Modeling a Multivariate Transaction Process

INGMAR NOLTE
University of Konstanz, CoFE

ABSTRACT

In this paper the dynamics of a joint transaction process are investigated. The transaction process is characterized by four marks: price changes, transaction volumes, bid–ask spreads and intertrade durations. Based on a copula approach, a model for their joint density is proposed, which avoids forcing a priori assumptions on the instantaneous causality relationships between the four variables as necessary in decomposition models, where the joint density is decomposed into its conditional and unconditional densities. The price change process is treated as a discrete process and specified with an integer count hurdle model and the transaction volumes, bid–ask spreads, and trade durations processes are modeled along the lines of fractionally integrated autoregressive conditional models, which are suited very well to capture the high persistency, empirically observed in these processes. The model is applied to three stocks traded at the New York Stock Exchange (NYSE) in May, 2001 and we investigate several market microstructure hypotheses in the empirical part of this paper.

KEYWORDS: copula functions, discrete price changes, fractionally integrated autoregressive conditional duration models, integer count hurdle model, market microstructure, transaction data.

1 Introduction

The empirical analysis of market microstructure (MMS) variables allows us to infer how market participants process information, submit orders with a specific volume at a specific time, interact with each other, and affect the price process of an asset. The results from these analyses are of utmost importance for the efficient design of the institutional setup and the trading systems in financial markets. Moreover, they are on one hand the basis for the development of theoretical MMS models and on the other they allow us to verify existing models empirically.

The importance being attributed to empirical MMS research is underpinned by an enormous amount of very well written papers in this field of research in the last two decades. Engle (1982), Bollerslev (1986), Hausman, Lo, and MacKinlay (1992), Hasbrouck and Sofianos (1993), O'Hara (1995), Engle and Russell (1998), Engle (2000), Dufour and Engle (2000), Ghysels, Gourieroux, and Jasiak (2004) represent only a tiny part thereof.

However, most of the existing studies do not analyze relationships between several MMS variables based on a joint model for these variables. Instead, the relationship between a set of MMS variables, such as price changes, volatility, transaction volumes, bid–ask spreads and intertrade durations is analyzed in univariate models; see Hausman, Lo, and MacKinlay (1992) and Dufour and Engle (2000) for example, or in decomposition models, see Russell and Engle (2002), Rydberg and Shephard (2003) and Manganelli (2005). In these models, one of the MMS variables is treated as the left-hand variable and the rest as covariates assuming an a priori instantaneous causality relationship between these variables or imposing a strict exogeneity assumption in the sense of Engle, Hendry, and Richard (1983). An exception is the study by Renault and Werker (2006), where the instantaneous causality effect between volatility and trade arrival times is disentangled from the Granger causality effects between these variables.

In this paper, we pick up the suggestion of Dufour and Engle (2000) and formalize a model for the joint system of price changes, transaction volumes, bid–ask spreads and intertrade durations, which, from our point of view, are four of the most important MMS variables. We rely on a copula approach to separate instantaneous causality effects from Granger causality effects and we avoid decomposing the joint process into conditional and marginal processes and postulating a particular instantaneous causality scheme between the four MMS variables a priori. Moreover, in the specification of the marginal processes, we try to take the most important MMS characteristics such as discreteness of price changes as well as the high persistency of transaction volumes, bid–ask spreads and intertrade durations into account. We use the Liesenfeld, Nolte, and Pohlmeier (2006) dynamic integer count hurdle (ICH) model for the specification of the discrete price change process, whereas the processes for transaction volumes, bid–ask spreads, and intertrade durations are modeled with fractionally integrated autoregressive conditional duration (FIACD) type models proposed by Jasiak (1999). The application of ACD-type models to transaction volumes and bid–ask spreads is unproblematic since by definition both variables have a positive outcome space. For transaction volumes,

ACD-type models have already been proposed by Manganelli (2005). Since, we are considering a four-dimensional process in which one component is discrete (price changes) and the remaining three components are continuous, we rely for the price change variable on the continuation approach suggested by Stevens (1950), Heinen and Rengifo (2007) and Denuit and Lambert (2005) during the specification of the copula function.

The empirical analysis is carried out, using data from the NYSE Trade and Quote (TAQ2) database, for three stocks: Black & Decker Corporation (BDK), International Business Machines (IBM), and Coca-Cola Company (KO), traded at the NYSE for the period from May 1, 2001 to May 31, 2001. We can show that the proposed model specification fits the dynamics of the joint process very well and we apply our model to verify several market microstructure hypotheses. We find, that the role of trade arrival times in explaining bid–ask spreads (Foster and Viswanathan (1990)) and price change volatility (Easley and O'Hara (1992), Dufour and Engle (2000)) is diminished and ambiguous, when controlling for the information conveyed by the transaction volume and the bid–ask spread processes.

The rest of the paper is structured in the following way: Section 2 presents the model in detail, Section 3 contains the empirical analysis as well as the application to market microstructure hypotheses, and Section 4 concludes the paper.

2 Multivariate Modeling

Let $P_t \in \varsigma \cdot \mathbb{Z}$ denote the transaction price at the tth trade, where $\varsigma \in \mathbb{R}^+$ denotes the tick size and $t = (0), 1, \ldots, T$. The associated (standardized) price change from the $(t-1)$th to the tth trade is then given by $C_t \equiv (P_t - P_{t-1})/\varsigma \in \mathbb{Z}$, the volume traded at the tth trade is denoted by $V_t \in \mathbb{R}^+$, the bid–ask spread at the tth trade is denoted by $S_t \in \mathbb{R}^+$, and the duration between the $(t-1)$th and the tth trade is denoted by $D_t \in \mathbb{R}^+$.

We collect these marks of the transaction process in the vector $M_t \in \mathbb{Z} \times \mathbb{R}^{+3}$:

$$M_t \equiv (C_t, V_t, S_t, D_t)',$$

we set $\mathfrak{F}_t \equiv \sigma(M_s | s \leq t)$ and denote θ as the generic overall parameter vector. Our aim is to model the conditional joint density of M_t denoted by $f_{M_t}(m_t | \mathfrak{F}_{t-1}; \theta)$ within a parametric framework. We want to avoid a specification, where the joint density is decomposed into a product of sequential conditional densities and a marginal density, since such a specification imposes a specific form on the instantaneous causality relationship of the variables under consideration. Therefore we decide to rely on a copula approach, which allows for a direct investigation of the instantaneous relationships in our system, rather than simply choosing that specification which seems to be the most reasonable according to our current market microstructure knowledge. Choosing the copula approach, we are, in particular, able to manage and investigate different instantaneous relationship patterns separately from their Granger causality effects between our variables across stocks.

The joint distribution function of M_t denoted by[1] $F_{M_t}(m_t|\mathfrak{F}_{t-1})$ can be related to its marginal distributions using the copula function $C : [0, 1]^4 \to [0, 1]$ as

$$F_{M_t}(m_t|\mathfrak{F}_{t-1}) = C\big(F_{C_t}(c_t|\mathfrak{F}_{t-1}), F_{V_t}(v_t|\mathfrak{F}_{t-1}), F_{S_t}(s_t|\mathfrak{F}_{t-1}), F_{D_t}(d_t|\mathfrak{F}_{t-1})\big),$$

where F_{C_t}, F_{V_t}, F_{S_t} and F_{D_t} denote the marginal distribution functions of the price change, the volume, the bid–ask spread and the duration, respectively. Sklar (1959) proved the existence of the copula function C and showed its uniqueness on $[0, 1]^k$ for the case where all k variables of the joint distribution function are continuous. In that case, one can take the derivative of the joint distribution function with respect to its components to obtain a valid specification for the joint density.

Also, according to Sklar (1959) in the case where certain components of the joint density are discrete, as C_t in our case, we do not achieve uniqueness of the copula function, in our case on $[0, 1]^4$, but only on Range(F_{C_t}) $\times [0, 1]^3$. The joint density function can then be obtained either by approximating the derivative of $F_{M_t}(m_t|\mathfrak{F}_{t-1})$ with respect to C_t by a finite difference as proposed by Meester and MacKay (1994) and Cameron, Li, Trivedi, and Zimmer (2004) or by artificially continuing C_t and computing the usual derivative based on the continued C_t as suggested by Stevens (1950), Heinen and Rengifo (2007) and Denuit and Lambert (2005).

In this paper, we rely on the latter strategy and we continue the discrete variable C_t, following Denuit and Lambert (2005) with the help of U_t being independent uniformly $U(0, 1)$ distributed, by setting

$$C_t^* \equiv C_t - U_t,$$

where $C_t^* \in \mathbb{R}$ denotes the continued price change variable. With M_t^* being the vector M_t, where C_t has been replaced by C_t^*, and F_{C^*} being the distribution function of C_t^*, we can obtain the joint density $f_{M_t^*}(m_t^*|\mathfrak{F}_{t-1}) = (\partial^4 F_{M_t^*}(m_t^*|\mathfrak{F}_{t-1}))/\partial C_t^*, \partial V_t, \partial S_t, \partial D_t$ as

$$f_{M_t^*}(m_t^*|\mathfrak{F}_{t-1}) = f_{C_t^*}(c_t^*|\mathfrak{F}_{t-1}) \cdot f_{V_t}(v_t|\mathfrak{F}_{t-1}) \cdot f_{S_t}(s_t|\mathfrak{F}_{t-1}) \cdot f_{D_t}(d_t|\mathfrak{F}_{t-1})$$
$$c\big(F_{C_t^*}(c_t^*|\mathfrak{F}_{t-1}), F_{V_t}(v_t|\mathfrak{F}_{t-1}), F_{S_t}(s_t|\mathfrak{F}_{t-1}), F_{D_t}(d_t|\mathfrak{F}_{t-1})\big), \qquad (1)$$

where c denotes the density of the copula function C. Since, we will specify a model for C_t, we need the relationship between the discrete distribution function of C_t and the continuous distribution function of C_t^* for the computation of the copula density. Since $C_t^* \le C_t$ a.s. and the integer part $\lfloor C_t^* \rfloor$ of the continuous variable C_t^* is given by $\lfloor C_t^* \rfloor = C_t - 1$ we obtain $F_{C_t^*}(c_t^*)$ as

$$
\begin{aligned}
F_{C_t^*}(c_t^*) = P(C_t^* \le c_t^*) &= P(C_t^* \le \lfloor c_t^* \rfloor) + P(\lfloor c_t^* \rfloor < C_t^* \le c_t^*) \\
&= P(C_t^* \le c_t - 1) + P(c_t - 1 < C_t^* \le c_t^*) \\
&= P(C_t \le c_t - 1) + P(c_t - 1 < C_t - U_t \le c_t - u_t) \\
&= P(C_t \le c_t - 1) + P(U_t \le u_t) \cdot P(C_t = c_t) \\
&= F_{C_t}(c_t - 1) + u_t \cdot f_{C_t}(c_t),
\end{aligned}
$$

---

[1]For the ease of notation we suppress the parameter vector θ.

where the last equation follows from the fact that for $U_t \sim U(0, 1)$, $P(U_t \leq u_t) = u_t$. Moreover, we used the relationship that $P(C_t^* \leq c_t - 1) = P(C_t \leq c_t - 1)$ since c_t is discrete. Please note, that the specification in Equation (1) is based on a fixed or unconditional copula C in that sense, that the copula function itself is not assumed to be time-varying or dependent on $\mathfrak{F}_t$. Such an extension, which has been proposed by Patton (2001 and 2006), would increase the computational burden in our model considerably.

2.1 Specification of the Marginal Densities and the Copula Function

We now address the specification of the marginal densities of the price change, the volume, the bid–ask spread, and the duration processes as well as the choice of the copula function.

