the next subsection, which is explicitly based on the left hand side of (5.43) and avoids the approximation by the partial derivative.

5.2. The impulse response function in nonlinear models. Although the concept of linear impulse response analysis is quite straightforward, the concise treatment of the subject can be quite involved in the context of nonlinear models. For nonlinear models some advances to the methodology of solving such a problem are provided by Gallant, Rossi, and Tauchen (1993), Koop, Pesaran, and Potter (1996), and Gourieroux and Jasiak (1999). Intuitively, the question is how does the dynamic respond to a shock on the system? The dynamic system is typically characterized by the conditional expectation of the dependent variable, e.g. $\mathrm{E}\left[y_{z,t}\right| \epsilon_{t+s}, \epsilon_{t+s-1}, \ldots, \epsilon_t, \bar{\epsilon}_{t-1}]$, given the history of innovations ϵ_t driving the system. This history of innovations can be seen to contain three distinct parts. First, the history of the system $\bar{\epsilon}_{t-1}$ until it is hit by the shock ϵ_t which is to be assessed (second). Third, the innovations driving the system until it reaches the state which is to be examined in the context of the impulse response $\epsilon_{t+1}, \epsilon_{t+2}, \ldots, \epsilon_{t+s}$.

A generic formulation of impulse responses serves as a starting point, where the impulse response is defined as the difference between two forecasts based on the same information set but deviating in the impulse ϵ_* inflicted on the system

$$\Gamma_{G,s} := \mathrm{E}\left[y_{z,t+s} \middle| \epsilon_{t+s}, \epsilon_{t+s-1}, \ldots, \epsilon_t, \bar{\epsilon}_{t-1} \right] - \mathrm{E}\left[y_{z,t+s} \middle| \epsilon_{t+s}, \epsilon_{t+s-1}, \ldots, \epsilon_t = \epsilon_*, \bar{\epsilon}_{t-1} \right] \quad (5.48)$$

Koop, Pesaran, and Potter (1996) isolate four distinct problems which need to be addressed

- History dependence, i.e. the state of the system at the time the system is excited by the shock.
- Composition dependence and
- shock dependence, i.e. the nature of the shock.
- The treatment of the future, i.e. the way the system evolves after the system is hit by the shock.

History dependence is essentially a problem of nonlinear models. To solve this problem Gallant, Rossi, and Tauchen (1993) suggest to condition on the history just as outlined in (5.48). Koop, Pesaran, and Potter (1996) on the other hand suggest either integrating $\bar{\epsilon}_{t-1}$ out or selecting characteristic sets of histories which correspond to particular economic settings which are of particular interest. To cope with history dependence in the context of market microstructure is somewhat eased by the fact that there is no clear cut definition of the state of the system as there is e.g. in the analysis of real business cycles where it makes a significant difference in which state of the economy the system is hit by a shock. In the given context, no significant information is lost, if the history of the system is integrated out in the analysis of impulses.

The treatment of future shocks might be solved in an analogue fashion to history dependence. Again the assumption of zero shocks, i.e. $\epsilon_{t+1} = \epsilon_{t+2} = \ldots = \epsilon_{t+n} = 0$, is clearly a non trivial assumption as in nonlinear models the zero shock assumption clearly has a different meaning from the linear context. Thus, the concept which appears to fit the given model set best is to consider the various history and future sets, $\bar{\epsilon}_{t-1}$ and $\epsilon_{t+s}, \epsilon_{t+s-1}, \ldots, \epsilon_{t+1}$, random and integrate them out by taking an appropriate expectation of the impulse function:

$$\Gamma_s = \mathrm{E}_\epsilon \left[\Gamma_{G,s}\right] \tag{5.49}$$

$$= \mathrm{E}\left[y_{z,t+s}\right] - \mathrm{E}\left[y_{z,t+s}\right| \epsilon_t = \epsilon_*\right]. \tag{5.50}$$

The expectation $\mathrm{E}_\epsilon\left[\cdot\right]$ is taken with respect to the innovations $\bar{\epsilon}_{t+s}$ of the system except the impulse ϵ_* in (5.48).

This solves the problems raised by the dependence of the impulse response function with respect to the history of the system and the future innovations. It remains to solve the shock dependence, i.e. the influence of the size of the shock ϵ_* on the form of the impulse response. This is also a problem specific to nonlinear models as standard linear models have the property of linearity and symmetry in the shocks, i.e. a shock of the size 2 has double the impact of a shock of size 1 and a shock of size -1 just has the opposite effect compared to a shock of +1. Thus, particular care needs to be devoted to the selection of ϵ_t in order not to generate impulse responses not representative for the dynamic system under consideration. Following the work of Koop, Pesaran, and Potter (1996) the emphasis will not lie on the analysis of a single shock ϵ_* but on a representative number drawn from the distribution specified for ϵ_t in the model. This allows to complement the analysis of individual paths Γ_s, $s = 1, \ldots, S$, by an analysis of the distribution of the system responses at various horizons, since the distribution of $\Gamma_s(\epsilon_*)$ can be evaluated. This is quite informative as it allows to assess how representative a certain impulse response is.

Composition dependence is a well known problem in the interpretation of VAR models. This problem relates to the multivariate distribution of the error term vector ϵ_t. The interpretation of an impulse response would go astray if it was based e.g. on the shock $\epsilon_* = \begin{bmatrix} 1 & 0 \end{bmatrix}$ and the covariance of the errors was not diagonal. Yet, in the absence of a priori knowledge on the causality between the equations some care needs to be taken in the interpretation of the results. Usually, an orthogonalization of the corresponding equations eases the analysis as this renders the covariance matrix diagonal. In this context an orthogonalization is directly available via (5.38).

In the analysis of the implications of estimation results, the latter concept of impulse responses for nonlinear models based on the explicit differences of conditional expectations will be employed to describe the dynamics of the transaction process. Evidence can thus be provided on the strength and the extension of impulses generated by innovations to the system.

Part 3

The relationship between transaction price changes and liquidity

CHAPTER 6

Empirical models of the process of price changes and time between transactions

1. Bund future trading at the DTB

1.1. The data set. The sample contains transaction data of the DTB Bund future trading. The Bund future was at the sampling time one of the most actively traded future contracts in Europe. The Bund future is a notional 6% German government bond of DEM 250.000 face value which matures in 8.5 to 10.5 years at contract expiration. There are four contract maturities per year, March, June, September and December. Prices are denoted in basis points of face value. In the sampling interval one tick was equivalent to a contract value of DEM 25. The Bund future was traded at the LIFFE, London, as well. The main difference between the Bund future trading at either exchange is that the DTB relies on an electronic trading system, whereas the LIFFE builds on floor trading. A thorough comparison of both trading places and a closer description of the data set is found in Franke and Hess (1995).

This study uses various samples up to data on 9 contracts and 563 trading days between 12/8/94 and 3/6/97. Thus the data set extends from the March 1995 to the March 1997 contract. Table 6.1 gives an overview over standard samples employed here. Although transaction evidence is the core of this study, some descriptive statistics are given for aggregates of the transaction process. The daily and 10 minute aggregates are used to round up the picture and make clear the changes due to the disaggregation of observations on the price process.

TABLE 6.1. Sample specification, *daily* and *10min* contain the aggregates indicated respectively. *Short*, *long*, and *composed* contain transaction data. No overnight observations are included in the samples *10min*, *short*, *long*, and *composed*.

Name	contract	days to maturity		trading days	date		observations
		start	end		start	end	
Daily	03/95-03/97	90	0	563	12/8/94	3/6/97	563
10min	03/95-03/97	90	0	563	12/8/94	3/6/97	31 730
Short	03/95	44	40	5	1/23/95	1/27/95	12 490
Long	03/95	44	12	25	1/23/95	2/24/95	55 970
Composed	03/95	44	40	5	1/23/95	1/27/95	12 490
	06/95	45	41	5	4/24/95	4/28/95	10 833
	09/95	45	41	5	7/24/95	7/28/95	9 122
	12/95	45	41	5	10/23/95	10/27/95	11 488
	03/96	45	41	5	1/22/96	1/26/96	14 020
	06/96	44	40	5	4/22/96	4/26/96	11 578
	09/96	46	42	5	7/22/96	7/26/96	9 017
	12/96	46	42	5	10/21/96	10/25/96	13 728
	03/97	45	41	5	1/20/97	1/24/97	20 464

One striking result of the empirical studies presented in the following chapter 7, is that the empirical relationships revealed by the analysis of volatility and liquidity are quite stable over the sample and that a rather small sample of five trading days is sufficient to capture these effects (sample *short* in table 6.1). To demonstrate this, an extended sample of 25 trading days (sample *long* in table 6.1) and a composed sample of a total of 9 times five consecutive trading days between the 1/23/95 and the 1/24/97 are used (sample *composed* in table 6.1). As the peculiarities of thin trading are not the focus of this analysis, only the last 90 trading days before the maturity of a contract are examined.

Table 6.2 gives a description of the unconditional distribution of the variables under consideration. There are mainly five variables used for the following empirical analysis:

- price changes
- absolute price changes
- time between transactions
- volume per transaction
- time of day

Concerning the observations one has to keep in mind, that they stem from an electronic trading system. This implies certain advantages and disadvantages. The main advantage is that they are by far more reliable than data originating from floor trading, since the latter suffer from the fact that data is not directly recorded as in an electronic trading system but gathered by price reporters on the floor and collected in an electronic information system. Data errors due to mistrades are supposedly reduced to a small level as there are some checks implemented in the trading system, e.g. it is not possible to carry out transactions which deviate too far from the present price level. One of the drawbacks of this type of data is, that the volume associated with an individual transaction is not necessarily the order volume, but the part of the original order which was executed at the given price. Possibly one order might trigger several transactions at different prices, if there are sufficient corresponding orders in the system. This aspect will be quite important, when the empirical results are interpreted. The time of day is not evaluated explicitly in the table but only via its first differences as the time between transactions. However, it will be used to generate regressors, which capture the well-known intraday seasonalities. This is outlined in the next subsection.

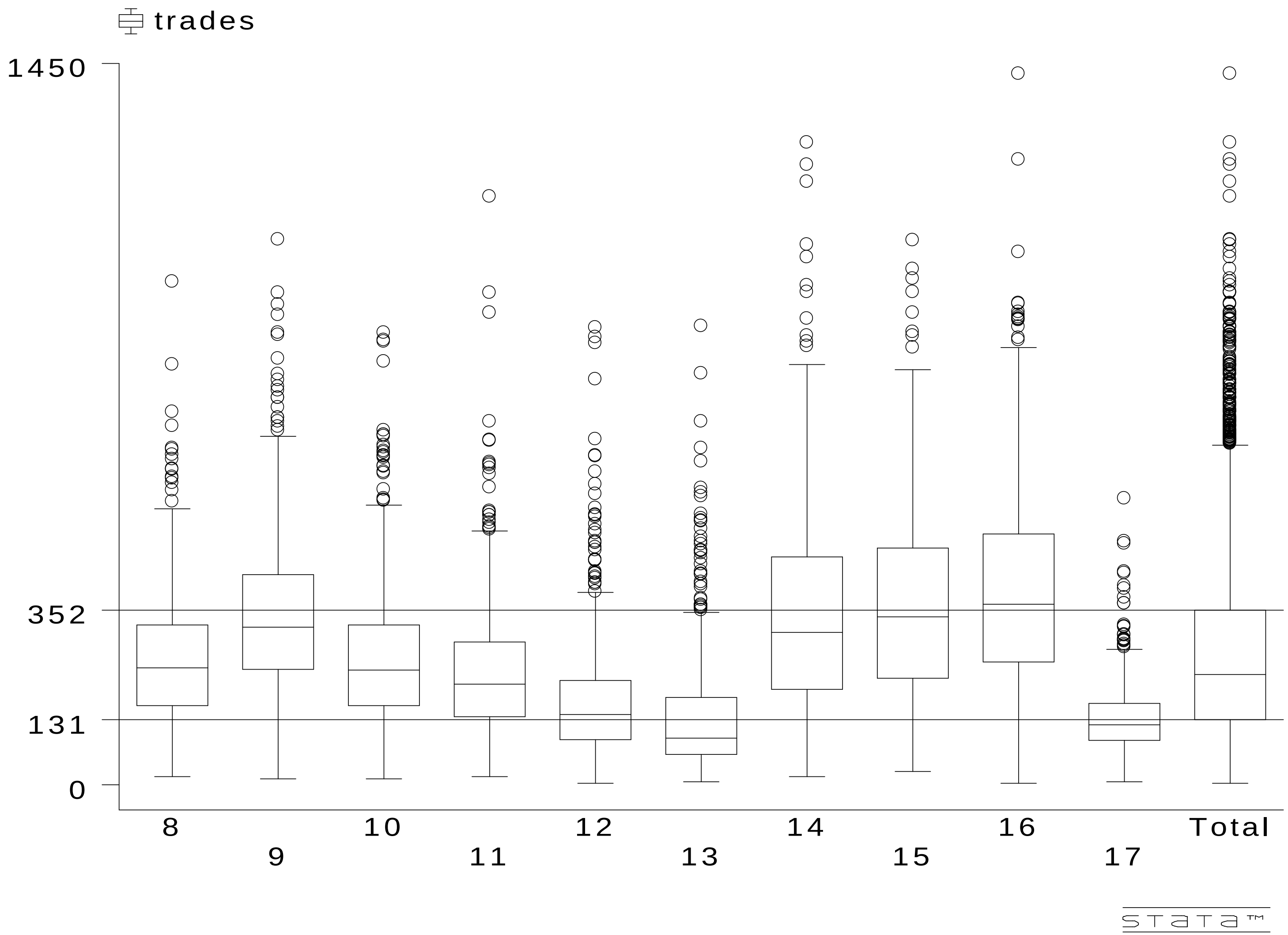
Price changes can be observed on various aggregation levels revealing a differing amount of information. 6.2 gives an impressive picture of the "increasing discreteness" of price changes when the aggregation level is decreased. By examining the percentiles, it becomes quite obvious, that the discreteness of price changes plays no role at the daily aggregation level. At the 10 minute aggregation level already 50% of the observations are made up by price changes of sizes plus and minus one and zero. At the transaction level more than 80% of the data have this characteristic. The exact value is 97% for the *long* sample, this cannot be determined from the table. Of the remaining 3% about two thirds are made up of price changes with an absolute value of two, the rest ranges between three and five. The number of transactions and the time between transactions give a clear picture of the liquidity of the contract under consideration. On average there are 14 seconds between successive transactions, amounting to an average of 46 transactions per 10 minute interval and 2600 observations per day. The average number of contracts exchanged per transaction is about 20 amounting to about 56 000 per day.

1.2. Intradaily seasonalities. The evidence on prominent intradaily seasonalities provided in the literature has already been addressed in chapter 2. In order to get an impression of the seasonalities prevailing in the BUND future trading, some descriptive evidence is presented in this section on the time between transactions or trade intensity and on the size of price changes which are both subject to some clear cut intradaily patterns. The examination of these patterns is on the one hand a value of its own, as it might give some preliminary evidence on the driving forces in the market and on the other hand it is important to correct for these seasonalities in the empirical models of the transaction process, as the coefficients of variables subject to these seasonalities might otherwise just reflect the latter.

TABLE 6.2. Descriptive statistics on the unconditional distribution of price changes, trade frequency, and volume on a daily and 10 minute aggregation level and descriptive statistics on disaggregated observations on the transaction level.

variable	sample	mean	variance	skewness	kurtosis	percentiles				
						10%	25%	50%	75%	90%
price changes	daily	3.5833	1069.6899	-0.5919	4.6649	-35	-16	4	23	39
	10min	0.5548	17.5507	-0.9088	42.3616	-4	-1	0	1	4
	long	-0.0230	0.4491	-0.0726	4.5377	-1	-1	0	1	1
number of transactions	daily	2598.5524	1197828.6	0.4485	3.5292	1335	1865	2525	3205	4035
	10min	46.6009	1531.0	2.3493	13.4166	10	19	36	60	93
time between transactions	long	14.5477	703.9567	7.2679	114.2439	1	1	5	15	34
volume	daily	56315.941	7.6656e+008	0.8231	4.3071	24875	36875	52375	70125	94125
	10min	999.664	1042212.9	3.1540	22.9742	140	340	700	1260	2140
	long	20.2925	665.6949	4.7484	50.1320	1	3	11	23	49

FIGURE 6.1. Box plot of the number of transactions per hour. Overall and for each full hour between 8.00 and 17.00. Boxes indicate the inter quartile range, the median is indicated by the line in the box.



The time between transactions regressor in the size of price changes regression might, e.g. just pick up the well known lunch time effect. To avoid this some corrections for intraday seasonalities are discussed at the end of this section. The box plots in figure 6.1 describe the distribution of the number of transactions per hour. They show a strong variation even if extreme observations as 1450 trades per hour are left apart. The interquartile range from the 25% percentile at 131 trades per hour to the 75% percentile at 352 implies a variation of 221 trades per hour. The typical pattern observed at the DTB is a trading peak in the morning between 9.00-9.59 with a steady decline just until 13.59. At two o'clock, more precisely at 14:30 (this cannot be determined from the figure), the trading intensity increases significantly and tops even the peak in the morning. This is a very stable pattern and goes back to the strong inflow of transactions originating in the U.S.

The distribution of price changes per hour is depicted in figure 6.2, i.e. all price changes observed in one trading hour are aggregated and the distribution over the entire sample is given for the indicated categories. The denomination is price change in basis points per hour.

The inter quartile range over the entire sample is the distance between -4 and 4 ticks per hour, which gives a pretty good impression about the standard speed of price changes in the market. In some particular fast trading phases price changes up to 75 or 50 basis points are observed. This is mainly due to adjustments of key interest rates and macro shocks which serve as indicators for upcoming changes in key interest rates. From this figure it is also evident that the inter quartile range of price changes, i.e. a proxy for volatility, is somewhat reduced around lunch, 12.00-13.59, and increased in the afternoon, 14.00-16.59.

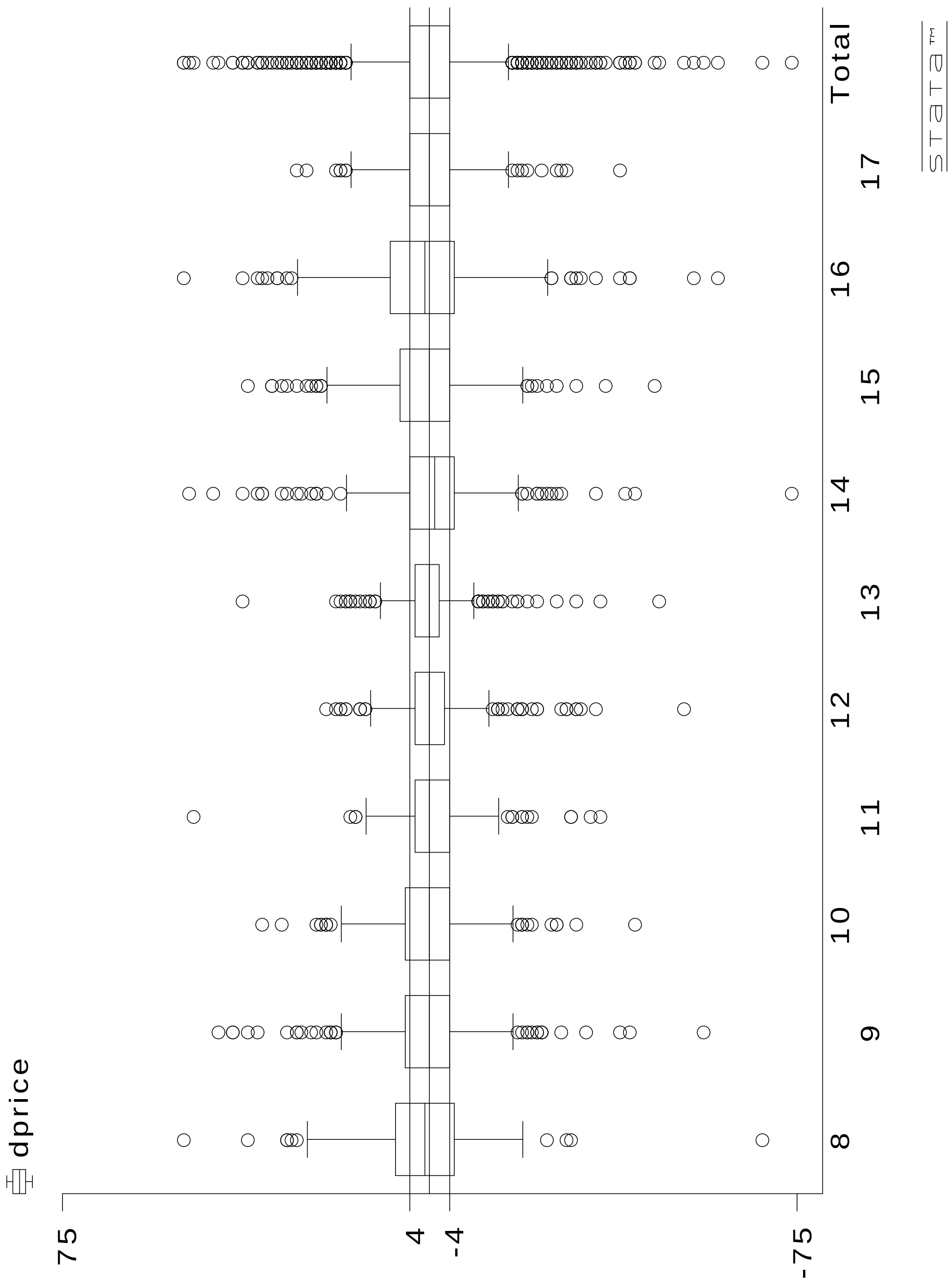


FIGURE 6.2. Box plot of aggregated price changes per hour measured in basis points. Overall and for each full hour between 8.00 and 17.00. Boxes indicate the inter quartile range, the median is indicated by the line in the box.

