Journal of Economic Dynamics & Control ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Economic Dynamics & Control

journal homepage: www.elsevier.com/locate/jedc

Capturing common components in high-frequency financial time series: A multivariate stochastic multiplicative error model

Nikolaus Hautsch *,[1]

Humboldt-Universität zu Berlin, CASE, CFS, QPL, Spandauer Str. 1, D-10099 Berlin, Germany

ARTICLE INFO

Article history:
Received 3 October 2006
Accepted 28 January 2008

JEL classification:
C15
C32
C52

Keywords:
Multiplicative error model
Common factor
Efficient importance sampling
Intra-day trading process

ABSTRACT

We model high-frequency trading processes by a multivariate multiplicative error model that is driven by component-specific observation driven dynamics as well as a common latent autoregressive factor. The model is estimated using efficient importance sampling techniques. Applying the model to 5 min return volatilities, trade sizes and trading intensities from four liquid stocks traded at the NYSE, we show that a subordinated common process drives the individual components and captures a substantial part of the dynamics and cross-dependencies of the variables. Common shocks mainly affect the return volatility and the trade size. Moreover, we identify effects that capture rather genuine relationships between the individual trading variables.

1. Introduction

Numerous empirical studies have documented a strong positive contemporaneous relationship between daily volume and volatility. This observation is consistent with the mixture-of-distribution hypothesis (MDH) pioneered by Clark (1973). The MDH relies on central limit arguments based on the assumption that daily returns consist of the sum of 'large' amounts of intra-daily logarithmic price changes associated with 'pseudo' intra-day equilibria. The assumption that these intra-day price changes are also accompanied by an increased trading volume leads to an extension of Clark's model and implies a positive contemporaneous correlation between daily volume and volatility.[2]

Whereas the employed central limit arguments provide a sensible framework for aggregated (daily) data, they are not applicable on a high-frequency level since the number of underlying 'pseudo' equilibria cannot be large, but converge to zero when we approach the transaction level. Nevertheless, under the assumption that daily volumes and returns, both consisting of intra-day aggregates, are driven by a subordinated common process, the latter should be identifiable also on an intra-day level. Actually, the idea of an underlying (unobservable) information process is consistent with typical asymmetric information-based market microstructure models such as, e.g., those introduced by Glosten and Milgrom (1985) and Easley and O'Hara (1992). In these settings, positive mutual correlations between trading volumes and

* Tel.: +49 30 20935711; fax: +49 30 20935712.
E-mail address: nikolaus.hautsch@wiwi.hu-berlin.de

[1] Institute for Statistics and Econometrics as well as CASE – Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Quantitative Products Laboratory (QPL), Berlin, and Center for Financial Studies (CFS), Frankfurt.

[2] See, e.g., Epps and Epps (1976), Tauchen and Pitts (1983), Lamoureux and Lastrapes (1990), Andersen (1996) or Liesenfeld (2001).

0165-1889/$ - see front matter © 2008 Published by Elsevier B.V.
doi:10.1016/j.jedc.2008.01.009

volatilities arise through interaction among asymmetrically informed market participants. However, market microstructure theory is typically relatively vague, if not silent, regarding the underlying time horizon and thus the frequency on which common information-induced effects should be observable. On the other hand, several empirical studies provide evidence for common movements and strong interdependencies in high-frequency volatilities and trading intensities, supporting the notion of an underlying common component jointly driving trading activity and volatility.[3]

In this study, we aim to analyze whether a common component in volatilities and trading volume is identifiable based not only on daily data but also on higher sampling frequencies. We associate this hypothesis with a 'micro-foundation' of the volume–volatility relationship. In this context, we will answer the following research questions: (i) To what extent do the interdependencies between volume and volatility reflect ('true') causal relationships, or rather spurious correlations due to the subordination to the same latent (information arrival) process? (ii) How strongly does the latent factor affect the individual trading components? In particular, are potential common movements with volatilities reflected in trade sizes, in trading intensities, or both? (iii) To what extent can we identify effects that are not driven by a common subordinated process but that reflect genuine trading-specific effects?

To address these questions, we propose modeling the return volatility, the average trade size, and the number of trades per time interval in terms of a new type of multiplicative error model (MEM) that is driven by two different dynamic processes: a common autoregressive latent factor with component-specific sensitivity and an observation-driven (VARMA type) dynamic capturing idiosyncratic effects given the latent factor. The resulting model is called a *stochastic multiplicative error model* (SMEM) and extends the multiplicative error structures proposed by Engle (2002) and Manganelli (2005) by a latent factor dynamic.

The proposed approach is motivated by two major aspects: First, a well-known result in the literature on tests of the MDH is that a single latent component is typically not sufficient to fully capture the short-run dynamic dependencies in both volume and volatility. As argued in Andersen (1996) and Bollerslev and Jubinski (1999), it is likely that different types of 'news', such as scheduled macroeconomic announcements, option expiration days or company-specific earnings announcements, affect volatility and volume processes differently. For instance, macroeconomic announcements lead to relatively short-lived jumps in volatility but to longer-lasting increases in trading volume. In contrast, earnings announcements are typically accompanied by strong price shifts combined with relatively little trading activity. Including such idiosyncratic effects requires accounting for additional factors. However, instead of allowing for multiple latent factors (e.g., in Liesenfeld, 2001), the SMEM captures these effects in terms of *observation-driven* dynamics. This idea has been suggested by Bauwens and Hautsch (2006) and leads to a still flexible, but computationally less burdensome specification, since only one factor is assumed to be unobservable and must be integrated out.

Second, combining a common latent dynamic with (multivariate) observation-driven components can be seen as a reduced-form description of the trading dynamics generated by a subordinated information process and asymmetrically informed market agents. In asymmetric information-based market microstructure models,[4] uninformed traders infer the existence of information by observing recent trading history. This leads to distinct (cross-) autocorrelation structures between price changes, volume, trading intensities and bid-ask spreads that are tested in a wide range of empirical market microstructure studies.[5] In the SMEM, the latent factor can be interpreted as a proxy for the underlying information process. It is associated with exogenous shocks simultaneously affecting the processes of volatilities, trade sizes and trading intensities. In addition, the observation-driven components capture conditional dynamics and cross-dependencies in volatilities, trade sizes and trading intensities given the latent process. That is, they represent effects that are attributed to genuine relationships between the individual trading components and are not driven by common factor movements.

Though the SMEM cannot be seen as a structural model, it nevertheless allows us to study trading processes in a more structural way than in reduced form descriptions of trading processes, as in Hasbrouck (1991), Dufour and Engle (2000), Engle (2000) or Manganelli (2005). Disentangling the trading dynamics in the proposed form makes it possible to better understand to what extent common information shocks are captured and processed in volatilities, trade sizes and the speed of trading. In this sense, our paper contributes to the literature analyzing the information content of individual trading components. For instance, based on daily data, Jones et al. (1994) show that the positive relationship between volatility and average trade size is statistically insignificant when the impact of the trading intensity on stock return volatility is taken into account. Ané and Geman (2000) show that the speed of trading, rather than the trade size, constitutes a stochastic clock driving a time change that allows recovering the normality of asset returns. In contrast, Xu and Wu (1999), Chan and Fong (2000) and Huang and Masulis (2003) find that the average trade size contains nontrivial information for return volatility, confirming the results by Blume et al. (1994) that trade size indicates the quality of information. Tran (2006) decomposes the return volatility into an erratic factor that is particularly sensitive to new information, as well as a persistent factor. He shows that both the trading intensity *and* the average trade size are positively correlated with the former.

[3] See, e.g., Ané and Geman (2000), Engle (2000), Grammig and Wellner (2002), Renault and Werker (2003), Manganelli (2005), Meddahi et al. (2006) or Bowsher (2007).

[4] See, e.g., Glosten and Milgrom (1985), Admati and Pfleiderer (1988), Easley and O'Hara (1992), Blume et al. (1994) and Easley et al. (1996) among others.

[5] See, e.g., Engle (2000) and Manganelli (2005) or the surveys by Bauwens and Giot (2001) or Hautsch (2004).

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

The SMEM is estimated by simulated maximum likelihood (SML). The computation of the likelihood requires to integrate the latent component out, leading to an integral of the dimension of the sample range. To approximate the likelihood function numerically we suggest using the efficient importance sampling (EIS) algorithm proposed by Richard (1998) and Richard and Zhang (2007). In the given setting, it turns out that the SML-EIS estimation of the SMEM works very efficiently and is computationally feasible.

In the empirical analysis, we use 5 min aggregates from four highly liquid stocks traded at the New York Stock Exchange (NYSE). Strong empirical evidence for the existence of an autoregressive common component is provided. We find that the unobservable factor jointly drives volatility, trade size and trading intensity in the same direction confirming the findings based on daily data (see, e.g., Tauchen and Pitts, 1983; Chan and Fong, 2000; Liesenfeld, 2001). It turns out that a substantial part of the (typically positive) causal effects between the individual variables are driven by a common factor confirming the notion of an underlying information process. Most interestingly, the latent component has a particularly strong effect on the return volatility as well as the average trade size, whereas the trading intensity is only relatively weakly affected. This provides evidence that underlying information seems to be reflected in trade sizes rather than in the speed of trading. Furthermore, we find that most of the simultaneity between volatility and trade size is captured by the joint component. In contrast, the latter can only explain a part of the positive mutual correlation between volatility and trading intensity. Complementing Jones et al. (1994) and Ané and Geman (2000), this illustrates that trading intensity is an important determinant of volatility but this relationship is obviously only partly driven by a subordinated common process. Furthermore, apart from influences of common factor dynamics, we can identify effects that are attributed to genuine relations between the trading components and hold true irrespective of the subordinated process. Finally, it is shown that the inclusion of the latent component clearly improves the goodness-of-fit as well as the dynamic and distributional properties of the model. This illustrates the usefulness of combining observation-driven and parameter-driven dynamics from a statistical viewpoint.

The remainder of the paper is organized in the following way: Section 2 presents the SMEM and Section 3 outlines its statistical properties. In Section 4, we present the statistical inference. Section 5 displays the data and discusses the estimation results. Finally, Section 6 concludes our paper.

2. The multivariate stochastic multiplicative error model

Define $\{Y_i, V_i, \rho_i\}$, $i = 1, \ldots, n$, as the three-dimensional time series associated with the intra-day process of returns, transaction volumes and trading intensities, respectively. In particular, Y_i corresponds to the log return measured over equidistant time intervals (i.e., 5 min intervals), V_i is the average volume per trade in the ith interval and ρ_i is the number of trades occurring during interval i. Furthermore, λ_i is defined as a common unobservable component that simultaneously influences Y_i, V_i and ρ_i and follows an autoregressive process that is updated in each interval i.

Define $W_i = \{w_j\}_{j=1}^{i}$ with $w_i = (Y_i, V_i, \rho_i)'$ and $\Lambda_i = \{\lambda_j\}_{j=1}^{i}$ and denote $\mathscr{F}_i := (W_i, \Lambda_i)$ as the history of the process up to period i. Following Engle (2000) and Manganelli (2005), we propose decomposing the joint conditional density given $\mathscr{F}_{i-1}$, $f(Y_i, V_i, \rho_i, \lambda_i | \mathscr{F}_{i-1})$, into the product of the corresponding conditional densities. Hence,

$$\begin{aligned} f(Y_i, V_i, \rho_i, \lambda_i | \mathscr{F}_{i-1}) &= f(Y_i | V_i, \rho_i, \lambda_i; \mathscr{F}_{i-1}) \cdot f(V_i, \rho_i, \lambda_i; \mathscr{F}_{i-1}) \\ &= f(Y_i | V_i, \rho_i, \lambda_i; \mathscr{F}_{i-1}) \cdot f(V_i | \rho_i, \lambda_i; \mathscr{F}_{i-1}) \cdot f(\rho_i | \lambda_i; \mathscr{F}_{i-1}) \\ &\quad \cdot f(\lambda_i; \Lambda_{i-1}), \end{aligned} \tag{1}$$

where it is assumed that λ_i depends only on its own history Λ_{i-1}. The chosen decomposition implies that V_i is predetermined for Y_i, whereas ρ_i is predetermined for both Y_i and V_i. Finally, ρ_i itself is assumed not to be affected by any contemporaneous variable. Of course, the order of the variables in the decomposition is arbitrary and depends on the research objective. Here, we proceed along the lines of Engle (2000) and Manganelli (2005), who are particularly interested in the volatility process given the contemporaneous volume and the contemporaneous trading intensity.

The basic idea of the SMEM is to combine *observation-driven* dynamics with *parameter-driven* dynamics in a multivariate multiplicative error framework as introduced by Engle (2002) and put forward by Engle and Gallo (2006) and Cipollini et al. (2007). Then, a three-dimensional SMEM for volatilities, trade sizes and trading intensities is given by

$$Y_i = \mathrm{E}[Y_i | \mathscr{F}_{i-1}] + \xi_i, \tag{2}$$

$$\xi_i = \sqrt{h_i \, \mathrm{e}^{\delta_1 \lambda_i} s_{h,i}} \eta_i, \quad \eta_i \sim \text{i.i.d. } N(0,1), \tag{3}$$

$$V_i = \Phi_i \, \mathrm{e}^{\delta_2 \lambda_i} s_{V,i} u_i, \quad u_i \sim \text{i.i.d. } \mathscr{GG}(p_2, m_2), \tag{4}$$

$$\rho_i = \Psi_i \, \mathrm{e}^{\delta_3 \lambda_i} s_{\rho,i} \varepsilon_i, \quad \varepsilon_i \sim \text{i.i.d. } \mathscr{GG}(p_3, m_3), \tag{5}$$

where h_i, Φ_i and Ψ_i denote observation-driven dynamic components, η_i, u_i and ε_i are process-specific innovation terms, which are assumed to be independent, and $s_{h,i}, s_{V,i}, s_{\rho,i} > 0$ capture deterministic time-of-day effects in volatilities, trade sizes, and trading intensities, respectively. We assume that the volatility innovations η_i follow a standard normal distribution whereas the volume and trading intensity innovations u_i and ε_i follow standard generalized gamma distributions with parameters p_2, m_2 and p_3, m_3, respectively. The generalized gamma distribution allows for a high

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

distributional flexibility including the cases of over-dispersion and under-dispersion as well as non-monotonic hazard shapes.[6]

The component $h_i e^{\delta_1 \lambda_i} s_{h,i}$ corresponds to the conditional return variance given $\mathscr{F}_{i-1}$, λ_i, and the time-of-day effect. Accordingly, up to a constant multiplicative factor,[7] $\Phi_i e^{\delta_2 \lambda_i} s_{V,i}$ and $\Psi_i e^{\delta_3 \lambda_i} s_{\rho,i}$ correspond to the conditionally expected volume and the conditionally expected trading intensity given $\mathscr{F}_{i-1}$, λ_i, and the time of day. Hence, the major idea of the SMEM is to model these conditional moments on the basis of a multiplicative interaction of the processes $\{h_i s_{h,i}, \Phi_i s_{V,i}, \Psi_i s_{\rho,i}\}$ and e^{λ_i}. Correspondingly, the parameters δ_1, δ_2 and δ_3 drive the process-specific impact of λ_i.