2.1.1 Price change process We apply the ICH model of Liesenfeld, Nolte, and Pohlmeier (2006), to model the discrete density of the price change

$$f_{C_t}(c_t|\mathfrak{F}_{t-1}).$$

The idea of the ICH model is to decompose the price change process into two components: (i) a direction process and (ii) a size or an absolute price change process given a nonzero price change.

Let π_{jt}, $j \in \{-1, 0, 1\}$ denote the conditional probability of a negative $P(C_t < 0|\mathfrak{F}_{t-1})$, a zero $P(C_t = 0|\mathfrak{F}_{t-1})$ or a positive price change $P(C_t > 0|\mathfrak{F}_{t-1})$ at time t. The conditional density of a price change is then specified as

$$f_{C_t}(c_t|\mathfrak{F}_{t-1}) = \pi_{-1t}{}^{\mathbf{1}_{\{C_t<0\}}} \cdot \pi_{0t}{}^{\mathbf{1}_{\{C_t=0\}}} \cdot \pi_{1t}{}^{\mathbf{1}_{\{C_t>0\}}} \cdot f_{|C_t|}(|c_t| \, |C_t \neq 0, \mathfrak{F}_{t-1})^{\left(1-\mathbf{1}_{\{C_t=0\}}\right)},$$

where $f_{|C_t|}(|c_t| \, |C_t \neq 0, \mathfrak{F}_{t-1})$ denotes the conditional density of an absolute price change, with support $\mathbb{N} \setminus \{0\}$.

To obtain a parsimoniously specified model, we adopt the simplification of Liesenfeld, Nolte, and Pohlmeier (2006), that the conditional density of an absolute price change stems from the same distribution irrespective of whether it is an upward or a downward price change.

2.1.1.1 Direction process In order to model the conditional probabilities of the direction process, we use the autoregressive conditional multinomial (ACM) model of Russell and Engle (2002) with a logistic link function, given by

$$\pi_{jt} = \frac{\exp(\Lambda_{jt})}{\sum_{j=-1}^{1} \exp(\Lambda_{jt})} \tag{2}$$

with normalizing constraint $\Lambda_{0t} = 0$, $\forall t$. The resulting vector of log-odds ratios $\Lambda_t \equiv (\Lambda_{-1t}, \Lambda_{1t})' = (\ln[\pi_{-1t}/\pi_{0t}], \ln[\pi_{1t}/\pi_{0t}])'$ is specified as a multivariate ARMA-type model:

$$(I_2 - \beta_p(L))(\Lambda_t - \zeta_r(L)\ln(Z_t)) = \mu + \alpha_q(L)\varepsilon_t + \gamma_s(L)|\varepsilon_t|, \tag{3}$$

where $\alpha_q(L)$, $\beta_p(L)$, and $\gamma_s(L)$ denote 2×2 matrix valued lag polynomials of order q, p, and s, respectively. $Z_t = (V_t, S_t, D_t)'$ denotes the vector of explanatory variables—volume, bid–ask spread, and duration—which are included statically and in lagged form through the 2×3 matrix valued lag polynomial $\zeta_r(L)$ of order r in this specification. μ denotes the 2×1 vector of constants, and the innovation vector of the ARMA model is specified as martingale differences given by

$$\varepsilon_t \equiv (\varepsilon_{-1t}, \varepsilon_{1t})', \quad \text{where} \quad \varepsilon_{jt} \equiv \frac{x_{jt} - \pi_{jt}}{\sqrt{\pi_{jt}(1 - \pi_{jt})}}, \quad j \in \{-1, 1\}, \tag{4}$$

and

$$x_t \equiv (x_{-1t}, x_{1t})' = \begin{cases} (1, 0)' & \text{if } C_t < 0 \\ (0, 0)' & \text{if } C_t = 0 \\ (0, 1)' & \text{if } C_t > 0, \end{cases} \tag{5}$$

denotes the 2×1 state vector indicating the direction of the price movement at time t. Thus, ε_t represents the standardized state vector x_t. Note, that we included also the absolute innovation term $|\varepsilon_t|$ to capture asymmetries in the news impact curve for the log odds ratios.

2.1.1.2 Absolute price change process The conditional density of the absolute price process is modeled with an at-zero-truncated Negative Binomial (NegBin) distribution, given by

$$f_{|C_t|}(|c_t| | C_t \neq 0, \mathfrak{F}_{t-1}) \equiv \frac{\Gamma(\kappa + |c_t|)}{\Gamma(\kappa)\Gamma(|c_t| + 1)} \left(\left[\frac{\kappa + \omega_t}{\kappa} \right]^\kappa - 1 \right)^{-1} \left(\frac{\omega_t}{\omega_t + \kappa} \right)^{|c_t|}, \tag{6}$$

where $|c_t| \in \mathbb{N} \setminus \{0\}$, $\kappa > 0$ denotes the dispersion parameter and ω_t is parameterized using the logarithmic link function with a generalized autoregressive moving average specification (GLARMA) of Shephard (1995) in the following way:

$$(1 - \tilde{\beta}_p(L))(\ln \omega_t - \tilde{\delta}' \tilde{D}_t - \tilde{\zeta}_r(L) \ln(Z_t)) = \tilde{\mu} + S(\nu, \tau, K) + \tilde{\alpha}_q(L)\tilde{\varepsilon}_t + \tilde{\gamma}_s(L)|\tilde{\varepsilon}_t|, \tag{7}$$

where $\tilde{D}_t \in \{-1, 1\}$ indicates a negative or a positive price change at time t with corresponding coefficient $\tilde{\delta}$. $Z_t = (\tilde{D}_t, V_t, S_t, D_t)'$ denotes again the vector of further explanatory variables, with associated 1×4 dimensional parameter lag polynomial $\zeta_r(L)$. $\tilde{\alpha}_q(L)$, $\tilde{\beta}_p(L)$, and $\tilde{\gamma}_s(L)$ denote scalar lag polynomials, $\tilde{\mu}$ the constant and

$$S(\nu, \tau, K) \equiv \nu_0 \tau + \sum_{k=1}^{K} \nu_{2k-1} \sin(2\pi(2k-1)\tau) + \nu_{2k} \cos(2\pi(2k)\tau) \tag{8}$$

a Fourier flexible form to capture intraday seasonality in the absolute prices changes, which can be considered as a measure for volatility, where τ is the intraday trading time standardized to $[0, 1]$ and ν is a $2K + 1$ dimensional parameter vector. In the spirit of Nelson (1991) for GARCH and Dufour and Engle (2000) for ACD models, we include again an absolute innovation term $|\tilde{\varepsilon}_t|$ to allow for an

asymmetric news response of $\ln \omega_t$. $\tilde{\varepsilon}_t$ constructed as

$$\tilde{\varepsilon}_t \equiv \frac{|C_t| - E(|C_t| \, |C_t \neq 0, \mathfrak{F}_{t-1})}{\mathrm{Var}(|C_t| \, |C_t \neq 0, \mathfrak{F}_{t-1})^{1/2}},$$

is the innovation term that drives the GLARMA model. For a more elaborate presentation of the ICH model and its components as well as for some stationarity considerations we refer the reader to Liesenfeld, Nolte, and Pohlmeier (2006).

2.1.2 Transaction volume, bid–ask spread and trade duration processes

Since we consider transaction volume, bid–ask spread, and trade duration as variables with a positive real domain, we specify their conditional densities in a similar way and we decide to present their models in a unified notation with $Y_t \in \{V_t, S_t, D_t\}$. Thus, we are concerned with the conditional density

$$f_{Y_t}(Y_t | \mathfrak{F}_{t-1}),$$

and we specify $f_{Y_t}(Y_t | \mathfrak{F}_{t-1})$ within the autoregressive conditional duration (ACD) framework introduced by Engle and Russell (1998). Manganelli (2005) already showed that the transaction volume process can be reasonably modeled using ACD specifications and we apply this approach to the bid–ask spread process as well. There are several extensions to the original ACD model, see for example Bauwens, Giot, Grammig, and Veredas (2004), Lunde (2000), Grammig and Maurer (2000), and Bauwens and Hautsch (2006). Here, we rely on the FIACD model of Jasiak (1999), which we augment by a multiplicative function to capture intraday seasonality.

Thus, we assume that the variable Y_t consists multiplicatively of a seasonality function $s(v, \tau, K)$, a conditional mean function $\varphi_t(\theta_Y | \mathfrak{F}_{t-1})$, and an error term ε_t:

$$Y_t = s(v, \tau, K) \cdot \varphi_t(\theta_Y | \mathfrak{F}_{t-1}) \cdot \varepsilon_t, \quad \varepsilon_t \sim i.i.d. \tilde{f}_{Y_t}(\cdot),$$

where $\tilde{f}_{Y_t}(\cdot)$ is an error term density with unit mean. Assuming $\tilde{f}_{Y_t}(\cdot)$ to be independent of the conditioning information $\mathfrak{F}_{t-1}$ facilitates modeling, since we do not need to be concerned with higher conditional moments. An extension that allows for separate dynamics in the conditional variance is presented by Ghysels, Gourieroux, and Jasiak (2004). Applying the transformation theorem[2] yields

$$Y_t \sim \frac{1}{s\varphi_t} \tilde{f}_{Y_t}\left(\frac{y_t}{s\varphi_t}\right), \tag{9}$$

where $s \equiv s(v, \tau, K)$ and $\varphi_t \equiv \varphi_t(\theta_Y | \mathfrak{F}_{t-1})$. To ensure a positive seasonality function we assume $s(v, \tau, K) = \exp(S(v, \tau, K))$ where $S(v, \tau, K)$ follows a Fourier flexible form as stated in Equation (8). We specify $\tilde{f}_{Y_t}(\cdot)$ as an exponential density with unit mean, i.e., $\tilde{f}_{Y_t}(\cdot) \sim \mathrm{Exp}(1)$ and the dynamics of the conditional mean function are

---

[2]See, e.g., Rohatgi (1976), p. 135, Theorem 6.

modeled according to Jasiak (1999) as

$$(1 - \beta_p(L))\varphi_t = \mu + \gamma_\infty(L)Y_t, \qquad (10)$$

with a constant μ and an infinite-dimensional scalar lag polynomial $\gamma_\infty(L)$, given by

$$\gamma_\infty(L) = [1 - \beta_p(L) - [1 - \alpha_q(L) - \beta_p(L)](1 - L)^d],$$

where $\alpha_q(L)$ and $\beta_p(L)$ denote scalar lag polynomials and $(1 - L)^d$, $0 < d < 1$ the fractional differencing operator given by

$$(1 - L)^d = \sum_{k=0}^{\infty} \varpi_k L^k,$$

with

$$\varpi_k = \frac{\Gamma(k - d)}{\Gamma(k + 1)\Gamma(-d)} = \prod_{0 < j \le k} \frac{j - 1 - d}{j}, \quad k = 0, 1, 2, \ldots$$

and $\Gamma(\cdot)$ the gamma function defined as

$$\Gamma(x) \equiv \begin{cases} \int\limits_0^\infty t^{x-1} \exp(-t)dt & \text{if } x > 0, \\ \infty & \text{if } x = 0, \\ x^{-1}\Gamma(1 + x) & \text{if } x < 0. \end{cases}$$

All coefficients in $\beta_p(L)$ and $\alpha_q(L)$ as well as μ have to be nonnegative to ensure positivity of the conditional mean function (out-of-sample). Jasiak (1999) shows that the FIACD(p, d, q) model is strictly stationary and ergodic for $0 \le d \le 1$, but not weakly stationary, since the first unconditional moment of Y_t is infinite, due to the fact that the fractional differencing operator evaluated at lag $L = 1$ is 0. The FIACD(p, d, q) class nests the classes of ACD(p, q) models for $d = 0$ and their integrated counterparts for $d = 1$.