In order to get a more concise picture of the intensity of price changes, some descriptive evidence is collected using a method suggested by Gourieroux, Jasiak, and Fol (1999) in order to estimate

- the intensity of the trade process given the time of the day,
- the intensity of price changes on the same basis, and
- the intensity of two tick price changes on the same basis.

The figure 6.3 gives the probability that one of the aforementioned three events occurs given a certain time of the day. In their model of intraday market activity Gourieroux, Jasiak, and Fol (1999) concentrate on the trade intensity λ which is assumed to depend exclusively on the time of the day, but not on other regressors as past time between transactions or other exogenous variables. The trade intensity λ maps the probability that there is one trade observed at the instant ϑ_t, i.e. between ϑ_t and $\vartheta_t + dt$. Note that ϑ_t is defined as the time in seconds since the start of trading.

They consider first the cumulated arrival rate of transactions

$$\Lambda(\vartheta_t) = \int_0^{\vartheta_t} \lambda(u)\, du \tag{6.1}$$

which is readily estimated on the basis of a sample of D trading days, if $N(\vartheta_t)$ gives the number of trades observed on a certain day up to time ϑ_t. The cumulated arrival rate equals the conditional expectation

$$\Lambda(\vartheta_t) = \mathrm{E}\left[N(\vartheta_t)\right] \tag{6.2}$$

for which the appropriate sample estimator is obviously

$$\hat{\Lambda}(\vartheta_t) = \frac{1}{D}\sum_{j=1}^{D} N(\vartheta_t). \tag{6.3}$$

Gourieroux, Jasiak, and Fol (1999) point out that the consistent estimator for the arrival rate $\lambda(\vartheta_t)$ needs to employ an appropriate smoothing over t as $\hat{\Lambda}$ is not differentiable with respect to its argument t. Thus, they propose

$$\hat{\lambda}(\vartheta^*) = \frac{1}{D}\sum_{j=1}^{D}\sum_{t=1}^{T_j} \frac{1}{h} K\left(\frac{\vartheta_{j,t} - \vartheta^*}{h}\right) \tag{6.4}$$

where K is the kernel employed for smoothing, h the bandwidth parameter, and $\vartheta_{j,t}$ the time observation t and day j is observed. See e.g. Pagan and Ullah (1999) for a thorough exposition of non parametric estimation methods. The estimation depicted in figure 6.3 was carried out using a Gaussian kernel with bandwidth $h = 20\text{min}$.

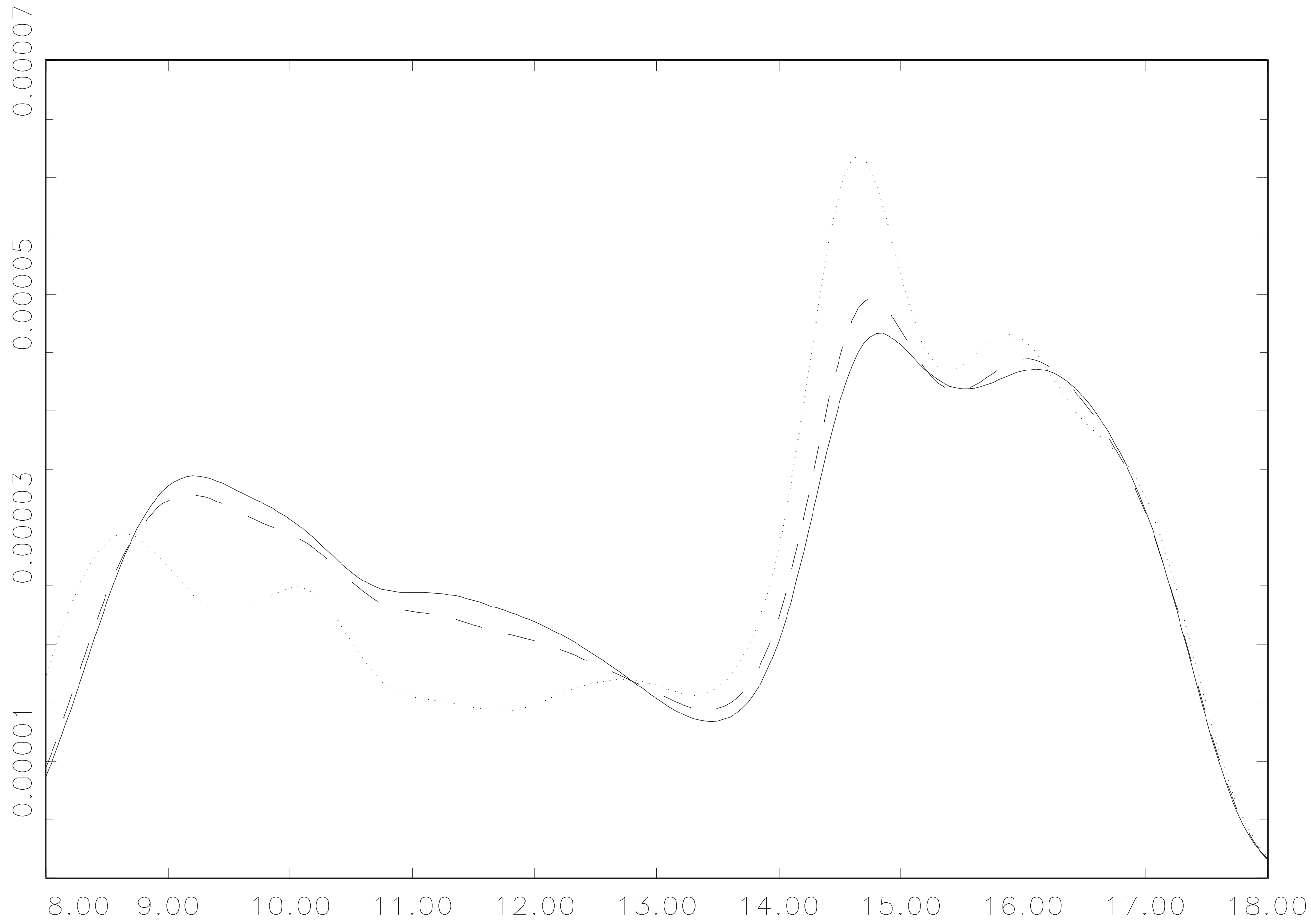
From the figure 6.3 it can be seen that there is a first peak of the trading intensity in the morning around 10.00 which is topped however by the strong increase in trading in the afternoon after 14.30, when U.S. traders are active in the market. Concerning the intensity of price changes larger than one tick, which have only a very small share in the sample, it can be resolved that they have a particularly high intensity around 14.30 and a particularly low intensity in the morning.

By the descriptive evidence provided above, it is clear that there are strong intraday seasonalities in both processes, in the process of time between transactions and in the process of the size of price changes. No evidence could be found that there are any intraday seasonalities in the sign of price changes, this however would not have been suspected in the first place.

There are different ways to correct for these seasonalities in the inference on the transaction process One solution is to generate seasonally adjusted series by partialling out the time-of-day effects. Engle and Russell (1997) regress the duration on a function of the time and obtain an estimator of the typical shape. By dividing the durations by their estimated typical shape they obtain the seasonally adjusted series. Another path to proceed is to include special regressors in the mean and variance function to capture seasonal effects.

These are usually semiparametric methods which allow for highly nonlinear deterministic functions of the time of day. This procedure is mainly advocated in Andersen and Bollerslev (1997) and further papers by Andersen and Bollerslev on intra-day heteroskedasticity models.

FIGURE 6.3. Transaction intensities over the trading day estimated according to Gourieroux, Jasiak, and LeFol (1999). Smoothing was done using a Gaussian kernel and a bandwidth of 20min. The solid line represents the trade intensity. The dashed line gives the intensity of trades with a price change. The dotted line gives the intensity of trades with price changes larger than one tick.



A third alternative is the use of structural time series models which are based on the Kalman filter and extract the seasonal components of the observed time series via a latent model, see e.g. the extended treatment in Harvey (1989). A particular application to the intra-day process of volatility was suggested by Bollerslev and Ghysels (1996) in the form of a time-varying coefficients model.

Here, a Fourier series approximation proposed by Andersen and Bollerslev (1998b) based on the work of Gallant (1981) is employed. This framework allows for a direct but parsimonious assessment of the volatility patterns based on different frequencies, like intraday seasonalities. Additional regressors of the form ϑ_t^*, $\cos(\vartheta_t^* \cdot 2\pi p)$, and $\sin(\vartheta_t^* \cdot 2\pi p)$ are included in the regression, where $\vartheta_t^* \in [0,1]$ and p is identical with the order of the term. The total effect of seasonalities with coefficients δ_t, $\delta_{c,p}$, and $\delta_{s,p}$ and order of approximation P is given by

$$s(\delta, t, P) = \delta_t \cdot \vartheta_t^* + \sum_{p=1}^{P} \left(\delta_{c,p} \cos(\vartheta_t^* \cdot 2\pi p) + \delta_{s,p} \sin(\vartheta_t^* \cdot 2\pi p) \right). \qquad (6.5)$$

The standardized position ϑ_t^* within the trading day is derived from the time ϑ_t, at which transaction t is carried out as

$$\vartheta_t^* = \frac{\vartheta_t}{\text{seconds between 8.30 and 17.15}}. \qquad (6.6)$$

Remember that ϑ_t is defined as the time in seconds since the trading start at 8.30.

This procedure has the advantage of being easy to implement, compared to a model of time-varying coefficients, since the set of regressors $s(\delta, t, P)$ can be directly included in the mean function of the size of transaction price changes equation and in the time between transactions duration. On the other hand by including the regressors directly into the specification, the two step procedure employed by Engle and Russell (1997) is avoided and thereby the results of the estimation can be directly interpreted including the seasonalities. Care has to be taken however when estimation results are to be interpreted as the trigonometric terms taken individually cannot be interpreted but have to be taken as an aggregate $s(\delta, t, P)$, which is a semiparametric approximation to an unknown function.

2. Empirical evidence on the sign of price changes

In this subsection determinants of the univariate process of the direction of price changes are scrutinized. Therefore exogenous explanatory variables for the process are examined first, in order to gain some descriptive evidence on the data. The dynamic model suggested in chapter 3 is then employed to analyse the process more closely, revealing its dynamic structure.

The empirical model employed in this subsection has the structure of an ordered probit, with the sign of price changes s_t as dependent variable. See the discussion on the decomposition of the price process for a motivation of the variable and for its relation to the process of price changes d_t. The observation equation maps the latent variable s_t^* into three observable categories:

$$s_t = \begin{cases} -1 & \text{, if} \quad s_t^* \leq \mu_{s,1} \\ 0 & \text{, if} \quad \mu_{s,1} < s_t^* \leq \mu_{s,2} \\ 1 & \text{, if} \quad \mu_{s,2} < s_t^* \end{cases} . \tag{6.7}$$

The full specification of the latent unobservable variable as it is developed in chapter 3 is given by

$$s_t^* = H_s m_{s,t}^{\xi} + w_{s,t}' \gamma_s + \epsilon_{s,t} \tag{6.8}$$

$$m_{s,t}^{\xi} = F_s(m_{s,t-1}^{\xi} + e_1 c_{s,t-1}) + e_1 x_{s,t}' \beta_s. \tag{6.9}$$

Here, $c_{s,t}^*$ denotes the generalised error of the limited dependent model of the sign process.

From an economic point of view it would certainly be an unexpected result if potential explanatory variables $w_{s,t}$ or $x_{s,t}$ for the size of price changes contained any information for the direction of price changes, at least considering the available information set on the transactions. Under the auspices of short sales constraints, see e.g. Diamond and Verrecchia (1987), this observation might find an economic interpretation. Short sales constraints are however not a very sensible assumption in the context of futures trading. Explanatory power of exogenous regressors might also be expected if trades could be classified into buyer and seller initiated transactions, the state of the market could be characterized more concisely, even including the sign of the transaction process. Given the available information set, findings in this spirit are very unlikely.

This a priori knowledge is clearly reflected by the regression results. Table 6.3 contains regression results on a model of the above type, without a dynamic, i.e. $H_s = 1$, $F_s = 0$, no regressors $x_{s,t}$ in the dynamic specification. These results need to be interpreted with care, as a dynamic is not accounted for. Nevertheless, the results derived by Poirier and Ruud (1988) on the robustness of Probit type models against dynamic misspecification clearly indicate that this evidence is informative.

Yet, it can be seen in table 6.3 that the explanatory variables volume and time between transactions are insignificant and the BIC clearly indicates that the regression without any regressors is favourable.

Before the dynamic model of the sign of price changes is specified, some descriptive evidence is gathered on the dynamics of the process. Therefor the empirical autocorrelation function (ACF) is analysed. The relationship between the observable and the latent ACF is discussed in section 3, so it is quite clear that the observed ACF and partial autocorrelation function (PACF) do not yield the same information as in the context of a fully observable model. These suffer also from the well known problems of statistics not fully adjusted to the count nature of the data, which are discussed in chapter 3. Thus, the observable ACF and PACF are only an approximation to the latent dynamic, which quality depends heavily on the thresholds $\mu_{s,i}$ and the unconditional moments of the latent variable of the model.

Yet, to gain a first impression of the serial dependency of observable price changes the ACF and the PACF are examined. See table 6.4 It can be seen that the transaction price changes show a dramatic serial dependency compared to daily aggregates. 10 min aggregates do not show such a high dependency but still deviate from the null hypothesis of no serial correlation. The descriptive evidence on the sign of price changes indicates, that either a MA(1) or a AR(1) dynamic might be appropriate, as the ACF and the PACF drop very fast to zero after the first lag.

TABLE 6.3. Ordered probit for the sign of price changes $s_t \in \{-1, 0, 1\}$ as described in (6.7), (6.8), and (6.9) with $H_s = 1$, $F_s = 0$ and no $x_{s,t}$. The regressors are log time between transactions, $\log \tau_t$, log volume, $\log v_t$, and trigonometric expansions of intraday seasonals $s(\delta, t, 4)$. Regression based on the sample *short*. The mean log likelihood (log. lik.) and the Bayes Information Criterion (BIC) are also provided for each regression. White standard errors of the coefficients are given in parentheses.

Specification	A		B		C		D	
Thresholds								
$\mu_{z,1}$	-0.9383	(0.0307)	-0.9381	(0.0333)	-0.9380	(0.0333)	-1.0423	(0.1073)
$\mu_{z,2}$	0.9005	(0.0303)	0.9012	(0.0329)	0.9012	(0.0329)	0.7982	(0.1067)
Static regressors $w_{s,t}$								
$\log \tau_{t-1}$					0.0098	(0.0336)	-0.0149	(0.0417)
$\log \tau_{t-2}$					0.0055	(0.0751)	-0.0132	(0.0422)
$\log \tau_{t-3}$					0.0179	(0.0517)	-0.0125	(0.0430)
$\log \tau_{t-4}$					-0.0087	(0.0429)	-0.0022	(0.0422)
$\log v_{t-1}$					0.0236	(0.0411)	-0.0581	(0.0437)
$\log v_{t-2}$					0.0061	(0.0403)	0.0044	(0.0451)
$\log v_{t-3}$					0.0058	(0.0375)	-0.0244	(0.0442)
$\log v_{t-4}$					-0.0066	(0.0357)	0.0087	(0.0453)
δ_t			0.0111	(0.0340)			0.0115	(0.0340)
$\delta_{s,1}$			0.0079	(0.0758)			0.0120	(0.0764)
$\delta_{s,2}$			0.0188	(0.0518)			0.0175	(0.0519)
$\delta_{s,3}$			-0.0059	(0.0444)			-0.0047	(0.0445)
$\delta_{s,4}$			0.0243	(0.0411)			0.0268	(0.0417)
$\delta_{c,1}$			0.0058	(0.0403)			0.0016	(0.0418)
$\delta_{c,2}$			0.0050	(0.0377)			0.0071	(0.0379)
$\delta_{c,3}$			-0.0042	(0.0367)			-0.0051	(0.0367)
$\delta_{c,4}$			-0.0106	(0.0362)			-0.0108	(0.0362)
log. lik.	-0.3909		-0.3908		-0.3908		-0.3906	
BIC	-4807		-4848		-4844		-4883	

A specification search on the basis of the direction of price changes s_t is depicted in table 6.5. It turns out that the MA(1) process for the latent dominates the other tested specifications based on the BIC. For the sake of brevity only specifications up to ARMA(3,3) are reported, as higher order specifications did not turn out to improve the estimation's information content.

The MA(1) process found in the data might be well attributed to the well-known bid-ask bounce, which is known to imply a dynamic structure as it is documented above. See e.g. the discussion in chapter 2. The observable ACF is straightforward to compute from a simple simulation exercise. The MA(1) parameter of 0.3531 in the ARMA(0,1) model in table 6.5 along with the corresponding thresholds implies an observable ACF of -0.2944 at the first lag. This finding is well in line with the descriptive evidence reported in table 6.4 where the ACF at the first lag shows a value of -0.2638.

TABLE 6.4. Estimates of the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the sign of price changes s_t of the *long* sample and of the price changes for the *daily* and *10min* sample. No corrections for day of the week, days to maturity, or overnight effects are employed. Effects of discreteness, irregular spacing of observations, and intraday seasonalities are not corrected. Column Q gives the Box-Ljung test up to the indicated lag. Corresponding p-values are also provided.

	daily aggregates				10min aggregates				transactions			
lag	ACF	PACF	Q	p-value	ACF	PACF	Q	p-value	ACF	PACF	Q	p-value
1	-0.0523	-0.0523	1.5489	0.2133	-0.0186	-0.0186	10.952	0.0009	-0.2638	-0.2639	4219.4	0.0000
2	-0.0075	-0.0104	1.5812	0.4536	0.0196	0.0193	23.152	0.0000	0.0065	-0.0678	4222	0.0000
3	0.0005	-0.0004	1.5813	0.6636	-0.0087	-0.0080	25.552	0.0000	0.0131	-0.0032	4232.4	0.0000
4	-0.0420	-0.0427	2.5847	0.6295	-0.0024	-0.0031	25.741	0.0000	0.0023	0.0060	4232.7	0.0000
5	0.0509	0.0472	4.0607	0.5407	0.0121	0.0124	30.403	0.0000	0.0073	0.0112	4235.9	0.0000
6	-0.0696	-0.0671	6.8256	0.3373	-0.0006	-0.0002	30.416	0.0000	0.0084	0.0146	4240.2	0.0000
7	-0.0085	-0.0146	6.8674	0.4428	-0.0075	-0.0080	32.185	0.0000	-0.0085	-0.0021	4244.5	0.0000
8	-0.0485	-0.0539	8.2167	0.4126	0.0002	0.0001	32.187	0.0001	0.0003	-0.0025	4244.5	0.0000
9	-0.0699	-0.0744	11.018	0.2745	0.0188	0.0192	43.456	0.0000	-0.0051	-0.0071	4246.1	0.0000
10	0.0509	0.0368	12.509	0.2524	0.0009	0.0013	43.481	0.0000	0.0016	-0.0020	4246.2	0.0000
11	-0.0546	-0.0502	14.229	0.2206	0.0004	-0.0003	43.487	0.0000	-0.0024	-0.0030	4246.6	0.0000
12	0.1152	0.1088	21.891	0.0388	0.0019	0.0024	43.605	0.0000	-0.0026	-0.0041	4247	0.0000
13	-0.0050	-0.0001	21.905	0.0568	-0.0087	-0.0085	45.985	0.0000	0.0001	-0.0017	4247	0.0000
14	0.0246	0.0321	22.256	0.0735	-0.0133	-0.0142	51.567	0.0000	-0.0005	-0.0009	4247	0.0000
15	0.0237	0.0086	22.582	0.0934	0.0158	0.0157	59.464	0.0000	-0.0064	-0.0071	4249.5	0.0000

TABLE 6.5. Specification search using an ordered probit for the sign of price changes $s_t \in \{-1, 0, 1\}$ as described in (6.7), (6.8), and (6.9) and no explanatory variables $w_{s,t}$ and $x_{s,t}$. Regression based on the sample *short*. The mean log likelihood (log. lik.) and the Bayes Information Criterion (BIC) are also provided for each regression. White standard errors of the coefficients are given in parentheses.

	ARMA(1,0)		ARMA(0,1)		ARMA(1,1)		ARMA(2,2)		ARMA(3,3)	
					Thresholds					
$\mu_{s,1}$	-0.9692	(0.0283)	-1.0285	(0.0313)	-1.0276	(0.0312)	-1.0303	(0.0319)	-1.0302	(0.0318)
$\mu_{s,2}$	0.9299	(0.0278)	0.9879	(0.0306)	0.9870	(0.0305)	0.9891	(0.0312)	0.9894	(0.0312)
					Parameters of the dynamic					
AR(1)	-0.2879	(0.0245)			0.0227	(0.0857)	0.4162	(0.4814)	-0.0522	(0.9067)
AR(2)							-0.1500	(0.0823)	-0.2018	(0.6558)
AR(3)									-0.1120	(0.1529)
MA(1)			0.3531	(0.0368)	0.3730	(0.0855)	0.7672	(0.4813)	0.2981	(0.9076)
MA(2)							-0.2733	(0.1827)	-0.1648	(0.8858)
MA(3)									-0.0684	(0.3379)
log. lik.	-0.37974		-0.37971		-0.37970		-0.37944		-0.37943	
BIC	-4674		-4674		-4678		-4685		-4694	

3. Empirical evidence on the size of price changes

3.1. Static effects. After having found that the sign of the price changes is mainly dominated by an MA(1) process which is well explained by the bid-ask bounce. The search of the determinants of the size of price changes promises to yield more interesting results from an economic perspective, since this process is closely related to the risk of price changes.