The common latent factor λ_i is assumed to follow a zero mean AR(1) process, given by

$$\lambda_i = a\lambda_{i-1} + v_i, \quad v_i \sim \text{i.i.d. } N(0,1), \tag{6}$$

where v_i is assumed to be independent of η_i, u_i and ε_i. Then, the process-specific impact of the latent factor is given by $\lambda_{ij} := \delta_j \lambda_i$ with $\lambda_{ij} = a\lambda_{i-1,j} + \delta_j v_i$, and thus $d\lambda_{ij}/dv_i > (<) 0$ for $\delta_j > (<) 0$ with $j = 1,2,3$.[8] Because of the symmetry of the distribution of v_i, the sign of the individual parameters δ_j are not identified without further restrictions. Hence, for instance, we cannot distinguish between the cases $\delta_1 > 0, \delta_2 < 0$ versus $\delta_1 < 0, \delta_2 > 0$. Nevertheless, we can identify whether the latent factor influences the two components in the same direction or in opposite directions. For that reason, we have to restrict the sign of one of the parameters δ_j. Then, the signs of all other coefficients δ_k with $k \neq j$ are identified.

The process-specific components h_i, Φ_i and Ψ_i represent 'genuine' trade-driven effects given the latent factor. They are assumed to follow a multivariate observation-driven dynamic that is parameterized in terms of a VARMA structure

$$\mu_i = \omega + A_0 z_{0,i} + \sum_{j=1}^{P} A_j z_{i-j} + \sum_{j=1}^{Q} B_j \mu_{i-j}, \tag{7}$$

where

$$\mu_i := (\ln h_i, \ln \Phi_i, \ln \Psi_i)', \tag{8}$$

$$z_{0,i} := (0, \ln V_i, \ln \rho_i)', \tag{9}$$

$$z_i := \left(\frac{|\xi_i|}{\sqrt{h_i s_{h,i}}}, \frac{V_i}{\Phi_i s_{V,i}}, \frac{\rho_i}{\Psi_i s_{\rho,i}} \right)' = \left(|\eta_i| e^{\delta_1 \lambda_i/2}, u_i e^{\delta_2 \lambda_i}, \varepsilon_i e^{\delta_3 \lambda_i} \right)', \tag{10}$$

$\omega = \{\omega_k\}$, $k = 1,2,3$, denotes a (3×1) vector, and $A_0 = \{\alpha_0^{kl}\}$ for $k,l = 1,2,3$ is a (3×3) triangular matrix where only the three upper right elements can be nonzero. Furthermore, $A_j = \{\alpha_j^{kl}\}$ and $B_j = \{\beta_j^{kl}\}$ for $k,l = 1,2,3$ are (3×3) matrices of innovation and persistence parameters, respectively. The triangular structure of A_0 reflects the imposed exogeneity assumptions underlying the decomposition of the joint density in (1). The log-linear form ensures the positiveness of the individual processes without imposing additional parameter restrictions. This property eases the estimation of the model particularly when $A_0 \neq 0$ or when additional explanatory variables are included.

According to (10) the process-specific dynamics in (7) are updated based on innovations z_i corresponding to the lagged (de-meaned) returns, volumes and trading intensities, standardized by their corresponding observation-driven components. We choose this specification for four reasons: First, the use of standardized left-hand variables as innovations is quite common in logarithmic multiplicative error specifications and ensures that the stationarity conditions of μ_i (given a) only depend on B_j. This form is chosen, for example, in logarithmic autoregressive conditional duration (Log-ACD) models (see Bauwens and Giot, 2000; Bauwens et al., 2008) or in exponential GARCH models (Nelson, 1991).[9] Secondly, standardizing only by observation-driven components ensures that the latter can be updated without integrating the latent factor out. Such a specification is in the spirit of an *observation*-driven process and clearly eases the estimation of the model since the computation of μ_i does not depend on λ_i. Third, since the latent variable is not integrated out of the innovations, it is clear that the latter implicitly depend on λ_i (see (10)). Hence, a shock in λ_i influences $\{h_i, \Phi_i, \Psi_i\}$ not only in period i, but (through z_i) also in the subsequent periods causing (cross-) autocorrelations between the individual processes. Because of this effect, the common component can generate cross-dependencies between the observation-driven processes h_i, Φ_i and Ψ_i even when $A_0 = A_j = 0$. This will be illustrated in more detail in Section 3 and is an important feature of the model allowing to parsimoniously capture cross-dependencies. Fourth, with this specification, we implicitly assume that conditional expectations of market participants, given the latent factor, are updated based on the *observable* history in volatilities, volumes and intensities. Consequently, the dynamics in μ_i capture (cross-) autocorrelations between volatilities, volumes and intensities, which are attributed to trading-driven effects and hold conditionally on the subordinated common process.

[6] In the given setting, we explicitly refer to the processes ξ_i, V_i and ρ_i. However, generally, the proposed structure can be used for any kind of positive-valued random variable including, for instance, bid-ask spreads or market depths.

[7] Note that the means of u_i and ε_i are not equal to one as long as $m_2, m_3, p_2, p_3 \neq 1$.

[8] Hence, in order to identify the parameters δ_j, $j = 1,2,3$, the latent variance $\text{Var}[v_i]$ is normalized to one.

[9] The specification can be easily extended to allow for nonlinear news responses as discussed in the context of ACD models by Fernandes and Grammig (2006) or Hautsch (2006) or in the context of GARCH models by Engle and Ng (1993) or Hentschel (1995). However, we find that the inclusion of nonlinearity effects does not qualitatively change our results but requires the estimation of significantly more parameters. Therefore, we decided to keep the basic specification.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

In order to illustrate the structure of the model in more detail, assume for simplicity $A_0 = 0$, $P = Q = 1$, $s_{h,i} = s_{V,i} = s_{\rho,i} = 1$, and diagonal parameterizations of A_1 and B_1. Then, the model is rewritten as

$$\xi_i = \sqrt{\tilde{h}_i}\eta_i, \quad \tilde{h}_i = h_i \mathrm{e}^{\delta_1 \lambda_i}, \tag{11}$$

$$V_i = \tilde{\Phi}_i u_i, \quad \tilde{\Phi}_i = \Phi_i \mathrm{e}^{\delta_2 \lambda_i}, \tag{12}$$

$$\rho_i = \tilde{\Psi}_i \varepsilon_i, \quad \tilde{\Psi}_i = \Psi_i \mathrm{e}^{\delta_3 \lambda_i}, \tag{13}$$

where

$$\ln \tilde{h}_i - \delta_1 \lambda_i = \omega_1 + \alpha_1^{11} \frac{|\xi_{i-1}|}{\sqrt{h_{i-1}}} + \beta_1^{11} (\ln \tilde{h}_{i-1} - \delta_1 \lambda_{i-1}), \tag{14}$$

$$\ln \tilde{\Phi}_i - \delta_2 \lambda_i = \omega_2 + \alpha_1^{22} \frac{V_{i-1}}{\Phi_{i-1}} + \beta_1^{22} (\ln \tilde{\Phi}_{i-1} - \delta_2 \lambda_{i-1}), \tag{15}$$

$$\ln \tilde{\Psi}_i - \delta_3 \lambda_i = \omega_3 + \alpha_1^{33} \frac{X_{i-1}}{\Psi_{i-1}} + \beta_1^{33} (\ln \tilde{\Psi}_{i-1} - \delta_3 \lambda_{i-1}). \tag{16}$$

Hence, it is evident that the latent component λ_i can be interpreted as an additional regressor entering the model statically and driven by its own dynamics according to (6).

The SMEM is an extension of the multiplicative error models of Engle (2002) and Manganelli (2005). Whereas Manganelli (2005) assumes process-specific innovations to be mutually uncorrelated, Cipollini et al. (2007) address the problem of specifying multivariate MEMs taking into account contemporaneous correlations between individual processes by means of copulas. The multivariate SMEM proposed in this paper can be seen as an alternative, which captures mutual correlations by the inclusion of a common latent factor. The usefulness of the combination of observation-driven and parameter-driven dynamics is also stressed by Blazsek and Escribano (2005), who propose applying the stochastic conditional intensity model introduced by Bauwens and Hautsch (2006) to model the intensity of patent activities of firms. Koopman et al. (2008) introduce an extension of the model and apply it to model credit rating transitions.

We call the first SMEM component a stochastic GARCH (SGARCH) model, whereas the second and third component is referred to as a stochastic ACD (SACD) model. These specifications nest several model classes. The SGARCH model encompasses a simplified EGARCH specification as well as the stochastic volatility (SV) model proposed by Taylor (1986) and permits both competing models to be tested against each other. In particular, for $\alpha_1^{11} = \beta_1^{11} = 0$, (14) can be rewritten as an SV model, while for $\delta_1 = 0$ it resembles a simplified form of Nelson's (1991) EGARCH model (without news impact asymmetries). Furthermore, for $\alpha_1^{11} \neq 0$ and $\beta_1^{11} = 0$ it can be interpreted as an SV model that is mixed with a further random variable. Accordingly, the SACD model as specified in (15) and (16) nests the SCD model (Bauwens and Veredas, 2004) for $\alpha_1^{22} = \beta_1^{22} = 0$ and $\alpha_1^{33} = \beta_1^{33} = 0$, respectively, the Log-ACD model (Bauwens and Giot, 2000) for $\delta_2 = 0$ and $\delta_3 = 0$, respectively, and, a mixed SCD model for $\alpha_1^{22} \neq 0$, $\beta_1^{22} = 0$ and $\alpha_1^{33} \neq 0$, $\beta_1^{33} = 0$, respectively.

In the univariate case, the SMEM corresponds to a two-factor model that might allow capturing dynamics not only in the first conditional moments but also in higher order conditional moments. In this sense, the SMEM could serve as a valuable alternative to the stochastic volatility duration model proposed by Ghysels et al. (2004). More detailed comparisons of both approaches merit further research.

3. Statistical properties of the model

The dynamic stability of the SMEM is ensured by the stability of the two underlying dynamic components. The strict stationarity of λ_i is guaranteed by $|a| < 1$. In this case, the innovations of the observation-driven dynamics, z_i, consist of products of i.i.d. variates and strictly stationary variables and thus are themselves strictly stationary. Then, the stability of the observation driven VAR(MA) dynamic characterized by (7) to (10) is ensured by the eigenvalues of the characteristic equation implied by B_j, $j = 1, \ldots, Q$, lying inside the unit circle.

The inclusion of the latent component in the model renders the analytical computation of unconditional moments and (cross-)autocorrelation functions relatively difficult. In the following, we analyze the statistical properties of the model based on several simulation studies. For different specifications of the SMEM, we generate 100 sets of 50,000 observations and analyze the distributional and dynamical properties. Tables 1 and 2 show the mean, standard deviation, minimum, maximum, kurtosis and selected quantiles as well as the Ljung and Box (1978) statistic for (univariate) SGARCH and SACD processes under different parameterizations.[10] Table 1 illustrates that the inclusion of a latent component has a strong influence on the standard deviation and the kurtosis as well as on the serial dependence in the second moments of the simulated return process. We observe that processes generated by high parameter values of a and δ_1 imply a high unconditional variance, overkurtosis and fat tails as well as a strong serial dependence in the conditional variance. Because of its mixture structure, the SGARCH process allows for a high distributional and dynamical flexibility and captures the well known statistical properties of typical financial return series.

[10] Since the distribution of returns under the SGARCH process is symmetric and the mean is set to zero, only the quantiles of the right tail of the distribution are shown.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

Table 1
Summary statistics of simulated SGARCH processes with $P = Q = 1$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Parameterization								
ω_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
α_1^1	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100
β_1^1	0.100	0.100	0.100	0.100	0.100	0.100	0.700	0.900
a	0.000	0.100	0.500	0.900	0.900	0.900	0.900	0.500
δ_1	0.000	0.100	0.100	0.100	0.200	0.300	0.300	0.500
Summary statistics								
S.D.	1.046	1.049	1.050	1.062	1.109	1.195	1.341	1.673
Max	4.472	4.539	4.539	5.016	6.477	9.556	11.944	11.615
quant75	0.705	0.705	0.704	0.702	0.695	0.684	0.751	0.997
quant90	1.340	1.341	1.342	1.346	1.368	1.406	1.557	2.012
quant95	1.721	1.725	1.726	1.743	1.809	1.918	2.141	2.706
quant99	2.435	2.454	2.454	2.515	2.741	3.116	3.533	4.255
Kurtosis	3.010	3.046	3.055	3.194	3.798	5.209	5.812	4.488
LB(20)	109.027	112.222	122.304	384.497	2277.973	6131.453	9835.528	1706.835
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Parameterization								
ω_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
α_1^1	0.100	0.100	0.100	0.100	0.100	0.100	0.200	0.500
β_1^1	0.900	0.950	0.950	0.950	0.950	0.950	0.700	0.500
a	0.100	0.100	0.500	0.700	0.900	0.900	0.900	0.900
δ_1	0.500	0.500	0.500	0.300	0.300	0.500	0.500	0.500
Summary statistics								
S.D.	1.617	2.457	2.554	2.416	2.978	6.860	3.206	48.175
Max	9.999	15.461	18.572	15.064	37.967	330.067	181.854	6993.980
quant75	0.999	1.507	1.509	1.499	1.497	1.536	0.836	0.949
quant90	1.980	2.996	3.053	2.961	3.226	4.026	2.078	2.518
quant95	2.634	3.995	4.126	3.928	4.581	6.544	3.255	4.202
quant99	4.036	6.158	6.558	6.014	8.137	15.854	7.237	11.448
Kurtosis	3.974	4.077	4.658	3.961	10.995	692.717	1422.349	15327.022
LB(20)	648.889	1422.047	2974.460	3837.337	29423.516	37688.351	18684.144	3758.646

The simulations are based on 100 sets of 50,000 observations. Evaluated statistics: Standard deviation, maximum, 75%-, 90%-, 95%- and 99%-quantile and kurtosis of the simulated return process as well as the Ljung–Box statistic (associated with 20 lags) for squared returns. The mean return is set to zero.

Similar findings are revealed for simulated SACD processes (Table 2). Again, an increase of the latent parameters a and δ_3 leads to a significant increase in the unconditional variance as well as in the autocorrelations of the resulting process. As for SGARCH processes, it is apparent that a high serial dependence in both the observation-driven component and the parameter-driven component generate distributions with strong fat tail behavior. These effects are amplified further when the Weibull parameter p_3 is greater than one.