An important point in the estimation is, that we have only a finite sample of data and therefore the "∞" in $(1 - L)^d = \sum_{k=0}^\infty \varpi_k L^k$ needs to be approximated and the preceding data points for the initialization need to be set. We set "$\infty = 1000$" and initiated the foregoing 1000 lags of Y_t with the unconditional mean of Y_t. Applying this approximation, we can consider the FIACD(p, d, q) models as ACD(p,1000+max(p,q)) models, with parameter restrictions of a specific functional form depending on d.

2.1.3 Copula function Let us recall Equation (1), which states the joint conditional density of the trading marks vector M_t^*:

$$f_{M_t^*}(m_t^*|\mathfrak{F}_{t-1}) = f_{C_t^*}(c_t^*|\mathfrak{F}_{t-1}) \cdot f_{V_t}(v_t|\mathfrak{F}_{t-1}) \cdot f_{S_t}(s_t|\mathfrak{F}_{t-1}) \cdot f_{D_t}(d_t|\mathfrak{F}_{t-1})$$
$$c\big(F_{C_t^*}(c_t^*|\mathfrak{F}_{t-1}), F_{V_t}(v_t|\mathfrak{F}_{t-1}), F_{S_t}(s_t|\mathfrak{F}_{t-1}), F_{D_t}(d_t|\mathfrak{F}_{t-1})\big).$$

In this equation, the copula density c is specified as a four-dimensional Gaussian copula density, given by:

$$c(y_{1t}, y_{2t}, y_{3t}, y_{4t}; \Sigma) = \det(\Sigma)^{-0.5} \exp\left(\frac{1}{2}q_t'(I_4 - \Sigma^{-1})q_t\right), \qquad (11)$$

where $\Sigma = (\rho_{ij})$ denotes the 4×4 correlation matrix of $q = (q_{1t}, q_{2t}, q_{3t}, q_{4t})'$ with $q_{it} = \Phi^{-1}(y_{it})$, $i = 1, \ldots, 4$. Thus, for example, the argument y_{1t} of the copula function c is the probability mass in the left tail of the conditional price change distribution $F_{C_t^*}(\cdot|\mathfrak{F}_{t-1})$ up to the observed (continued) price change c_t^* at time t; and q_{1t} is the quantile of the standard normal distribution associated with that left tail probability. In that sense q_{1t} represents a kind of "normalized" price change observation c_t^* where the ordering of the observations is not interchanged since $F_{C_t^*}(\cdot|\mathfrak{F}_{t-1})$ as well as $\Phi(\cdot)$ are strictly monotonically increasing. The observations v_t, s_t, and d_t are "normalized" to q_{2t}, q_{3t}, and q_{4t} analogously and the correlation matrix Σ represents the correlation between these "normalized" variables. We redefine Σ in the following way:

$$\Sigma = (\rho_{ij}) = \begin{pmatrix} 1 & \rho_{CV} & \rho_{CS} & \rho_{CD} \\ \rho_{CV} & 1 & \rho_{VS} & \rho_{VD} \\ \rho_{CS} & \rho_{VS} & 1 & \rho_{SD} \\ \rho_{CD} & \rho_{VD} & \rho_{SD} & 1 \end{pmatrix}, \qquad i, j = C, V, S, D, \qquad (12)$$

to facilitate an intuitive interpretation of the parameters in Σ with respect to the relations between price changes (C), transaction volumes (V), bid–ask spreads (S), and intertrade durations (D).

3 Empirical Analysis

3.1 Database

In the empirical analysis we use tick-by-tick data from May 1, 2001 to May 31, 2001 of three stocks traded at the NYSE: BDK, IBM, and KO. The data stems from the Trade and Quote (TAQ2) database, which is separated into two files: the trade database and the quote database.

The trade database contains all transaction prices and volumes and the quote database consists of all bid and ask quotes and depths, timestamped to the second. To determine, which bid and ask quotes (bid–ask spread) were valid at a certain trade observation and whether this trade was a buy or a sell, one has to merge the two databases. The common algorithm that has been applied predominantly in the literature is the Lee and Ready (1991) procedure, which relies on a so-called "five-seconds rule," which means that each trade is assigned to the quotes posted at least 5 seconds before. The identification of a buy or sell is done in the following way: If the transaction price is above (below) the midquote, the trade is defined as a buy (sell); for transaction prices at the midquote, the tick rule applies, i.e., if the

transaction takes place at a higher (lower) price than the price of the most recent trade with a different price, the trade is characterized as a buy (sell).

In the last few years this algorithm has been criticized, e.g., by Boehmer, Grammig, and Theissen (2007) and Vergote (2005) concerning the time span of 5 seconds, which does not seem to be appropriate anymore, due to advances in computer technology. Therefore, we check for which lag value in seconds (between 0 and 5) the number of the transaction prices corresponding exactly to previous quotes is maximized and it turns out that for all three stocks, this occurs at 1 second. Thus, we apply the Lee and Ready (1991) algorithm in a modified way using a delay of 1, instead of 5 seconds only.

A further problem arises due to the fact that in some cases there are several trades at exactly the same timestamp. This can happen due to an automatic matching of different orders on one side of the specialist's book against a larger order on the other side (split-transaction). Moreover, such transactions can also result from different market participants, who posted their orders (electronically) within 1 second or, as pointed out by Veredas, Rodriguez-Poo, and Espasa (2002), by limit orders of different market participants with exactly the same limit, e.g., at round prices. Unfortunately, these differences cannot be identified with the TAQ2 database. We have treated all trades as split transactions, which were recorded at the same second. In this case, we have simply aggregated their volume to one transaction and assigned the last price in the sequence to the aggregated transaction. Furthermore, we have removed all trades outside the regular trading hours as well as each day's first trade, to circumvent contamination due to the opening call auction at the NYSE.

3.2 Descriptive Statistics

In Figure 1 we have plotted the histograms and the autocorrelograms up to lag 50 for the tick-by-tick price changes of BDK, KO, and IBM. With a mean duration of 53.1, 14.8, and 6.9 seconds, respectively, a lag of 50 corresponds approximately to 44, 12, and 6 minutes.

The histograms show a fairly large support; most of the mass is concentrated between -7 and 7 ticks, but for BDK and IBM even the classes ± 8, ± 9, and ± 10 still possess a mass of around 1% each. The 0 tick classes have frequencies between 40% and 50% and ± 1 tick classes take frequencies between 10% and 15%. The large support in combination with the high concentration at the 0 tick class, justifies the application of the discrete ICH model. An alternative approach to model the discrete price change process would be the decomposition model of Rydberg and Shephard (2003), which is also capable of modeling discrete outcomes with a fairly large support. The ordered probit model of Hausman, Lo, and MacKinlay (1992) or the multinomial model of Russell and Engle (2002), however, suffer from the drawback that they can only model reliably discrete outcomes with a bounded support. The autocorrelograms of IBM and KO exhibit the usual negative first-order autocorrelation coefficient, which can be explained by the bid–ask bounce
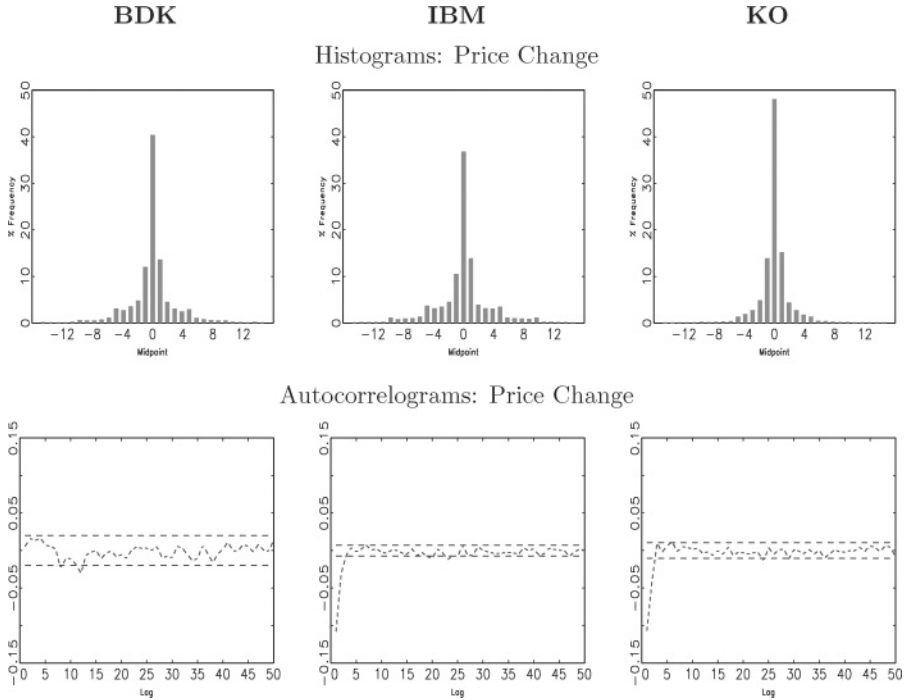
Figure 1 Histograms (first row) and autocorrelograms (second row) of BDK, IBM, and KO price changes. Confidence bands: asymptotic 95%.

effect examined by Roll (1984). For BDK, however, such an effect cannot be observed and all autocorrelation coefficients are basically insignificant.

Figure 2 depicts the autocorrelograms, again up to lag 50, of transaction volumes, bid–ask spreads, and intertrade durations for all three stocks. Except the transaction volume of BDK, all autocorrelation functions have in common that even the autocorrelation coefficient at lag 50 is still significantly different from zero. The decay of the empirical autocorrelation functions for transaction volumes and intertrade durations is very slow, whereas for the bid–ask spread it is quite fast until, say, lag 10, but from thereon again very slow. Because of these observations, we do not consider the decay of these empirical autocorrelation functions to be exponential and therefore we rely on the fractionally integrated specification for the dynamics of the mean function of transaction volume, bid–ask spread, and duration.