The empirical model employed in this subsection is very similar to the model of the sign of price changes, the difference stems mainly from the values observed for the dependent variable. Just like the model for the sign of price changes, the model of the size of price changes z_t has the structure of an ordered probit. This variable models the size of the observed price change, see again the discussion on the decomposition of the price process into sign and size in section 2 of chapter 5. The observation equation maps the latent variable z_t^* again into three observable categories:

$$z_t = \begin{cases} 0 & \text{, if} \quad z_t^* \leq \mu_{z,1} \\ 1 & \text{, if} \quad \mu_{z,1} < z_t^* \leq \mu_{z,2} \\ 2 & \text{, if} \quad \mu_{z,2} < z_t^* \end{cases} . \tag{6.10}$$

Price changes of an absolute value larger than two are recoded as two. This is necessary for estimation purposes, as the thresholds of the larger categories would be rendered virtually unidentified. Note that less than 0.5% of the data set are affected by this recoding and that the largest observed price change is five ticks. The approximation error incurred by the recoding of the dependent variable is discussed in chapter 2 for the analysis of price changes, yet the arguments given there translate in a straightforward fashion to the present analysis. See also the discussion of the recoding effect and possible remedies in Pohlmeier and Gerhard (2000).

The full specification of the latent unobservable size of price changes z_t^* as it is developed in chapter 3 is given by

$$z_t^* = H_z m_{z,t}^{\xi} + w_{z,t}' \gamma_z + \epsilon_{z,t} \tag{6.11}$$

$$m_{z,t}^{\xi} = F_z(m_{z,t-1}^{\xi} + e_1 c_{z,t-1}) + e_1 x_{z,t}' \beta_z. \tag{6.12}$$

The conditional probabilities implied by the above model and the distributional assumptions given in chapter 3, could not only serve as conditional probabilities to model the conditional probability distribution of price changes d_t as outlined in chapter 5, section 4.1, but also have a value of their own concerning the risk of the price process, as it is amply discussed by Granger and Ding (1995). The latter aspect will dominate the discussion in this section. The determinants of the risk of transaction price changes in particular its relation to the trade intensity are in the focus of this analysis.

In table 6.6 regression results are reported in a static model context, i.e. no ARMA terms of the latent variable are included. Again some careful interpretation is in place, as the presence of an unaccounted dynamic might invalidate the results. The BIC indicates that the model without regressors including only thresholds would be preferred. This is identical to the model which analyses the sign of price changes s_t. In this context however, the first four lags of the time between transactions and the first lag of volume have a significant impact on the dependent variable. This was clearly not the case for the sign of price changes. Both volume and time between transactions enter the equation with a negative sign.

As a consequence of these negative signs an increased time between transactions in the past, i.e. a rather slow trading, will yield lower probabilities to observe large price changes or any price changes at all. Since the regression reported in column D accounts for intraday seasonalities the estimated coefficients for volume and time between transactions cannot be suspected to capture the type of seasonality effects due to the lunch break and the opening of other exchanges. Therefore the conclusion holds, under the mentioned limitations of this regression, that in periods of thin trading no sudden price changes are to be expected or put differently, the last price is still a valuable approximation of the value of the future. This is well in line with theoretical models like Easley and O'Hara (1992) who show that an increased (decreased) information flow comes along with a higher (lower) trade frequency and a increased (decreased) probability to observe large price changes. The underlying

TABLE 6.6. Ordered probit for the size of price changes $z_t \in \{0, 1, 2\}$ as described in (6.10), (6.11), and (6.12) with $H_s = 1$, $F_s = 0$ and no $x_{s,t}$. The regressors are log time between transactions, $\log \tau_t$, log volume, $\log v_t$, and trigonometric expansions of intraday seasonals $s(\delta, t, 4)$. Regression based on the sample *short*. The mean log likelihood (log. lik.) and the Bayes Information Criterion (BIC) are also provided for each regression. White standard errors of the coefficients are given in parentheses.

Specification	A		B		C		D	
Thresholds								
$\mu_{z,1}$	0.3639	(0.0119)	0.3819	(0.0130)	0.0215	(0.0626)	0.0399	(0.0476)
$\mu_{z,2}$	1.9208	(0.0234)	1.9422	(0.0239)	1.5903	(0.0662)	1.6106	(0.0519)
Static regressors $w_{s,t}$								
τ_{t-1}					-0.0381	(0.0205)	-0.0316	(0.0206)
τ_{t-2}					-0.0641	(0.0192)	-0.0583	(0.0196)
τ_{t-3}					-0.0486	(0.0196)	-0.0432	(0.0194)
τ_{t-4}					-0.0484	(0.0199)	-0.0430	(0.0194)
v_{t-1}					-0.1488	(0.0206)	-0.1505	(0.0214)
v_{t-2}					-0.0348	(0.0266)	-0.0365	(0.0213)
v_{t-3}					-0.0307	(0.0227)	-0.0319	(0.0216)
v_{t-4}					0.0146	(0.0477)	0.0133	(0.0200)
δ_t			0.0112	(0.0067)			0.0120	(0.0949)
$\delta_{s,1}$			-0.0399	(0.0583)			-0.0226	(0.1773)
$\delta_{s,2}$			0.0474	(0.0515)			0.0425	(0.1246)
$\delta_{s,3}$			-0.0007	(0.0152)			0.0011	(0.0167)
$\delta_{s,4}$			-0.0098	(0.0171)			0.0023	(0.1184)
$\delta_{c,1}$			0.0117	(0.0196)			-0.0083	(0.0357)
$\delta_{c,2}$			-0.0380	(0.0195)			-0.0297	(0.0191)
$\delta_{c,3}$			-0.0044	(0.0231)			-0.0081	(0.0218)
$\delta_{c,4}$			0.0057	(0.0123)			0.0056	(0.0292)
log. lik.	-0.3252		-0.3247		-0.3234		-0.3231	
BIC	-4000		-4037		-4016		-4055	

reasoning is based on the assumption that some traders only initiate transactions if they receive an information indicating that the value of the asset has changed.

The negative sign of past volume might be attributed to the electronic trading system of the DTB. Large transaction volumes can only be observed if there are buy or sell offers of a significant size in the system for one price otherwise a large transaction is split into several smaller ones which are carried out at increasing or decreasing prices to fulfil the total volume. The consequences of this forced order split are also in accordance with the findings for time between transactions since a small time between transactions, which is a consequence of the forced order split, leads to an increased probability of price changes.

Although the individual coefficients of the trigonometric expansion are insignificant, due to the limited estimation sample, the intra-day seasonal pattern is still examined briefly, as quite some descriptive evidence is presented in the preceding sections that intraday seasonals are present in the Bund future trading. Yet, it should be kept in mind that these regressors have no significant explanatory power.

To gain some insight into the seasonal pattern of the size of price changes, the influence of intraday seasonalities on the latent variable $s(\delta, t, 4)$ are depicted in the following. In the upper picture of figure 6.4, and in the lower picture probabilities are provided to observe price changes of a certain size. The latter gives the conditional probability given the time of day, to observe a price change larger than one tick, i.e. Prob $[z_t = 2\,|\vartheta_t]$, equal to one tick, i.e. Prob $[z_t = 1\,|\vartheta_t]$, and the probability that the price does not change at all, i.e. Prob $[z_t = 0\,|\vartheta_t]$, from bottom to top. From the second image it can be resolved that the intraday seasonalities play no major role for the discrimination between 2 tick price changes and smaller ones but is of significant importance for the probability to observe a price change at all. These reflect the well known picture already found in the context of the descriptive statistics of the preceding subsection.

FIGURE 6.4. Influence of intradaily seasonalities. The top graph give the effects of trigonometric terms on the mean of the latent variable. The bottom graph depict the probability to observe a price change of zero ticks (top line), of one tick (middle line), and two ticks or more (bottom line).

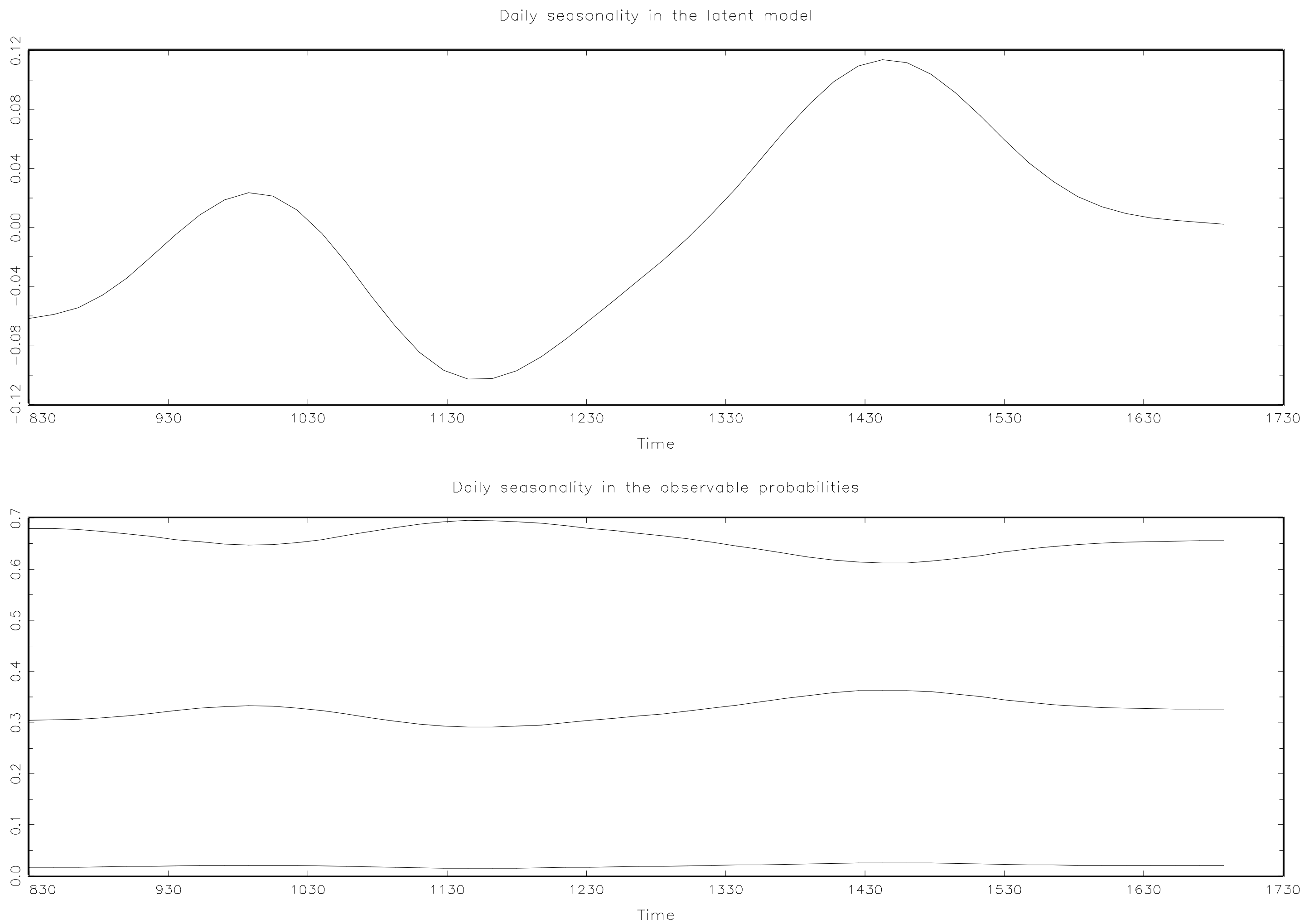


FIGURE 6.5. Iso-probability plot of trade frequency versus volume. The computed probabilities are based on the regression D in table 6.6. Different volume/time between transactions combinations are plotted which yield the same probability to observe no price change. Probabilities 0.75, 0.70, 0.65, and 0.60 are provided. All four lags of the regressors are set to the indicated values. The circle indicates a volume/duration combination of 30/10, the triangle 30/20, and the square 10/60. The corresponding loci are marked in figure 6.6 by a vertical line. The seasonals are fixed at 20% of the trading day, i.e. around 10.15.

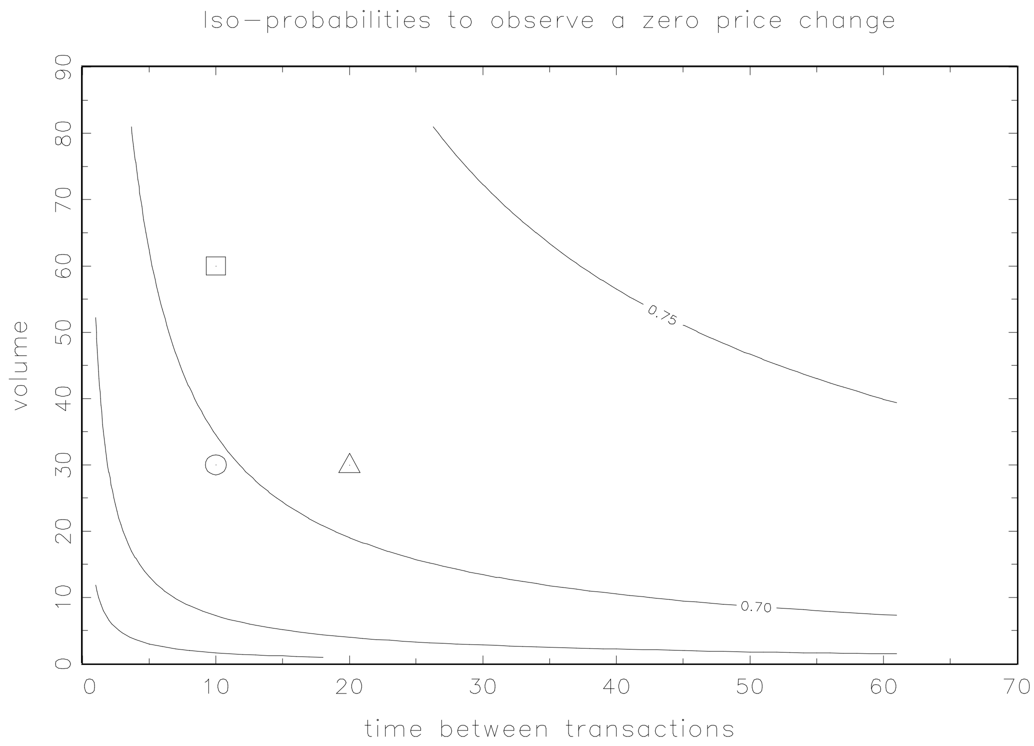
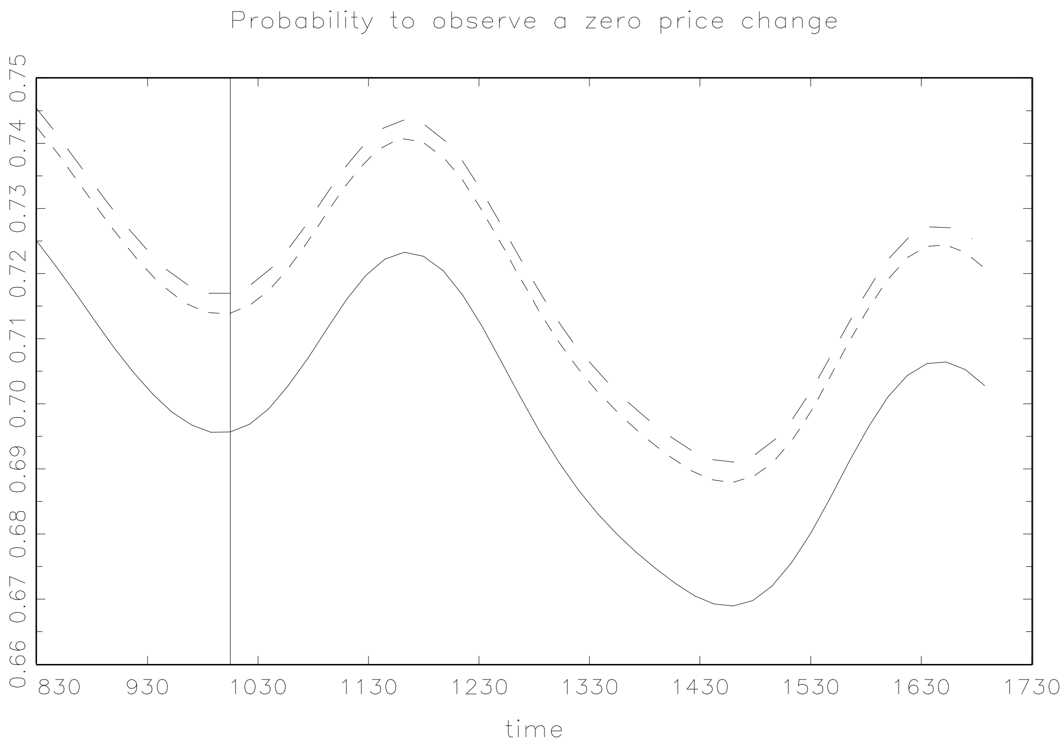


FIGURE 6.6. Plot of the probability to observe no price change per transaction versus the time of the day. The computed probabilities are based on the regression D in table 6.6. The solid line indicates a volume/time between transactions combination of 30/10, the dashed line 30/20, the small dashes 10/60. All four lags of the regressors are set to the indicated values. The point in time used as a basis for figure 6.5 is marked by a vertical line. The three points indicated in figure 6.5 are at the intersection of this line with the respective graphs.



Although the volume-trade frequency relationship is not the focus of this analysis some interesting conclusions can be drawn from the regression results in table 6.6. If the volume throughput of a market over a certain time span is under consideration, there are two ways to increase market turnover, either by an increased volume per transaction or by an increased trade frequency with a constant volume. The striking result is that a low trading frequency observed in the past as well as high past volumes lead to a higher probability to observe a zero price change compared to one or two tick price changes. The figure 6.6 gives the probability to observe no price change for different values of traded volume per transaction and time between transactions over the course of the trading day. To clarify the relationship between past trade frequency and past traded volume with respect to the probability to observe a certain price change, figure 6.5 gives an iso-probability plot of volume versus time between transactions. The figure 6.6 gives the probability to observe no price change for different values of traded volume per transaction and time between transactions over the course of the trading day.

There are several interpretations available for this result. On the one hand this striking result fits well in with the hypothesis of Admati and Pfleiderer and Foster and Viswanathan that traders who have some discretion with respect to the timing of their trades will trade when they expect to trade at low costs. Costs for a potential trader in this sense are not only the bid-ask spread considered by the authors but also the probability that the prices change adversely if the intended order is submitted. Considering that parts of the order book are visible to traders, the empirical results on the relationship of volume and trade frequency are quite reasonable. On the other hand some attention should be devoted to the fact that lags of volume and time between transactions are included in this regression. One might be tempted to claim that the market is just dried out by past trading volume or trading intensity. This however cannot be true, as a transaction was indeed carried out, remarkably with a lower probability of a price change. Nevertheless, it would be quite interesting to examine the contemporaneous effects. To include the contemporaneous volume and time between transactions is however not appropriate because of a possible simultaneity bias which might affect the coefficients. The next subsection will analyse the simultaneity relationship of the size of price changes and time between transactions more closely.

3.2. The dynamics of absolute price changes. In order to assess the serial dependency of the size of price changes z_t, table 6.7 gives the descriptive statistics on the serial dependency of the size of price changes. It indicates that there seems to be some longer range dependence in the size of price changes. Yet, this descriptive instrument needs to be interpreted again with some care, as has been already pointed out in the context of the sign of price changes. The results are such that even at a daily aggregation level some serial dependency is found which increases however dramatically, when intraday aggregates are examined and even more so for transaction data.

From the estimation results in table 6.8 it can be seen that the descriptive statistics presented in table 6.7 for the transaction price changes were indeed misleading. This is well in line with the findings in chapter 3, that the observable and the latent ACF are quite close if there is enough variation in the dependent observable variable. If however the relative position of the thresholds is very skewed, as it is the case for the size of price changes, the standard ACF gives only a very poor picture of the true dynamic of the latent variable, this is also reflected by the long memory like appearance of the ACF of the transaction data, which is not found at all in the 10 minute aggregates.

As a matter of fact the specification search based on the BIC yields an AR(1) process as the favourable model, if seasonality terms are included. For the sake of brevity results without the seasonality components are omitted, yet those would have shown that an ARMA model of higher order would be appropriate. This is another reason, why the descriptive evidence in the preceding table is to be handled with care.

TABLE 6.7. Estimates of the autocorrelation function (ACF) and partial autocorrelation function (PACF) for the size of transaction price changes for the *long* sample and for the size of price changes for the *daily* and *10min* sample. No corrections for day of the week, days to maturity, or overnight effects are employed. Effects of discreteness, irregular spacing of observations, and intraday seasonalities are not corrected. Column Q gives the Box-Ljung test up to the indicated lag. Corresponding p-values are also provided.

	daily aggregates				10min aggregates				transactions			
lag	ACF	PACF	Q	p-value	ACF	PACF	Q	p-value	ACF	PACF	Q	p-value
1	0.0453	0.0453	1.1592	0.2816	0.2270	0.2270	1634.3	0.0000	0.7163	0.7163	31100	0.0000
2	0.0499	0.0484	2.5699	0.2767	0.1700	0.1250	2551.6	0.0000	0.6844	0.3519	59492	0.0000
3	0.0620	0.0582	4.7507	0.1910	0.1735	0.1193	3506.8	0.0000	0.6720	0.2395	86862	0.0000
4	0.0588	0.0520	6.7191	0.1515	0.1614	0.0927	4333.6	0.0000	0.6602	0.1706	1.1e+05	0.0000
5	0.0775	0.0687	10.144	0.0713	0.1442	0.0680	4993.8	0.0000	0.6520	0.1337	1.4e+05	0.0000
6	0.0951	0.0839	15.314	0.0180	0.1316	0.0529	5543.4	0.0000	0.6432	0.1034	1.6e+05	0.0000
7	0.0513	0.0340	16.821	0.0186	0.1208	0.0417	6006.5	0.0000	0.6309	0.0714	1.9e+05	0.0000
8	0.0752	0.0565	20.061	0.0101	0.1226	0.0459	6483.2	0.0000	0.6242	0.0649	2.1e+05	0.0000
9	0.0351	0.0123	20.77	0.0137	0.1157	0.0376	6908.2	0.0000	0.6100	0.0355	2.3e+05	0.0000
10	0.0582	0.0363	22.717	0.0118	0.1201	0.0440	7366.0	0.0000	0.6110	0.0588	2.6e+05	0.0000
11	0.0655	0.0412	25.19	0.0085	0.0998	0.0192	7682.1	0.0000	0.5971	0.0247	2.8e+05	0.0000
12	0.0125	-0.0152	25.281	0.0135	0.1077	0.0334	8050.4	0.0000	0.5893	0.0245	3.0e+05	0.0000
13	0.0155	-0.0101	25.421	0.0203	0.0886	0.0103	8299.6	0.0000	0.5823	0.0227	3.2e+05	0.0000
14	0.1030	0.0840	31.572	0.0046	0.0845	0.0123	8526.1	0.0000	0.5713	0.0094	3.4e+05	0.0000
15	0.0618	0.0405	33.788	0.0036	0.0973	0.0299	8826.5	0.0000	0.5635	0.0105	3.6e+05	0.0000

As lags of explanatory variables are found to be significant for the size of price changes in a non-dynamic model, it seems worthwhile to examine more parsimonious model structures which allow for a longer dependence while the number of parameters involved is held constant or even lowered. A feasible solution involving an infinite lag structure is outlined in chapter 3, i.e. the inclusion of regressors as $x_{z,t}$ in (6.12). The regressions in table 6.9 extend the results from 6.6 in column A as an ARMA error term is considered. The other columns in table 6.9 examine infinite lag structures, where the regressors for volume and the time between transactions are included in the dynamics.