Below we study the dynamic properties of multivariate SMEMs. For brevity we concentrate mainly on the impact of the common latent factor on the dynamic properties of the multivariate process, whereas the influence of the observation-driven VARMA component is less interesting. Figs. 1–6[11] show the autocorrelation and cross-autocorrelation functions implied by a two-dimensional SMEM(1,1) for the volatility and the intensity process.[12] Figs. 1–3 show SMEM processes where the latent factor is strongly (positively) autocorrelated ($a = 0.9$). Moreover, we impose diagonal specifications of A_1 and B_1 implying no direct dependencies between h_i and Ψ_i and assume that the autocorrelations of the processes h_i and Ψ_i to be quite weak ($\alpha_1^{ii} = \beta_1^{ii} = 0.1$ for $i = 1, 3$). Nevertheless, we observe that the existence of the latent factor causes distinct autocorrelations in h_i and Ψ_i as well as in Y_i^2 and ρ_i. As described in Section 2, this is because h_i and Ψ_i are updated by innovations z_{i-1}, which jointly depend on an autocorrelated common component λ_{i-1} (recall Eq. (10)). This induces significant serial dependencies in $\{h_i, \Psi_i, Y_i^2, \rho_i\}$, even for values of A_1 and B_1 close to zero. Similarly, the latent factor also

[11] Figs. 1–6 show autocorrelation functions (ACFs) and cross-autocorrelation functions (CACFs) implied by bivariate SMEM(1,1) processes for the return volatility and the trading intensity. The model is specified as a two-dimensional version of the processes as given by (1)–(10) with $s_{h,i} = s_{V,i} = s_{\rho,i} = 1$. From left to right: ACF of λ_i, ACFs of h_i (solid line) and Ψ_i (broken line), ACFs of Y_i^2 (solid line) and ρ_i (broken line), CACFs of Y_i^2 and ρ_i (solid line) as well as of h_i and Ψ_i (broken line). The CACF graphs show the plot of $\mathrm{Corr}(x_{1,i}, x_{2,i-j})$ versus j for $x_{1,i} \in \{Y_i^2, h_i\}$ and $x_{2,i} \in \{\rho_i, \Psi_i\}$. The conditional mean return is set to zero. The simulations are based on 100 sets of 50,000 observations.

[12] Since the volume component is parameterized similarly, it reveals the same interactions with the other processes. For this reason, we refrain from showing the results for the three-dimensional processes.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

Table 2
Summary statistics of simulated SACD processes with $P = Q = 1$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Parameterization								
ω_3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
α_1^3	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100
β_1^3	0.100	0.100	0.100	0.100	0.100	0.700	0.900	0.900
p_3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
m_3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
a	0.000	0.100	0.500	0.900	0.900	0.900	0.500	0.100
δ_3	0.000	0.100	0.100	0.100	0.200	0.200	0.500	0.500
Summary statistics								
Mean	1.124	1.131	1.133	1.167	1.310	1.795	4.585	3.771
S.D.	1.138	1.158	1.166	1.267	1.846	3.095	14.255	6.032
Max	14.371	15.559	15.732	21.927	90.964	185.843	1689.343	327.620
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
quant01	0.011	0.011	0.011	0.011	0.010	0.012	0.025	0.026
quant05	0.057	0.057	0.057	0.056	0.051	0.063	0.130	0.135
quant10	0.117	0.117	0.117	0.115	0.106	0.130	0.272	0.280
quant25	0.320	0.319	0.319	0.314	0.296	0.368	0.782	0.791
quant50	0.773	0.774	0.772	0.769	0.755	0.953	2.105	2.052
quant75	1.552	1.559	1.558	1.577	1.651	2.144	5.029	4.627
quant90	2.591	2.611	2.617	2.709	3.076	4.161	10.433	8.915
quant95	3.387	3.421	3.436	3.618	4.357	6.103	16.006	12.908
quant99	5.255	5.359	5.399	5.920	8.192	12.530	36.328	25.277
LB(20)	615.291	648.092	751.600	2203.408	10739.092	26168.753	22413.452	10965.638
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Parameterization								
ω_3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
α_1^3	0.100	0.100	0.100	0.200	0.500	0.100	0.100	0.100
β_1^3	0.950	0.950	0.950	0.700	0.500	0.700	0.700	0.700
p_3	1.000	1.000	1.000	1.000	1.000	0.800	1.500	5.000
m_3	1.000	1.000	1.000	1.000	1.000	1.200	0.500	0.500
a	0.700	0.100	0.100	0.100	0.500	0.900	0.900	0.900
δ_3	0.300	0.500	0.300	0.300	0.300	0.200	0.200	0.200
Summary statistics								
Mean	11.946	12.470	9.145	2.239	6.210	1.794	1.806	1.788
S.D.	24.246	20.691	11.628	2.939	75.995	3.129	3.614	3.035
Max	1494.324	959.318	292.183	123.230	6500.553	206.091	314.030	175.815
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
quant01	0.071	0.078	0.073	0.018	0.022	0.012	0.012	0.012
quant05	0.365	0.402	0.374	0.094	0.117	0.063	0.063	0.063
quant10	0.760	0.837	0.773	0.194	0.242	0.131	0.131	0.130
quant25	2.170	2.387	2.158	0.539	0.683	0.368	0.369	0.367
quant50	5.753	6.328	5.445	1.350	1.808	0.953	0.955	0.950
quant75	13.495	14.716	11.748	2.879	4.335	2.146	2.153	2.140
quant90	27.540	29.405	21.505	5.208	9.519	4.159	4.182	4.150
quant95	41.859	43.777	30.103	7.253	15.926	6.091	6.128	6.074
quant99	93.118	91.284	54.793	13.248	55.442	12.537	12.607	12.451
LB(20)	49335.797	26192.140	19098.324	8604.206	255.547	25235.396	25247.540	26940.587

The simulations are based on 100 sets of 50,000 observations. Evaluated statistics: Mean, standard deviation, maximum, minimum, 1%-, 5%-, 10%-, 25%-, 50%-, 75%-, 90%-, 95%- and 99%-quantile as well as the Ljung–Box statistic (associated with 20 lags) of ρ_i.

causes distinct *cross*-autocorrelations between both h_i and Ψ_i as well as between Y_i^2 and ρ_i even for diagonal specifications of A_1 and B_1. Clearly, the magnitude of the (cross-) autocorrelations increases with the parameters δ_1 and δ_3 as well as with the persistence of the latent process as driven by a. These illustrations show that a persistent latent component can be the major source of the observed cross-dependencies in the multivariate process. This is one of the main features of the model.

Fig. 3 shows the effect of the parameters δ_1 and δ_3 which have opposite signs. Since a, A_1 and B_1 are unchanged, the autocorrelations in $\{h_i, \Psi_i, Y_i^2, \rho_i\}$ are identical to those shown in Fig. 2. However, since the latent factor influences the two processes in opposite directions, we observe negative CACFs between h_i and Ψ_i as well as between Y_i^2 and ρ_i. Hence, it is shown that under certain parameter constellations, the latent factor can cause positive serial dependencies in Y_i^2 and ρ_i while simultaneously inducing negative *cross*-autocorrelations between the two processes.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

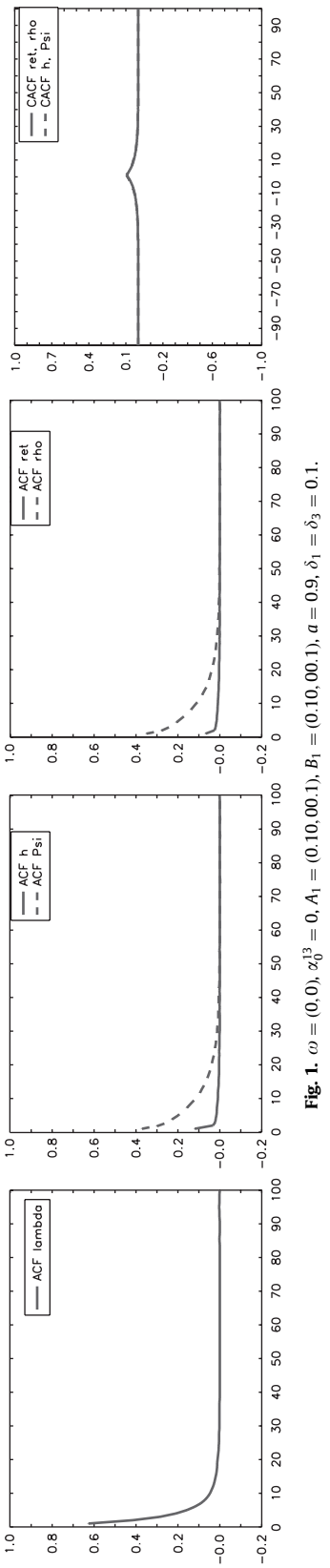

Fig. 1. $\omega = (0,0)$, $\alpha_0^{13} = 0$, $A_1 = (0.10, 00.1)$, $B_1 = (0.10, 00.1)$, $a = 0.9$, $\delta_1 = \delta_3 = 0.1$.

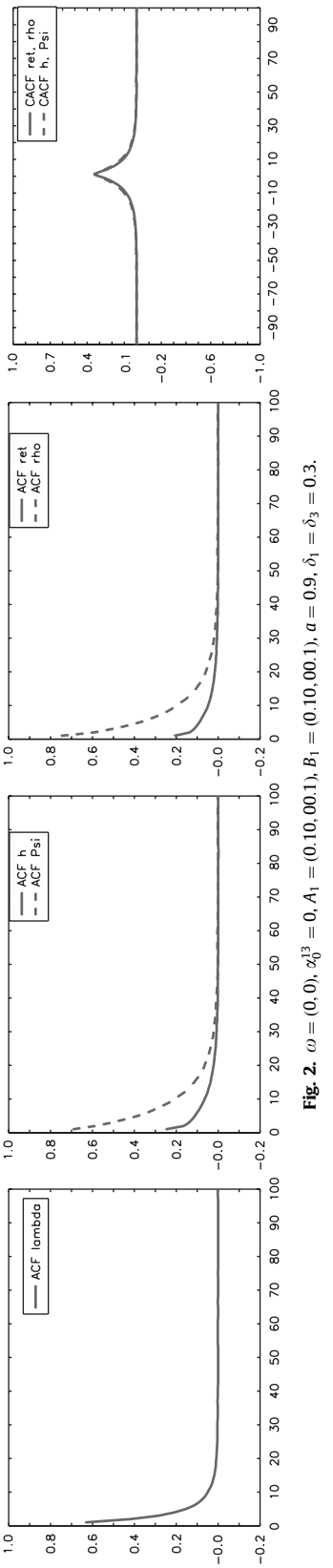

Fig. 2. $\omega = (0,0)$, $\alpha_0^{13} = 0$, $A_1 = (0.10, 00.1)$, $B_1 = (0.10, 00.1)$, $a = 0.9$, $\delta_1 = \delta_3 = 0.3$.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

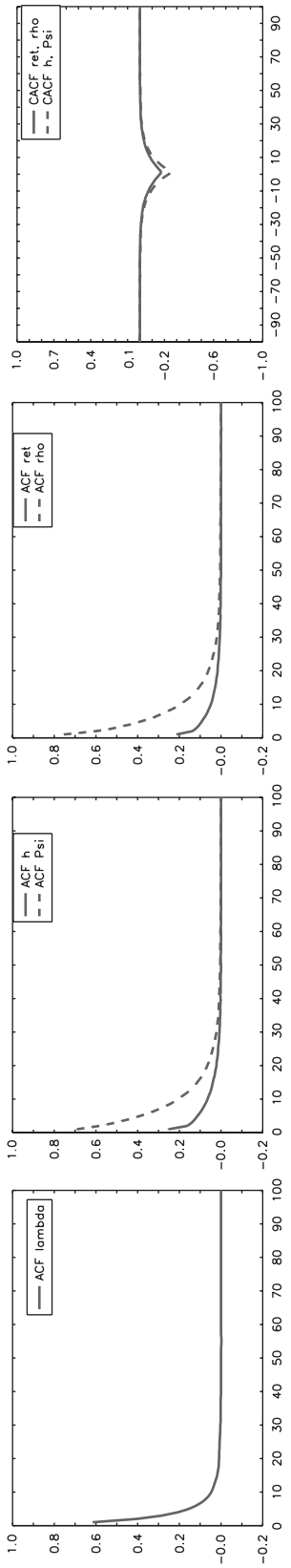

Fig. 3. $\omega = (0,0)$, $\alpha_0^{13} = 0$, $A_1 = (0.10, 00.1)$, $B_1 = (0.10, 00.1)$, $a = 0.9$, $\delta_1 = 0.3$, $\delta_3 = -0.3$.

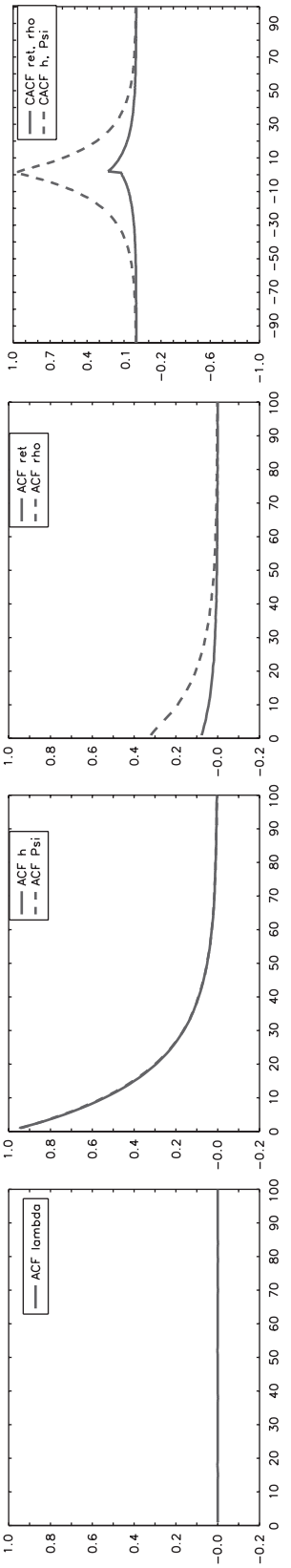

Fig. 4. $\omega = (0,0)$, $\alpha_0^{13} = 0.1$, $A_1 = (0.10.1, 0.10.1)$, $B_1 = (0.70.2, 0.20.7)$, $a = 0$, $\delta_1 = \delta_3 = 0$.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

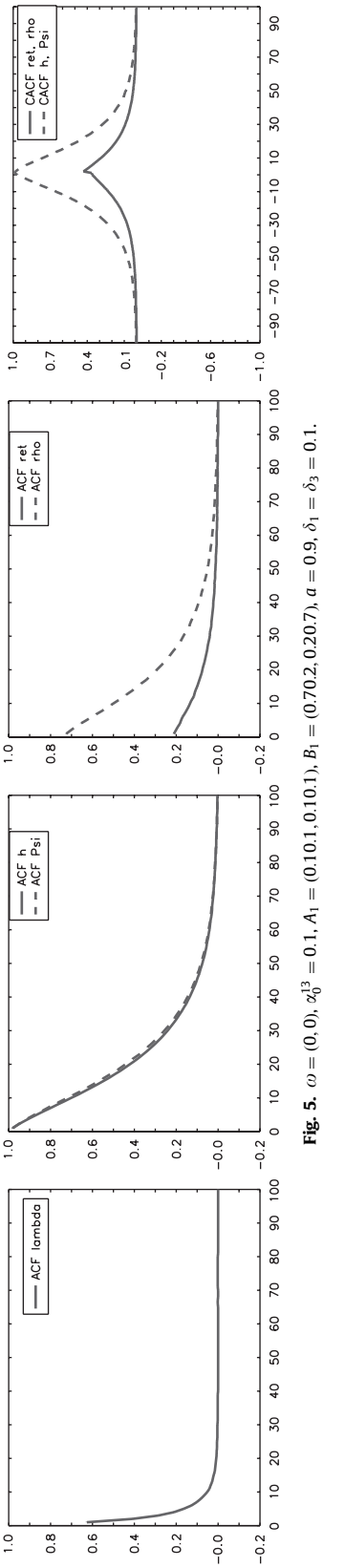

Fig. 5. $\omega = (0,0)$, $\alpha_0^{13} = 0.1$, $A_1 = (0.10.1, 0.10.1)$, $B_1 = (0.70.2, 0.20.7)$, $a = 0.9$, $\delta_1 = \delta_3 = 0.1$.