3.3 Estimation Results

All estimation results are obtained by jointly maximizing the log-likelihood implied by Equation (1). After a careful model selection procedure, we decide to model the conditional mean functions for all four variables of all three stocks
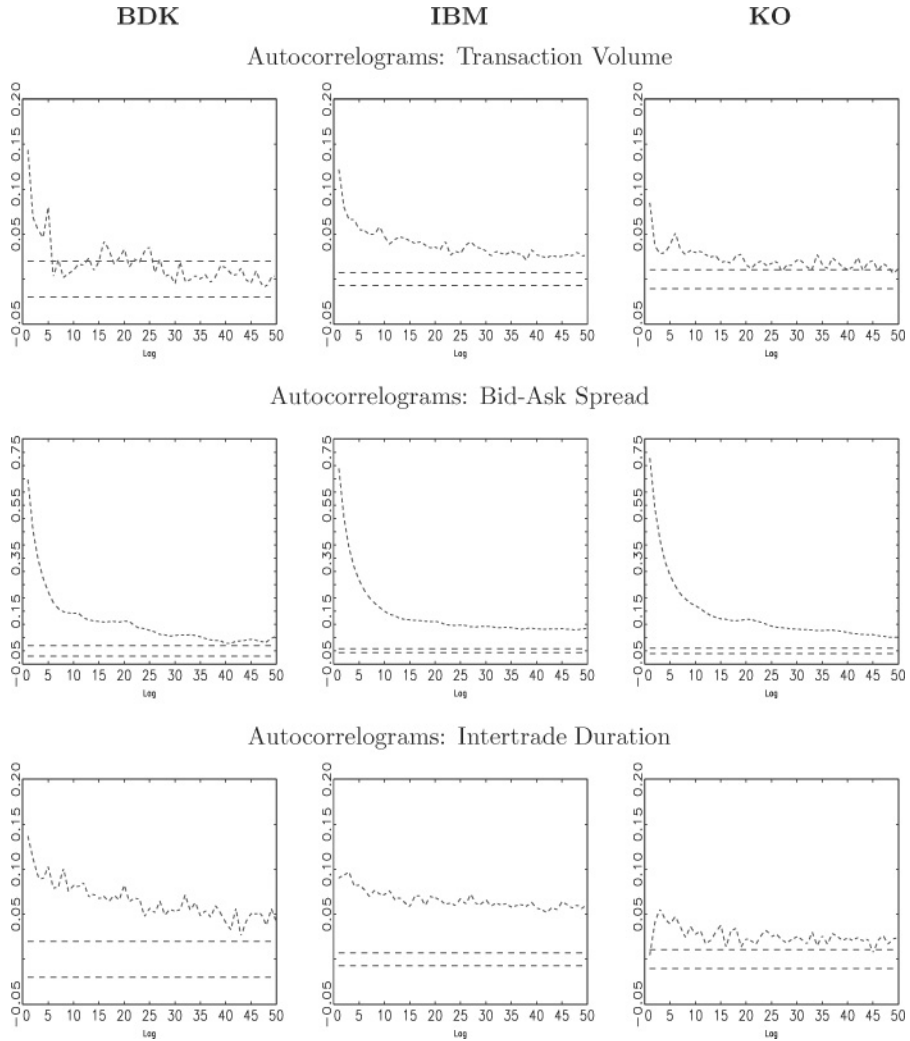
Figure 2 Autocorrelograms of BDK, IBM, and KO transaction volumes (first row), bid–ask spreads (second row), and intertrade durations (third row). Confidence bands: asymptotic 95%.

with specifications that possess a lag length of 1 for the autoregressive and the innovation variables. We allow the explanatory variables to be lagged up to lag 3.

In detail, we specify the conditional mean function in the ACM model for the price change direction (see Equation (3)) as:

$$(I_2 - \beta_1(L))(\Lambda_t - \zeta_3(L)\ln(Z_t)) = \mu + \alpha_1(L)\varepsilon_t + \gamma_1(L)|\varepsilon_t|, \tag{13}$$

where μ, α_1, β_1, γ_1, and ζ_i for $i = 1, 2, 3$ are assumed to be symmetric. We impose these symmetry constraints on the parameters to ensure a parsimonious model

specification. $Z_t = (V_t, S_t, D_t)'$ denotes the vector of explanatory variables. To clarify the notation and to ease the interpretation of ζ_i with respect to the explanatory variables, we denote the components of ζ_i in the following way:

$$\zeta_i = \begin{pmatrix} \zeta_{i1}^V & \zeta_{i1}^S & \zeta_{i1}^D \\ \zeta_{i2}^V & \zeta_{i2}^S & \zeta_{i2}^D \end{pmatrix}. \tag{14}$$

The conditional mean function in the GLARMA model, see Equation (7), for the size of the price change is specified as:

$$(1 - \tilde{\beta}_1(L))(\ln \omega_t - \tilde{\delta} \tilde{D}_t - \tilde{\zeta}_3(L) \ln(Z_t)) = \tilde{\mu} + S(\nu, \tau, K) + \tilde{\alpha}_1(L)\tilde{\varepsilon}_t + \tilde{\gamma}_1(L)|\tilde{\varepsilon}_t|, \tag{15}$$

where $\tilde{D}_t \in \{-1, 1\}$ indicates the direction of the contemporaneous price change at time t, and $Z_t = (\tilde{D}_t, V_t, S_t, D_t)'$ denotes again the vector of further explanatory variables. The components of the parameter vector ζ_i are now denoted as $\tilde{\zeta}_i = (\tilde{\zeta}_i^{\tilde{D}}, \tilde{\zeta}_i^V, \tilde{\zeta}_i^S, \tilde{\zeta}_i^D)$ and K is set to 2 in the specification of the Fourier flexible form.

In the fractionally integrated models, the conditional mean function (see Equation (10)) is specified as

$$(1 - \beta_1(L))(\varphi_t - \zeta_3(L) \ln(Z_t)) = \mu + \gamma_\infty(L)Y_t, \tag{16}$$

with

$$\gamma_\infty(L) = [1 - \beta_1(L) - [1 - \alpha_1(L) - \beta_1(L)](1 - L)^d],$$

where i) for transaction volume, i.e. $Y_t = V_t$, the vector of further explanatory variables is specified as $Z_t = (C_t, S_t, D_t)'$ with parameter vector $\zeta_i = (\zeta_i^C, \zeta_i^S, \zeta_i^D)$, ii) for the bid–ask spread, i.e., $Y_t = S_t$, the vector of further explanatory variables is specified as $Z_t = (V_t, C_t, D_t)'$ with parameter vector $\zeta_i = (\zeta_i^V, \zeta_i^C, \zeta_i^D)$, and iii) for intertrade duration, i.e., $Y_t = D_t$, the vector of further explanatory variables is specified as $Z_t = (V_t, S_t, C_t)'$ with parameter vector $\zeta_i = (\zeta_i^V, \zeta_i^S, \zeta_i^C)$. The parameter K in the exponential Fourier flexible form $s(\nu, \tau, K)$ in Equation (9) is set to 2, again.

In Tables 1–3 we report the estimation results of our model for BDK, IBM, and KO, respectively. Before interpreting the results in Section 3.4 in the light of market microstructure applications, we first consider the general pattern of the parameter estimates and the goodness of fit of our models across stocks.

For all three stocks, we find a clear diurnal seasonality pattern (represented by the joint significance of ν) in the duration, the spread, the volume, and the price changes size processes. Since, these patterns coincide with the usual U- and J-shape patterns, see, e.g., Bauwens and Giot (2001), for financial tick-by-tick variables, we do not present their graphs here.

The ACM models responsible for the price change direction process are characterized by a moderate degree of persistence: the parameter estimates for the components of β_1 lie between 0.3 and 0.6; and, they reflect the bid–ask bounce effect observed for IBM and KO (Figure 1) as well as the absence of it for BDK, through the parameter matrices α_1 and γ_1, which represent the effect of a

Table 1 Estimation results for Black & Decker (BDK)

Parameter	Model	
	Price change: Direction	
	ACM model	
# Parameters: 16	Estimates	Robust t-statistics
μ	$\begin{pmatrix} -0.9526 \\ -0.9526 \end{pmatrix}$	$\begin{pmatrix} -0.6790 \\ -0.6790 \end{pmatrix}$
β_1	$\begin{pmatrix} 0.4269 & 0.5003 \\ 0.5003 & 0.4269 \end{pmatrix}$	$\begin{pmatrix} 1.9655 & 2.3050 \\ 2.3050 & 1.9655 \end{pmatrix}$
α_1	$\begin{pmatrix} 0.1182 & -0.0132 \\ -0.0132 & 0.1182 \end{pmatrix}$	$\begin{pmatrix} 2.4149 & -0.2674 \\ -0.2674 & 2.4149 \end{pmatrix}$
γ_1	$\begin{pmatrix} -0.1643 & 0.2215 \\ 0.2215 & -0.1643 \end{pmatrix}$	$\begin{pmatrix} -1.3556 & 1.8504 \\ 1.8504 & -1.3556 \end{pmatrix}$
$\zeta_1 = \left(\zeta_1^V \vert \zeta_1^S \vert \zeta_1^D \right)$	$\begin{pmatrix} -0.0172 & 0.2287 & 0.0143 \\ -0.0172 & 0.2287 & 0.0143 \end{pmatrix}$	$\begin{pmatrix} -0.8534 & 6.9484 & 0.8897 \\ -0.8534 & 6.9484 & 0.8897 \end{pmatrix}$
$\zeta_2 = \left(\zeta_2^V \vert \zeta_2^S \vert \zeta_2^D \right)$	$\begin{pmatrix} 0.0033 & 0.0297 & 0.0151 \\ 0.0033 & 0.0297 & 0.0151 \end{pmatrix}$	$\begin{pmatrix} 0.1627 & 0.7875 & 0.9349 \\ 0.1627 & 0.7875 & 0.9349 \end{pmatrix}$
$\zeta_3 = \left(\zeta_3^V \vert \zeta_3^S \vert \zeta_3^D \right)$	$\begin{pmatrix} -0.0511 & -0.0227 & -0.0127 \\ -0.0511 & -0.0227 & -0.0127 \end{pmatrix}$	$\begin{pmatrix} -2.5023 & -0.6815 & -0.7880 \\ -2.5023 & -0.6815 & -0.7880 \end{pmatrix}$
	Price change: Size	
	GLARMA model	
# Parameters: 23	Estimates	Robust t-statistics
$\tilde{\mu}$	-0.0060	-0.8899
$\tilde{\beta}_1$	0.9821	176.20
$\tilde{\alpha}_1$	0.0133	3.4380
$\tilde{\gamma}_1$	-0.0003	-0.0677
ν_0	0.0392	6.0416
ν_1	0.0018	1.8374
ν_2	0.0022	1.7509
ν_3	0.0118	5.0956
ν_4	0.0068	3.5527
$\tilde{\delta}$	-0.0285	-0.9614
$\tilde{\xi}_1 = \left(\tilde{\xi}_1^{\tilde{D}}, \tilde{\xi}_1^V, \tilde{\xi}_1^S, \tilde{\xi}_1^D \right)$	$(0.0668\ 0.0959\ 0.3993\ 0.0062)$	$(2.3134\ 4.9202\ 11.375\ 0.3627)$
$\tilde{\xi}_2 = \left(\tilde{\xi}_2^{\tilde{D}}, \tilde{\xi}_2^V, \tilde{\xi}_2^S, \tilde{\xi}_2^D \right)$	$(0.0314\ 0.0257\ -0.0056\ -0.0117)$	$(1.1459\ 1.2446\ -0.1413\ -0.7212)$
$\tilde{\xi}_3 = \left(\tilde{\xi}_3^{\tilde{D}}, \tilde{\xi}_3^V, \tilde{\xi}_3^S, \tilde{\xi}_3^D \right)$	$(-0.0215\ -0.0134\ 0.0710\ -0.0502)$	$(-0.7391\ -0.7072\ 2.2129\ -2.8998)$
$\kappa^{-0.5}$	1.5510	20.835
	Transaction volume	
	FIACV model	
# Parameters: 18	Estimates	Robust t-statistics
μ	0.1474	16.197
β_1	0.4920	26.892
α_1	0.1055	14.731
ν_0	0.2992	16.728
ν_1	0.0122	3.1381
ν_2	0.0195	5.5857
ν_3	0.0628	10.463
ν_4	0.0476	9.4162
d	0.0290	5.1464
$\zeta_1 = \left(\zeta_1^C, \zeta_1^S, \zeta_1^D \right)$	$(0.0072\ -0.0255\ -0.0337)$	$(5.0719\ -2.9003\ -11.678)$
$\zeta_2 = \left(\zeta_2^C, \zeta_2^S, \zeta_2^D \right)$	$(0.0010\ -0.0216\ -0.0027)$	$(0.8519\ -2.1163\ -0.7827)$
$\zeta_3 = \left(\zeta_3^C, \zeta_3^S, \zeta_3^D \right)$	$(-0.0006\ 0.0298\ -0.0298)$	$(-0.4027\ 4.9261\ -9.0408)$

(*Continued*)

Table 1 (*Continued*)