A closer inspection of the lag structure yields however the result that the infinite lag structure is not really needed. The median lag is 1 for both volume and the time between transactions in all of the specifications B-E. The calculation of the median lag in the given model as well as description of the infinite lag structure are given in section 2.3.3 of chapter 3. Apparently, the inclusion of the first lag suffices to capture the information contained in the history of durations and volume. This can also be seen from the long run effects on the latent variable which is -0.1296 for the durations and -0.2163 for the volume. Those values are both very close to the values of the parameters themselves. Note that the lag length of the time between transactions variable is clearly reduced, once the dynamics in the size of price changes are accounted for.

TABLE 6.8. Specification search using an ordered probit for the size of price changes $z_t \in \{0, 1, 2\}$ as described in (6.10), (6.11), and (6.12) and no explanatory variables $w_{s,t}$. The regressors are trigonometric expansions of intraday seasonals $s(\delta, t, 4)$. Regression based on the sample *short*. The mean log likelihood (log. lik.) and the Bayes Information Criterion (BIC) are also provided for each regression. White standard errors of the coefficients are given in parentheses.

	ARMA(0,0)		ARMA(1,0)		ARMA(0,1)		ARMA(1,1)		ARMA(2,2)		ARMA(3,3)	
						Thresholds						
$\mu_{z,1}$	0.3818	(0.0301)	0.3846	(0.0351)	0.3846	(0.0338)	0.3839	(0.0363)	0.3836	(0.0426)	0.3828	(0.0422)
$\mu_{z,2}$	1.9422	(0.0574)	1.9799	(0.0636)	1.9759	(0.0622)	1.9811	(0.0648)	1.9825	(0.0690)	1.9824	(0.0688)
						Parameters of the dynamic						
AR(1)			0.2099	(0.0287)			0.4056	(0.1270)	1.1446	(0.2634)	0.6590	(2.2539)
AR(2)									-0.1969	(0.2142)	0.4147	(2.4148)
AR(3)											-0.1504	(0.3853)
MA(1)					-0.1960	(0.0295)	0.2088	(0.1374)	0.9462	(0.2671)	0.4606	(2.2542)
MA(2)									-0.0338	(0.1939)	0.4842	(1.9662)
MA(3)											-0.0714	(0.2024)
						Static regressors $w_{s,t}$						
δ_t	0.0112	(0.0355)	0.0102	(0.0422)	0.0101	(0.0405)	0.0104	(0.0437)	0.0133	(0.0494)	0.0123	(0.0490)
$\delta_{s,1}$	-0.0399	(0.0804)	-0.0423	(0.0953)	-0.0427	(0.0916)	-0.0420	(0.0987)	-0.0369	(0.1128)	-0.0393	(0.1117)
$\delta_{s,2}$	0.0474	(0.0558)	0.0472	(0.0657)	0.0474	(0.0632)	0.0467	(0.0680)	0.0505	(0.0786)	0.0494	(0.0779)
$\delta_{s,3}$	-0.0009	(0.0482)	-0.0025	(0.0566)	-0.0026	(0.0545)	-0.0020	(0.0586)	-0.0016	(0.0681)	-0.0024	(0.0675)
$\delta_{s,4}$	-0.0097	(0.0447)	-0.0116	(0.0526)	-0.0114	(0.0506)	-0.0129	(0.0545)	-0.0083	(0.0633)	-0.0105	(0.0627)
$\delta_{c,1}$	0.0117	(0.0437)	0.0099	(0.0512)	0.0102	(0.0493)	0.0102	(0.0530)	0.0101	(0.0623)	0.0092	(0.0617)
$\delta_{c,2}$	-0.0381	(0.0409)	-0.0375	(0.0481)	-0.0375	(0.0463)	-0.0377	(0.0498)	-0.0354	(0.0585)	-0.0362	(0.0579)
$\delta_{c,3}$	-0.0044	(0.0400)	-0.0067	(0.0471)	-0.0064	(0.0453)	-0.0070	(0.0488)	-0.0052	(0.0570)	-0.0060	(0.0565)
$\delta_{c,4}$	0.0057	(0.0393)	0.0048	(0.0462)	0.0050	(0.0444)	0.0041	(0.0478)	0.0043	(0.0558)	0.0053	(0.0553)
log. lik.	-0.3247		-0.3201		-0.3204		-0.3200		-0.3197		-0.3196	
BIC	-4037		-3985		-3989		-3989		-3994		-4002	

TABLE 6.9. Explanatory variables for the size of price changes. Ordered probit for the size of price changes $z_t \in \{0, 1, 2\}$ as described in (6.10), (6.11), and (6.12). The regressors are log time between transactions, $\log \tau_t$, log volume, $\log v_t$, and trigonometric expansions of intraday seasonals $s(\delta, t, 4)$. Regression based on the sample *short*. The mean log likelihood (log. lik.) and the Bayes Information Criterion (BIC) are also provided for each regression. White standard errors of the coefficients are given in parentheses.

	A		B		C		D		E	
					Thresholds					
μ_1	-0.0245	(0.1319)	0.2921	(0.0563)	0.1589	(0.0751)	0.0552	(0.0892)	0.0661	(0.0852)
μ_2	1.5858	(0.1385)	1.8934	(0.0735)	1.7621	(0.0901)	1.6636	(0.0987)	1.6736	(0.0956)
					Parameters of the dynamic					
AR(1)	0.4544	(0.0262)	0.3802	(0.1247)	0.4043	(0.1237)	0.3821	(0.1211)	0.2256	(0.0288)
MA(1)			0.1716	(0.1344)	0.2018	(0.1340)	0.1678	(0.1308)		
					Static regressors $w_{z,t}$					
τ_{t-1}	-0.0704	(0.0456)								
τ_{t-2}	-0.0703	(0.0454)								
τ_{t-3}	-0.0436	(0.0456)								
τ_{t-4}	-0.0379	(0.0456)								
v_{t-1}	-0.1697	(0.0479)								
v_{t-2}	-0.0454	(0.0489)								
v_{t-3}	-0.0351	(0.0487)								
v_{t-4}	0.0142	(0.0487)								
δ_t	0.0125	(0.0443)	0.0096	(0.0438)	0.0126	(0.0439)	0.0116	(0.0440)	0.0111	(0.0427)
$\delta_{s,1}$	-0.0171	(0.1008)	-0.0259	(0.0991)	-0.0464	(0.0992)	-0.0305	(0.0997)	-0.0315	(0.0968)
$\delta_{s,2}$	0.0413	(0.0694)	0.0433	(0.0682)	0.0468	(0.0685)	0.0430	(0.0687)	0.0434	(0.0668)
$\delta_{s,3}$	0.0009	(0.0599)	-0.0040	(0.0587)	0.0032	(0.0591)	0.0013	(0.0592)	0.0005	(0.0575)
$\delta_{s,4}$	0.0041	(0.0560)	-0.0010	(0.0548)	-0.0151	(0.0549)	-0.0042	(0.0552)	-0.0042	(0.0536)
$\delta_{c,1}$	-0.0166	(0.0561)	-0.0104	(0.0540)	0.0193	(0.0536)	-0.0018	(0.0546)	-0.0020	(0.0530)
$\delta_{c,2}$	-0.0254	(0.0511)	-0.0302	(0.0501)	-0.0396	(0.0501)	-0.0315	(0.0505)	-0.0312	(0.0490)
$\delta_{c,3}$	-0.0123	(0.0499)	-0.0094	(0.0490)	-0.0065	(0.0491)	-0.0095	(0.0493)	-0.0095	(0.0479)
$\delta_{c,4}$	0.0052	(0.0489)	0.0051	(0.0480)	0.0041	(0.0482)	0.0045	(0.0483)	0.0046	(0.0469)
					Dynamic regressors $x_{z,t}$					
τ_{t-1}			-0.0922	(0.0428)			-0.0971	(0.0428)	-0.1004	(0.0428)
v_{t-1}					-0.1643	(0.0472)	-0.1677	(0.0472)	-0.1675	(0.0473)
log. lik.	-0.3181		-0.3195		-0.3188		-0.3184		-0.3185	
BIC	-3998		-3987		-3979		-3978		-3975	

4. Empirical evidence on the time between transactions

In the analysis of the process of the size of price changes it turns out that the time between transactions has a significant negative impact on the probability that a large price change occurs. Now the question is considered which factors determine the process of trade intensity τ_t, which might be interpreted as one observable dimension of liquidity in this market. This component of the transaction process is of particular interest, because it allows to link two risk factors. First, the size of price changes, as a building block for the volatility of the process and second the liquidity of the market, which indicates in this context, whether a transaction occurs, given the state of the market. Two alternative hypotheses can be thought of. On the one hand, high price volatility could go along with a high liquidity of the market, i.e. phases of intense trading are characterized by large price changes per transaction. On the other hand, phases of slow trading might be characterized by larger price changes per transaction. The latter would imply a considerable risk for investors, as the last price observation might only serve as a very crude approximation of the value of an asset, and if an investor would actually want to buy or sell the asset given a large price change in the past, it might not be possible.

This view of liquidity as the mere occurrence of trades is somewhat limited, as the discussion in the literature, e.g. Kyle (1985), shows. Most of the variables discussed in economics under the heading of liquidity are however not observable, as they concern the attitudes of the market participants. Some variables, which might give an approximation to these attitudes, like the order book and its dynamic behaviour are not observable in the particular context. Thus liquidity is taken in this framework as the occurrence of trades, or more concisely, the time between transactions.

TABLE 6.10. Estimates of the autocorrelation function (ACF) and partial autocorrelation function (PACF) for the number of transactions per interval for the *daily* and *10min* sample and for the time between transactions based on the *long* sample. No corrections for day of the week, days to maturity, or overnight effects are employed. Effects of intraday seasonalities are not corrected. Column Q gives the Box-Ljung test up to the indicated lag. Corresponding p-values are also provided.

	daily aggregates				10min aggregates				transactions			
lag	ACF	PACF	Q	p-value	ACF	PACF	Q	p-value	ACF	PACF	Q	p-value
1	0.5792	0.5836	189.88	0.0000	0.6434	0.6435	13138	0.0000	0.2152	0.2152	2808.3	0.0000
2	0.4114	0.1162	285.84	0.0000	0.4972	0.1421	20984	0.0000	0.2067	0.1682	5397.9	0.0000
3	0.3277	0.0774	346.84	0.0000	0.4194	0.0933	26567	0.0000	0.1999	0.1367	7820.6	0.0000
4	0.2944	0.0806	396.16	0.0000	0.3530	0.0386	30521	0.0000	0.1903	0.1100	10016	0.0000
5	0.3559	0.1903	468.37	0.0000	0.2998	0.0238	33374	0.0000	0.1919	0.1014	12248	0.0000
6	0.2832	-0.0350	514.17	0.0000	0.2646	0.0293	35596	0.0000	0.1873	0.0872	14375	0.0000
7	0.2315	0.0007	544.84	0.0000	0.2373	0.0246	37383	0.0000	0.1687	0.0594	16100	0.0000
8	0.2277	0.0609	574.55	0.0000	0.2200	0.0297	38919	0.0000	0.1690	0.0585	17832	0.0000
9	0.2244	0.0462	603.46	0.0000	0.2258	0.0601	40538	0.0000	0.1622	0.0489	19427	0.0000
10	0.2453	0.0537	638.08	0.0000	0.1917	-0.0195	41705	0.0000	0.1643	0.0508	21063	0.0000
11	0.1644	-0.0853	653.65	0.0000	0.1766	0.0149	42694	0.0000	0.1504	0.0332	22435	0.0000
12	0.1716	0.0654	670.65	0.0000	0.1765	0.0320	43683	0.0000	0.1567	0.0422	23923	0.0000
13	0.2251	0.1140	699.96	0.0000	0.1544	-0.0091	44440	0.0000	0.1457	0.0289	25211	0.0000
14	0.2360	0.0503	732.24	0.0000	0.1412	0.0073	45074	0.0000	0.1439	0.0283	26466	0.0000
15	0.2377	0.0208	765.02	0.0000	0.1377	0.0174	45675	0.0000	0.1376	0.0221	27614	0.0000

The joint process of the components of price changes and liquidity is treated more in depth in the next section. Beforehand, the properties of the univariate process of time between transactions τ_t are analysed in this section. For the sake of brevity, static regression results are omitted and the analysis concentrates right away on the dynamic structure of time between transactions.

For daily and 10min aggregates the dynamic structure of the number of transactions per interval is described in table 6.10. For the transaction data the time between transactions is analysed. Thus, some care needs to be applied to the interpretation of the results if the dynamics are compared. It can be observed that the serial dependency increases, going from daily aggregates to 10 minute aggregates. Another fact concerning the transaction data is the higher persistency in the process, compared e.g. to the sign of price changes. A comparison to the ACF and PACF of the size of price changes is not feasible, as it was noted before, that those are too heavily distorted by the discrete nature of price changes.

Descriptive results on the dynamics of volume per time and per transaction are given in table 6.11, for completeness and because volume is often used as a proxy for trade frequency when the latter is not available. See e.g. the early work of Clark (1973), but also many others. Here, the drop from the 10 minute aggregates to the results on the transaction data is even more severe. Intuitively, this reflects the fact that most of the serial dependency in the aggregates of volume is due to the dynamic properties of the duration process and not so much due to serial dependencies in the volume per transaction.

TABLE 6.11. Estimates (PACF) for the volume traded per interval for the *daily* and *10min* sample of the autocorrelation function (ACF) and partial autocorrelation function (PACF) for the volume per transaction based on the *long* sample the volume traded per interval for the *daily* and *10min* sample and for the volume per transaction based on the *long* sample. No corrections for day of the week, days to maturity, or overnight effects are employed. Effects of intraday seasonalities are not corrected. Column Q gives the Box-Ljung test up to the indicated lag. Corresponding p-values are also provided.

	daily aggregates				10min aggregates				transactions			
lag	ACF	PACF	Q	p-value	ACF	PACF	Q	p-value	ACF	PACF	Q	p-value
1	0.5679	0.5711	182.52	0.0000	0.5972	0.5972	11318	0.0000	0.0535	0.0535	173.64	0.0000
2	0.4223	0.1489	283.65	0.0000	0.4548	0.1527	17883	0.0000	0.0320	0.0292	235.67	0.0000
3	0.3560	0.1014	355.64	0.0000	0.3881	0.1062	22662	0.0000	0.0277	0.0246	282.17	0.0000
4	0.3417	0.1127	422.08	0.0000	0.3237	0.0405	25987	0.0000	0.0248	0.0212	319.37	0.0000
5	0.4198	0.2308	522.56	0.0000	0.2768	0.0312	28418	0.0000	0.0179	0.0141	338.88	0.0000
6	0.3334	-0.0276	586.03	0.0000	0.2451	0.0302	30326	0.0000	0.0169	0.0134	356.24	0.0000
7	0.2843	0.0108	632.26	0.0000	0.2189	0.0235	31846	0.0000	0.0161	0.0125	371.88	0.0000
8	0.2936	0.0870	681.66	0.0000	0.2081	0.0372	33221	0.0000	0.0185	0.0150	392.69	0.0000
9	0.2830	0.0392	727.63	0.0000	0.2127	0.0537	34657	0.0000	0.0121	0.0082	401.5	0.0000
10	0.2951	0.0402	777.74	0.0000	0.1788	-0.0129	35672	0.0000	0.0135	0.0100	412.47	0.0000
11	0.2076	-0.0809	802.58	0.0000	0.1659	0.0153	36545	0.0000	0.0120	0.0085	421.27	0.0000
12	0.2134	0.0580	828.86	0.0000	0.1653	0.0278	37413	0.0000	0.0171	0.0137	439.04	0.0000
13	0.2634	0.1061	868.98	0.0000	0.1470	0.0003	38099	0.0000	0.0145	0.0106	451.84	0.0000
14	0.2601	0.0380	908.19	0.0000	0.1376	0.0113	38700	0.0000	0.0229	0.0190	483.67	0.0000
15	0.2878	0.0858	956.26	0.0000	0.1370	0.0207	39296	0.0000	0.0167	0.0119	500.63	0.0000

The empirical model employed for the time between transaction process is developed in chapter 4 and takes the form of a log ACD model in state space form

$$\log \tau_t = H_\tau \xi_{\tau,t} + w'_{\tau,t}\gamma_\tau \tag{6.13}$$

$$\xi_{\tau,t} = F_\tau \xi_{\tau,t-1} + e_1 x'_{\tau,t}\beta_\tau + e_1 \epsilon_{\tau,t}. \tag{6.14}$$

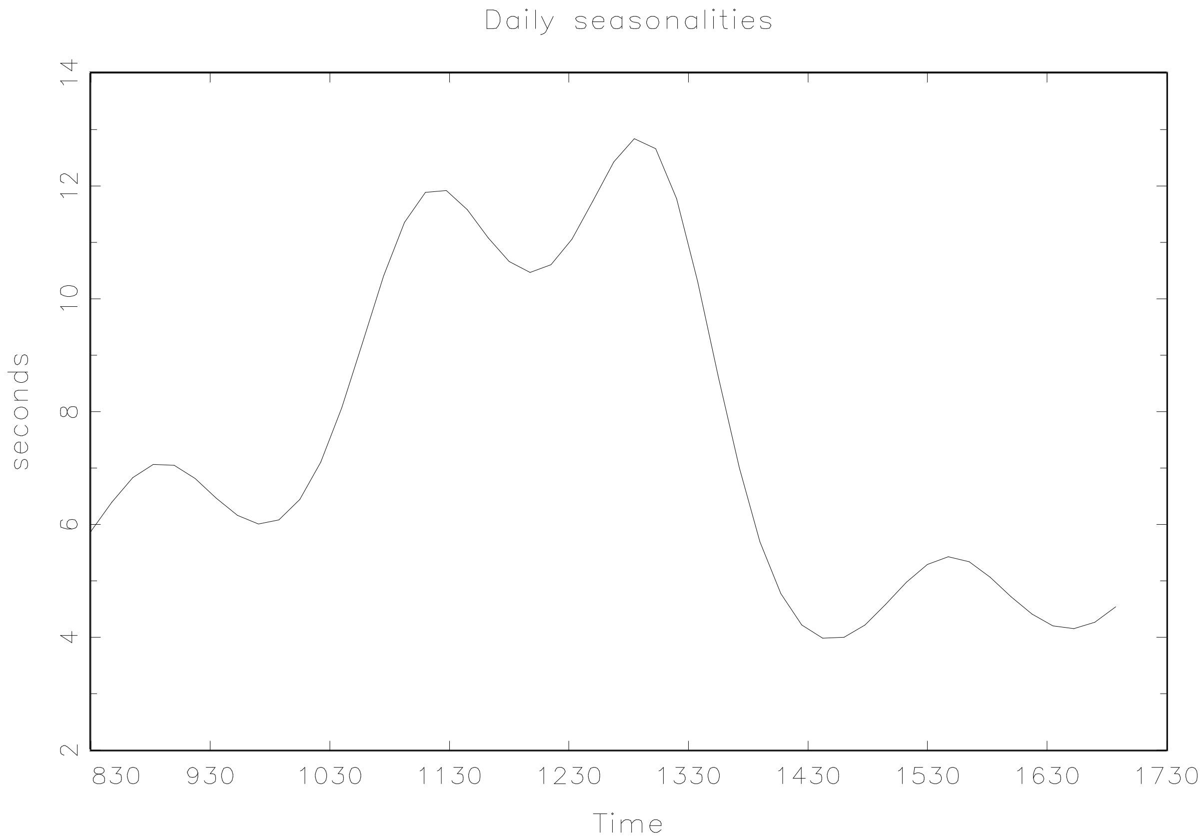
The results of a specification search for the process of time between transactions favors an ARMA(1,1) based on the BIC. The stronger serial dependency already noted in the descriptive statistics shows also in the considerably improved fit of the ARMA(1,1) or the ARMA(2,2) specification compared to the pure AR(1) or MA(1) process. See table 6.12. It is also quite noteworthy that some terms capturing intraday seasonalities are indeed significant. This was not found for the components of the process of price changes.

The intraday seasonalities implied by the Fourier series expansions $s(\delta, t, 4)$ are depicted in figure 6.7. The difference between the largest expected duration around 12:30 and the smallest is about 9 seconds, i.e. the expected duration is decreased by this amount all other influences fixed moving from 12.30 to 14.30. This is well in line with the descriptive evidence provided in the beginning of this chapter on intraday seasonalities. There is a phase in the morning around 10.00 with a rather intensive trading which is again topped by the active trading after 14.30 when the U.S. market participants are active. Expected time between transactions drops from 6 seconds down to about 4 seconds, increasing thereby the number of transactions per minute by one third.