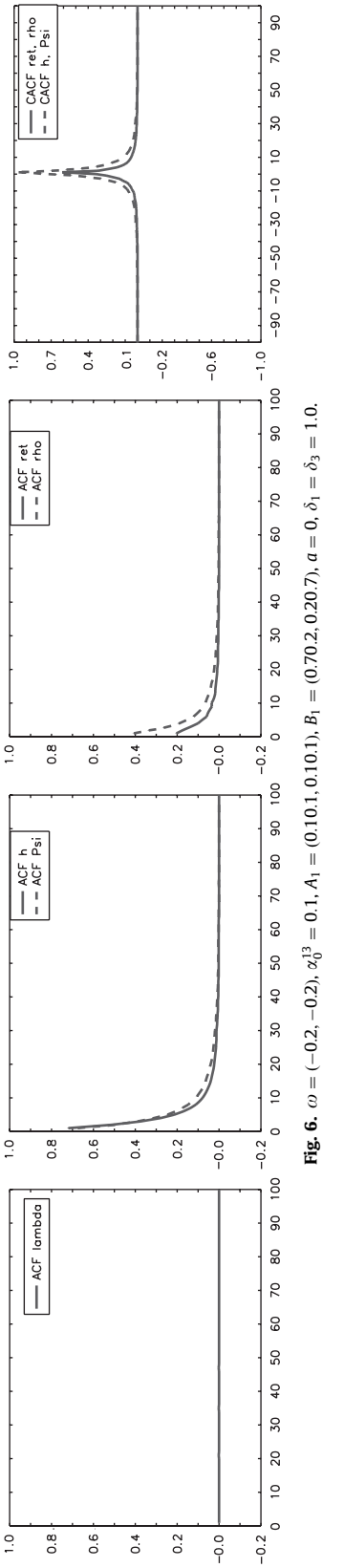

Fig. 6. $\omega = (-0.2, -0.2)$, $\alpha_0^{13} = 0.1$, $A_1 = (0.10.1, 0.10.1)$, $B_1 = (0.70.2, 0.20.7)$, $a = 0$, $\delta_1 = \delta_3 = 1.0$.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

In contrast, Figs. 4 and 5 show SMEM processes where the observation-driven dynamics also reveal distinct cross-dependencies. In Fig. 4, λ_i is set to zero, whereas in Fig. 5, λ_i follows a persistent process with $a = 0.9$ and $\delta_1 = \delta_3 = 0.1$. A comparison of Figs. 4 and 5 demonstrates that the inclusion of the latent factor in Fig. 5 induces a significant increase in the ACF of h_i and Ψ_i, and also in the CACF between Y_i^2 and ρ_i.[13] Hence, if the latent factor is sufficiently strong, it can completely overlay and dominate the multivariate dynamics. Clearly, the strength of this effect depends on the process-specific sensitivities d_1 and d_3.

Finally, Fig. 6 illustrates the effects when the latent factor reveals no serial dependence at all ($a = 0$) but a strong impact on the individual components ($\delta_1 = \delta_3 = 1$). Then, the complete process is effectively overlaid by a white noise variable that clearly reduces the persistence in h_i, Φ_i, Y_i^2 and ρ_i and drives the cross-autocorrelations toward zero.

In summary, we observe that the SMEM is able to capture a wide range of multivariate dynamics arising either from a common underlying component and/or idiosyncratic observation-driven dependencies. Most importantly, we show that the existence of a persistent common latent factor can be a source of distinct (cross-)autocorrelations and contemporaneous correlations in the multivariate process even when there are no (or weak) multivariate observation-driven dynamics. This reflects the idea that an underlying component may indeed be the major driving force of the observed serial (cross-)dependencies in multivariate trading processes. This will be empirically tested in Section 5.

4. Statistical inference

Let W denote the data matrix with $W_i := \{w_j\}_{j=1}^{i}$ and define θ to be the vector of SMEM parameters. The conditional likelihood, given the realizations of the latent variable Λ_i, is given by

$$\begin{aligned}\mathscr{L}(W;\theta|\Lambda_n) = \prod_{i=1}^{n} &\frac{1}{\sqrt{2\tilde{h}_i\pi}}\exp\left[-\frac{\xi_i^2}{2\tilde{h}_i}\right]\frac{p_2 V_i^{p_2 m_2-1}}{\Gamma(m_2)\tilde{\Phi}_i^{p_2 m_2}}\exp\left[-\left(\frac{V_i}{\tilde{\Phi}_i}\right)^{p_2}\right]\\ &\times\frac{p_3\rho_i^{p_3 m_3-1}}{\Gamma(m_3)\tilde{\Psi}_i^{p_3 m_3}}\exp\left[-\left(\frac{\rho_i}{\tilde{\Psi}_i}\right)^{p_3}\right],\end{aligned} \tag{17}$$

where

$$\begin{aligned}\tilde{h}_i &= h_i\,\mathrm{e}^{\delta_1\lambda_i}s_{h,i},\\ \tilde{\Phi}_i &= \Phi_i\,\mathrm{e}^{\delta_2\lambda_i}s_{V,i},\\ \tilde{\Psi}_i &= \Psi\,\mathrm{e}^{\delta_3\lambda}s_{\rho,i}.\end{aligned}$$

Since the latent process is not observable, the conditional likelihood function must be integrated with respect to λ_i using the assumed normal distribution of the latter. Hence, the integrated log likelihood function is given by

$$\begin{aligned}\mathscr{L}(W;\theta) &= \int\prod_{i=1}^{n}\frac{1}{\sqrt{2\tilde{h}_i\pi}}\exp\left[-\frac{\xi_i^2}{2\tilde{h}_i}\right]\frac{p_2 V_i^{p_2 m_2-1}}{\Gamma(m_2)\tilde{\Phi}_i^{p_2 m_2}}\exp\left[-\left(\frac{V_i}{\tilde{\Phi}_i}\right)^{p_2}\right]\\ &\quad\times\frac{p_3\rho_i^{p_3 m_3-1}}{\Gamma(m_3)\tilde{\Psi}_i^{p_3 m_3}}\exp\left[-\left(\frac{\rho_i}{\tilde{\Psi}_i}\right)^{p_3}\right]\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(\lambda_i-\mu_{0,i})^2\right]\mathrm{d}\Lambda\\ &= \int\prod_{i=1}^{n} g(w_i|\lambda_i,W_{i-1};\theta)p(\lambda_i|\Lambda_{i-1};\theta)\,\mathrm{d}\Lambda = \int\prod_{i=1}^{n} f(w_i,\lambda_i|W_{i-1},\Lambda_{i-1};\theta)\,\mathrm{d}\Lambda,\end{aligned} \tag{18}$$

where $\mu_{0,i} := \mathrm{E}[\lambda_i|\Lambda_{i-1}]$, $g(\cdot)$ denotes the conditional density of w_i given (λ_i, W_{i-1}) and $p(\cdot)$ denotes the conditional density of λ_i given Λ_{i-1}. The computation of the n-dimensional integral in (18) is performed numerically using the efficient importance sampling (EIS) method proposed by Richard and Zhang (2007). This algorithm was shown to work quite well in the context of the class of latent factor models (see, e.g., Liesenfeld and Richard, 2003 or Bauwens and Hautsch, 2006).

To implement the EIS algorithm, the integral (18) is rewritten as

$$\mathscr{L}(W;\theta) = \int\prod_{i=1}^{n}\frac{f(w_i,\lambda_i|W_{i-1},\Lambda_{i-1};\theta)}{m(\lambda_i|\Lambda_{i-1},\phi_i)}\prod_{i=1}^{n} m(\lambda_i|\Lambda_{i-1},\phi_i)\,\mathrm{d}\Lambda, \tag{19}$$

where $\{m(\lambda_i|\Lambda_{i-1},\phi_i)\}_{i=1}^{n}$ denotes a sequence of auxiliary importance samplers indexed by auxiliary parameters ϕ_i. Then, the importance sampling estimate of the likelihood is obtained by

$$\mathscr{L}(W;\theta) \approx \hat{\mathscr{L}}_R(W;\theta) = \frac{1}{R}\sum_{r=1}^{R}\prod_{i=1}^{n}\frac{f(w_i,\lambda_i^{(r)}(\phi_i)|W_{i-1},\Lambda_{i-1}^{(r)}(\phi_{i-1});\theta)}{m(\lambda_i^{(r)}(\phi_i)|\Lambda_{i-1}^{(r)}(\phi_{i-1}),\phi_i)}, \tag{20}$$

[13] The asymmetric cross-autocorrelations between Y_i^2 and ρ_i are because the return innovation ξ_i is driven by the square root of h_i, whereas ρ_i is driven by Ψ_i itself.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

where $\{\lambda_i^{(r)}(\phi_i)\}_{i=1}^n$ denotes a trajectory of random draws from the sequence of auxiliary importance samplers m and R such trajectories are generated.

The idea of the EIS approach is to choose a sequence of samplers for $m(\lambda_i|\Lambda_{i-1},\phi_i)$ that exploits the sample information on the λ_is revealed by the observable data. As shown by Richard and Zhang (2007), the EIS principle is to choose the auxiliary parameters $\{\phi_i\}_{i=1}^n$ in a way that provides a good match between $\Pi_{i=1}^n m(\lambda_i|\Lambda_{i-1},\phi_i)$ and $\Pi_{i=1}^n f(w_i,\lambda_i|W_{i-1},\Lambda_{i-1};\theta)$ in order to minimize the Monte Carlo sampling variance of $\hat{\mathscr{L}}_R(W;\theta)$. Richard and Zhang (2007) illustrate that the resulting high-dimensional minimization problem can be split up into solvable low-dimensional subproblems. This makes the approach tractable even for very high dimensions. The detailed EIS procedure is described in Appendix A.

An important advantage that facilitates the computation of the function $f(\cdot)$ is that the time series recursion of the observation-driven components h_i, Φ_i and Ψ_i can be computed without the latent factor being known. As discussed in Section 2, this is because $\{h_i,\Phi_i,\Psi_i\}$ are driven based on innovations z_i that are observable given the history of $\{\xi_i,V_i,\rho_i\}$ and $\{h_i,\Phi_i,\Psi_i\}$. Then, h_i, Φ_i and Ψ_i can be computed in a first step according to the VARMA structure given by (7)–(10) and can be used in a second step to evaluate the sampler $\{m(\lambda_i|\Lambda_{i-1},\phi_i\}_{i=1}^n$.

Filtered estimates of an arbitrary function of λ_i, $\vartheta(\lambda_i)$, given the observable information set up to t_{i-1} are given by

$$\mathrm{E}[\vartheta(\lambda_i)|W_{i-1}]=\frac{\int\vartheta(\lambda_i)p(\lambda_i|W_{i-1},\Lambda_{i-1},\theta)f(W_{i-1},\Lambda_{i-1}|\theta)\,\mathrm{d}\Lambda_i}{\int f(W_{i-1},\Lambda_{i-1}|\theta)\,\mathrm{d}\Lambda_{i-1}}. \tag{21}$$

The integral in the denominator corresponds to the marginal likelihood function of the first $i-1$ observations, $\mathscr{L}(W_{i-1};\theta)$, and can be evaluated on the basis of the sequence of auxiliary samplers $\{m(\lambda_j|\Lambda_{j-1},\hat{\phi}_j^{i-1})\}_{j=1}^{i-1}$ where $\{\hat{\phi}_j^{i-1}\}$ denotes the value of the EIS auxiliary parameters associated with the computation of $\mathscr{L}(W_{i-1};\theta)$ and θ is set equal to its corresponding maximum likelihood estimate. Correspondingly, the numerator is computed by

$$\frac{1}{R}\sum_{r=1}^{R}\left\{\vartheta(\lambda_i^{(r)}(\theta))\prod_{j=1}^{i-1}\left[\frac{f(w_j,\lambda_j^{(r)}(\hat{\phi}_j^{i-1})|W_{j-1},\Lambda_{j-1}^{(r)}(\hat{\phi}_{j-1}^{i-1}),\theta)}{m(\lambda_j^{(r)}(\hat{\phi}_j^{i-1})|\Lambda_{j-1}^{(r)}(\hat{\phi}_{j-1}^{i-1}),\hat{\phi}_j^{i-1})}\right]\right\}, \tag{22}$$

where $\{\lambda_j^{(r)}(\hat{\phi}_j^{i-1})\}_{j=1}^{i-1}$ denotes a trajectory drawn from the sequence of importance samplers associated with $\mathscr{L}(W_{i-1};\theta)$, and $\lambda_i^{(r)}(\theta)$ is a random draw from the conditional density $p(\lambda_i|W_{i-1},\Lambda_{i-1}^{(r)}(\hat{\phi}_{i-1}^{i-1}),\theta)$. The computation of the sequence of filtered estimates $\mathrm{E}[\vartheta(\lambda_i)|W_{i-1}]$, $i=1,\ldots,n$, requires a rerun of the EIS algorithm for every i (from 1 to n). Then, the filtered residuals are given by

$$\hat{\eta}_i=\frac{\hat{\xi}_i}{\sqrt{\hat{h}_i\mathrm{E}[\mathrm{e}^{\delta_1\lambda_i}|W_{i-1}]\hat{s}_{h,i}}}, \tag{23}$$

$$\hat{u}_i=\frac{V_i}{\hat{\Phi}_i\mathrm{E}\,[\mathrm{e}^{\delta_2\lambda_i}|W_{i-1}]\hat{s}_{V,i}}, \tag{24}$$

$$\hat{\varepsilon}_i=\frac{\rho_i}{\hat{\Psi}_i\mathrm{E}[\mathrm{e}^{\delta_3\lambda_i}|W_{i-1}]\hat{s}_{\rho,i}}. \tag{25}$$

5. Empirical results

5.1. Data and descriptive statistics

The empirical study uses transaction data from AOL, Boeing, IBM and JP Morgan stocks traded at the New York Stock Exchange (NYSE). The data is extracted from the Trade and Quote (TAQ) database released by the NYSE and covers a period of 5 months between 02.01.2001 and 31.05.2001.