Parameter	Model	
	Bid–ask spread	
	FIACS model	
# Parameters: 18	Estimates	Robust t-statistics
μ	0.1288	4.3317
β_1	0.0602	0.8634
α_1	0.3487	6.1900
ν_0	−0.0737	−0.9020
ν_1	0.0131	0.5493
ν_2	0.0040	0.1828
ν_3	−0.0061	−0.1796
ν_4	−0.0220	−0.8267
d	0.3090	4.0275
$\zeta_1 = \left(\zeta_1^V, \zeta_1^C, \zeta_1^D\right)$	(0.0746 0.0052 −0.0176)	(8.1029 1.8138 −3.0696)
$\zeta_2 = \left(\zeta_2^V, \zeta_2^C, \zeta_2^D\right)$	(0.0125 −0.0009 0.0029)	(1.4603 −0.2711 0.4417)
$\zeta_3 = \left(\zeta_3^V, \zeta_3^C, \zeta_3^D\right)$	(−0.0043 −0.0020 −0.0078)	(−0.5314 −0.5778 −1.1885)
	Intertrade duration	
	FIACD model	
# Parameters: 18	Estimates	Robust t-statistics
μ	0.0988	7.3331
β_1	0.6097	19.707
α_1	−0.1967	−8.0732
ν_0	−0.0550	−2.3109
ν_1	−0.0649	−8.2725
ν_2	−0.0046	−0.9933
ν_3	−0.0017	−0.1860
ν_4	−0.0287	−4.3971
d	0.2954	11.536
$\zeta_1 = \left(\zeta_1^V, \zeta_1^S, \zeta_1^C\right)$	(−0.0321 0.1452 −0.0036)	(−4.9485 13.878 −1.5489)
$\zeta_2 = \left(\zeta_2^V, \zeta_2^S, \zeta_2^C\right)$	(−0.0182 0.0090 −0.0034)	(−2.8899 0.7428 −1.6379)
$\zeta_3 = \left(\zeta_3^V, \zeta_3^S, \zeta_3^C\right)$	(−0.0131 0.0011 0.0032)	(−2.0157 0.1090 1.3560)
	Copula function	
	4-Dimensional Gaussian	
# Parameters: 6	Estimates	Robust t-statistics
$\Sigma = (\rho_{ij}), i, j = C, V, S, D$	$\begin{pmatrix} 1 & 0.0469 & -0.0273 & -0.0134 \\ 0.0469 & 1 & 0.0519 & -0.0398 \\ -0.0273 & 0.0519 & 1 & -0.1160 \\ -0.0134 & -0.0398 & -0.1160 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 5.0394 & -1.7924 & -1.3200 \\ 5.0394 & 1 & 3.4278 & -4.1950 \\ -1.7924 & 3.4278 & 1 & -7.4671 \\ -1.3200 & -4.1950 & -7.4671 & 1 \end{pmatrix}$

previous positive or negative price change on the current value of the log odds ratio vector Λ_t.

In the GLARMA specifications for the size of the price change process, $\kappa^{-0.5}$ is significantly different from zero for all three stocks, so that we observe a clearly overdispersed size process and thus have to reject the null of an at-zero truncated Poisson in favor of an at-zero truncated NegBin distribution. Furthermore, the price change size process is characterized by a high degree of persistence and for IBM we observe a significantly negative parameter $\tilde{\delta}$, which measures the current influence of the price change direction process on the size process. Thus, we observe that a negative price change implies a higher volatility (size) of the price change

Table 2 Estimation results for International Business Machines (IBM)

Parameter	Model	
	Price change: Direction ACM model	
# Parameters: 16	Estimates	Robust t-statistics
μ	$\begin{pmatrix} 1.5786 \\ 1.5786 \end{pmatrix}$	$\begin{pmatrix} 6.1901 \\ 6.1901 \end{pmatrix}$
β_1	$\begin{pmatrix} 0.3747 & 0.4030 \\ 0.4030 & 0.3747 \end{pmatrix}$	$\begin{pmatrix} 22.530 & 24.559 \\ 24.559 & 22.530 \end{pmatrix}$
α_1	$\begin{pmatrix} 0.0262 & 0.3504 \\ 0.3504 & 0.0262 \end{pmatrix}$	$\begin{pmatrix} 1.6420 & 23.901 \\ 23.901 & 1.6420 \end{pmatrix}$
γ_1	$\begin{pmatrix} -0.0694 & -0.3506 \\ -0.3506 & -0.0694 \end{pmatrix}$	$\begin{pmatrix} -1.8401 & -9.8543 \\ -9.8543 & -1.8401 \end{pmatrix}$
$\zeta_1 = \left(\zeta_1^V \middle| \zeta_1^S \middle| \zeta_1^D \right)$	$\begin{pmatrix} -0.0834 & 0.2329 & 0.0150 \\ -0.0834 & 0.2329 & 0.0150 \end{pmatrix}$	$\begin{pmatrix} -13.017 & 18.776 & 1.5663 \\ -13.017 & 18.776 & 1.5663 \end{pmatrix}$
$\zeta_2 = \left(\zeta_2^V \middle| \zeta_2^S \middle| \zeta_2^D \right)$	$\begin{pmatrix} 0.0030 & -0.0190 & -0.0226 \\ 0.0030 & -0.0190 & -0.0226 \end{pmatrix}$	$\begin{pmatrix} 0.4725 & -1.3422 & -2.3538 \\ 0.4725 & -1.3422 & -2.3538 \end{pmatrix}$
$\zeta_3 = \left(\zeta_3^V \middle| \zeta_3^S \middle| \zeta_3^D \right)$	$\begin{pmatrix} -0.0103 & 0.0206 & -0.0169 \\ -0.0103 & 0.0206 & -0.0169 \end{pmatrix}$	$\begin{pmatrix} -1.7026 & 1.6886 & -1.7790 \\ -1.7026 & 1.6886 & -1.7790 \end{pmatrix}$
	Price change: Size GLARMA model	
# Parameters: 23	Estimates	Robust t-statistics
$\tilde{\mu}$	0.0294	7.4077
$\tilde{\beta}_1$	0.9714	327.80
$\tilde{\alpha}_1$	0.0193	14.484
$\tilde{\gamma}_1$	0.0074	5.9731
ν_0	0.0010	0.8097
ν_1	0.0017	4.4310
ν_2	0.0006	1.8417
ν_3	0.0012	2.3565
ν_4	0.0013	3.1661
$\tilde{\delta}$	-0.0598	-9.8300
$\tilde{\zeta}_1 = \left(\tilde{\zeta}_1^{\tilde{D}}, \tilde{\zeta}_1^V, \tilde{\zeta}_1^S, \tilde{\zeta}_1^D \right)$	$(\ 0.0838 \ \ 0.0438 \ \ 0.5392 \ \ 0.0093)$	$(\ 12.782 \ \ 9.7196 \ \ 64.339 \ \ 1.5731)$
$\tilde{\zeta}_2 = \left(\tilde{\zeta}_2^{\tilde{D}}, \tilde{\zeta}_2^V, \tilde{\zeta}_2^S, \tilde{\zeta}_2^D \right)$	$(\ 0.0440 \ \ 0.0186 \ \ 0.0399 \ \ -0.0195)$	$(\ 6.5811 \ \ 4.1366 \ \ 4.8163 \ \ -2.7538)$
$\tilde{\zeta}_3 = \left(\tilde{\zeta}_3^{\tilde{D}}, \tilde{\zeta}_3^V, \tilde{\zeta}_3^S, \tilde{\zeta}_3^D \right)$	$(0.0053 \ \ 0.0088 \ \ 0.0249 \ \ -0.0119)$	$(0.7817 \ \ 2.0849 \ \ 3.5276 \ \ -1.6628)$
$\kappa^{-0.5}$	1.0330	113.81
	Transaction volume FIACV model	
# Parameters: 18	Estimates	Robust t-statistics
μ	0.1083	37.093
β_1	0.5308	37.656
α_1	0.0548	18.899
ν_0	0.0770	10.141
ν_1	0.0364	19.884
ν_2	0.0317	18.594
ν_3	0.0141	5.6875
ν_4	0.0095	5.9042
d	0.1349	44.670
$\zeta_1 = \left(\zeta_1^C, \zeta_1^S, \zeta_1^D \right)$	$(0.0053 \ \ 0.0717 \ \ 0.0691)$	$(14.152 \ \ 35.266 \ \ 45.263)$
$\zeta_2 = \left(\zeta_2^C, \zeta_2^S, \zeta_2^D \right)$	$(-0.0019 \ \ -0.0476 \ \ 0.0208)$	$(-4.4966 \ \ -17.888 \ \ 14.258)$
$\zeta_3 = \left(\zeta_3^C, \zeta_3^S, \zeta_3^D \right)$	$(-0.0026 \ \ -0.0363 \ \ 0.0180)$	$(-6.3633 \ \ -15.523 \ \ 12.508)$

(*Continued*)

Table 2 (*Continued*)

Parameter	Model	
	Bid–ask spread	
	FIACS model	
# Parameters: 18	Estimates	Robust *t*-statistics
μ	0.0959	10.660
β_1	0.0972	5.3372
α_1	0.3389	17.998
ν_0	−0.0371	−1.3525
ν_1	−0.0206	−2.6091
ν_2	0.0008	0.1096
ν_3	−0.0131	−1.2203
ν_4	−0.0027	−0.3176
d	0.3105	11.941
$\zeta_1 = \left(\zeta_1^V, \zeta_1^C, \zeta_1^D\right)$	(0.0210 0.0026 0.0085)	(8.9077 7.0798 2.5481)
$\zeta_2 = \left(\zeta_2^V, \zeta_2^C, \zeta_2^D\right)$	(0.0130 0.0012 −0.0107)	(5.3322 1.2668 −3.2167)
$\zeta_3 = \left(\zeta_3^V, \zeta_3^C, \zeta_3^D\right)$	(−0.0013 −0.0016 −0.0071)	(−0.5581 −1.7400 −2.1539)
	Intertrade duration	
	FIACD model	
# Parameters: 18	Estimates	Robust *t*-statistics
μ	0.1224	12.447
β_1	0.5160	26.561
α_1	−0.1843	−19.234
ν_0	0.0175	1.3051
ν_1	−0.0399	−10.558
ν_2	0.0104	3.6792
ν_3	0.0091	1.7568
ν_4	−0.0002	−0.0586
d	0.2172	23.362
$\zeta_1 = \left(\zeta_1^V, \zeta_1^S, \zeta_1^C\right)$	(0.0087 0.0716 −0.0010)	(3.2677 12.366 −1.1771)
$\zeta_2 = \left(\zeta_2^V, \zeta_2^S, \zeta_2^C\right)$	(0.0054 0.0097 0.0007)	(1.9157 1.5207 0.7586)
$\zeta_3 = \left(\zeta_3^V, \zeta_3^S, \zeta_3^C\right)$	(−0.0059 0.0038 0.0000)	(−2.1155 0.7453 0.0000)
	Copula function	
	4-Dimensional Gaussian	
# Parameters: 6	Estimates	Robust *t*-statistics
$\Sigma = (\rho_{ij}), i, j = C, V, S, D$	$\begin{pmatrix} 1 & 0.0409 & 0.0555 & -0.0074 \\ 0.0409 & 1 & 0.1148 & 0.1484 \\ 0.0555 & 0.1148 & 1 & 0.4168 \\ -0.0074 & 0.1484 & 0.4168 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 18.426 & 11.486 & -1.5160 \\ 18.426 & 1 & 23.156 & 33.835 \\ 11.486 & 23.156 & 1 & 94.097 \\ -1.5160 & 33.835 & 94.097 & 1 \end{pmatrix}$

process, which is in line with the well-known leverage effect. For BDK and KO $\tilde{\delta}$ is insignificant.