TABLE 6.12. Specification search using a log ACD model of time between transactions as described in (6.13) and (6.14) and no explanatory variables $x_{\tau,t}$. The regressors are trigonometric expansions of intraday seasonals $s(\delta, t, 4)$. Regression based on the sample *short*. The mean log likelihood (log. lik.) and the Bayes Information Criterion (BIC) are also provided for each regression. White standard errors of the coefficients are given in parentheses.

	ARMA(1,0)		ARMA(0,1)		ARMA(1,1)		ARMA(2,2)		ARMA(3,3)	
	Parameters of the dynamic									
AR(1)	0.3088	(0.0231)			0.9985	(0.0020)	1.2728	(0.0486)	0.8417	(0.1939)
AR(2)							-0.2735	(0.0486)	0.4837	(0.3221)
AR(3)									-0.3261	(0.1393)
MA(1)			-0.2399	(0.0070)	0.9084	(0.0121)	1.1243	(0.0473)	0.6951	(0.1930)
MA(2)							-0.1814	(0.0444)	0.5022	(0.2912)
MA(3)									-0.2597	(0.1139)
	Static regressors $w_{\tau,t}$									
δ_t	-0.0202	(0.0285)	-0.0190	(0.0100)	-0.0153	(0.0552)	-0.0202	(0.0173)	-0.0218	(0.0674)
$\delta_{s,1}$	0.1019	(0.0642)	0.1095	(0.0221)	-0.2291	(0.2100)	-0.2337	(0.0965)	-0.2325	(0.1840)
$\delta_{s,2}$	-0.0424	(0.0430)	-0.0416	(0.0154)	-0.0289	(0.1572)	-0.0287	(0.0520)	-0.0247	(0.2772)
$\delta_{s,3}$	-0.0179	(0.0367)	-0.0204	(0.0136)	0.0902	(0.1447)	0.0669	(0.0620)	0.0603	(0.0637)
$\delta_{s,4}$	0.0631	(0.0350)	0.0680	(0.0125)	0.0968	(0.1097)	0.0963	(0.0499)	0.0948	(0.0483)
$\delta_{c,1}$	-0.1050	(0.0323)	-0.1135	(0.0118)	0.0613	(0.1635)	0.0380	(0.0858)	0.0272	(0.0762)
$\delta_{c,2}$	0.2294	(0.0227)	0.2337	(0.0088)	0.0404	(0.0292)	0.0414	(0.0095)	0.0421	(0.0096)
$\delta_{c,3}$	0.2003	(0.0209)	0.2084	(0.0078)	-0.0344	(0.0292)	-0.0318	(0.0096)	-0.0314	(0.0105)
$\delta_{c,4}$	0.1694	(0.0224)	0.1727	(0.0088)	-0.0225	(0.0295)	-0.0201	(0.0093)	-0.0196	(0.0094)
log. lik.	-0.1878		-0.1925		-0.1595		-0.1588		-0.1587	
BIC	-2342		-2399		-2000		-2001		-2008	

FIGURE 6.7. Influence of intradaily seasonalities on the expected time between transactions. The estimated model is specification ARMA(1,1) from table 6.12. The regressor $s(\delta, t, 4)$ is included in the static regressors $w_{\tau,t}$. To obtain the plot, H_τ is set to zero and the regressors varied over the trading day.



An analysis of the effects of past volume is included in table 6.13. Specification B is favored by the BIC criterion, which includes an ARMA(1,1) dynamic and volume with an infinite lag structure, as opposed to explicit lags of volume, as in specification A or an AR(1) specification. The result that higher trading volumes per transaction imply a lower time between transactions is quite stable across all specifications. A comparison between the use of explicit lags versus the use of a distributed lags model favors the ARMA(1,1) model with the distributed lag structure. A closer examination of the lag structure yields quite different results compared to the regression of the size of price changes on volume. The median lag is 11, which implies that volume ranging back for about 2 minutes makes up about half of the information content of this regressor, if the mean duration of about 14 seconds is used. Again, see section 2.3.3 in chapter 3 for a discussion of the lag structure employed to include volume here.[1] This of course is only a rough approximation but gives a good impression of the influence of volume on the durations compared to the influence of traded volume on price changes. The long run effect of volume on time between transactions is about -0.1796 which is much smaller than the value of the coefficient itself.

The sign and the lag structure of past volume is quite understandable if one accepts volume as a proxy for trading activity, due to an information inflow into the market. If there is need for an increased throughput of the market, both determinants of this size adjust, volume per transaction increases, and time between transactions decreases, or put differently, illiquid market phases do not go along with high volumes per transaction. This does not contradict the argumentation used to explain the determinants of the size of price changes, where it was argued that small volumes raise the probability for larger price changes, because one order triggers several transactions at different prices, because this argument would relate to the contemporaneous volume. Here however, only past volumes are considered, in order to avoid the inclusion of a potentially simultaneous regressor.

[1] Note that this section refers actually to the dynamic ordered probit, yet the specification of both models goes along the same line, once the discreteness of the dependent variable is accounted for.

TABLE 6.13. Log ACD model for the time between transactions as described in (6.13) and (6.14). The regressors are log volume, $\log v_t$, and trigonometric expansions of intraday seasonals $s(\delta, t, 4)$. Regression based on the sample *short*. The mean log likelihood (log. lik.) and the Bayes Information Criterion (BIC) are also provided for each regression. White standard errors of the coefficients are given in parentheses.

	A		B		C	
	Parameters of the dynamic					
AR(1)	0.9778	(0.0024)	0.9755	(0.0021)	0.0751	(0.0034)
AR(2)						
MA(1)	0.9384	(0.0035)	0.9339	(0.0032)		
MA(2)						
	Static regressors $w_{\tau,t}$					
$\log v_{t-1}$	-0.066	(0.0097)				
$\log v_{t-2}$	-0.053	(0.0094)				
$\log v_{t-3}$	-0.026	(0.0115)				
$\log v_{t-4}$	-0.012	(0.0107)				
δ_c	0.9097	(0.0235)	0.9219	(0.0121)	0.8453	(0.0131)
δ_t	-0.0100	(0.0165)	-0.0093	(0.0130)	-0.0121	(0.0066)
$\delta_{s,1}$	0.1008	(0.0408)	0.1060	(0.0283)	0.1126	(0.0155)
$\delta_{s,2}$	-0.0237	(0.0281)	-0.0238	(0.0255)	-0.0357	(0.0121)
$\delta_{s,3}$	-0.0298	(0.0191)	-0.0169	(0.0232)	-0.0155	(0.0092)
$\delta_{s,4}$	0.0851	(0.0171)	0.0809	(0.0170)	0.0793	(0.0097)
$\delta_{s,1}$	-0.1654	(0.0182)	-0.1535	(0.0184)	-0.1588	(0.0101)
$\delta_{s,2}$	0.0612	(0.0163)	0.0688	(0.0161)	0.0620	(0.0094)
$\delta_{s,3}$	-0.0203	(0.0152)	-0.0222	(0.0201)	-0.0228	(0.0092)
$\delta_{s,4}$	-0.0050	(0.0281)	-0.0024	(0.0193)	0.0057	(0.0080)
	Dynamic regressors $x_{\tau,t}$					
$\log v_{t-1}$			-0.0666	(0.0045)	-0.0860	(0.0098)
log. lik.	-0.4274		-0.4276		-0.4311	
BIC	-5321		-5310		-5347	

5. The joint transaction process

After having resolved the properties of the individual processes making up the transaction process considered in this work, i.e. the process of time between transactions τ_t, and the two components of the process of price changes d_t, namely the sign of price changes s_t and the size of price changes z_t, in this section the joint transaction process is to be analysed in the framework developed in section 4.1 of chapter 5. There, it is shown that the transaction process can be decomposed into two joint processes $y_{s,t}$ and $y_{z,t}$ under the assumption that the sign and the size of transaction price changes are independent of each other apart from the occurrence of a price change. Concisely, under this assumption the observation of a price change implies trivially that the sign of the price change is either plus one or minus one, but the observation of the exact size of the price change, e.g. equal to one tick or equal to two ticks, does not influence the probability to observe a price change upwards or downwards. These two processes $y_{s,t}$ and $y_{z,t}$ are in the focus of this section.

Several questions of economic relevance are tied to the analysis of these processes. The first one, was already raised in the context of the liquidity analysis, which is limited in this work to the analysis of time between transactions. It boils down to the question whether the risk of the asset under consideration due to the variability of the market price is linked to the liquidity of the asset. Several distinct cases can be thought of. First, liquidity and volatility are independent processes, governed by different determinants and thus need to be considered as two distinct risk factors. Second, liquidity and volatility might jointly depend on certain determinants, where a high liquidity, i.e. low time between transactions, implies a high volatility per transaction. This would more or less characterize the ideal relationship since investors have indeed the opportunity to adapt their positions to a changed market price. The third possibility is quite similar to the second one, in the respect that liquidity and volatility still depend jointly on certain determinants, yet a low liquidity implies a high volatility per transaction. This might be caused by the fact that the price formation is dependent on the observation of past trades and thus the adjustment of the market price to the changing value of the asset is occurring quite erratic. The latter might however raise substantial problems for an investor, as the last price which is observable might not be a reliable approximation of the underlying value of the asset and it might not be possible for him to change his portfolio position in the desired way after substantial price changes. These different hypotheses are linked to the analysis of the process of the size of price changes and the time between transactions $y_{z,t}$.

The second question relates to the relationship of time between transactions and the sign of price changes. It has been argued before in this work that the time between transactions should have no explanatory power for the process of the direction of price changes, due to the absence of short sale constraints and similar market inefficiencies. Yet, the question remains, whether this is indeed the case, or whether there might be other effects present not taken into consideration so far, which cause a simultaneity of the process $y_{s,t}$.

The third question of interest relates to the dynamic implications the joint model has for the individual components of the process. It is shown that the process of time between transactions has a serial dependency which is quite different from the pattern observed for the size of price changes. The question under consideration is, how an innovation in the liquidity equation, i.e. the time between transactions, influences the future trade frequency and the contemporaneous and future size of price changes. More intuitively, how long does a shock on the time between transactions prevail in the market, especially in the volatility and the liquidity component. The answer to this question will involve the impulse response function implied by the estimations of the joint process $y_{z,t}$.

The first model under consideration is the joint process of the sign of price changes and the time between transactions $y_{s,t}$. Therefore, the specification suggested in chapter 5 section 4.2 for the size of price changes z_t is slightly modified and applied to the analysis of the sign of price changes s_t. The modification concerns only the nature of the observable values, i.e. the mapping of the latent unobservable to the observable discrete variable, which was already used in this chapter, see (6.7). The latent sign of price changes is modelled by

$$s_t^* = H_s m_{s,t}^{\xi} + w_{s,t}' \gamma_s + \epsilon_{s,t} \tag{6.15}$$

$$m_{s,t}^{\xi} = F_s(m_{s,t-1}^{\xi} + e_1 c_{s,t-1}) + e_1 x_{s,t}' \beta_s. \tag{6.16}$$

The log ACD model is given by the usual specification as

$$\log \tau_t = H_\tau \xi_{\tau,t} + w_{\tau,t}' \gamma_\tau + \kappa_\tau \mathrm{E}\left[s_t^* | \mathcal{F}_{t-1}\right] \tag{6.17}$$

$$\xi_{\tau,t} = F_\tau \xi_{\tau,t-1} + e_1 (x_{\tau,t}' \beta_\tau + \epsilon_{\tau,t}). \tag{6.18}$$

As already described in chapter 5, the assumption of joint normality is imposed on the error terms $\epsilon_t = \begin{bmatrix} \epsilon_{s,t} & \epsilon_{\tau,t} \end{bmatrix}$:

$$\epsilon_t \sim N(0, \Sigma_s) \tag{6.19}$$

$$\text{with } \Sigma_s = \begin{bmatrix} 1 & \rho_s \\ \rho_s & 1 \end{bmatrix} \tag{6.20}$$

Although simultaneity of price changes and time between transactions is highly questionable from an economic point of view as the trading of futures is considered in this work, it will be examined for the sake of completeness in this section. The rather questionable implications of a non-zero correlation between the error terms can be explained on the basis of the linear impulse response function Γ_s^* in conjunction with a positive correlation of the error terms ($\rho > 0$). From the conditional expectation of the latent dependent variable and the time between transactions given the corresponding contemporaneous observation the consequences for the impulse response function can be evaluated at $t = 0$ assuming $\rho > 0$. From 5.46 in chapter 5 we have that $\frac{\partial s_{t+s}^*}{\partial \epsilon_{\tau,t}} > 0$, i.e. the latent variable is shifted upwards. The sign of the impulse response Γ_s^* depends on the sign of the factors $(\phi(\nu_{j-1,t}) - \phi(\nu_{j,t}))$ for the inner categories, $j = 2, \dots, J-1$. The lowest category will always decrease in probability as the corresponding factor is $-\phi(\nu_{1,t}) < 0$. The probability of the upper category will certainly increase due to the factor $\phi(\nu_{2,t}) > 0$. Thus it can be determined that the impulse has a positive effect on the sign of a price change $\Gamma_0^* > 0$ which is not quite plausible in a market without short sales restrictions.

This would be only plausible if there was some systematic discrimination between buyer and seller initiated trades, as e.g. short sales constraints. Because this would imply that a waiting time between transactions which is unexpectedly high would cause a higher probability for a price change upwards. If ρ is negative, the reverse argument holds and a positive duration shock would imply a higher probability for a price change downwards.

The empirical evidence reported in table 6.14 is quite plausible under this light. The results given in column A on the simultaneous process of the direction of price changes and the time between transactions $y_{s,t}$ show that neither the contemporaneous nor the past time between transactions is found to be significant. The latter is also the case in the univariate model. More important however is the fact that ρ is found insignificant and thereby a simultaneity between the duration and the price equation has to be rejected. See Smith and Blundell (1986) or Pohlmeier (1989, Chap. 2.4) for the derivation of the test. As a first important result, it can be concluded, that innovations on the duration process τ_t have no impact on the direction of price changes s_t.

TABLE 6.14. Simultaneous estimation of the Log ACD model for the time between transactions and the sign of price changes as described in (6.15)-(6.18), reported in column A, and simultaneous estimation of the Log ACD model for the time between transactions and the size of price changes reported column B-E. The regressors are log volume, $\log v_t$, and trigonometric expansions of intraday seasonals $s(\delta, t, 4)$. Regression based on the sample *short*. The mean log likelihood (log. lik.) and the Bayes Information Criterion (BIC) are also provided for each regression. White standard errors of the coefficients are given in parentheses.

Specification	A		B		C		D		E	
Dependent variable					Time between transactions					
equation one	τ_t		τ_t		τ_t		τ_t		τ_t	
					Parameters of the dynamic					
AR(1)	0.9708	(0.0122)	0.9592	(0.0130)	0.9587	(0.0164)	0.9674	(0.0119)	0.9660	(0.0133)
MA(1)	0.8906	(0.0220)	0.8450	(0.0295)	0.8776	(0.0250)	0.8863	(0.0216)	0.8824	(0.0237)
					Static regressors $w_{\tau,t}$					
z^*_{t-1}			-0.4225	(0.1539)	-0.3180	(0.3187)			-0.5053	(0.3918)
δ_c	0.7639	(0.0678)	0.6096	(0.0823)	0.6787	(0.0819)	0.7523	(0.0572)	0.7496	(0.0595)
δ_t	-0.0044	(0.0504)	-0.0072	(0.0463)	-0.0059	(0.0476)	-0.0045	(0.0503)	-0.0059	(0.0514)
$\delta_{s,1}$	0.1315	(0.1300)	0.1026	(0.1106)	0.1163	(0.1155)	0.1291	(0.1249)	0.1276	(0.1283)
$\delta_{s,2}$	-0.0191	(0.0975)	-0.0297	(0.0793)	-0.0273	(0.0838)	-0.0253	(0.0930)	-0.0259	(0.0965)
$\delta_{s,3}$	-0.0271	(0.0863)	-0.0203	(0.0698)	-0.0205	(0.0741)	-0.0236	(0.0826)	-0.0220	(0.0862)
$\delta_{s,4}$	0.1064	(0.0794)	0.0793	(0.0654)	0.0887	(0.0696)	0.1046	(0.0761)	0.0994	(0.0791)
$\delta_{c,1}$	-0.1775	(0.0888)	-0.1377	(0.0687)	-0.1494	(0.0735)	-0.1726	(0.0825)	-0.1640	(0.0866)
$\delta_{c,2}$	0.0644	(0.0819)	0.0470	(0.0626)	0.0535	(0.0672)	0.0562	(0.0757)	0.0572	(0.0793)
$\delta_{c,3}$	-0.0168	(0.0770)	-0.0167	(0.0603)	-0.0139	(0.0645)	-0.0158	(0.0719)	-0.0106	(0.0751)
$\delta_{c,4}$	-0.0213	(0.0731)	0.0100	(0.0588)	0.0017	(0.0627)	0.0018	(0.0695)	-0.0058	(0.0727)

Specification	A		B		C		D		E	
Dependent variable	Sign of price changes		Size of price changes							
equation two	s_t		z_t		z_t		z_t		z_t	
					Thresholds					
μ_1	-1.0290	(0.0541)	0.1077	(0.1527)	0.5690	(0.0644)	0.0723	(0.0694)	0.3842	(0.0360)
μ_2	0.8863	(0.0526)	1.7528	(0.1571)	2.2094	(0.0829)	1.7188	(0.0856)	1.9801	(0.0642)
					Parameters of the dynamic					
AR(1)			0.2332	(0.0284)	0.2521	(0.0285)	0.2374	(0.0285)	0.2109	(0.0288)
MA(1)	0.3268	(0.0257)								
					Contemporaneous dependent variable of equation one κ_t					
τ_t	-0.0623	(0.1014)	-0.2245	(0.1825)	0.3859	(0.0447)	-0.2208	(0.0448)		
					Static regressors $w_{\tau,t}$					
τ_{t-1}	-0.0219	(0.0684)	-0.1253	(0.0477)	-0.1166	(0.0429)	-0.1634	(0.0468)		
δ_t			0.0062	(0.0471)	0.0136	(0.0445)	0.0066	(0.0459)	0.0099	(0.0427)
$\delta_{s,1}$			-0.0019	(0.1085)	-0.0721	(0.1006)	0.0054	(0.1048)	-0.0437	(0.0966)
$\delta_{s,2}$			0.0380	(0.0735)	0.0580	(0.0692)	0.0382	(0.0724)	0.0467	(0.0668)
$\delta_{s,3}$			-0.0116	(0.0633)	0.0008	(0.0595)	-0.0129	(0.0626)	-0.0025	(0.0577)
$\delta_{s,4}$			0.0172	(0.0610)	-0.0348	(0.0555)	0.0226	(0.0583)	-0.0140	(0.0535)
$\delta_{c,1}$			-0.0502	(0.0657)	0.0497	(0.0552)	-0.0589	(0.0582)	0.0095	(0.0526)
$\delta_{c,2}$			-0.0141	(0.0551)	-0.0521	(0.0508)	-0.0130	(0.0537)	-0.0348	(0.0493)
$\delta_{c,3}$			-0.0160	(0.0529)	-0.0035	(0.0495)	-0.0152	(0.0525)	-0.0093	(0.0481)
$\delta_{c,4}$			0.0058	(0.0517)	0.0021	(0.0486)	0.0066	(0.0513)	0.0055	(0.0472)
					Correlation					
ρ	-0.0614	(0.1120)	0.6521	(0.1859)			0.5921	(0.1559)		
log. lik.	-0.5324		-0.4665		-0.4679		-0.4671		-0.4751	
BIC	-6618		-5862		-5879		-5865		-5948	

The simultaneity of the time between transactions process τ_t and the size of price changes z_t is assessed on the basis of the model outlined in section 4.2 of chapter 5. The latent process of the size of price changes corresponding to the observable price changes z_t is thus given by

$$z_t^* = H_z m_{z,t}^{\xi} + w_{z,t}' \gamma_z + \kappa_z \log \tau_t + \epsilon_{z,t} \tag{6.21}$$

$$m_{z,t}^{\xi} = F_z(m_{z,t-1}^{\xi} + e_1 c_{z,t-1}) + e_1 x_{z,t}' \beta_z. \tag{6.22}$$

The corresponding observation rule is given in (6.10) in this chapter. For the sake of completeness the equation describing the time between transactions is also given as

$$\log \tau_t = H_\tau \xi_{\tau,t} + w_{\tau,t}' \gamma_\tau + \kappa_\tau \mathrm{E}\left[z_t^* \middle| \mathcal{F}_{t-1} \right] \tag{6.23}$$

$$\xi_{\tau,t} = F_\tau \xi_{\tau,t-1} + e_1(x_{\tau,t} \beta_\tau + \epsilon_{\tau,t}). \tag{6.24}$$

along with the distributional assumption on the joint error term $\epsilon_t^{\dagger} = \begin{bmatrix} \epsilon_{z,t} & \epsilon_{\tau,t} \end{bmatrix}$.

$$\epsilon_t^{\dagger} \sim N(0, \Sigma_z) \tag{6.25}$$

$$\text{with } \Sigma = \begin{bmatrix} 1 & \rho_z \\ \rho_z & 1 \end{bmatrix} \tag{6.26}$$

The intensity of the trading process is often considered weakly exogenous for the process of price changes, see e.g. Russell and Engle (1998). The results on the sign of price changes reported in specification A of table 6.14 provided some evidence on this account, yet, the weak exogeneity does not necessarily hold for the process of absolute price changes z_t and time between transactions τ_t.