We choose an aggregation level of 5 min as a trade-off between utilizing a maximum amount of intra-day information on one hand, while ending up with tractable sample sizes on the other hand. In addition, this reduces the influence of too much noise induced by market microstructure effects (like effects due to price-discreteness, split-transactions, liquidity-induced price impacts or irregular spacing in time). Consequently, the resulting time series consist of 8,008 observations of 5-min log midquote returns, the average 5-min trading volume per transaction and the number of trades occurring in each interval. Table 3 shows the mean, standard deviation, minimum, maximum, different quantiles and kurtosis as well as univariate and multivariate Ljung–Box statistics associated with the individual time series. The latter is computed according to Hosking (1980) and is given by

$$MLB(s):=n(n+2)\sum_{j=1}^{s}\frac{1}{n-j}\mathrm{trace}\left(\hat{C}_j'\hat{C}_0^{-1}\hat{C}_j\hat{C}_0^{-1}\right)\overset{a}{\sim}\chi^2_{ks},$$

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

Table 3
Descriptive statistics of log returns (multiplied by 100), squared log returns, average volumes per trade as well as the number of transactions based on 5 min intervals for AOL, Boeing, JP Morgan, and IBM stocks traded at the NYSE. 8008 observations

	AOL				Boeing			
	Returns	Sq. ret.	Avg. vol.	Trades	Returns	Sq. ret.	Avg. vol.	Trades
Mean	0.005	0.143	7084.337	26.424	0.000	0.060	1829.393	19.735
S.D.	0.378	0.403	5979.153	9.537	0.245	0.153	1683.701	8.283
Min	−2.973	0.000	1.000	1.000	−1.680	0.000	1.000	1.000
Max	3.000	9.000	84250.000	75.000	1.854	3.437	24766.666	63.000
q05	−0.556	0.000	1645.000	12.000	−0.391	0.000	450.000	9.000
q10	−0.404	0.001	2131.818	15.000	−0.272	0.000	561.765	10.000
q50	0.000	0.034	5383.333	26.000	0.000	0.013	1327.273	19.000
q90	0.406	0.335	13876.471	39.000	0.269	0.150	3600.000	31.000
q95	0.593	0.581	17995.000	43.000	0.380	0.257	4900.000	35.000
Kurtosis	8.968	–	–	–	7.423	–	–	–
LB(20)	25.129	1132.700	14754.230	9868.857	42.988	1767.233	2878.382	18609.976
MLB(20)			41942.224				35931.744	
	JP Morgan				IBM			
	Returns	Sq. ret.	Avg. vol.	Trades	Returns	Sq. ret.	Avg. vol.	Trades
Mean	0.002	0.099	2960.285	33.070	0.001	0.073	2375.869	41.962
S.D.	0.315	0.374	2685.456	11.204	0.270	0.186	2076.318	12.175
Min	−2.355	0.000	1.000	1.000	−1.668	0.000	1.000	1.000
Max	3.994	15.950	59153.332	78.000	2.000	4.000	45540.000	101.000
q05	−0.476	0.000	747.826	16.000	−0.430	0.000	696.774	24.000
q10	−0.334	0.000	920.000	19.000	−0.310	0.000	841.509	27.000
q50	0.000	0.021	2233.333	32.000	0.000	0.018	1794.595	41.000
q90	0.338	0.229	5729.412	48.000	0.289	0.182	4470.371	58.000
q95	0.479	0.411	7358.824	53.000	0.425	0.308	5906.667	64.000
Kurtosis	15.197	–	–	–	7.464	–	–	–
LB(20)	55.335	1401.054	9011.564	12520.965	26.568	2590.101	19751.120	18070.49
MLB(20)			43873.529				70348.102	

Extracted from the 2001 TAQ database. The sample period was from 02.01.01 to 31.05.01.
The following descriptive statistics are shown: Number of observations, mean, standard deviation, minimum, maximum, 5%-, 10%-, 50%-, 90%-, and 95%-quantile, kurtosis, univariate and multivariate Ljung–Box statistic (computed for squared log returns, volumes and number of trades per interval) associated with 20 lags.

where $k = 3$ denotes the dimension of the process, s the number of lags taken into account, and $\hat{C}_j$ is the jth residual autocovariance matrix. The high Ljung–Box statistics in Table 3 indicate that the 5-min trading data reveal strong serial (cross-)dependencies.

Since we are not particularly interested in the conditional mean function of returns, we reduce the complexity of the model by estimating ξ_i in a separate step as the residuals of an ARMA(1,1) process for the Y_i series.[14] In the next step, we estimate the intra-day seasonality components $s_{h,i}, s_{V,i}, s_{\rho,i}$. A simultaneous estimation of seasonality effects in the SMEM is theoretically possible, however, it considerably increases the computational burden because of the high number of parameters. For this reason, we exploit the multiplicative structure in (3)–(5) and estimate $s_{h,i}, s_{V,i}, s_{\rho,i}$ in a separate step based on cubic spline functions using 30-min nodes.[15] Finally, we use $\xi_i/\sqrt{\hat{s}_{h,i}}$, $V_i/\hat{s}_{V,i}$ and $\rho_i/\hat{s}_{\rho,i}$ to estimate the SMEM.[16]

Figs. 7–16[17] show the empirical autocorrelation and cross-autocorrelation functions of Y_i^2, V_i, and ρ_i as well as $Y_i^2/\hat{s}_{h,i}$, $V_i/\hat{s}_{V,i}$, and $\rho_i/\hat{s}_{\rho,i}$. It turns out that all processes reveal significantly positive autocorrelations with a relatively high persistence. The highest serial dependence is observed for volumes and trading intensities, whereas for the volatility

[14] However, since for all return series the ARMA component is very close to zero, and thus ξ_i is very similar to Y_i, we refrain from showing the estimates here.

[15] The component $s_{h,i}$ is estimated based on squared log returns.

[16] The resulting estimates of $s_{h,i}, s_{V,i}, s_{\rho,i}$ reveal the well-known U-shaped pattern, which is typically found in intra-day trading variables. For brevity, we do not include them in the paper but they are available upon request from the author.

[17] The upper plots are based on the plain series, whereas the lower plots are based on the seasonally adjusted series. The pictures on the left show the ACFs of squared log returns (solid line), average volumes (broken line) and the number of trades (dotted line). The pictures on the right show the CACFs of squared log returns and average volumes (solid line), of squared log returns and the number of trades (broken line), and of average volumes and the number of trades (dotted line). Data was extracted from the 2001 TAQ database. The sample period was from 02.01.01 to 31.05.01.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

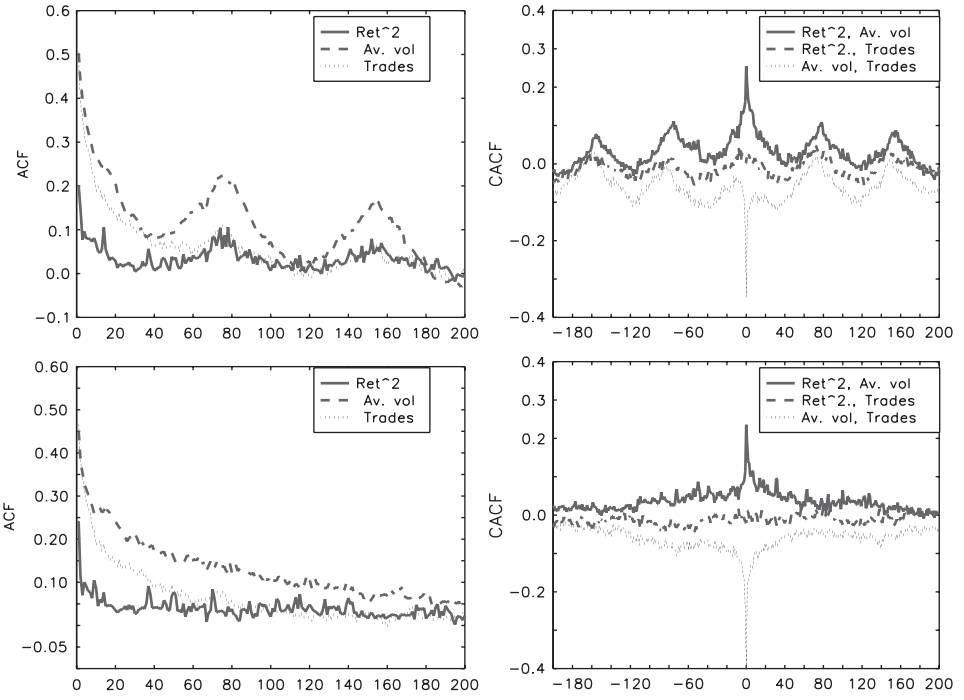

Fig. 7. (Cross-)autocorrelation functions for the AOL stock.

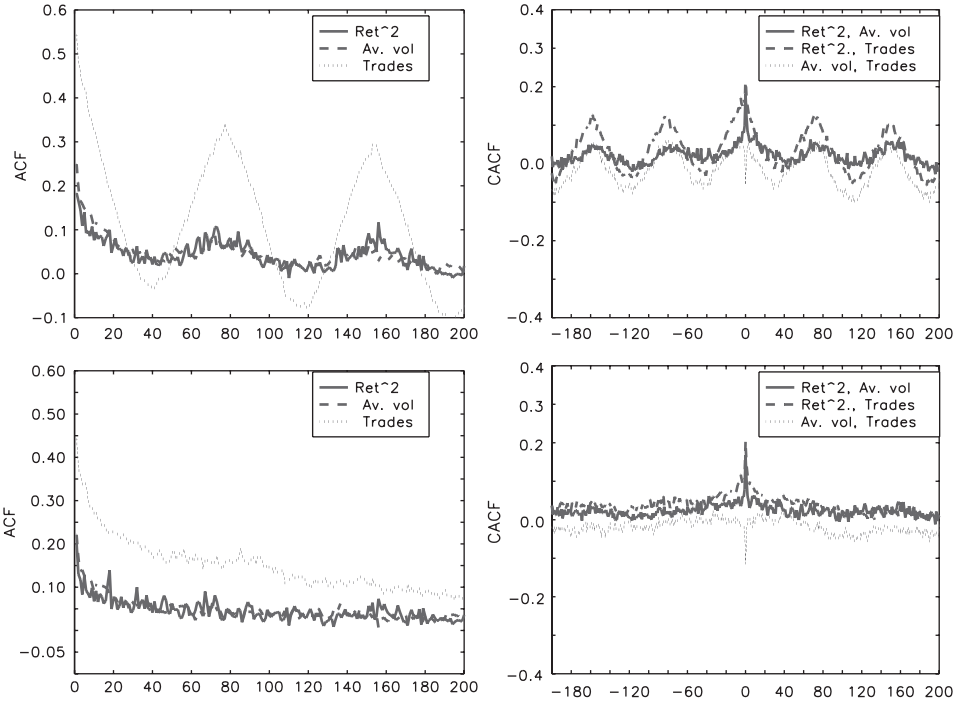

Fig. 8. (Cross-)autocorrelation functions for the Boeing stock.

process, lower autocorrelations are found. Moreover, we find significantly positive cross-autocorrelations between the return volatility and the trading volume, whereas the interdependencies between volatility and trading intensity are relatively weak. In contrast, significantly negative cross-autocorrelations between trading volume and trading intensity are found. Hence, higher volumes obviously enter the market with a lower speed. In Section 5.3, we analyze whether and to what extent these effects are driven by a subordinated common factor and/or by genuine trade-specific effects.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

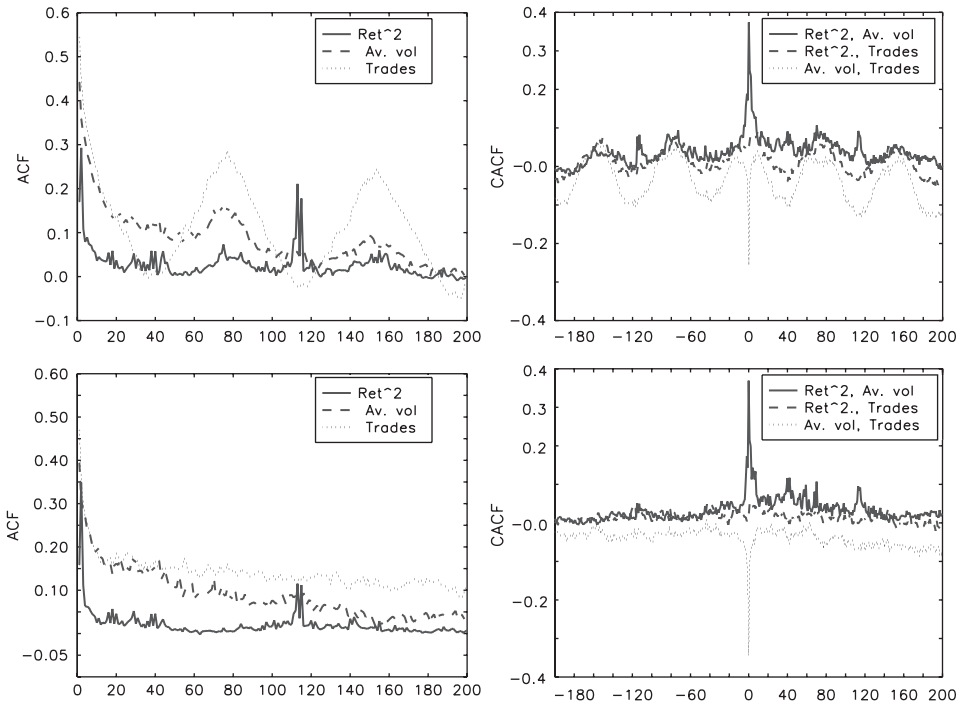

Fig. 9. (Cross-)autocorrelation functions for the JP Morgan stock.

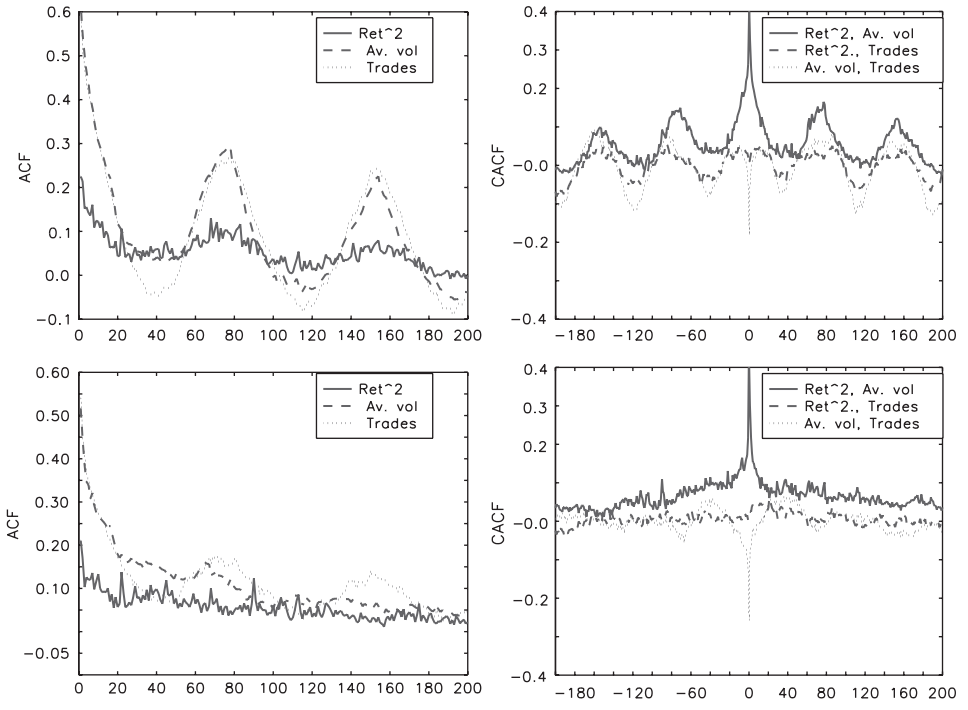

Fig. 10. (Cross-)autocorrelation functions for the IBM stock.