Comparing the (multivariate) Ljung–Box (LB) statistics of the raw with the residual series for the price change direction and the price change size processes in Table 4, shows that both the ACM and the GLARMA models are able to explain a large part of the underlying dynamics very well. However, for IBM, which is the most liquidly traded stock in our sample, the simple ACM(1,1) specification seems unable to capture the dynamic structure completely. A similar picture emerges by comparing the (multivariate) autocorrelograms for the raw with the residual direction and size series. Since there is no new information in these autocorrelograms

Table 3 Estimation results for Coca Cola (KO)

Parameter	Model	
	Price change: Direction	
	ACM model	
# Parameters: 16	Estimates	Robust t-statistics
μ	$\begin{pmatrix} 1.1789 \\ 1.1789 \end{pmatrix}$	$\begin{pmatrix} 2.2977 \\ 2.2977 \end{pmatrix}$
β_1	$\begin{pmatrix} 0.3413 & 0.5600 \\ 0.5600 & 0.3413 \end{pmatrix}$	$\begin{pmatrix} 17.653 & 30.449 \\ 30.449 & 17.653 \end{pmatrix}$
α_1	$\begin{pmatrix} -0.2755 & 0.5357 \\ 0.5357 & -0.2755 \end{pmatrix}$	$\begin{pmatrix} -11.916 & 28.100 \\ 28.100 & -11.916 \end{pmatrix}$
γ_1	$\begin{pmatrix} 0.4799 & -0.6922 \\ -0.6922 & 0.4799 \end{pmatrix}$	$\begin{pmatrix} 10.641 & -17.414 \\ -17.414 & 10.641 \end{pmatrix}$
$\zeta_1 = \left(\zeta_1^V \middle| \zeta_1^S \middle| \zeta_1^D\right)$	$\begin{pmatrix} -0.0738 & 0.0796 & 0.0085 \\ -0.0738 & 0.0796 & 0.0085 \end{pmatrix}$	$\begin{pmatrix} -9.0091 & 4.0511 & 0.9074 \\ -9.0091 & 4.0511 & 0.9074 \end{pmatrix}$
$\zeta_2 = \left(\zeta_2^V \middle| \zeta_2^S \middle| \zeta_2^D\right)$	$\begin{pmatrix} -0.0139 & 0.0098 & 0.0116 \\ -0.0139 & 0.0098 & 0.0116 \end{pmatrix}$	$\begin{pmatrix} -1.6861 & 0.4229 & 1.2382 \\ -1.6861 & 0.4229 & 1.2382 \end{pmatrix}$
$\zeta_3 = \left(\zeta_3^V \middle| \zeta_3^S \middle| \zeta_3^D\right)$	$\begin{pmatrix} -0.0146 & 0.0100 & 0.0184 \\ -0.0146 & 0.0100 & 0.0184 \end{pmatrix}$	$\begin{pmatrix} -1.8126 & 0.5236 & 1.9950 \\ -1.8126 & 0.5236 & 1.9950 \end{pmatrix}$
	Price change: Size	
	GLARMA model	
# Parameters: 23	Estimates	Robust t-statistics
$\tilde{\mu}$	-0.0072	-6.6569
$\tilde{\beta}_1$	0.9981	1638.3
$\tilde{\alpha}_1$	0.0055	6.4415
$\tilde{\gamma}_1$	0.0055	4.9863
ν_0	0.0095	9.3744
ν_1	0.0003	2.0814
ν_2	0.0003	1.7109
ν_3	0.0027	7.4451
ν_4	0.0014	5.1365
δ	0.0248	1.4426
$\tilde{\zeta}_1 = \left(\tilde{\zeta}_1^D, \tilde{\zeta}_1^V, \tilde{\zeta}_1^S, \tilde{\zeta}_1^D\right)$	$(0.0433\ 0.0594\ 0.6496\ 0.0805)$	$(2.5143\ 5.8216\ 30.577\ 7.0254)$
$\tilde{\zeta}_2 = \left(\tilde{\zeta}_2^D, \tilde{\zeta}_2^V, \tilde{\zeta}_2^S, \tilde{\zeta}_2^D\right)$	$(0.0208\ -0.0012\ 0.0108\ -0.0076)$	$(1.1733\ -0.1189\ 0.4502\ -0.6170)$
$\tilde{\zeta}_3 = \left(\tilde{\zeta}_3^D, \tilde{\zeta}_3^V, \tilde{\zeta}_3^S, \tilde{\zeta}_3^D\right)$	$(0.0242\ 0.0009\ 0.0502\ -0.0295)$	$(1.3506\ 0.0944\ 2.4628\ -2.5389)$
$\kappa^{-0.5}$	1.5158	28.958
	Transaction volume	
	FIACV model	
# Parameters: 18	Estimates	Robust t-statistics
μ	0.0423	31.715
β_1	0.7985	153.60
α_1	-0.0153	-8.0018
ν_0	0.1013	31.889
ν_1	0.0173	22.651
ν_2	0.0095	15.473
ν_3	0.0245	25.928
ν_4	0.0173	23.381
d	0.1159	51.378
$\zeta_1 = \left(\zeta_1^C, \zeta_1^S, \zeta_1^D\right)$	$(-0.0001\ 0.0757\ 0.0625)$	$(-0.0794\ 21.665\ 46.251)$
$\zeta_2 = \left(\zeta_2^C, \zeta_2^S, \zeta_2^D\right)$	$(0.0067\ -0.0414\ 0.0198)$	$(9.4855\ -10.426\ 13.356)$
$\zeta_3 = \left(\zeta_3^C, \zeta_3^S, \zeta_3^D\right)$	$(0.0008\ -0.0378\ -0.0107)$	$(0.8064\ -12.762\ -6.9014)$

(*Continued*)

Table 3 (*Continued*)

Parameter	Model	
	Bid–ask spread FIACS model	
# Parameters: 18	Estimates	Robust *t*-statistics
μ	0.1330	8.9947
β_1	0.1179	3.3698
α_1	0.3948	10.792
ν_0	−0.1358	−3.7532
ν_1	0.0073	0.6593
ν_2	0.0020	0.1869
ν_3	−0.0378	−2.4447
ν_4	−0.0150	−1.1720
d	0.3157	6.4453
$\zeta_1 = \left(\zeta_1^V, \zeta_1^C, \zeta_1^D\right)$	(0.0296 0.0034 0.0102)	(7.7483 1.2044 2.3167)
$\zeta_2 = \left(\zeta_2^V, \zeta_2^C, \zeta_2^D\right)$	(−0.0014 0.0034 −0.0001)	(−0.3614 1.1775 −0.0321)
$\zeta_3 = \left(\zeta_3^V, \zeta_3^C, \zeta_3^D\right)$	(−0.0001 0.0020 −0.0058)	(−0.0268 0.5581 −1.3161)
	Intertrade duration FIACD model	
# Parameters: 18	Estimates	Robust *t*-statistics
μ	0.2772	14.195
β_1	0.4041	15.222
α_1	−0.1706	−18.214
ν_0	−0.0019	−0.1098
ν_1	−0.0550	−10.655
ν_2	0.0110	2.7698
ν_3	0.0102	1.4867
ν_4	−0.0022	−0.4405
d	0.1677	18.804
$\zeta_1 = \left(\zeta_1^V, \zeta_1^S, \zeta_1^C\right)$	(−0.0430 0.1114 0.0025)	(−12.458 13.251 1.2250)
$\zeta_2 = \left(\zeta_2^V, \zeta_2^S, \zeta_2^C\right)$	(0.0113 0.0503 0.0027)	(3.0341 4.8096 1.6251)
$\zeta_3 = \left(\zeta_3^V, \zeta_3^S, \zeta_3^C\right)$	(0.0143 −0.0368 −0.0019)	(4.0697 −4.2543 −0.7112)
	Copula function 4-Dimensional Gaussian	
# Parameters: 6	Estimates	Robust *t*-statistics
$\Sigma = (\rho_{ij}), i, j = C, V, S, D$	$\begin{pmatrix} 1 & 0.0295 & -0.1456 & 0.0334 \\ 0.0295 & 1 & 0.0577 & 0.1574 \\ -0.1456 & 0.0577 & 1 & -0.1727 \\ 0.0334 & 0.1574 & -0.1727 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 6.8352 & -18.063 & 5.6428 \\ 6.8352 & 1 & 7.4651 & 35.529 \\ -18.063 & 7.4651 & 1 & -16.996 \\ 5.6428 & 35.529 & -16.996 & 1 \end{pmatrix}$

in addition to the Ljung–Box statistics, we do not present them here for the price change subprocesses (direction and size); instead we only present the residual autocorrelogram of the complete price change process in the first row of Figure 3. Comparing these autocorrelograms with those of the raw series in Figure 1 for all three stocks, demonstrates that the bid–ask bounce effects have been explained by the proposed model specifications.

Let us consider Tables 1–3 again and address the fractionally integrated specifications for the transaction volume, the bid–ask spread, and the intertrade durations. For all three stocks, the fractionally differencing parameter d is smallest for the volume, second smallest for the duration, and largest for the spread process.

Table 4 Model evaluation for Black & Decker (BDK) upper panel, International Business Machines (IBM) middle panel, Coca Cola (KO) lower panel

	BDK					
	Price change					
	Direction		Size		Change	
	Raw	Residuals	Raw	Residuals	Raw	Residuals
LB(10)	218.46	60.058	296.22	11.381	15.955	28.520
LB(20)	300.04	99.909	348.38	20.823	30.484	38.570
LB(30)	342.02	132.81	368.70	27.446	34.907	41.658
LB(40)	380.08	173.49	380.03	34.940	41.947	52.311
LB(50)	420.30	213.35	387.31	42.427	46.786	58.275
Mean	$\begin{pmatrix} 0.2963 \\ 0.3030 \end{pmatrix}$	$\begin{pmatrix} -0.0027 \\ 0.0082 \end{pmatrix}$	3.1235	-0.0003	0.0042	0.0044
Variance	$\begin{pmatrix} 0.2085 & -0.0897 \\ -0.0897 & 0.2112 \end{pmatrix}$	$\begin{pmatrix} 0.8104 & 0.0646 \\ 0.0646 & 1.2533 \end{pmatrix}$	11.021	1.3137	12.449	1.1718

	Transaction volume		Bid–ask spread		Intertrade duration	
	Raw	Residuals	Raw	Residuals	Raw	Residuals
LB(10)	363.64	17.055	9575.2	5.8768	909.24	10.625
LB(20)	425.44	21.400	10882	19.647	1429.6	16.066
LB(30)	464.27	30.668	11471	24.359	1758.1	20.788
LB(40)	472.62	35.249	11710	30.361	2054.7	27.799
LB(50)	477.07	44.947	11867	37.145	2256.9	33.918
Mean	710.03	1.0064	0.0670	1.0072	53.077	0.9813
Variance	2537815	3.9570	0.0029	0.5480	5525.6	1.5751

	IBM					
	Price change					
	Direction		Size		Change	
	Raw	Residuals	Raw	Residuals	Raw	Residuals
LB(10)	3260.0	617.19	4280.4	50.870	954.47	57.414
LB(20)	3379.3	681.53	5418.7	58.781	977.38	62.454
LB(30)	3464.4	755.82	6137.6	65.356	1009.8	74.402
LB(40)	3513.2	790.01	6647.8	76.419	1017.4	81.319
LB(50)	3567.5	840.42	7038.7	82.239	1028.2	90.472
Mean	$\begin{pmatrix} 0.3072 \\ 0.3277 \end{pmatrix}$	$\begin{pmatrix} -0.0089 \\ 0.0295 \end{pmatrix}$	3.5603	0.0015	-0.0052	0.0309
Variance	$\begin{pmatrix} 0.2128 & -0.1007 \\ -0.1007 & 0.2203 \end{pmatrix}$	$\begin{pmatrix} 0.7803 & 0.0829 \\ 0.0829 & 1.2980 \end{pmatrix}$	14.987	1.6244	17.580	1.2620