For the sign of the impulse response function Γ_s^* in the model of absolute price changes a similar argument as in the preceding discussion of the process $y_{s,t}$ can be used. The implications simultaneity has are much more plausible in the context of absolute price changes. Assume for the moment a positive ρ. Then, a positive shock on the duration equation, i.e. a waiting time between transactions which is longer than expected, would imply an increased probability to observe a price change versus the alternative of observing a transaction without a price change. Note that a positive correlation ($\rho > 0$) of the error terms leads to an increased probability of price changes. Thus an increase in the contemporaneous unexpected component of time between transactions increases the conditional volatility. Considering e.g. the model of Easley and O'Hara (1992) it is the exogenous but unobservable event of information accrual which influences both simultaneously, the process of absolute price changes and the intensity of the trading process.

The empirical evidence reported in table 6.14 supports the hypothesis of simultaneity. The correlation ρ of the error terms has a significantly positive sign, i.e. a shock on the duration equation implies actually a higher probability to observe a price change. See specification B. The severity of the simultaneous equation bias can be assessed from the erroneously significantly positive coefficient of τ_t reported in column C. In this specification the correlation is restricted to zero. This coefficient is either insignificant as in column B or significantly negative as in column D. The latter are both equations which do not suffer from the simultaneous equation bias since the correlation between both equations is explicitly accounted for in the maximum likelihood estimation of the system through the assumption of joint normality imposed on the error terms. From the analysis of the BIC it can be seen that specification B is actually favourable. Specification E, which excludes explicitly the contemporaneous time between transactions and the correlation term suffers from a heavy loss in fit, compared to specifications B-D, which provides additional evidence in favor of the simultaneous specification.

The empirical evidence provided above gives a clear answer to the question of the relationship between the size of price changes and the time between transactions. The two processes are not independent and the result is such that a large innovation on the time between transaction equation implies a large innovation on the size of price changes and thus an increased probability to observe larger price changes. Note that the contemporaneous τ_t has an insignificant influence in regression B. Yet, some care has to be taken in the interpretation of this result, because past time between transactions enter the specification with a negative sign.

This leaves the question open, how the dynamic effects within the simultaneous model $y_{z,t}$ have to be assessed.[2] To scrutinize the dynamic properties of the joint model the concept of nonlinear impulse responses outlined in chapter 5 section 5.2 is employed. Remember that it relies on the difference of conditional expectations as

$$\Gamma_s = \mathrm{E}\left[y_{z,t+s}\right] - \mathrm{E}\left[y_{z,t+s}\right| \epsilon_t = \epsilon_*\right], \tag{6.27}$$

where ϵ_* is a particular impulse on the system.

The conditional expectation $\mathrm{E}\left[y_{z,t+s}\right| \epsilon_*\right]$ will be estimated by Monte Carlo methods. To obtain the above conditional expectation a set of N replications for the conditional expectation of $y_{z,t+s}$ using draws $\zeta_{j,i}$ and $\bar{\zeta}_i$, $i = 1,\ldots,N$, from the distribution of ϵ_t

$$y_{z,t+s,i} = \mathrm{E}\left[y_{z,t+s}\right| \epsilon_{t+s} = \zeta_{1,i}, \epsilon_{t+s-1} = \zeta_{2,i}, \ldots, \epsilon_{t+1} = \zeta_{s,i}, \epsilon_t = \epsilon_*, \bar{\epsilon}_{t-1} = \bar{\zeta}_i\right]. \tag{6.28}$$

The history of the system $\bar{\epsilon}_{t-1}$ is assumed to consist of 30 periods in the practical application. The conditional expectation can thus be estimated from the ordinary sample mean of the above given generated conditional expectations as

$$\hat{\mathrm{E}}\left[y_{z,t+s}\right| \epsilon_t = \epsilon_*\right] = \frac{1}{N}\sum_{i=1}^{N} y_{z,t+s,i}. \tag{6.29}$$

The procedure outlined by 6.28 and 6.29 can be repeated for a total of M different values of ϵ_* thus obtaining a whole range of sample paths $\hat{\mathrm{E}}\left[y_{z,t+s}\right| \epsilon_t = \epsilon_*\right]$, $s = 1,\ldots,S$. The impulse ϵ_* is usually drawn as well from the distribution of ϵ_t, so that the distribution of sample paths gives a representative picture of the system's dynamic.

To sum up, the individual steps carried out to analyse the impulse responses are

(1) Historical and future realizations of shocks are drawn according to the distributional assumptions of ϵ_t $N = 500$ times.

[2] An analysis of the process $y_{s,t}$ does not seem worthwhile as simultaneity has to be rejected.

(2) The conditional expectation $\hat{\mathrm{E}}\left[y_{z,t+s} \middle| \epsilon_t = \epsilon_* \right]$ is evaluated from the estimates of specification B in table 6.14 as the sample mean of the replications.

(3) Steps 1 and 2 are carried out for a sample of $M = 50$ draws of the shock ϵ_* drawn from the distribution of ϵ_t.

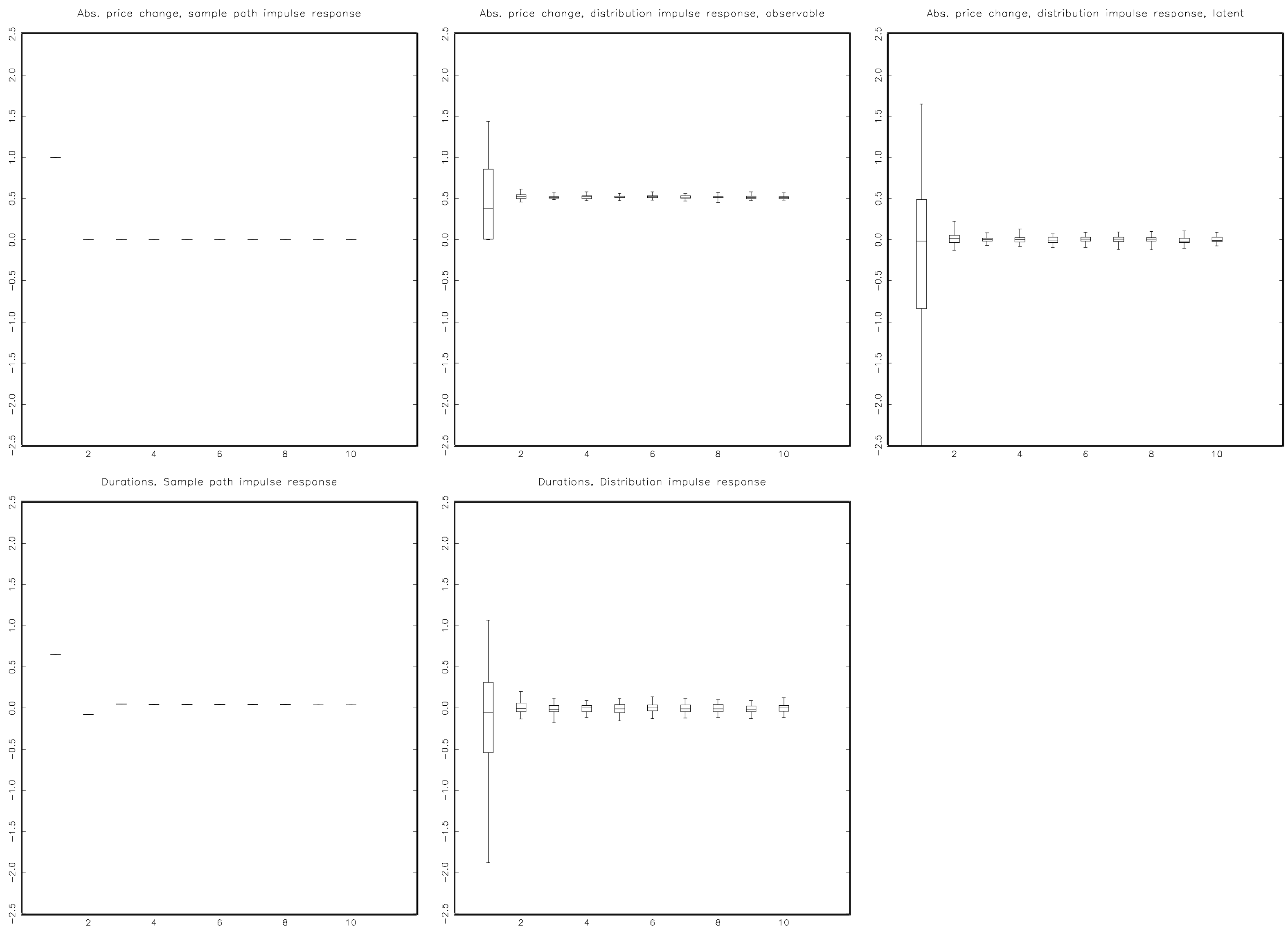
(4) The distribution of conditional expectations $\hat{\mathrm{E}}\left[y_{z,t+s} \middle| \epsilon_t = \epsilon_{*,i} \right]$, $i = 1, \ldots, 100$, is depicted for the forecast horizons $s = 0, \ldots, 10$.

Practically, the shock is inflicted on the time between transactions equation τ_t at $s = 0$, via the orthogonal decomposition already used in chapter 5 in (5.38)

$$\epsilon_{z,t} = \rho \epsilon_{\tau,t} + \nu_{z,t} \tag{6.30}$$

the simultaneous effect on the size of price changes is computed. The regressors not relevant to the dynamic are set to zero in this case. Note that the impulse responses given in figures 6.8 and 6.9 are not given as the difference denoted in (6.27), but depict only the conditional expectation $\hat{\mathrm{E}}\left[y_{z,t+s} \middle| \epsilon_t = \epsilon_* \right]$. Thus, the impulse response of the observable size of price changes z_t does not return to zero, as s increases but returns to the unconditional expectation $\mathrm{E}\left[y_{z,t+s} \right]$. For the other variables, the latent z_t^* and the time between transactions τ_t, this is zero, due to the omission of static regressors.

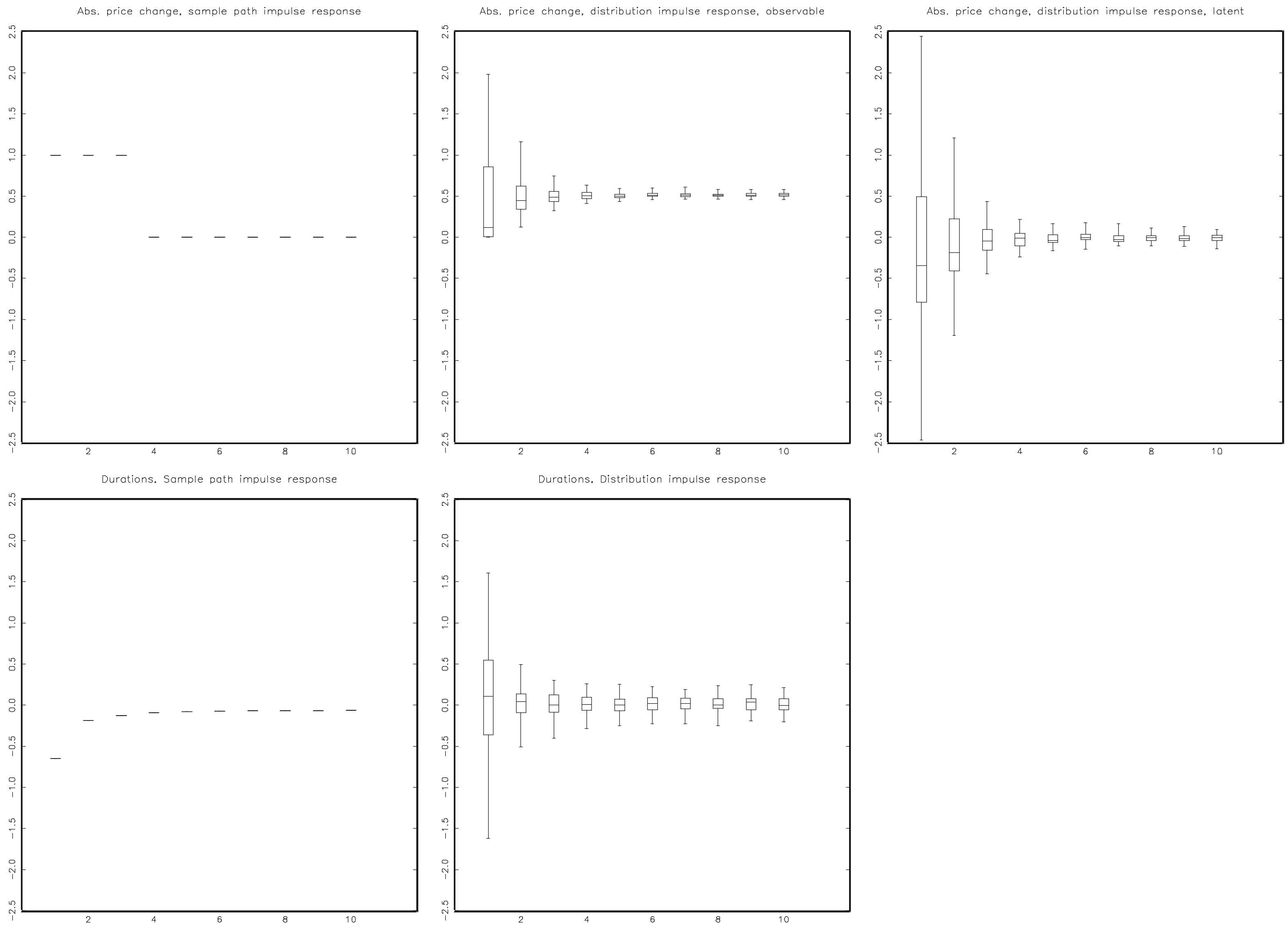
FIGURE 6.8. Impulse responses for the estimation of the simultaneous Log ACD model for the time between transactions and the size of price changes as described in (6.21)-(6.24), reported in column B of table 6.14. The coefficients of intraday seasonals are set to zero.



A clear cut picture of the implied dynamics of the system can be gained from an inspection of the impulse response function depicted in figure 6.8. The two left hand panels give the conditional expectation of one particular impulse $\hat{\mathrm{E}}\left[y_{z,t+s}\right|\epsilon_t=\epsilon_*]$. The two middle panels describe the distribution of the observable conditional expectations $\hat{\mathrm{E}}\left[y_{z,t+s}\right|\epsilon_t=\epsilon_{*,i}]$, $i=1,\ldots,M$. The right hand panel gives the distribution of the usually unobservable conditional expectation of the latent size of price changes $\hat{\mathrm{E}}\left[z^*_{t+s}\right|\epsilon_t=\epsilon_{*,i}\left.\right]$, $i=1,\ldots,M$. The estimated values for the dynamic of $y_{z,t}$ imply a very quick dampening of the impulse inflicted on the time between transaction equation. The positive correlation ρ in conjunction with the significant negative influence of past values of the latent variable z^*_t on the duration process leads to a dampening of the impulse. This is quite plausible from an intuitive point of view, it implies that trades, which occur unexpectedly quick, have no significant influence on future price changes, the influence is limited to the present transaction.

To demonstrate that it is indeed the positive correlation ρ, which has a substantial influence on the dynamic properties of the process, a little experiment is carried out in figure 6.9. The sign of the estimated correlation is exchanged so that all the parameters are identical, on which 6.8 and 6.9 are based, but the correlation which is changed to $\rho_z=-0.6521$. The effect is quite dramatic. The shock inflicted on the duration equation lasts for another two observations after the event.

FIGURE 6.9. Impulse responses for the estimation of the simultaneous Log ACD model for the time between transactions and the size of price changes as described in (6.21)-(6.24), reported in column B of table 6.14. The coefficients of intraday seasonals are set to zero. Compared to the indicated estimation, the sign of the correlation is switched, i.e. $\rho_z = -0.6521$.



CHAPTER 7

The stability of the price process

1. The econometric problem

The instability of empirical findings is a grave problem when it comes to the question whether the results of an empirical study are just an artefact attributed to the particular properties of the chosen sample. On the other hand there is a good chance that appropriate samples are quite representative for the economic system under consideration. However, the finding that the data set in conjunction with the employed model are indeed subject to structural breaks also carries the potential to learn more about the economic background of the market under consideration. This is the approach chosen by Gerhard, Hess, and Pohlmeier (1998). They try to isolate macro shocks of particular importance for the price process of Bund future trading. In the present context the focus is not on the effects of macro shocks but on the question how representative the chosen sample is for the analysis of the financial market under consideration. Therefore, the methodology suggested by Gerhard, Hess, and Pohlmeier (1998) is employed making a strong use of the large amount of data available for this contract.

The econometric problem under consideration is described by assuming that the density of the data x_t, $t = 1, \ldots, T$, is given by $f(x_t, \theta)$. The general setting consists of a density $f(x_t, \theta)$, with a parameter vector θ which may change at a certain observation index t^* which is a nuisance parameter only present under the alternative hypothesis H1.

$$f(x_t, \theta_t) = f(x_t, \theta_1) 1_{(t<t^*)} + f(x_t, \theta_2) 1_{(t \geq t^*)}, \tag{7.1}$$

$$t = 1, \ldots, T \qquad\qquad t^* \in [1, T]$$

$$\theta_t \in \{\theta_1, \theta_2\}$$

There are two different paths to go by if one suspects to find structural breaks in the empirical model under consideration. Particularly, if one considers the possibility that there might be more than one switch in the parameters involved, then it might be appropriate to choose a model for time varying structures in the first place. State space models are natural candidates as appropriate model structures be it that they employ a discrete number of states as in the Markov switching class proposed by Hamilton (1990) or be it that they employ a continuous state space as in the standard Kalman filter models. For the latter see e.g. Harvey (1989) and for a brief description of both see Hamilton (1994).

On the other hand, one might back up one step and try to find out whether the data/model combination reflects any structural breaks at all, and how these breaks can be motivated economically at all.

There are different approaches to solve this problem. First, one could test parameter instability given a known change point. The null hypothesis may be formulated as:

> **H0:** No change in the parameters θ_t of the density $f(x_t, \theta_t)$ has occurred in the sequence $x_1, \ldots, x_T$ at $t = t^*$, where θ_1 and θ_2 are unknown.

The most prominent representative of these tests is Chow (1960).

On the other hand, not all candidates for structural breakpoints are as obvious as the German reunification. Particularly, in the context of intra day data it is a formidable task to keep track of all possible macro-shocks, which hit the system. Thus, the null hypothesis may be restated as:

> **H0:** No change in parameters θ_t of the density $f(x_t, \theta_t)$ has occurred in the sequence $x_1, \ldots, x_T$, where θ_0, θ_1, and t^* are unknown.

A general discussion is found in Andrews (1993) or Hawkins (1987). If one decides to construct a standard likelihood ratio test, with test statistic LR, for all possible points t^*, $t^* \in [1, T]$, then the test statistics are clearly non standard. They involve the supremum over all possible points in time which might be a candidate for a structural breakpoint.

$$\sup_{t^* \in [1,T]} LR(t^*) \tag{7.2}$$

This complicates matters considerably as the distribution of test statistics of the above form is quite complicated. As can be seen from the discussion in the literature. See e.g. Bai (1999) and further references given there.

For transaction data however there is a modified approach feasible. It is suggested by Gerhard, Hess, and Pohlmeier (1998) and involves a two stage estimator. The idea behind the estimator and the corresponding test is quite straightforward and rests on the large amount of data available for estimation. If one considers the index t in (7.1) as an index over trading days or over groups of trading days in the sample, then the θ_t can be estimated in a first step by N individual regressions. This is only possible because of the abundance of transaction data compared to lower sampling frequencies. To provide just a brief example, consider the 2.200 observations of a typical trading day. It would take about 10 years of daily observations or 40 years of weekly observations to gather the same amount of data and it would be questionable to assume that these 10 or 40 years could be treated as an entity. The information content of the intraday observations compared to aggregates is in a sense quite reduced but it allows to estimate a coefficient for each trading day. Certainly, this rests on the assumption that the intraday structural breaks show in the parameter estimates of one trading day. Considering the nature of transaction data this seems not unreasonable.

To be more concise, the estimator and the associated test work as follows:

(1) The first step estimates $\hat{\theta}_i$ and the associated covariance matrices $\widehat{\text{Var}}\left[\hat{\theta}_i\right]$ are obtained from a sequence of standard ML estimations of the model under consideration.[1]

(2) The joint parameter θ is estimated from the individual parameters based on the assumption that all θ_i are equal, i.e. $\theta = \theta_i$ should hold for all

[1]Note that the index of trading days is changed to i instead of t, which is used to index intraday observations.

$i = 1, \ldots, N$. This assumption leads to a restriction function

$$g_i = \theta - \hat{\theta}_i. \tag{7.3}$$

On this basis the minimum distance estimator for θ is given by

$$\begin{aligned}\hat{\theta} &= \arg\min_{\theta \in \Theta} \quad \sum_{i=1}^{N} g_i' \left(\widehat{\text{Var}}\left[\hat{\theta}_i\right]\right)^{-1} g_i \\ &= \left(\sum_{i=1}^{N} \left(\widehat{\text{Var}}\left[\hat{\theta}_i\right]\right)^{-1}\right)^{-1} \sum_{i=1}^{N} \left(\widehat{\text{Var}}\left[\hat{\theta}_i\right]\right)^{-1} \hat{\theta}_i.\end{aligned}$$

For a general treatment of minimum distance estimators or asymptotic least squares estimators as they are sometimes called, see e.g. the extended discussion in Gourieroux and Monfort (1995, chap. 9.1).

(3) The estimator for the corresponding covariance matrix is readily available as

$$\widehat{\text{Var}}\left[\hat{\theta}\right] = \left(\sum_{i=1}^{N} \left(\widehat{\text{Var}}\left[\hat{\theta}_i\right]\right)^{-1}\right)^{-1} \tag{7.4}$$

(4) The test statistic associated with the null hypothesis that the restriction (7.3) holds given by the target function at the minimum.

$$\xi_1 = \sum_{i=1}^{N} (\hat{\theta} - \hat{\theta}_i)' \left(\widehat{\text{Var}}\left[\hat{\theta}_i\right]\right)^{-1} (\hat{\theta} - \hat{\theta}_i) \tag{7.5}$$

The test statistic is χ^2 distributed with $(N - 1) \cdot m$ degrees of freedom, where m is the dimension of θ_i.