5.2. Estimation results for univariate SMEMs

Tables 4–6 show the estimation results of univariate SGARCH as well as SACD models for 5-min volatilities, trading volumes and trading intensities. To restrict the computational effort, we restrict the class of considered models to

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

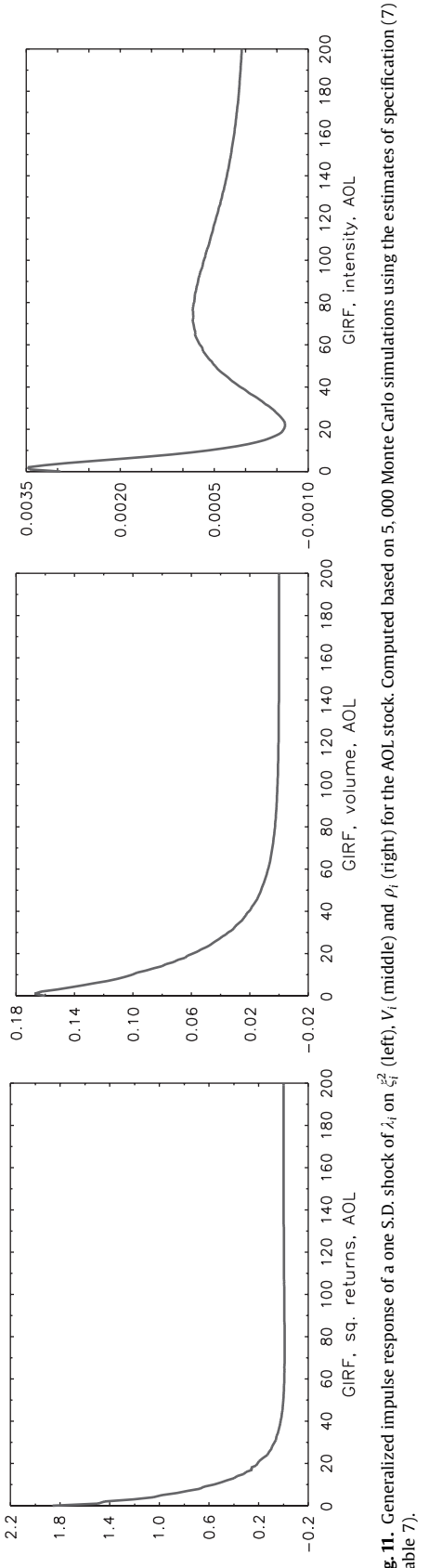

Fig. 11. Generalized impulse response of a one S.D. shock of λ_i on ξ_i^2 (left), V_i (middle) and ρ_i (right) for the AOL stock. Computed based on 5,000 Monte Carlo simulations using the estimates of specification (7) (Table 7).

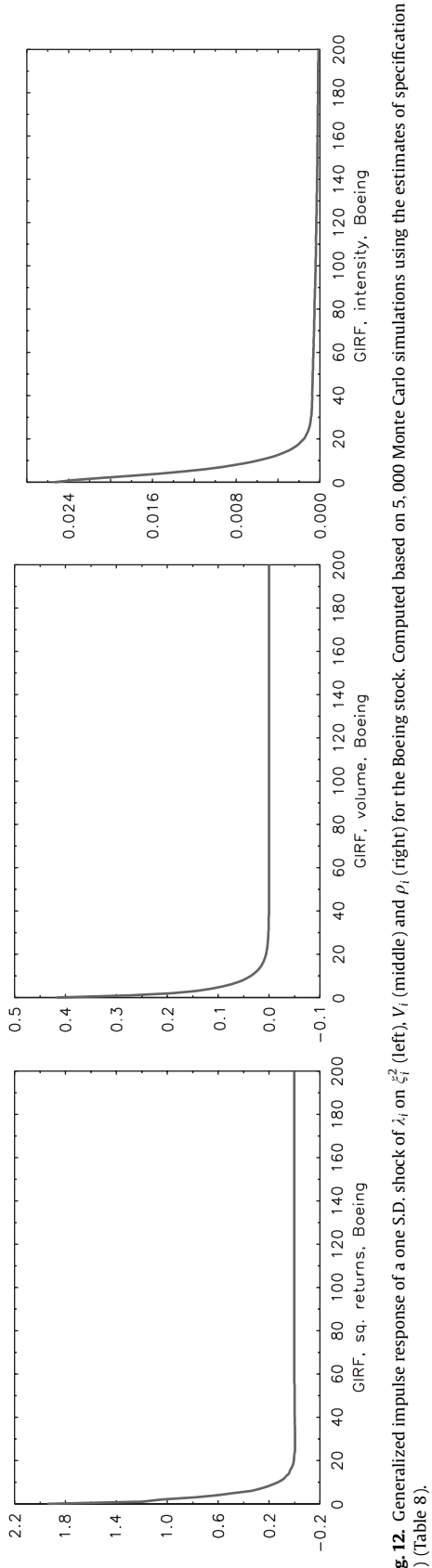

Fig. 12. Generalized impulse response of a one S.D. shock of λ_i on ξ_i^2 (left), V_i (middle) and ρ_i (right) for the Boeing stock. Computed based on 5,000 Monte Carlo simulations using the estimates of specification (7) (Table 8).

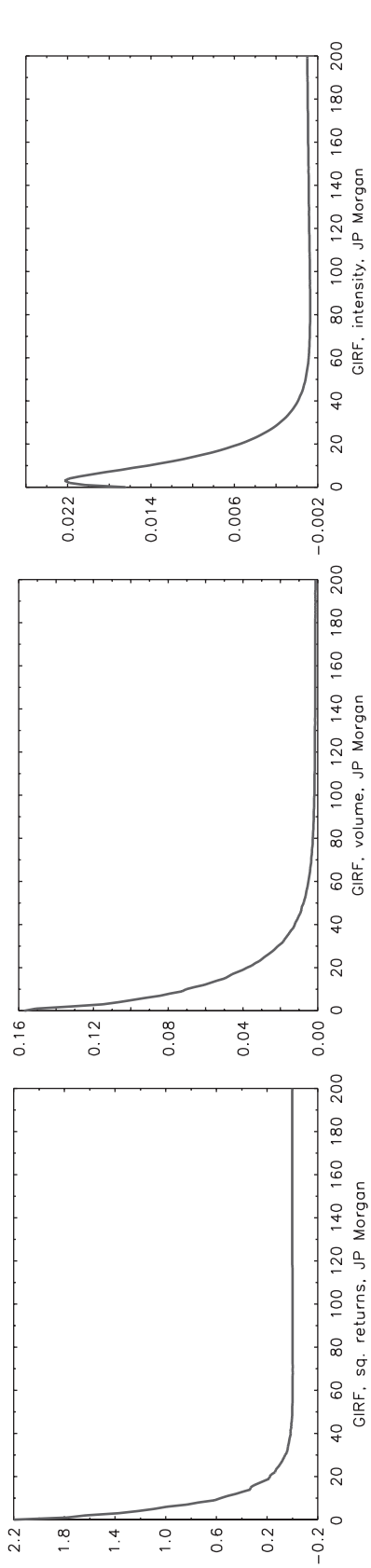

Fig. 13. Generalized impulse response of a one S.D. shock of λ_i on ξ_i^2 (left), V_i (middle) and ρ_i (right) for the JP Morgan stock. Computed based on 5,000 Monte Carlo simulations using the estimates of specification (7) (Table 9).

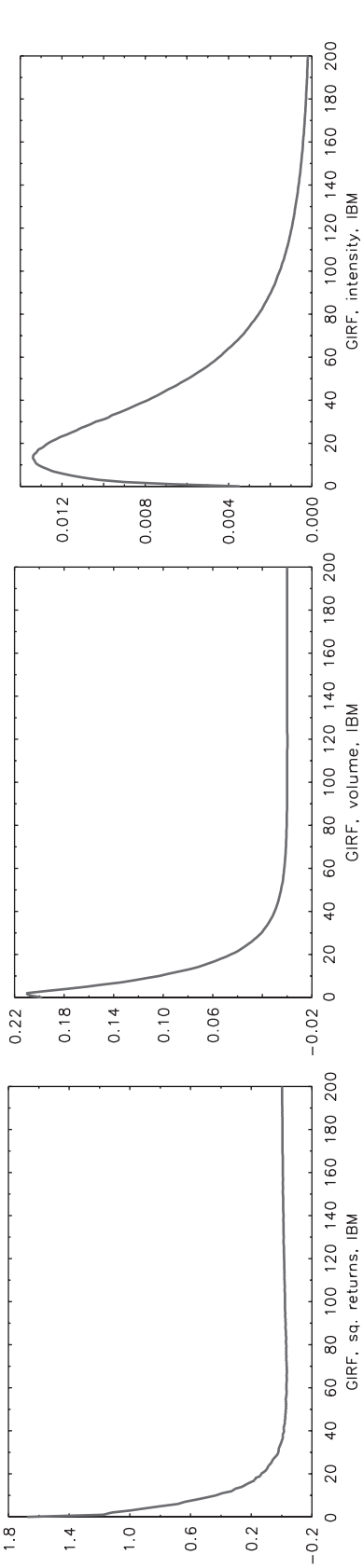

Fig. 14. Generalized impulse response of a one S.D. shock of λ_i on ξ_i^2 (left), V_i (middle) and ρ_i (right) for the IBM stock. Computed based on 5,000 Monte Carlo simulations using the estimates of specification (7) (Table 10).

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

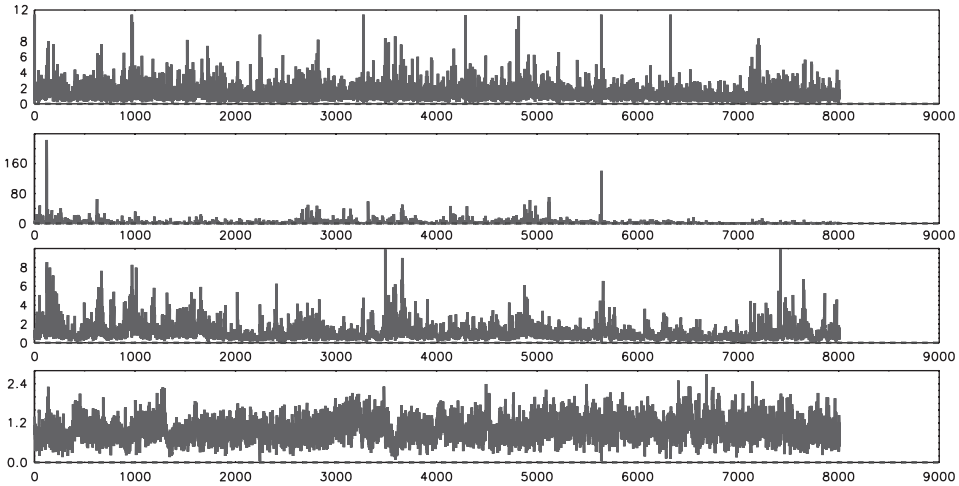

Fig. 15. Plotted (filtered) estimates of $\exp(\lambda_i)$, Y_i^2, V_i and ρ_i (from top to down) for the AOL stock. Computed based on the estimates of specification (7) (Table 7).

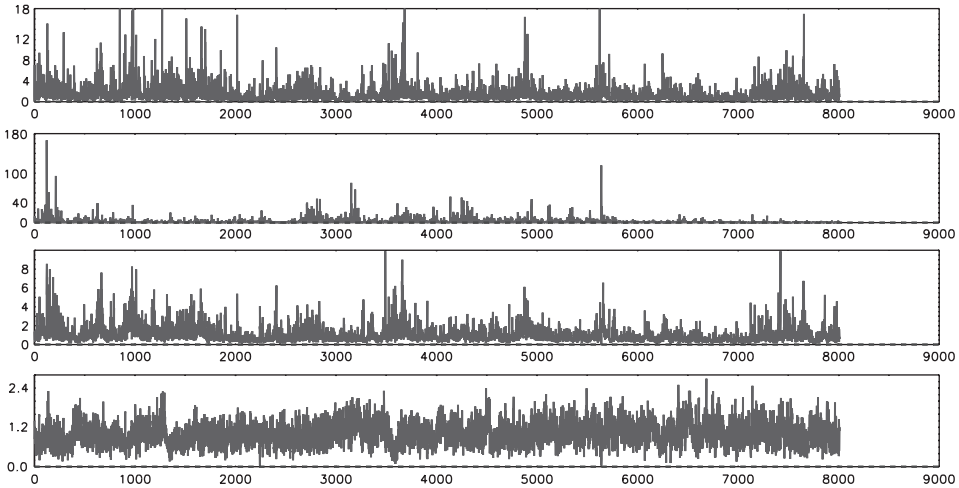

Fig. 16. Plotted (filtered) estimates of $\exp(\lambda_i)$, Y_i^2, V_i and ρ_i (from top to down) for the IBM stock. Computed based on the estimates of specification (7) (Table 10).

specifications with a maximal lag order of two. For all processes and all stocks, we find significant evidence for the existence of a persistent latent component. As revealed by the estimates of the parameter a, the strongest serial dependence in the latent component is observed for the volatility and trading intensity, whereas it is lower for trading volumes. It turns out that the parameter-driven dynamic and the observation-driven dynamic interact. In particular, a declines when observation-driven dynamics are included. Accordingly, in the observation-driven component, the innovation parameter declines and the persistence parameter is driven toward one when the latent factor is taken into account. Hence, news enter the model primarily through the latent component, which is in line with the idea that the latter serves as a proxy for the unobserved information process. Furthermore, it is shown that the inclusion of the unobservable factor increases the goodness-of-fit as well as the dynamical properties of the model. Actually, for the volatility and the volume processes, a pure parameter-driven dynamic in the form of an SV or SCD specification (column (3)), respectively, outperforms a pure observation-driven dynamic in the form of an EGARCH or Log-ACD specification (column (2)), respectively. Nevertheless, we observe that neither the parameter-driven component nor the observation-driven component can be rejected. Hence, for nearly all time series, the best goodness-of-fit is obtained by specifications (4) or (5), which include *both* types of dynamics. This result illustrates that the dynamics in volatilities, volumes and trading intensities are not sufficiently captured by a one-factor model but rather by a two-factor model. This observation is in line with the findings by Ghysels et al. (2004) on the basis of a stochastic volatility duration model.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