	Transaction volume		Bid–ask spread		Intertrade duration	
	Raw	Residuals	Raw	Residuals	Raw	Residuals
LB(10)	3599.5	38.732	91591.5	21.322	5016.6	40.528
LB(20)	4896.8	58.703	102299	52.631	8417.3	56.191
LB(30)	5807.5	66.184	109206	66.393	11485	76.715
LB(40)	6419.2	74.193	114782	73.519	14192	86.081
LB(50)	6926.6	82.433	119813	91.328	16728	103.82
Mean	1744.9	1.0263	0.0710	1.0026	6.8712	1.0257
Variance	222993	4.6370	0.0030	0.4456	42.716	0.8251

(*Continued*)

Table 4 (*Continued*)

	KO					
	Price change					
	Direction		Size		Change	
	Raw	Residuals	Raw	Residuals	Raw	Residuals
LB(10)	1792.0	168.72	2145.2	50.497	467.32	47.373
LB(20)	1955.5	210.67	2309.4	56.986	474.61	52.915
LB(30)	2034.7	255.95	2415.1	62.715	489.16	65.152
LB(40)	2094.6	300.41	2492.5	72.452	495.40	71.206
LB(50)	2181.9	348.52	2599.1	84.397	502.76	81.091
Mean	$\begin{pmatrix} 0.2558 \\ 0.2638 \end{pmatrix}$	$\begin{pmatrix} -0.0100 \\ 0.0091 \end{pmatrix}$	2.1369	−0.0029	0.0033	0.0002
Variance	$\begin{pmatrix} 0.1904 & -0.0675 \\ -0.0675 & 0.1942 \end{pmatrix}$	$\begin{pmatrix} 0.8678 & 0.0344 \\ 0.0344 & 1.1463 \end{pmatrix}$	4.2193	1.4900	4.5672	1.1628
	Transaction volume		Bid–ask spread		Intertrade duration	
	Raw	Residuals	Raw	Residuals	Raw	Residuals
LB(10)	636.54	12.040	48575	16.208	505.85	8.5313
LB(20)	837.67	17.154	54208	24.113	756.74	29.092
LB(30)	936.63	24.574	57453	33.122	960.70	38.590
LB(40)	1046.1	30.875	59495	46.422	1148.0	51.200
LB(50)	1130.4	39.117	60681	61.834	1305.7	59.327
Mean	1996.0	1.0037	0.0398	1.0064	14.754	0.9216
Variance	37479314	7.3450	0.0012	0.4880	367.30	1.4014

*LB denotes the (multivariate) Ljung Box statistic.

This pattern reflects the different degrees of persistency already depicted by the autocorrelograms in Figure 2. Analyzing the Ljung–Box statistics of the raw and the residual series for these three processes for all three stocks in Table 4 and considering the autocorrelograms of the residual series in Figure 3 (second to fourth row), lead to the conclusion that the fractional integrated model specifications capture the underlying dynamic behavior extremely well.

3.4 Market Microstructure Implications

Market microstructure research analyzes how market participants interact with each other, process information, place orders of a specific size at a certain time and affect the price process within a given institutional framework. Although, theoretical market microstructure models have been available for almost 40 years, Demsetz (1968), Bagehot (1971), and Smidt (1971), for example, attempt to model how market makers set bid and ask quotes and therefore determine the price process, the availability of high-frequency data sets, and the advances in computer technology have spurred the theoretical and especially the empirical market microstructure research within the last decade. The proposed modeling framework in this paper is not based on an explicit theoretical market microstructure model and the estimated relationships should therefore not be interpreted as structural economic relations.
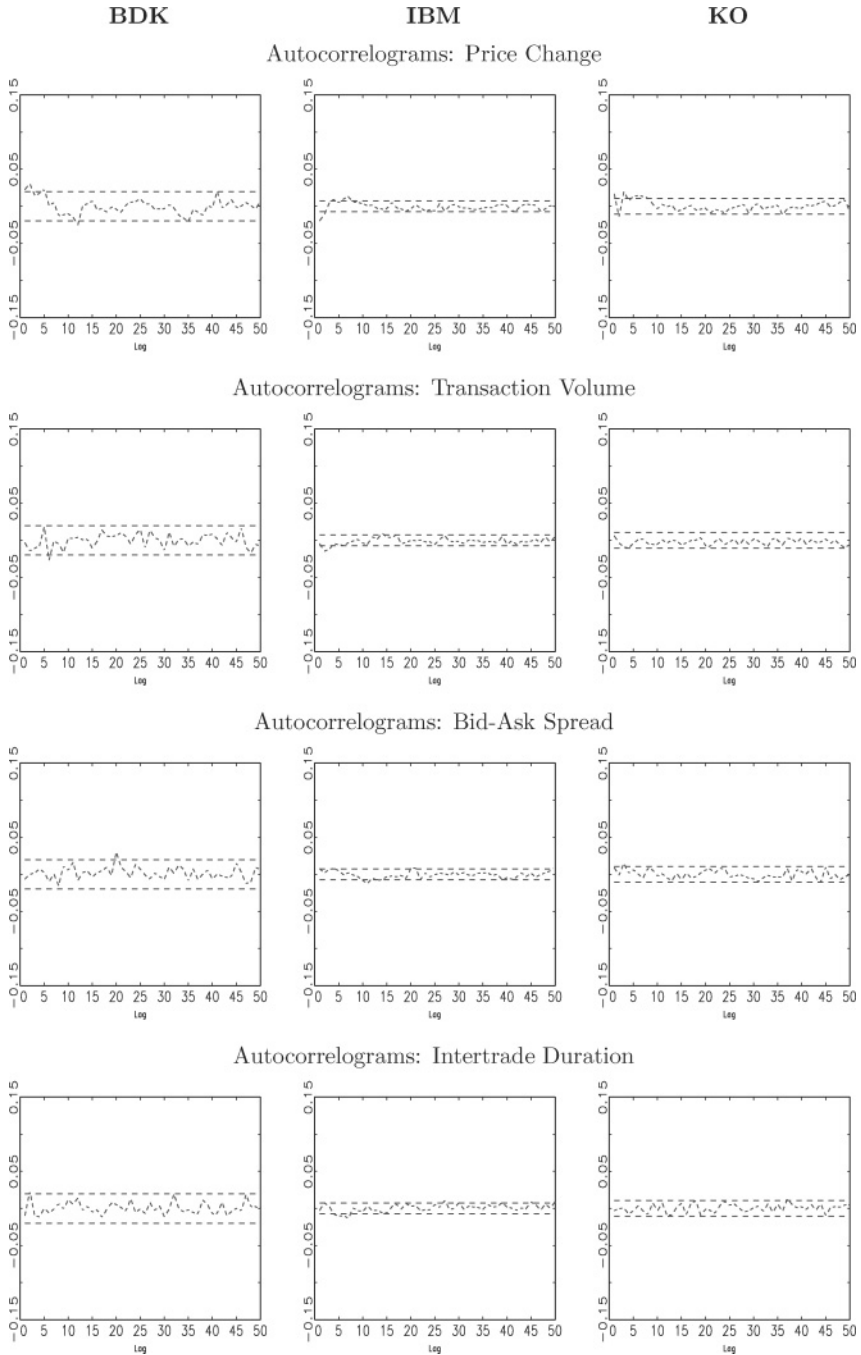
Figure 3 Residual autocorrelograms of BDK, IBM, and KO price changes (first row), transaction volumes (second row), bid-ask spreads (third row) and intertrade durations (fourth row). Confidence Bands: Asymptotic 95%.

Nevertheless, our model tries to explain the joint process of four of the most relevant market microstructure variables: price changes, transaction volumes, bid–ask spreads, and intertrade durations, and therefore it allows us to investigate how these variables interact with each other and to test well-known implications of theoretical market microstructure models. The use of the copula functions makes the model particularly attractive, since it enables us to analyze the instantaneous relations between our four variables directly and separately from the effects of the lagged explanatory variables (Granger causality relations). This distinction between contemporaneous and lagged effects is often diluted, when implications of market microstructure models are tested with aggregated data. The time scale in our model is defined through the arrival of trades, therefore "instantaneous" refers exactly to the time of a specific trade, which is the highest possible disaggregation level for our data.

The instantaneous causality relations between different market microstructure variables have attracted a lot of attention. Diamond and Verrecchia (1987), Easley and O'Hara (1992), Dufour and Engle (2000), and Renault and Werker (2006) examine how intertrade duration and return volatilities interact. Whereas, the first three papers postulate an a priori causality relationship from duration to volatility, Renault and Werker (2006) take a structural approach to identify and quantify instantaneous effects in addition to Granger causality effects. Diamond and Verrecchia (1987) hypothesize, assuming that short selling is prohibited, that bad news are reflected by longer durations which should cause negative price reactions and increasing volatility. Contrarily, Easley and O'Hara (1992) and Dufour and Engle (2000) consider smaller durations as a sign for a larger share of informed investors being active in the market, which is anticipated by less informed traders and thus causes more uncertainty and therefore higher volatility. The latter point of view has also been confirmed by Engle (2000) as well as by Renault and Werker (2006) for the instantaneous relation between durations and volatility.

Another strand of the literature focuses on the relationship between trading volume and the bid–ask spread. In the early inventory models, see Demsetz (1968) for example, the market maker is seen as providing a service for which he has to be compensated by letting him earn the bid–ask spread. Smidt (1971), Garman (1976), Stoll (1978), Ho and Stoll (1981), Hasbrouck and Sofianos (1993), and Madhavan and Smidt (1993), however, assume that the market maker, who might be allowed to trade actively in the market on his own account, holds an optimal inventory and deviations from this optimal inventory, induced by large trades, cause an inventory risk which is mirrored into the bid–ask spread. A different explanation for the connection between transaction volumes and bid–ask spreads is given by asymmetric information based models, see Bagehot (1971), Copeland and Galai (1983), Glosten and Milgrom (1985), or Easley and O'Hara (1987). In these models the market maker is assumed to make profits when trading with uninformed investors and assumed to make losses when trading with informed investors. Thus, based on historical order flow, which includes trading speed and trading volume, the market maker tries to figure out whether he faces an informed investor, which would force him to increase the bid–ask spread.

Let us consider Tables 1–3 again and examine the instantaneous effects among our variables captured by the copula parameters ρ_{ij}, i, $j = \{C, V, S, D\}$, as well as the Granger causality effects represented by the parameters of the lagged explanatory variables ζ_i. For all three stocks, the copula parameter ρ_{VS} is significantly positive, representing a positive instantaneous relationship between transaction volume and bid–ask spread. Moreover, within the FIACS model, we observe for all three stocks only significant positive parameters ζ_i^V, $i = 1, \ldots, 3$, reflecting that higher trading volumes Granger cause higher bid–ask spreads. These observations are clearly in line with the findings from the inventory and asymmetric information models cited above.

The effect of time on the bid–ask spread and on the price change volatility, which is modeled by the GLARMA specification is not that clear-cut but shows an interesting pattern. First of all, the instantaneous effect between intertrade duration and the bid–ask spread, ρ_{SD}, is negative for BDK and KO but positive for IBM, which means that a higher trading intensity (smaller durations) is accompanied by a higher bid–ask spread for BDK and KO, but by a smaller one for IBM. Moreover, in the FIACS model, ζ_1^D is also significantly negative for BDK strengthening the instantaneous effect, positive for KO compensating the instantaneous effect and positive for IBM supporting the instantaneous effect, which is then weakened by ζ_2^D and ζ_3^D, which are significantly negative.