Alternatively, one might be tempted to consider the two one step ML estimators to which the above described MD estimator is asymptotically equivalent under certain conditions as it is discussed in Kodde, Palm, and Pfann (1990).

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \quad \sum_{i=1}^{N} \sum_{t=1}^{T_i} \log f(x_{i,t}, \theta) \tag{7.6}$$

$$\hat{\theta^*} = \arg\max_{\theta^* \in \Theta^*} \quad \sum_{i=1}^{N} \sum_{t=1}^{T_i} \log f(x_{i,t}, \theta_i) \tag{7.7}$$

Here, T_i is the number of observations of day i and θ^* the $(N \cdot m \times 1)$ joint parameter vector containing all θ_i. An alternative to the test statistic ξ_1 might be the likelihood ratio statistic based on (7.6) and (7.7). Yet, at least the second estimator (7.7) is clearly unfeasible for practical purposes. Even if the number of observations $\sum_{i=1}^{N}\sum_{t=1}^{T_i} 1$ is still manageable taken for itself, the number of intermediate parameters $N \cdot m$ would certainly render this estimation unfeasible.

2. Empirical evidence concerning the price process

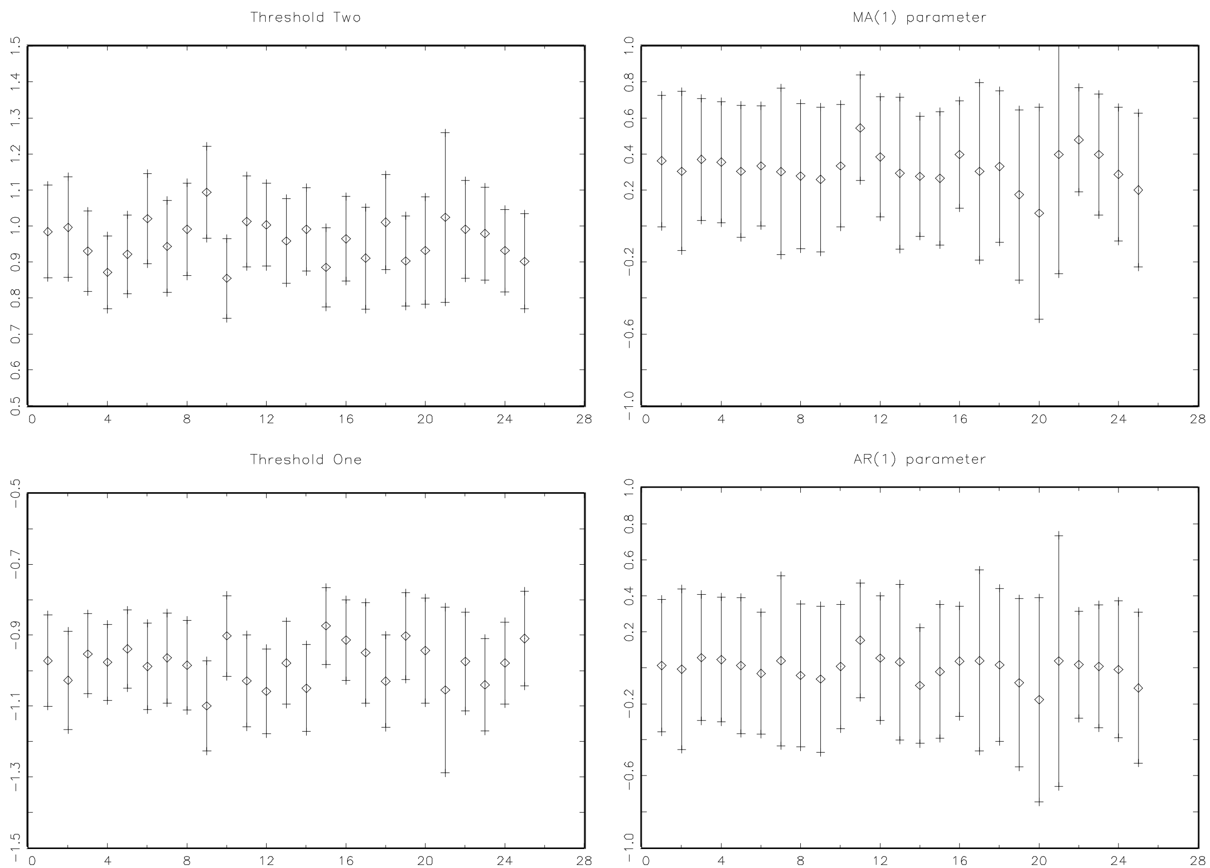
2.1. The stability of empirical results. There are various paths to proceed in order to assess the structural stability of regression results in this context. The importance of this assessment can be illustrated by examining figures 7.1-7.4 which depict estimation results for individual regressions.

To assess the stability of the empirical findings in the preceding chapter the samples *long* and *composed* are utilized, which are described in section 1.1 of chapter 6. *Long* is made of 25 trading days of the same contract which is already used for the sample *short* in chapter 6. The sample *composed* is made of blocks of five trading days selected from a sequence of nine futures contracts. The five days start approximately around 45 days to maturity in each contract. This is right in between the roll over phase around 90 days to maturity and the expiration of the contract.

The estimated models are based on the by now well-known limited dependent dynamic of the ordered probit. For the sign of price changes s_t, the observation equation takes on the following form, already known from the empirical results in chapter 6 and the model description in chapter 3

$$s_t = \begin{cases} -1 & \text{, if} \quad s_t^* \leq \mu_{s,1} \\ 0 & \text{, if} \quad \mu_{s,1} < s_t^* \leq \mu_{s,2} \\ 1 & \text{, if} \quad \mu_{s,2} < s_t^* \end{cases} . \tag{7.8}$$

FIGURE 7.1. Estimates for the process of the direction of price changes, s_t. Individual estimates are based on individual trading days of the *long* sample. Confidence intervals are given on the 5% level, based on the normal distribution. The point estimates are characterized by diamonds.



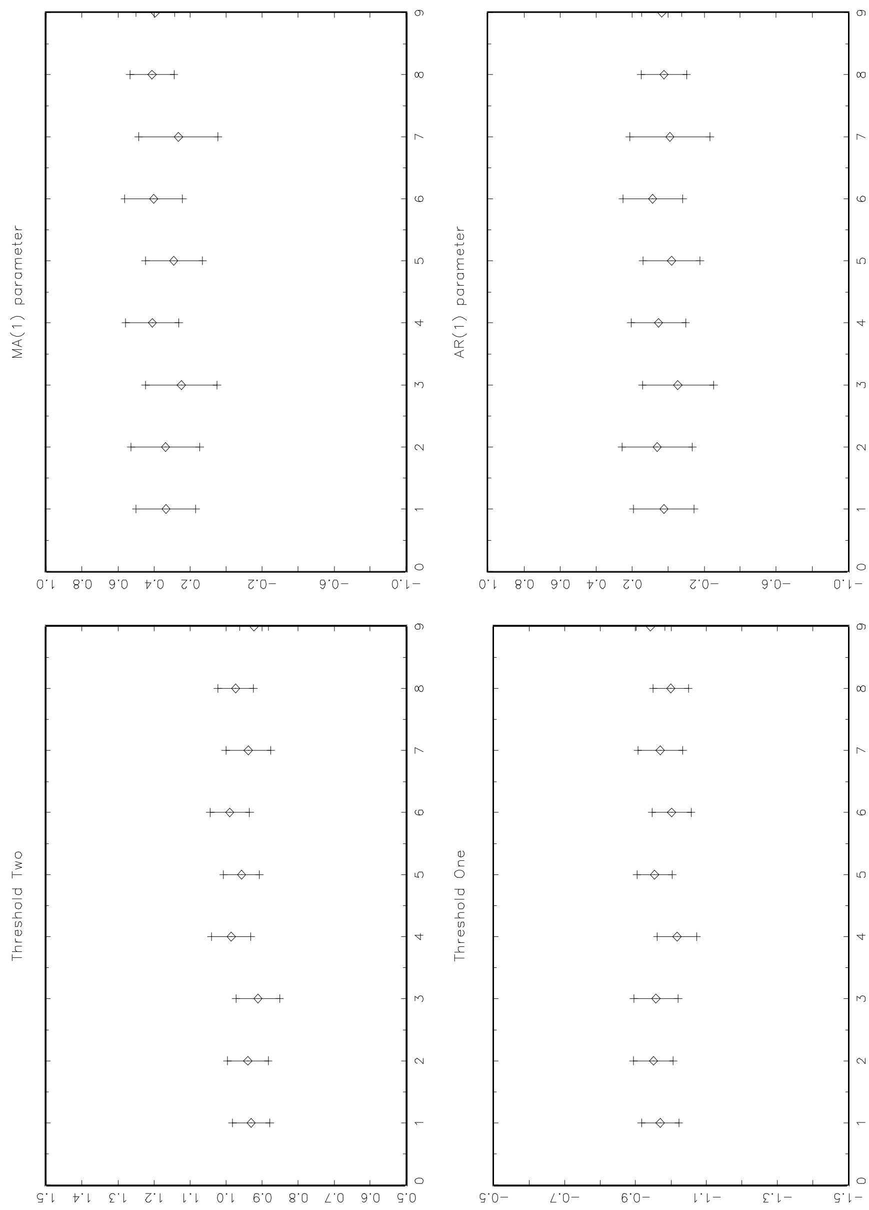


FIGURE 7.2. Estimates for the direction of price changes, s_t. Individual estimates are based on blocks of five trading days of the *composed* sample. Confidence intervals are given on the 5% level, based on the normal distribution. The point estimates are characterized by diamonds.

In this context only the dynamic of the latent variable is included and an ARMA(1,1) model is estimated for each of the sub samples, which takes on the form

$$s_t^* = \begin{bmatrix} 1 & \theta \end{bmatrix} m_{s,t}^{\xi} + \epsilon_{s,t} \tag{7.9}$$

$$m_{s,t}^{\xi} = \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix} (m_{s,t-1}^{\xi} + e_1 c_{s,t-1}). \tag{7.10}$$

The graphs give the point estimates and the variation of the estimated parameters, i.e. the 5% confidence intervals based on the assumption of asymptotic normality and the estimated parameter variance. They include the estimates for all four parameters, the two thresholds and the two parameters of the dynamic, the MA-Parameter θ and the AR parameter ϕ.

Evidence on the parameters describing the direction of price changes s_t is gathered in figures 7.1 and 7.2. Figure 7.1 is based on the *long* sample. Each mark corresponds to one estimated parameter based on the observations of one trading day. It is quite notable that size and position of the intervals vary considerably over the trading days. Figure 7.2 on the other hand gives estimation results based on the *composed* sample, which is made up of five day sub samples, stretching over nine different contracts. The increased size of the sub samples lead to a much more stable behaviour of successive estimates and to confidence intervals which are considerably smaller.

Figures 7.3 and 7.4 give details on the estimates for the size of price changes z_t. For the sake of brevity the specification of the model is not repeated explicitly as it is quite similar to (7.9) and (7.10). The observation rule is given explicitly in chapter 6 in (6.10). Again, a latent ARMA(1,1) model with three categories is estimated. Considering figure 7.3 based on the *long* sample, the variation of parameters is somewhat stronger, than it is observed for the sign of price changes in figure 7.1. The effect of the increased sample size can again be observed in figure 7.4 compared to figure 7.3, which is based on the *composed* sample.

FIGURE 7.3. Estimates for the size of price changes, z_t. Individual estimates are based on individual trading days of the *long* sample. Confidence intervals are given on the 5% level, based on the normal distribution. The point estimates are characterized by diamonds.

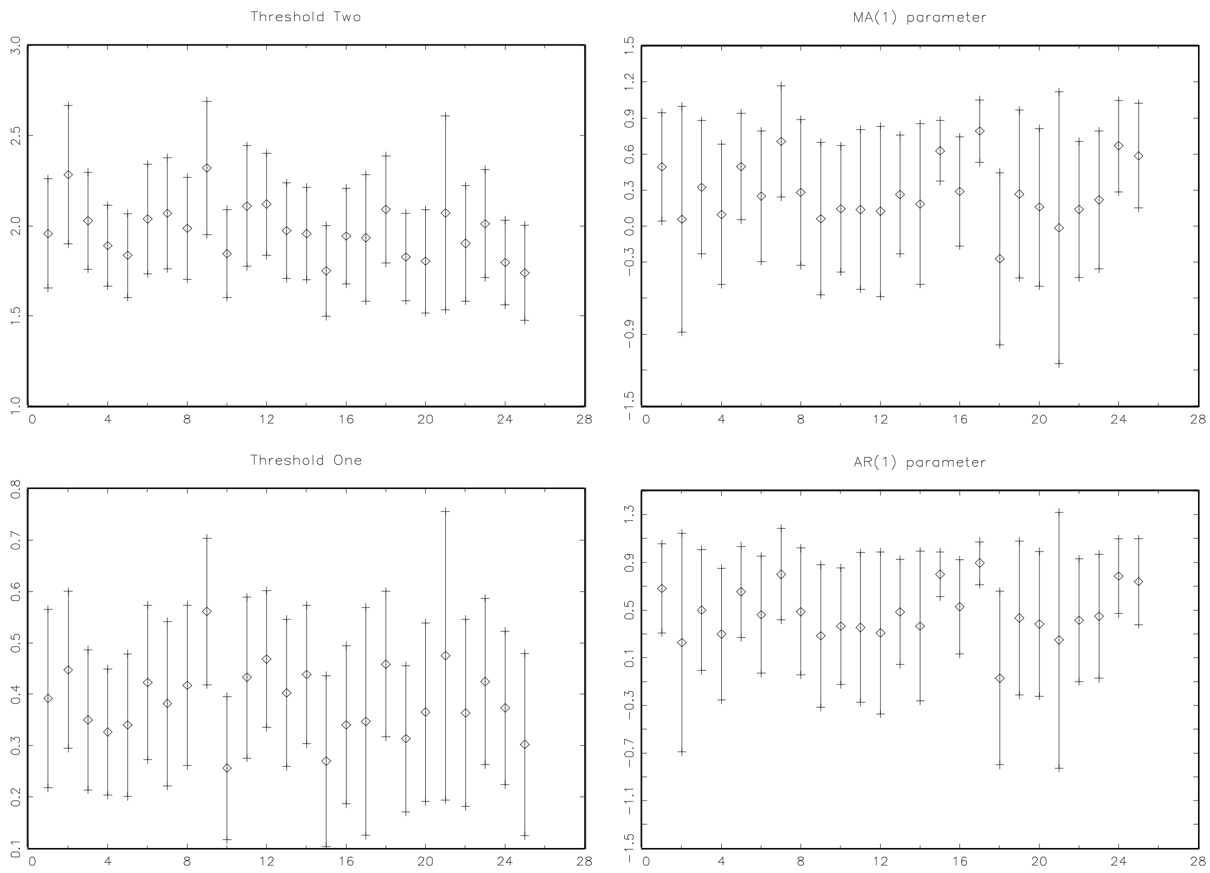


FIGURE 7.4. Estimates for the process of the size of price changes, z_t. Individual estimates are based on blocks of five trading days of the *composed* sample. Confidence intervals are given on the 5% level, based on the normal distribution. The point estimates are characterized by diamonds.

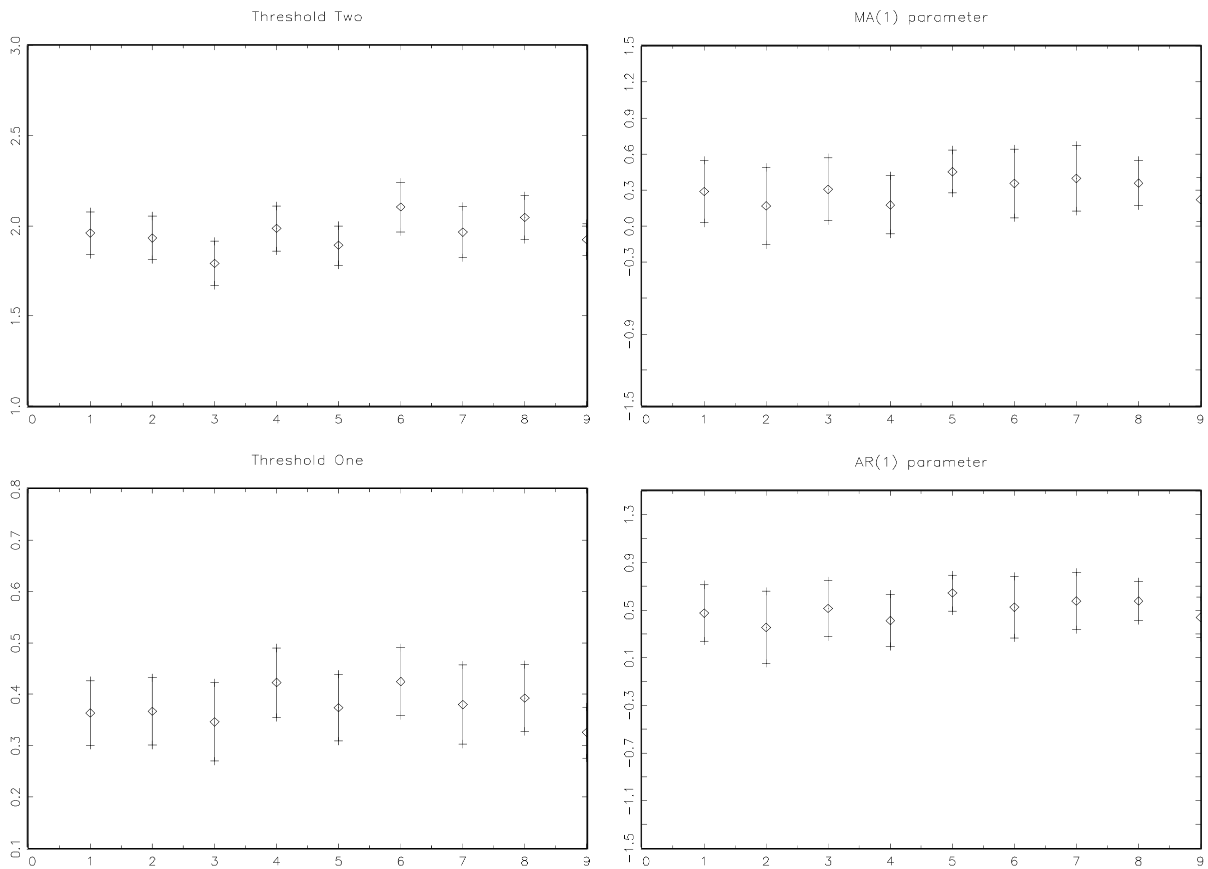


TABLE 7.1. Minimum distance estimation of the direction and the size of price changes based on the *long* and *composed* sample.

	s_t Long		s_t Composed		z_t Long		z_t Composed	
	Coeff.	Std.dev.	Coeff.	Std.dev.	Coeff.	Std.dev.	Coeff.	Std.dev.
Thresholds								
μ_1	-0.9717	(0.0127)	-0.9705	(0.0089)	0.38801	(0.0156)	0.3737	(0.0110)
μ_2	0.9518	(0.0126)	0.9482	(0.0088)	1.95060	(0.0284)	1.9478	(0.0199)
Parameters of the dynamic								
AR(1)	0.0149	(0.0385)	0.0340	(0.0264)	0.65466	(0.0388)	0.5256	(0.0343)
MA(1)	0.3344	(0.0379)	0.3604	(0.0258)	0.46970	(0.0453)	0.3268	(0.0385)
ξ_1	54.49		35.72		70.71		33.25	
p-val	0.99		0.48		0.98		0.59	

Overall, it is apparent that there is some variation in the coefficients. By first inspection this appears to be insignificant, i.e. due to the variation of estimated coefficients. A second feature worth mentioning is that the quality of estimated coefficients varies substantially over the trading days.

The evidence presented so far in the form of descriptive statistics based on a first step estimator can be made more concise using the minimum distance estimator introduced before. First of all, it allows to consistently estimate a set of parameters in a second estimation step based on the regressions of individual trading days, which describes the total sample. This estimator is asymptotically equivalent to a maximum likelihood estimator based on the entire sample.

A striking result of the minimum distance estimation step reported in table 7.1 is that the stability of the coefficients cannot be rejected in any of the four models, i.e. the equality restriction holds. From this it can be concluded that estimations based on the sub samples are indeed representative for the market. The differences between the p-values of the test on the *long* and on the *composed* sample can well be attributed to the larger parameter variances in the first step estimates based on daily observations, which in turn lowers the probability of a rejection of the null hypothesis of equal parameters.

This implies that the reduction of the sample to five trading days certainly limited the quality of the estimates but valid inferences can be drawn for the financial market under consideration. Another surprising fact is that the parameters are even stable over an extended horizon of nine different contracts, with only a very slight variation in the estimated coefficients. The latter fact encourages indeed the search for structural properties of exchanges, which are constant over time and can be attributed to the particular exchange mechanism and the traded asset.

CHAPTER 8

Conclusion

After a sketch of the peculiarities of the intraday transaction process in chapter 2, the main innovation of this work was the introduction of a latent dynamic for quantal response models from the linear exponential family. This new model has the advantage that it combines an ARMA type dynamic with a dependent variable which is only observed discretely. This model is based on the well-known concept of generalized residuals and does not employ ad-hoc assumptions on the functional form, with which past observations enter the dynamic of the process, as it is often the case in the context of GLM which were compared to the new specification. Compared to the approach involving generalizations of Markov chains, this approach is clearly more parsimonious, because the dynamic is not modelled via a state transition but via the dynamic of a continuous, but unobservable variable. The possible applications of this model include not only empirical market microstructure models, but also macroeconomic models involving e.g. central bank behaviour or financial models analysing rating transitions of bonds.