Table 4
Maximum likelihood efficient importance sampling (ML-EIS) estimates of different parameterizations of (S)GARCH models up to a lag order of $P = Q = 2$ for 5 min log returns based on the AOL, Boeing, JP Morgan and IBM stocks traded at the NYSE

	AOL					Boeing				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
ω_1	−0.098***	−0.001***	0.314*	−0.028***	−0.006	−0.091***	−0.154***	0.3092***	−0.035***	−0.011**
α_1^1	0.140***	0.197***		0.034***	−0.035	0.136***	0.166***		0.042***	−0.064***
α_2^1		−0.195***			0.043*		0.060***			0.077***
β_1^1	0.987***	1.917***		0.997***	1.718***	0.982***	0.143***		0.996***	1.585***
β_2^1		−0.918***			−0.719***		0.830***			−0.586***
Latent Component										
a			0.961***	0.768***	0.830***			0.941***	0.675***	0.775***
δ_1			0.229***	0.428***	0.403***			0.291***	0.553***	0.513***
Diagnostics										
LL	−13343	−13278	−13115	−13060	−13057	−13456	−13449	−13145	−13217	−13151
BIC	−13357	−13300	−13129	−13082	−13088	−13469	−13471	−13177	−13230	−13173
Mean	−0.002	−0.004	−0.003	−0.005	−0.004	0.008	0.009	0.008	0.008	0.008
S.D.	1.000	1.001	1.030	1.021	1.022	1.000	1.000	1.037	1.019	1.020
LB	16.047	16.075	18.404	15.957	16.191	27.207	27.38	24.4236	24.664	24.046
LB2	58.640***	37.563**	23.961	35.679**	37.821***	61.011***	56.599***	28.739*	29.936*	38.530***
	JP Morgan					IBM				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
ω_1	−0.149***	−0.139***	0.283**	−0.013***	−0.003***	−0.121***	−0.009***	0.323	−0.030***	−0.045**
α_1^1	0.224***	0.260***		0.016***	−0.031***	0.168***	0.228***		0.037***	0.003
α_2^1		−0.052			0.035***		−0.215***			0.051*
β_1^1	0.967***	0.972***		0.998***	1.729***	0.986***	1.826***		0.997***	0.593
β_2^1		−0.000			−0.730***		−0.827***			0.402
Latent component										
a			0.951***	0.860***	0.876***			0.981***	0.860***	0.847***
δ_1			0.275***	0.379***	0.384***			0.167***	0.302***	0.320***
Diagnostics										
LL	−13351	−13349	−13071	−13028	−13023	−13040	−12998	−12883	−12849	−12848
BIC	−13365	−13371	−13085	−13050	−13055	−13054	−13021	−12896	−12871	−12879
Mean	−0.004	−0.004	−0.005	−0.006	−0.006	0.001	0.000	0.001	0.000	0.000
S.D.	1.000	1.000	1.029	1.025	1.020	1.000	1.002	1.017	1.013	1.013
LB	24.978	25.265	25.793	26.419	26.311	26.761	24.070	27.162	24.154	24.141
LB2	21.383	21.585	21.984	27.806	26.033	51.424***	17.027	26.663	11.659	14.762

Data extracted from the 2001 TAQ database. The sample period was from 02.01.01 to 31.05.01. Overnight returns are excluded. The models are re-initialized on each trading day. Standard errors are computed based on the inverse of the estimated Hessian. The ML-EIS estimates are computed using $R = 50$ Monte Carlo replications based on 5 EIS iterations.
Diagnostics: Log likelihood function (LL), Bayes Information Criterion (BIC), mean, standard deviation and Ljung–Box statistic (LB) of the filtered residuals as well as Ljung–Box statistic (LB2) of the squared filtered residuals. The Ljung–Box statistics are computed based on 20 lags.

5.3. Estimation results for multivariate SMEMs

Tables 7–10 give the estimation results for multivariate SMEMs including all three trading components. In order to identify the sign of the parameters δ_j, we restrict δ_1 to be positive. As in the univariate models, we ensure model parsimony by restricting the maximal lag order to two. In addition, we restrict A_2 and B_2 to be diagonal matrices. The major findings can be summarized as follows:

(i) We find significant evidence for the existence of a latent common component with an autoregressive parameter that is on average around $\hat{a} \approx 0.94$. Hence, common shocks are relatively persistent over time which is in accordance with corresponding results based on daily data (see, e.g., Bollerslev and Jubinski, 1999). Obviously, the latent factor seems to

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

Table 5
Maximum likelihood efficient importance sampling (ML-EIS) estimates of different parameterizations of (S)ACD models up to a lag order of $P = Q = 2$ for 5 min average trading volumes per trade based on the AOL, Boeing, JP Morgan and IBM stocks traded at the NYSE

	AOL					Boeing				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
ω_2	−0.384***	−0.180***	−3.291***	−0.080***	−0.037***	−0.403***	−0.416***	−5.841***	−0.070***	−0.057***
α_1^2	0.007***	0.013***		0.018***	−0.013***	0.001***	0.002***		0.011***	−0.006**
α_2^2		−0.007***			0.021***		−0.000			0.014***
β_1^2	0.943***	1.216***		0.986***	1.340***	0.928***	0.761***		0.981***	1.074***
β_2^2		−0.240***			−0.349***		0.165			−0.091
p_2	0.636***	0.682***	0.715***	1.315***	1.342***	0.517***	0.519***	0.450***	1.202***	1.188***
m_2	7.916***	6.974***	9.380***	4.609***	4.672***	8.032***	7.991***	12.753***	4.276***	4.321***
Latent component										
a			0.934***	0.500***	0.654***			0.954***	0.260***	0.391***
δ_2			0.180***	0.373***	0.363***			0.113***	0.506***	0.494***
Diagnostics										
LL	−4896	−4859	−4665	−4594	−4568	−6211	−6200	−6019	−5954	−5943
BIC	−4919	−4890	−4688	−4626	−4608	−6234	−6232	−6041	−5985	−5983
Mean	1.003	1.002	1.009	1.009	1.015	1.011	1.010	1.017	1.024	1.027
S.D.	0.655	0.648	0.661	0.650	0.656	0.871	0.870	0.890	0.901	0.909
LB	78.440***	24.061	69.079***	30.044*	19.679	39.166***	23.256	38.896***	19.486	15.183
	JP Morgan					IBM				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
ω_2	−0.403***	−0.307***	−3.420***	−0.072***	−0.139***	−0.435***	−0.206***	−2.440***	−0.087***	−0.060***
α_1^2	0.004***	0.007***		0.018***	0.021***	0.008***	0.015***		0.022***	−0.002
α_2^2		−0.002***			0.020***		−0.008***			0.018***
β_1^2	0.932***	0.915***		0.984***	−0.016***	0.929***	1.242***		0.980***	1.194***
β_2^2		0.034			0.983***		−0.273***			−0.210
p_2	0.591***	0.614***	0.666***	1.388***	1.475***	0.692***	0.737***	0.888***	1.462***	1.531***
m_2	7.844***	7.314***	8.772***	3.950***	3.458***	8.696***	7.740***	7.894***	4.445***	4.371***
Latent component										
a			0.919***	0.398***	0.419***			0.932***	0.512***	0.590***
δ_2			0.182***	0.429***	0.424***			0.159***	0.314***	0.317***
Diagnostics										
LL	−5454	−5429	−5260	−5164	−5158	−4229	−4201	−4038	−3973	−3956
BIC	−5477	−5460	−5282	−5196	−5199	−4251	−4232	−4060	−4004	−3996
Mean	1.005	1.005	1.012	1.014	1.019	1.001	1.001	1.0062	1.009	1.007
S.D.	0.741	0.738	0.752	0.751	0.745	0.570	0.563	0.5704	0.569	0.570
LB	37.492**	10.981	31.336*	28.924*	26.131	74.995***	24.832	73.2686***	35.089**	24.806

Data extracted from the 2001 TAQ database. The sample period was from 02.01.01 to 31.05.01. Overnight observations are excluded. The models are re-initialized on each trading day. Standard errors are computed based on the inverse of the estimated Hessian. The ML-EIS estimates are computed using $R = 50$ Monte Carlo replications based on 5 EIS iterations.
Diagnostics: Log likelihood function (LL), Bayes Information Criterion (BIC), mean, standard deviation and Ljung–Box statistic (LB) of the filtered residuals. The Ljung–Box statistics are computed based on 20 lags.

capture common long-run dependence, which is not easily covered even by highly parameterized observation-driven dynamics. This result is remarkably robust over all individual specifications.

(ii) The estimated parameters δ_1, δ_2 and δ_3 are significantly positive indicating that common (information) shocks simultaneously increase volatility, average trading volume and trading intensity. It turns out that the joint factor influences primarily the volatility and trade size, whereas its impact on the trading intensity is comparably weak. Hence, based on high-frequency data, we find that a subordinated common (information) process is obviously more strongly reflected in the average trade size rather than in the trading intensity. In this sense our findings confirm the results by Xu and Wu (1999), Chan and Fong (2000) and Huang and Masulis (2003) as well as those by Blume et al. (1994) showing that trade size is an important indicator for the quality of news.

(iii) The inclusion of the latent factor leads to a significant decline of the magnitude of the parameters α_0^{12} and α_0^{13}. This indicates that the *conditional* contemporaneous correlations between volatilities, volumes and trading intensities

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

Table 6
Maximum likelihood efficient importance sampling (ML-EIS) estimates of different parameterizations of (S)ACD models up to a lag order of $P = Q = 2$ for the number of trades in 5 min intervals based on AOL, Boeing, JP Morgan and IBM stocks traded at the NYSE

	AOL					Boeing				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
ω_3	−0.307***	−0.060***	−0.226***	−0.055***	−0.047***	−0.202***	−0.025***	−0.426***	−0.035***	−0.042***
α_1^3	0.146***	0.177***		0.053***	0.104***	0.077***	0.105***		0.026***	0.046***
α_2^3		−0.143***			−0.061***		−0.092***			−0.018
β_1^3	0.892***	1.554***		0.972***	0.940***	0.953***	1.704***		0.991***	0.562***
β_2^3		−0.571***			0.035		−0.708***			0.427*
p_3	1.910***	2.016***	2.800***	3.695***	3.241***	1.632***	1.752***	2.170***	2.563***	2.366***
m_3	2.970***	2.707***	2.002***	1.326***	1.524***	3.674***	3.237***	2.639***	2.100***	2.317***
Latent component										
a			0.913***	0.731***	0.793***			0.957***	0.766***	0.821***
δ_3			0.097***	0.128***	0.103***			0.067***	0.110***	0.092***
Diagnostics										
LL	−1707	−1680	−1697	−1661	−1652	−1982	−1957	−1975	−1925	−1921
BIC	−1729	−1711	−1720	−1693	−1693	−2005	−1988	−1998	−1956	−1961
Mean	1.000	0.999	1.006	1.002	1.001	0.999	0.999	1.002	1.001	1.002
S.D.	0.308	0.307	0.311	0.309	0.308	0.323	0.322	0.324	0.322	0.322
LB	69.340***	37.517**	74.685***	40.070***	28.066	55.018***	33.285**	54.158***	36.095**	24.503
	JP Morgan					IBM				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
ω_3	−0.339***	−0.013***	−0.018	−0.013***	−0.011**	−0.372***	−0.167***	−0.226***	−0.4510***	−0.022**
α_1^3	0.184***	0.212***		0.014***	0.144***	0.150***	0.183***		0.1223***	0.056***
α_2^3		−0.203***			−0.134***		−0.107***			−0.081***
β_1^3	0.847***	1.629***		0.997***	1.252***	0.908***	1.282***		0.3125***	0.811***
β_2^3		−0.631***			−0.254***		−0.316***			−0.009
p_3	2.234***	2.429***	4.174***	4.561***	3.307***	2.121***	2.220***	3.758***	2.8989***	3.436***
m_3	2.718***	2.386***	1.316***	1.214***	1.702***	4.762***	4.406***	2.492***	3.2868***	2.758***
Latent component										
a			0.873***	0.739***	0.783***			0.913***	0.9506***	0.951***
δ_3			0.105***	0.126***	0.082***			0.083***	0.0464***	0.068***
Diagnostics										
LL	−966	−869	−969	−877	−854	948	987	985	1021	1022
BIC	−988	−901	−991	−909	−894	926	955	962	989	981
Mean	1.000	1.000	1.003	1.002	0.998	1.000	0.999	1.002	1.000	1.003
S.D.	0.277	0.273	0.279	0.276	0.274	0.218	0.217	0.219	0.217	0.218
LB	139.049***	46.393***	138.346***	69.345***	30.356*	87.077 ***	10.400	92.165***	10.655	20.401

Data extracted from the 2001 TAQ database. The sample period was from 02.01.01 to 31.05.01. Overnight observations are excluded. The models are re-initialized on each trading day. Standard errors are computed based on the inverse of the estimated Hessian. The ML-EIS estimates are computed using $R = 50$ Monte Carlo replications based on 5 EIS iterations.
Diagnostics: Log likelihood function (LL), Bayes Information Criterion (BIC), mean, standard deviation and Ljung–Box statistic (LB) of the filtered residuals. The Ljung–Box statistics are computed based on 20 lags.

given the latent component are lower than the corresponding *unconditional* ones. Consequently, a significant part of the mutual relations between the conditional return variance and the average trade size as well as the trading intensity actually do arise because of the existence of a common component. This finding confirms the well-known volume–volatility relation found for daily data and can be seen as a 'micro-foundation' of the latter. In particular, the positive relation between return volatility and trade size is nearly completely captured by the common component. Nevertheless, we find a significantly positive correlation between the trading intensity and the volatility, which is only partly explained by the subordinated process. Hence, while we can confirm that trading intensity and intra-day volatility are closely connected and support the results reported by Ané and Geman (2000), our results also indicate that this relationship is obviously only partly due to a common (information) process. Actually, we also observe genuine trade-specific effects, which remain valid irrespective of the latent factor. They reflect that a high liquidity