These ambiguous effects (especially the instantaneous ones) between durations and the bid–ask spread can be interpreted in the light of the theoretical models of Admati and Pfleiderer (1988) and Foster and Viswanathan (1990). Whereas, in the model of Admati and Pfleiderer (1988) both informed and uninformed investors have a high incentive to trade and therefore a high trading intensity when trading costs are low, Foster and Viswanathan (1990) assume that high trading intensity is only caused by informed investors, preventing uninformed investors from trading at these times and causing higher bid–ask spreads being set by the market maker. Our results do not allow us to favor one of these models when the instantaneous effect between duration and bid–ask spread is examined. However, the Granger causality effects of the bid–ask spread on durations (ζ_i^S $i = 1, \ldots, 3$ in the FIACS model) show that for all three stocks a higher bid–ask spread at the previous trades leads to increasing intertrade durations and thus to lower trading activity. This observation supports the Admati and Pfleiderer (1988) model. Interpreting the Foster and Viswanathan (1990) model in the Granger causality sense, we can still not find support for this model since, as mentioned above, also the lagged effects of durations on the bid–ask spread are indistinct and seem to be dominated by the effects of trading volume.

The ambiguity of the effect from durations on the bid–ask spreads can be combined with the observation that for all three stocks, price change volatility (GLARMA model) is significantly increasing in lagged trading volumes ($\tilde{\zeta}_i^V$, $i = 1, \ldots, 3$) and lagged bid–ask spreads ($\tilde{\zeta}_i^S$, $i = 1, \ldots, 3$), but again there is no clear influence of lagged durations on volatility. Furthermore, we also observe no clear effect of instantaneous (ρ_{VD}) and lagged durations (ζ_i^D, $i = 1, \ldots, 3$) on transaction volume (FIACV model) for all three stocks. Taking these observations

together an explanation for this finding is, that information on potential asymmetric information and informed trading is better conveyed by the bid–ask spread process and the transaction volume process than by the duration process. This finding does not contradict the conclusions of Diamond and Verrecchia (1987), Easley and O'Hara (1992), Engle (2000), and Renault and Werker (2006) where only the bivariate link between durations and volatility is examined, but it shows that through the correlation between the spread, the volume and the duration processes, the information content of the duration process is diminished and dominated by the information contained in the bid–ask spread and the transaction volume processes. A similar observation is made by Dufour and Engle (2000), who also observe that taking bid–ask spreads and trading volumes into account reduces and sometimes reverses the effect of intertrade durations on price change volatility. The advantage of our model specification is that we consider the interactions within a joint system and we do not have to impose their exogeneity assumptions.

4 Conclusion

In this paper, we propose a model for the joint process of price changes, transaction volumes, bid–ask spreads, and intertrade durations on the tick-by-tick time scale. We rely on a copula approach in specifying the joint density of our four-dimensional system, which enables us to avoid a decomposition framework, where an instantaneous causality scheme between the four variables has to be imposed a priori. In our model specification, we regard discrete price changes and high persistency patterns in the empirical autocorrelation functions of bid–ask spreads, transaction volumes, and intertrade durations, through the application of the ICH model of Liesenfeld, Nolte, and Pohlmeier (2006) and the FIACD-type models of Jasiak (1999), respectively. Since, the price change process is discrete whereas the remaining three processes are continuous, we rely on continued price change variables in the specification of the copula function as proposed by Denuit and Lambert (2005).

We apply our model to tick-by-tick data from three stocks: BDK, IBM, and KO for a period of 1 month (May, 2001). Although the proposed model specifications are able to account for the dynamics of the underlying series very well, some need for fine tuning arises in the specification of the ICH model for the very frequently traded IBM stock. The *ad hoc* choice of the four-dimensional Gaussian copula is mainly driven by the fact that its six correlation parameters can be easily interpreted and allow for a very flexible model specification. Dynamic extensions of the copula function as suggested by Patton (2006) are not considered here, since they would increase the computational burden in the model estimation considerably.

The model is used to verify several market microstructure theories. Whereas, we observe the usual relationships between bid–ask spreads and transaction volumes, as postulated by inventory and asymmetric information models, as well as their clear positive influence on price volatility and trade arrival times, we find ambiguous effects for the influence of intertrade durations on price volatility and bid–ask spreads. This observation shows that the information captured by the transaction volume and the bid–ask spread processes dominates the information

captured by the trade arrival process, when not being used as a stand-alone regressor. This finding does not invalidate the role of trade arrival times in reflecting order flow information, but the direction of its influence might need a deeper investigation using models in which the processes of several market microstructure variables are specified jointly.

REFERENCES

Admati, A., and P. Pfleiderer. (1988). "A Theory of Intra-Day Patterns: Volume and Price Variability." *Review of Financial Studies 1*, 3–40.

Bagehot, W. (1971). "The Only Game in Town." *Financial Analysts Journal 27*, 12–14, 22.

Bauwens, L., and P. Giot. (2001). *Econometric Modelling of Stock Market Intraday Activity*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Bauwens, L., P. Giot, J. Grammig, and D. Veredas. (2004). "A Comparison of Financial Duration Models Via Density Forecasts." *International Journal of Forecasting 20(4)*, 589–609.

Bauwens, L., and N. Hautsch. (2006). "Stochastic Conditional Intensity Processes." *Journal of Financial Econometrics 4*, 450–493.

Boehmer, E., J. Grammig, and E. Theissen. (2007). "Estimating the Probability of Informed Trading – Does Trade Misclassification Matter?." *Journal of Financial Markets 20*, 26–47.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics 31*, 307–327.

Cameron, C., T. Li, P. Trivedi, and D. Zimmer. (2004). "Modelling the Differences in Counted Outcomes Using Bivariate Copula Models with Application to Mismesured Counts." *Econometrics Journal 7*, 566–584.

Copeland, T. E., and D. Galai. (1983). "Information Effects and the Bid-Ask Spread." *Journal of Finance 38*, 1457–1469.

Demsetz, H. (1968). "The Cost of Transacting." *Quarterly Journal of Economics 82*, 33–53.

Denuit, M., and P. Lambert. (2005). "Constraints on Concordance Measures in Bivariate Discrete Data." *Journal of Multivariate Analysis 93*, 40–57.

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.

Dufour, A., and R. F. Engle. (2000). "Time and the Impact of a Trade." *Journal of Finance 55*, 2467–2498.

Easley, D., and M. O'Hara. (1987). "Price, Trade Size, and Information in Securities Markets." *Journal of Financial Economics 19*, 69–90.

Easley, D., and M. O'Hara. (1992). "Time and Process of Security Price Adjustment." *Journal of Finance 47(2)*, 577–605.

Engle, R. F. (1982). "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation." *Econometrica 50*, 987–1008.

——— (2000). "The Econometrics of Ultra-High-Frequency Data." *Econometrica 68*, 1, 1–22.

Engle, R. F., D. F. Hendry, and J.-F. Richard. (1983). "Exogeneity." *Econometrica, 51*, 2, 277–304.

Engle, R. F., and J. R. Russell. (1998). "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data." *Econometrica 66*, 1127–1162.

Foster, D. F., and S. Viswanathan. (1990). "A Theory of the Intraday Variations in Volumes, Variances and Trading Costs in Security Markets." *Review of Financial Studies 3*, 593–624.

Garman, M. B. (1976). "Market Microstructure." *Journal of Financial Economics 3*, 257–275.

Ghysels, E., C. Gourieroux, and J. Jasiak. (2004). "Stochastic Volatility Duration Models." *Journal of Econometrics 119(2)*, 413–433.

Glosten, L. R., and P. R. Milgrom. (1985). "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders." *Journal of Financial Economics 14*, 71–100.

Grammig, J., and K.-O. Maurer. (2000). "Non-Monotonic Hazard Functions and the Autoregressive Conditional Duration Model." *Econometrics Journal 3*, 16–38.

Hasbrouck, J., and G. Sofianos. (1993). "The Trades of Market Makers: An Empirical Analysis of NYSE Specialists." *Journal of Finance 48*, 1565–1593.

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.

Heinen, A., and E. Rengifo. (2007). "Multivariate Autoregressive Modelling of Time Series Count Data Using Copulas." *Journal of Empirical Finance 14(4)*, 564–583.

Ho, T., and H. Stoll. (1981). "Optimal Dealer Pricing Under Transactions and Return Uncertainty." *Journal of Financial Economics 9*, 47–73.

Jasiak, J. (1999). "Persistence in Intertrade Durations." Discussion paper, York University.

Lee, C. M. C., and M. J. Ready. (1991). "Inferring Trade Direction from Intraday Data." *Journal of Finance 46*, 733–746.

Liesenfeld, R., I. Nolte, and W. Pohlmeier. (2006). "Modelling Financial Transaction Price Movements: A Dynamic Integer Count Data Model." *Empirical Economics 30*, 795–825.

Lunde, A. (2000). "A Generalized Gamma Autoregressive Conditional Duration Model." Discussion paper, Aarlborg University.

Madhavan, A., and S. Smidt. (1993). "An Analysis of Daily Changes in Specialists' Inventories and Quotations." *Journal of Finance 48*, 1595–1628.

Manganelli, S. (2005). "Duration, Volume and Volatility Impact of Trades." *Journal of Financial Markets 8(4)*, 377–399.

Meester, S., and J. MacKay. (1994). "A Parametric Model for Cluster Correlated Categorical Data." *Biometrics 50*, 954–963.

Nelson, D. B. (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Journal of Econometrics 43*, 227–251.

O'Hara, M. (1995). *Market Microstructure*. Oxford: Basil Blackwell.

Patton, A. (2001). "Modelling Time-Varying Exchange Rate Dependence Using the Conditional Copula." Discussion Paper, University of California, San Diego, Department of Economics.

——— (2006). "Estimation of Multivariate Models for Time Series of Possibly Different Lengths." *Journal of Applied Econometrics 21(2)*, 147–173.

Renault, E., and B. Werker. (2006). "Causality Effects in Return Volatility Measures with Random Times." Working Paper, Tilburg University, Center for Economic Research.

Rohatgi, V. K. (1976). *An Introduction to Probability Theory and Mathematical Statistics*. New York: John Wiley.

Roll, R. (1984). "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market." *The Journal of Finance 39(4)*, 1127–1139.

Russell, J. R., and R. F. Engle. (2002). "Econometric Analysis of Discrete-Valued Irregularly-Spaced Financial Transactions Data." revised Version of Discussion Paper 98-10, University of California, San Diego.

Rydberg, T., and N. Shephard. (2003). "Dynamics of Trade-by-Trade Price Movements: Decomposition and Models." *Journal of Financial Econometrics 1*, 2–25.

Shephard, N. (1995). "Generalized Linear Autoregressions." Nuffield College, Oxford.

Sklar, A. (1959). "Fonctions de répartition à n dimensions et leurs marges." *Public Institute of Statistics at the University of Paris 8*, 229–231.

Smidt, S. (1971). "Which Road to an Efficient Stock Market: Free Competition or Regulated Monopoly." *Financial Analysts Journal 27*, 18–20, 64–69.

Stevens, W. (1950). "Fiducial Limits of the Parameter of a Discontinuous Distribution." *Biometrika 37*, 117–129.

Stoll, H. (1978). "The Supply of Dealer Services in Securities Markets." *Journal of Finance 33*, 1133–1151.

Veredas, D., J. Rodriguez-Poo, and A. Espasa. (2002). "On the (Intradaily) Seasonality and Dynamics of a Financial Point Process: A Semiparametric Approach." CORE Discussion Paper 2002/23.

Vergote, O. (2005). "How to Match Trades and Quotes for NYSE Stocks?" Working paper, K.U. Leuven.