It was shown that this new estimator fits in well with a modified autoregressive conditional duration model suggested in chapter 4 to form a simultaneous estimator to assess the transaction process. Here, the advantage of using a classical ARMA structure in the form of a state space model became evident, as it allowed to build analogue models for the time between transactions, the sign and the size of the process of price changes. Along the lines an alternative decomposition of the price process was developed in chapter 5, which combines the advantages of Russell and Engle (1998) and Rydberg and Shephard (1998), since it employs an ACD model to map trade frequency, which is somewhat more parsimonious than the discrete time framework proposed by Rydberg and Shephard. It employs the decomposition of the price process suggested by Rydberg and Shephard, which allows to efficiently analyse the sign and the size process separately from each other, while being able to recover the distribution of price changes from the two processes. A simultaneous estimator for the processes of the sign of price changes and of the size of price changes combined with the time between transactions was developed in chapter five.

The empirical results found on the basis of a data set on Bund future trading at the DTB in Frankfurt shed some light on the transaction process. It was found that the direction of price changes is indeed independent from the duration process, which is quite plausible from an economic point of view considering the nature of the traded asset. The process of the size of price changes and the duration process were found to be simultaneously dependent with a positive correlation of their error terms. This implies that an unexpectedly long time between transactions goes along with a higher contemporaneous price change. This is however compensated by the influence of lags of the dependent variables on the dynamic. It was resolved, that an impulse on the time between transactions equation was damped almost instantly. This has been shown in an impulse response analysis for nonlinear models. Concerning the stability of empirical results it was found that even a small sample of one trading week, i.e. five days, suffices to draw valid conclusions on the process of price changes and absolute price changes.

The modeling framework presented here is open for several extensions three of which should be mentioned explicitly. First, the volume and transactions costs could be included as additional dimensions of the transaction process. Empirical results on the univariate series seem to indicate that this might be of particular interest for the size of price changes process. The inclusion of volume is quite straightforward, as it would only necessitate to expand the dynamic linear duration model to a bivariate linear model. The distribution of the size of price changes would thus be conditioned on both variables and the appropriate estimates of the residuals could be included in the conditional maximum likelihood estimation of the ordered probit. The inclusion of transaction costs would be somewhat more involved as it would lead to the estimation of a bivariate quantal response model, which is manageable but certainly not as straightforward as the inclusion of volume.

The second major issue for an extension of the present work is the analysis of aggregates based on the estimated dynamics. In particular it would be interesting to assess in which respect the simultaneity and the dynamic structure influences the involved aggregates, especially the number of transactions and the volatility over a certain time span. This would allow to assess questions of the optimal aggregation of intraday transaction data for the use of risk estimators. Also it would allow to answer the important question, in which respect the simultaneity and the dynamic properties found on the transaction level carry over to an aggregate level.

The third extension leads to weakening of the distributional assumptions involved in the estimation of the ordered probit. It was shown that this specification relies on the assumed distribution in contrast to the duration model which is open to a PML interpretation. Thus, it rests only on the conditional moment restriction to ensure the consistency of the estimated parameters and not on the distributional assumptions on the error term. The inference on the mean function in the model would thus not be invalidated by a misspecification of the distribution. It would be quite interesting to test on the appropriateness of the assumptions made and then to weaken the imposed structures.

Bibliography

Admati, A., and P. Pfleiderer (1988): "A Theory of Intra-Day Patterns: Volume and Price Variability," *Review of Financial Studies*, 1, 3–40.

Amemiya, T. (1985): *Advanced Econometrics*. Basil Blackwell, Oxford.

Andersen, T. G., and T. Bollerslev (1997): "Intraday Periodicity and Volatility Persistence in Financial Markets," *Journal of Empirical Finance*, 4, 115–158.

——— (1998a): "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885–905.

——— (1998b): "Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies," *Journal of Finance*, 53, 219–265.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (1999): "The Distribution of the Exchange Rate Volatility," Discussion Paper 6961, NBER Working Paper Series.

Anderson, B. D. O., and J. B. Moore (1979): *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, N.J.

Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

Bai, J. (1999): "Likelihood Ratio Tests for Multiple Structural Changes," *Journal of Econometrics*, 91, 299–323.

Bai, X., J. R. Russell, and G. C. Tiao (1999): "Beyond Merton's Utopia; Effects of Non-Normality and Dependence on the Precision of Variance Estimates Using High-Frequency Financial Data," Discussion paper, Graduate School of Business, University of Chicago.

Ball, Torous, and Tschoegl (1985): "The Degree of Price Resolution: The Case of the Gold Market," *Journal of Futures Markets*, 5, 29–43.

Ball, C. (1988): "Estimation Bias Induced by Discrete Security Prices," *Journal of Finance*, 43, 841–865.

Bauwens, L., and P. Giot (1997): "The Logarithmic ACD Model: An Application to Market Microstructure and NASDAQ," Discussion Paper 9789, CORE.

Bauwens, L., and D. Veredas (1999): "The Stochastic Conditional Duration Model: A Latent Factor Model for the Analysis of Financial Durations," Discussion Paper 99xx, CORE.

Bernard, H., and S. Gerlach (1996): "Does the Term Structure Predict Recessions? The International Evidence," Discussion paper, Bank for International Settlements.

Blundell, R. W., and R. J. Smith (1993): "Simultaneous Microeconometric Models with Censored and Qualitative Dependent Variables," in *Handbook of Statistics, Vol. 11*, ed. by G. S. Maddala, C. R. Rao, and H. D. Vinod, pp. 117–143. Elsevier Science Publishers.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bollerslev, T., and E. Ghysels (1996): "Periodic Autoregressive Conditional Heteroscedasticity," *Journal of Business and Economic Statistics*, 14, 139–151.

Broseta, B. (1993): "Estimation of a Game-Theoretic Model of Learning: An Autoregressive Conditional Heteroskedasticity Approach," Discussion Paper 93-35, University of California, San Diego.

Cameron, A. C., and P. K. Trivedi (1996): "Count Data Models for Financial Data," in *Handbook of Statistics, Vol. 14*, ed. by G. S. Maddala, and C. R. Rao, chap. 12, pp. 363–391. Elsevier Science.

Cameron, A. C., and P. K. Trivedi (1998): *Regression Analysis of Count Data*. Cambridge University Press, Cambridge.

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

Chib, S., and E. Greenberg (1998): "Analysis of Multivariate Probit Models," *Biometrika*, 85, 347–361.

Cho, D. C., and E. W. Frees (1988): "Estimating the Volatility of Discrete Stock Prices," *Journal of Finance*, 43, 451–466.

Chow, G. C. (1960): "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28, 591–605.

Clark, P. K. (1973): "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica*, 41, 135–155.

Cox, D. R. (1972): "Regression Models and Life-Tables," *Journal of the Royal Statistical Society, Series B*, 34, 187–200.

Cox, D. R. (1981): "Statistical Analysis of Time Series: Some Recent Developments," *Scandinavian Journal of Statistics*, 8, 93–115.

Cox, D. R., and E. J. Snell (1989): *Analysis of Binary Data*. Chapman and Hall, New York, 2nd. edn.

DACOROGNA, M. M., C. L. GAUVREAU, U. A. MÜLLER, R. B. OLSEN, AND O. V. PICTET (1996): "Changing Time Scale for Short-term Forecasting in Financial Markets," *Journal of Forecasting*, 15, 203–227.

DAVIDSON, J. (1994): *Stochastic Limit Theory*. Oxford University Press, Oxford, New York.

DAVUTYAN, N., AND W. R. PARKE (1995): "The Operations of the Bank of England 1890-1908: A Dynamic Ordered Probit Approach," *Journal of Money, Credit, and Banking*, 27, 1099–1112.

DEMPSTER, A. P., N. M. LAIRD, AND D. B. RUBIN (1977): "Maximum Likelihood from Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society Series B*, 39, 1–38.

DIAMOND, D. W., AND R. E. VERRECCHIA (1987): "Constraints on Short-Selling and Asset Price Adjustment to Private Information," *Journal of Financial Economics*, 18, 277–311.

DROST, F. C., AND T. E. NIJMAN (1993): "Temporal Aggregation of GARCH Processes," *Econometrica*, 61, 909–927.

DUEKER, M. (1999): "Conditional Heteroskedasticity in Quantile Response Models of Time Series: A Gibbs-Sampling Approach to the Bank Prime Rate," *Journal of Business Economics and Statistics*, 17, 466–472.

DUFOUR, A., AND R. F. ENGLE (2000): "Time and the Price Impact of a Trade," *Journal of Finance*, 32, 2467–2498.

DUNSMUIR, W., AND E. J. HANNAN (1976): "Vector Linear Time Series Models," *Advances in Applied Probability*, 8, 339–364.

EASLEY, D., N. M. KIEFER, AND M. O'HARA (1997): "The Information Content of the Trading Process," *Journal of Empirical Finance*, 4, 159–186.

EASLEY, D., N. M. KIEFER, M. O'HARA, AND J. B. PAPERMAN (1996): "Liquidity, Information, and Infrequently Traded Stocks," *Journal of Finance*, 51, 1405–1436.

EASLEY, D., AND M. O'HARA (1992): "Time and the Process of Security Price Adjustment," *Journal of Finance*, 47, 577–607.

EICHENGREEN, B., M. W. WATSON, AND R. S. GROSSMAN (1985): "Bank Rate Policy under the Interwar Gold Standard," *Economic Journal*, 95, 725–45.

EICHENGREEN, B., M. W. WATSON, AND C. WYPLOSZ (1995): "Exchange Rate Mayhem. The Antecedents and Aftermath of Speculative Attack.," *Economic Policy*, 21, 249–312.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987–1008.

——— (1996): "The Econometrics of Ultra-High Frequency Data," Discussion paper, Department of Economics, UCSD.

——— (2000): "The Econometrics of Ultra-High-Frequency Data," *Econometrica*, 68, 1–22.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983): "Exogeneity," *Econometrica*, 51, 277–304.

Engle, R. F., and J. R. Russell (1997): "Forecasting the Frequency of Changes in Quoted Foreign Exchange Prices with the Autoregressive Conditional Duration Model," *Journal of Empirical Finance*, 4, 187–212.

——— (1998): "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.

Epps, T. W., and M. L. Epps (1976): "The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis," *Econometrica*, 44, 305–321.

Estrella, A., and F. S. Mishkin (1997): "The Predictive Power of the Term Structure of Interest Rates in Europe and the United States: Implications for the European Central Bank," *European Economic Review*, 41, 1375–1401.

Foster, F. D., and S. Viswanathan (1990): "A Theory of the Interday Variations in Volume, Variance, and Trading Costs in Security Markets," *Review of Financial Markets*, 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.

Franke, G., and D. Hess (1995): "Anonymous Electronic Trading Versus Floor Trading," *Discussion Paper, University of Konstanz.*

French, K. R., and R. Roll (1986): "Stock Return Variances – The Arrival of Information and the Reaction of Traders," *Journal of Financial Economics*, 17, 5–26.

Gallant, A. R. (1981): "On the Bias in Flexible Functional Forms and an Essential Unbiased Form: The Fourier Flexible Form," *Journal of Econometrics*, 15, 211–245.

Gallant, A. R., P. E. Rossi, and G. Tauchen (1993): "Nonlinear Dynamic Structures," *Econometrica*, 61, 871–907.

Gerhard, F., and N. Hautsch (1999): "Volatility Estimation on the Basis of Price Intensity," Discussion paper, CoFE, Konstanz.

——— (2000a): "Determinants of Inter-Trade Durations and Hazard Rates Using Proportional ARMA Models," Discussion paper, CoFE.

——— (2000b): "The Proportional Hazard ARMA Model," Discussion paper, Center of Finance and Econometrics.

Gerhard, F., D. Hess, and W. Pohlmeier (1998): "What a Difference a Day Makes: On the Common Market Microstructure of Trading Days," Discussion Paper 98/01, CoFE, Konstanz.

Ghysels, E., C. Gourieroux, and J. Jasiak (1998): "Stochastic Volatility Duration Models," Discussion paper, CIRANO.

Ghysels, E., and J. Jasiak (1998): "GARCH for Irregularly Spaced Financial Data: The ACD-GARCH Model," *Studies in Nonlinear Economics and Econometrics*, 2, 133–149.

Goodhart, C. A. E., and M. O'Hara (1997): "High Frequency Data in Financial Markets: Issues and Applications," *Journal of Empirical Finance*, 4, 73–114.

Gottlieb, G., and A. Kalay (1985): "Implications of the Discreteness of Observed Stock Prices," *Journal of Finance*, 40, 135–154.

Gourieroux, C., and J. Jasiak (1999): "Non-Linear Innovations and Impulse Responses," Discussion paper, York University.

Gourieroux, C., J. Jasiak, and G. L. Fol (1999): "Intra-Day Market Activity," *Journal of Financial Markets*, 2, 193–226.

Gourieroux, C., and A. Monfort (1995): *Statistics and Econometric Models*. Cambridge University Press, Cambridge.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon (1987): "Generalised Residuals," *Journal of Econometrics*, 34, 5–32.

Gourieroux, C., A. Monfort, and A. Trognon (1984): "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 52, 681–700.

Grammig, J., R. Hujer, S. Kokot, and K.-O. Maurer (1998): "Modeling the Deutsche Telekom IPO Using a New ACD Specification. An Application of the Burr-ACD Model Using High-Frequency Ibis Data," Discussion paper, Humboldt-Universität, Berlin.

Grammig, J., and K.-O. Maurer (2000): "Non-Monotonic Hazard Functions and the Autoregressive Conditional Duration Model," *Econometrics Journal*, 3(50), 16–38.

Grammig, J., and M. Wellner (1999): "Modeling the Interdependence of Volatility and Inter-Transaction Duration Processes," Discussion Paper 21, Sonderforschungsbereich 373, Humboldt-Universität, Berlin.

Granger, C. W. J., and Z. Ding (1995): "Some Properties of Absolute Return - an Alternative Measure of Risk," *Annales d'Economie et de Statistique*, 40, 67–91.

Hamilton, J. D. (1990): "Analysis of Time Series Subject to Changes in Regime," *Journal of Econometrics*, 45, 39–70.

——— (1994): *Time Series Analysis*. Princeton University Press.

Hamilton, J. D., and O. Jorda (2000): "A Model for the Federal Funds Rate Target," Discussion paper, University of California, San Diego.

Han, A., and J. A. Hausman (1990): "Flexible Parametric Estimation of Duration and Competing Risk Models," *Journal of Applied Econometrics*, 5, 1–28.

Harris, L. (1990): "Estimation of Stock Variances and Serial Covariances from Discrete Observations," *Journal of Financial and Quantitative Analysis*, 25, 291–306.

——— (1991): "Stock Price Clusterings and Discreteness," *Review of Financial Studies*, 4, 389–415.

Harris, L. E. (1994): "Minimum Price Variations, Discrete Bid-Ask Spreads, and Quotation Sizes," *Review of Financial Studies*, 7, 149–178.

Harvey, A., E. Ruiz, and N. Shephard (1994): "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264.

Harvey, A. C. (1989): *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

——— (1990): *The Econometric Analysis of Time Series*. Philip Allan, Hemel Hempstead, 2nd. edn.

Harvey, A. C., and C. Fernandes (1989): "Time Series Models for Count or Qualitative Observations," *Journal of Business and Economic Statistics*, 7, 407–417.

Hasbrouck, J. (1991): "Measuring the Information Content of Stock Trades," *Journal of Finance*, 66, 179–206.

Hausman, J. A., A. W. Lo, and A. C. MacKinlay (1992): "An Ordered Probit Analysis of Transaction Stock Prices," *Journal of Financial Economics*, 31, 319–379.

Hautsch, N. (1999): "Analyzing the Time Between Trades with a Gamma Compounded Hazard Model. An Application to LIFFE Bund Future Transctions," Discussion Paper 99/03, CoFE, Konstanz.

Hawkins, D. L. (1987): "A Test for a Change Point in a Parametric Model Based on a Maximal Wald-Type Statistic," *Sankhya*, 49, 368–376.

Hendry, D. F. (1995): *Dynamic Econometrics*. Oxford University Press, Oxford.

Horowitz, J. L. (1999): "Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity," *Econometrica*, 67, 1001–1028.

Jasiak, J. (1999): "Persistence in Intertrade Durations," Discussion paper, York University.

Jazwinski, A. H. (1970): *Stochastic Processes and Filtering Theory*. Academic Press, New York.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T.-C. Lee (1985): *The Theory and Practice of Econometrics*. Wiley, Chichester.

Karpoff, J. (1987): "The Relation Between Price Changes and Trading Volume," *Journal of Quantitative Analysis*, 22, 109–126.

Kedem, B. (1980): "Estimation of the Parameters in Stationary Autoregressive Processes After Hard Limiting," *Journal of the American Statistical Association*, 75, 146–153.

Keenan, D. M. (1982): "A Time Series Analysis of Binary Data," *Journal of the American Statistical Association*, 77, 816–821.

KODDE, D. A., F. C. PALM, AND G. A. PFANN (1990): "Asymptotic Least-Squares Estimation Efficiency Considerations and Applications," *Journal of Applied Econometrics*, 5, 229–243.

KOOP, G., M. H. PESARAN, AND S. M. POTTER (1996): "Impulse Response Analysis in Nonlinear Multivariate Models," *Journal of Econometrics*, 74, 119–147.

KYLE, A. (1985): "Continous Auctions and Insider Trading," *Econometrica*, 22, 477–498.

LANCASTER, T. (1994): *The Econometric Analysis of Transition Data*. Cambridge University Press, New York.

LEE, M.-J. (1992): "Median Regression for Ordered Discrete Response," *Jounral of Econometrics*, 51, 59–77.

LEE, S., AND B. HANSEN (1994): "Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator," *Econometric Theory*, 10, 29–52.

LO, A. W., AND A. C. MACKINLAY (1999): *A Non-Random Walk Down Wall Street*. Princeton University Press, Princeton, New Jersey.

LOMNICKI, Z. A., AND S. K. ZAREMBA (1955): "Some Applications of Zero-One Processes," *Journal of the Royal Statistical Society, Series B*, 17, 243–255.

LUMSDAINE, R. (1996): "Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models," *Econometrica*, 64, 575–596.

LÜTKEPOHL, H. (1991): *Introduction to Multiple Time Series Analysis*. Springer, New York.

MADDALA, G. S. (1983): *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge.

MADHAVAN, A. (2000): "Market Microstructure: A Survey," Discussion paper, Marshall School of Business, University of Southern California.

MANRIQUE, A., AND N. SHEPHARD (1998): "Simulation-Based Likelihood Inference for Limited Dependent Processes," *Econometrics Journal*, 1, C174–C202.

MCCULLAGH, P., AND J. A. NELDER (1989): *Generalized Linear Models*. Chapman and Hall, London.

MCKELVEY, R., AND W. ZAVOINA (1975): "A Statistical Model for the Analysis of Ordinal Level Dependent Variables," *Journal of Mathematical Sociology*, 4, 103–120.

MERTON, R. C. (1980): "On Estimating the Expected Return on the Market - an Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.

MEYER, B. (1990): "Unemployment Insurance and Unemployment Spells," *Econometrica*, 58, 757–782.

NELDER, J. A., AND R. W. M. WEDDERBURN (1972): "Generalized Linear Models," *Journal of the Royal Statistical Society, Series A*, 135, 370–384.

NEWEY, W. K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837.

——— (1993): "Efficient Estimation of Models with Conditional Moment Restrictions," in *Handbook of Statistics*, ed. by G. S. Maddala, C. R. Rao, and H. D. Vinod, pp. 419–454. Elsevier Science.

NEWEY, W. K., AND K. D. WEST (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703–708.

NIEDERHOFER, V. (1965): "Clustering of Stock Prices," *Operations Research*, 13, 258–265.

——— (1966): "A New Look at Clustering of Stock Prices," *Journal of Business*, 39, 309–313.

O'HARA, M. (1995): *Market Microstructure Theory.* Blackwell Publishers, Oxford.

PAGAN, A., AND A. ULLAH (1999): *Nonparametric Econometrics.* Cambridge University Press, Cambridge.

POHLMEIER, W. (1989): *Simultane Probit- und Tobitmodelle.* Springer, Berlin, Heidelberg.

POHLMEIER, W., AND F. GERHARD (2000): "Identifying Intraday Volatility," Discussion paper, Center of Finance and Econometrics.

POIRIER, D. L., AND P. A. RUUD (1988): "Probit with Dependent Observations," *Review of Economic Studies*, 55, 593–614.

RESNICK, S. I. (1992): *Adventures in Stochastic Processes.* Birkhäuser.

ROLL, R. (1984): "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market," *Journal of Finance*, 39, 1127–1139.

RUIZ, E. (1994): "Quasi-Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Econometrics*, 63, 289–306.

RUSSELL, J. R., AND R. F. ENGLE (1998): "Econometric Analysis of Discrete-Valued Irregularly-Spaced Financial Transactions Data Using a New Autoregressive Conditional Multinominal Model," *Working Paper of the University of California.*

RYDBERG, T. H., AND N. SHEPHARD (1998): "Dynamics of Trade-by-Trade Price Movements: Decomposition and Models," Discussion paper, Nuffield College, Oxford.

SHEPHARD, N. (1995): "Generalized Linear Autoregressions," Discussion paper, Nuffield College, Oxford.

SMITH, R. J., AND R. W. BLUNDELL (1986): "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply," *Econometrica*, 54, 679–685.

TAUCHEN, G. E., AND M. PITTS (1983): "The Price Variability-Volume Relationship on Speculative Markets," *Econometrica*, 51, 485–505.

WATSON, M. W. (1989): "Recursive Solution Methods for Dynamic Linear Rational Expectations Models," *Journal of Econometrics*, 41, 65–89.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

——— (1994): *Estimation, Inference and Specification Analysis*. Cambridge University Press, Cambridge.

Wood, R. A., T. H. McInish, and J. K. Ord (1985): "An Investigation of Transactions Data for NYSE Stocks," *Journal of Finance*, 40, 723–741.

Zeger, S. L., and B. Qaqish (1988): "Markov Regression Models for Time Series: A Quasi-Likelihood Approach," *Biometrics*, 44, 1019–1031.