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

Table 7
Maximum likelihood efficient importance sampling (ML-EIS) estimates of different parameterizations of SMEM specifications up to a lag order of $P = Q = 2$ models for the log return volatility, the average volume per trade and the number of trades per 5 min interval for the AOL stock traded on the NYSE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ω_1	−0.419***	−0.041	−0.540***	0.592***	−0.062***	0.271***	0.308*
ω_2	−0.933***	−2.236***	−1.116***	−2.091***	−1.627***	−1.879***	−1.877***
ω_3	−0.265***	−0.062***	−0.023***	−0.286***	−0.307***	−0.057***	−0.076***
α_0^{12}	0.082***	0.789***	0.255***	0.365***	−0.038***	0.080***	0.064***
α_0^{13}	−0.144***	0.714***	0.201***	0.692***	−0.032***	0.112***	0.130**
α_0^{23}	−0.739***	−0.669***	−0.762***	−0.589***	−0.778***	−0.785***	−0.784***
α_1^{11}	0.136***	0.331***	0.165***		0.064***	0.110***	0.091***
α_1^{12}	0.010*	0.047***	0.030***			0.002	−0.004
α_1^{13}	0.010***	0.000	−0.004***			0.000	0.002
α_1^{21}	−0.001***	0.007***	0.001			−0.030***	−0.026***
α_1^{22}	0.018	0.014***	0.011***		0.003*	0.005***	0.009***
α_1^{23}	0.000***	0.000**	0.000***			0.000	0.000
α_1^{31}	0.156	0.028	−0.121***			−0.065*	−0.025
α_1^{32}	0.328***	−0.109***	0.020			−0.008	−0.015
α_1^{33}	0.177***	0.183***	0.228***		0.146***	0.176***	0.170***
α_2^{11}		0.293***	0.134***			0.053***	0.032*
α_2^{22}		0.010***	0.006***			0.001	−0.001
α_2^{33}		−0.151***	−0.212***			−0.144***	−0.145***
β_1^{11}	0.974***	−0.137***	0.055***		0.994***	0.508***	0.516***
β_1^{12}	0.050***		0.111***				−0.239***
β_1^{13}	−0.011***		0.000				−0.005*
β_1^{21}	−0.058***		−0.174***				0.051
β_1^{22}	0.840***	0.079*	0.298***		0.223***	0.233	0.193***
β_1^{23}	0.021***		0.001				−0.012
β_1^{31}	0.113***		−0.188***				−0.005
β_1^{32}	0.700***		0.838***				0.003
β_1^{33}	0.904***	1.574***	1.699***		0.892***	1.567***	1.551***
β_2^{11}		0.290***	0.864***			0.450***	0.473***
β_2^{22}		0.371***	0.394***			−0.025***	−0.004
β_2^{33}		−0.589***	−0.706***			−0.582***	−0.578***
p_2	0.771***	0.684***	0.732***	0.965***	0.989***	0.916***	0.898***
m_2	7.081***	8.233***	8.031***	6.248***	6.153***	6.770***	6.775***
p_3	2.020***	2.000***	2.231***	2.103***	1.910***	2.009***	2.008***
m_3	2.686***	2.746***	2.227***	2.052***	2.970***	2.728***	2.730***
Latent component							
a				0.933***	0.936***	0.950***	0.943***
δ_1				0.176***	0.231***	0.263***	0.279***
δ_2				0.156***	0.143***	0.119***	0.111***
δ_3				−0.032***	0.000	0.002	0.003
Diagnostics							
LL	−18856	−19100	−18730	−19818	−18359	−18278	−18250
BIC	−18981	−19226	−18883	−19881	−18449	−18422	−18421
MLB	182.992***	391.119***	354.426***	14629.940***	135.187***	88.525***	95.614***
Diagnostics for the return process							
Mean	−0.004	−0.016	−0.006	−0.004	−0.003	−0.005	−0.004
S.D.	1.000	1.000	0.999	1.025	1.039	1.046	1.050
LB	15.472	16.682	16.279	18.568	18.195	19.094	19.861
LB2	40.032***	322.974***	40.399***	141.275***	13.632	12.503	12.941

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

Table 7 (*continued*)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Diagnostics for the volume process							
Mean	1.000	1.000	1.000	1.003	1.007	1.007	1.005
S.D.	0.536	0.554	0.531	0.554	0.537	0.534	0.528
LB	116.756***	1306.748***	67.612***	95.032***	49.266***	26.084	26.617
Diagnostics for the trading intensity process							
Mean	0.999	0.999	0.999	1.000	0.999	0.999	0.999
S.D.	0.308	0.307	0.308	0.348	0.308	0.307	0.307
LB	67.939***	32.587***	138.980***	7589.183***	69.350***	37.790***	37.616***

Data extracted from the 2001 TAQ database. The sample period was from 02.01.01 to 31.05.01. Overnight observations are excluded. The models are re-initialized on each trading day. Standard errors are computed based on the inverse of the estimated Hessian. The ML-EIS estimates are computed using $R = 50$ Monte Carlo replications based on 5 EIS iterations.
Diagnostics: Log likelihood function (LL), Bayes Information Criterion (BIC), mean, standard deviation and Ljung–Box statistics of the filtered residuals (LB) and squared filtered residuals (LB2, only for the return process) as well as multivariate Ljung–Box statistic (MLB). The Ljung–Box statistics are computed based on 20 lags.

Table 8
Maximum likelihood efficient importance sampling (ML-EIS) estimates of different parameterizations of SMEM specifications up to a lag order of $P = Q = 2$ models for the log return volatility, the average volume per trade and the number of trades per 5 min interval for the Boeing stock traded on the NYSE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ω_1	1.829***	0.532***	0.733***	0.625***	−0.121***	0.460***	0.712
ω_2	−1.087***	−0.809***	−0.610***	−6.611***	−2.197***	−2.531***	−2.124***
ω_3	−0.224***	−0.030***	−0.033***	−0.408***	−0.188***	−0.155***	−0.128***
α_0^{12}	0.554***	0.576***	0.565***	0.594***	−0.087***	0.020	0.004
α_0^{13}	1.037***	1.040***	1.081***	0.682***	0.034***	0.259***	0.828***
α_0^{23}	−0.235***	−0.042***	−0.214***	−0.422***	−0.381***	−0.442***	−0.257***
α_1^{11}	0.257***	0.252***	0.224***		0.134***	0.083***	0.083***
α_1^{12}	−0.005	0.032***	−0.025***			−0.032***	−0.043***
α_1^{13}	0.003	−0.001	−0.001**			−0.006**	−0.002***
α_1^{21}	−0.001***	−0.001***	−0.001***			−0.005***	−0.004***
α_1^{22}	0.000***	0.001***	0.001***		−0.001**	−0.000	−0.001**
α_1^{23}	0.000	0.000**	0.000***			−0.000***	0.000**
α_1^{31}	−0.167***	−0.173***	−0.207***			−0.236***	−0.269***
α_1^{32}	0.048***	0.016	0.078***			−0.064**	0.035
α_1^{33}	0.076***	0.109***	0.099***		0.077***	0.090***	0.085***
α_2^{11}		0.193***	0.158***			0.140***	0.116***
α_2^{22}		0.001***	−0.001***			0.000	−0.001*
α_2^{33}		−0.095***	−0.093***			−0.002	−0.084***
β_1^{11}	0.386***	0.256***	0.337***		0.948***	0.560***	0.560***
β_1^{12}	0.112***		0.066***				−0.066**
β_1^{13}	0.002		0.002***				−0.005**
β_1^{21}	0.227***		0.058				0.176
β_1^{22}	0.808***	0.221**	1.090***		0.169***	0.163	0.423**
β_1^{23}	−0.003		−0.004***				−0.035***
β_1^{31}	−0.029		−0.306***				−0.742***
β_1^{32}	0.032		0.104***				0.187**
β_1^{33}	0.950***	1.680***	1.767***		0.948***	0.606***	1.778***
β_2^{11}		0.117***	0.121***			0.301***	0.343***
β_2^{22}		0.652***	−0.197***			0.092	−0.051
β_2^{33}		−0.685***	−0.773***			0.343***	−0.783***

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

Table 8 (*continued*)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
p_2	0.491***	0.500***	0.505***	0.396***	0.774***	0.625***	0.622***
m_2	9.167***	8.640***	8.698***	12.650***	6.750***	8.313***	8.520***
p_3	1.628***	1.786***	1.739***	2.164***	1.726***	1.879***	1.663***
m_3	3.696***	3.128***	3.293***	2.551***	3.424***	3.084***	3.626***
Latent component							
a				0.950***	0.656***	0.804***	0.827***
δ_1				0.111***	0.625***	0.524***	0.453***
δ_2				0.058***	0.407***	0.305***	0.283***
δ_3				0.069***	0.046***	0.070***	0.025***
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Diagnostics							
LL	−21087	−21140	−21015	−21744	−21022	−20906	−20819
BIC	−21213	−21265	−21168	−21807	−21112	−21050	−20990
MLB	356.292***	159.818***	208.131***	5425.582***	345.815***	120.257***	86.327**
Diagnostics for the return process							
Mean	−0.000	−0.002	−0.001	0.001	0.007	0.005	0.000
S.D.	0.999	1.000	0.999	1.005	1.033	1.027	1.020
LB	46.321***	41.711***	42.410***	46.833***	29.763*	29.401*	31.776**
LB2	192.847***	154.040***	116.050***	408.055***	46.259***	60.415***	40.050***
Diagnostics for the volume process							
Mean	1.010	1.010	1.010	1.012	1.016	1.017	1.019
S.D.	0.851	0.862	0.849	0.875	0.857	0.840	0.855
LB	44.154***	39.175***	24.091	1883.849***	226.193***	93.834***	60.773***
Diagnostics for the trading intensity process							
Mean	0.999	0.999	0.999	1.000	1.001	0.997	1.019
S.D.	0.323	0.321	0.321	0.329	0.323	0.322	0.328
LB	54.691***	33.924**	33.712**	201.586***	44.662***	22.152	34.923**

Data extracted from the 2001 TAQ database. The sample period was from 02.01.01 to 31.05.01. Overnight observations are excluded. The models are re-initialized on each trading day. Standard errors are computed based on the inverse of the estimated Hessian. The ML-EIS estimates are computed using $R = 50$ Monte Carlo replications based on 5 EIS iterations.
Diagnostics: Log likelihood function (LL), Bayes Information Criterion (BIC), mean, standard deviation and Ljung–Box statistics of the filtered residuals (LB) and squared filtered residuals (LB2, only for the return process) as well as multivariate Ljung–Box statistic (MLB). The Ljung–Box statistics are computed based on 20 lags.

Table 9
Maximum likelihood efficient importance sampling (ML-EIS) estimates of different parameterizations of SMEM specifications up to a lag order of $P = Q = 2$ models for the log return volatility, the average volume per trade and the number of trades per 5 min interval for the JP Morgan stock traded on the NYSE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ω_1	2.394***	0.217***	1.996***	0.529***	−0.097***	0.167***	0.209
ω_2	−1.584***	−2.635***	−1.410***	−2.240***	−1.471***	−2.035***	−2.122***
ω_3	−0.292***	−0.018***	−0.016***	−0.225***	−0.337***	−0.008***	−0.005
α_0^{12}	0.817***	0.979***	0.859***	0.492***	−0.078***	0.023*	0.050***
α_0^{13}	0.841***	1.112***	0.910***	0.365***	−0.091***	0.022	−0.033
α_0^{23}	−0.885***	−0.785***	−0.882***	−1.158***	−0.953***	−0.941***	−0.958***
α_1^{11}	0.149***	0.225***	0.138***		0.097***	0.085***	0.072***
α_1^{12}	0.025***	0.068***	0.009			−0.004	−0.010
α_1^{13}	0.020***	0.000	0.001***			0.000	0.001
α_1^{21}	−0.001	0.005***	−0.003			−0.022***	−0.029***
α_1^{22}	0.003***	0.010***	0.005***		0.005***	0.005***	0.010***
α_1^{23}	0.000**	0.000	0.000**			0.000***	0.000**
α_1^{31}	0.554***	0.008	0.550***			−0.037	0.096

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

Table 9 (*continued*)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
α_1^{32}	0.320***	−0.110***	0.388***			−0.054	−0.101**
α_1^{33}	0.208***	0.216***	0.216***		0.188***	0.212***	0.214***
α_2^{11}		0.263***	0.117***			0.053**	0.049**
α_2^{22}		0.007***	−0.002***			−0.001	−0.004**
α_2^{33}		−0.205***	−0.207***			−0.206***	−0.210***
β_1^{11}	0.007	−0.107***	0.051		0.972***	0.861***	0.789***
β_1^{12}	0.191***		0.170***				−0.246***
β_1^{13}	−0.006		0.000				−0.001
β_1^{21}	0.677***		0.646***				0.066
β_1^{22}	0.646***	0.057	0.708***		0.178***	0.115**	0.083
β_1^{23}	0.025***		0.000				−0.001
β_1^{31}	1.320***		1.166***				0.180**
β_1^{32}	0.472***		0.527***				−0.427**
β_1^{33}	0.873***	1.609***	1.623***		0.836***	1.625***	1.655***
β_2^{11}		0.089***	−0.068**			0.086	0.177
β_2^{22}		0.305***	−0.017			0.025	−0.077
β_2^{33}		−0.612***	−0.625***			−0.626***	−0.657***
p_2	0.687***	0.634***	0.720***	0.863***	1.027***	0.854***	0.890***
m_2	7.633***	8.066***	6.947***	5.800***	4.917***	6.537***	5.869***
p_3	2.308***	2.393***	2.438***	2.414***	2.287***	2.454***	2.579***
m_3	2.577***	2.451***	2.373***	1.986***	2.620***	2.364***	2.173***
Latent component							
a				0.951***	0.907***	0.941***	0.930***
δ_1				0.165***	0.339***	0.339***	0.350***
δ_2				0.122***	0.176***	0.136***	0.132***
δ_3				0.024***	0.009***	0.011***	0.015***
Diagnostics							
LL	−18403	−18784	−18306	−19398	−18201	−18040	−18009
BIC	−18529	−18910	−18458	−19461	−18291	−18184	−18180
MLB	769.914***	1830.305***	665.637***	12670.643***	288.422***	171.290***	197.625***
Diagnostics for the return process							
Mean	−0.018	−0.021	−0.019	−0.011	−0.004	−0.005	−0.006
S.D.	0.999	0.999	0.999	1.026	1.048	1.067	1.056
LB	36.093**	37.518**	35.911**	31.627**	26.265	25.315	24.722
LB2	376.349***	406.581***	355.220***	35.564**	40.740***	6.513	11.903
Diagnostics for the volume process							
Mean	1.001	1.001	1.001	1.003	1.009	1.009	1.006
S.D.	0.598	0.617	0.600	0.609	0.598	0.597	0.592
LB	27.975	1393.848***	22.132	90.389***	46.434***	11.269	17.671
Diagnostics for the trading intensity process							
Mean	1.000	1.000	1.000	0.999	1.000	0.999	1.000
S.D.	0.276	0.274	0.273	0.306	0.276	0.273	0.273
LB	138.865***	38.987***	43.659***	6337.353***	140.305***	50.553***	53.280***

Data extracted from the 2001 TAQ database. The sample period was from 02.01.01 to 31.05.01. Overnight observations are excluded. The models are re-initialized on each trading day. Standard errors are computed based on the inverse of the estimated Hessian. The ML-EIS estimates are computed using $R = 50$ Monte Carlo replications based on 5 EIS iterations.

Diagnostics: Log likelihood function (LL), Bayes Information Criterion (BIC), mean, standard deviation and Ljung–Box statistics of the filtered residuals (LB) and squared filtered residuals (LB2, only for the return process) as well as multivariate Ljung–Box statistic (MLB). The Ljung–Box statistics are computed based on 20 lags.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

Table 10
Maximum likelihood efficient importance sampling (ML-EIS) estimates of different parameterizations of SMEM specifications up to a lag order of $P = Q = 2$ models for the log return volatility, the average volume per trade and the number of trades per 5 min interval for the IBM stock traded on the NYSE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ω_1	0.057***	0.547***	0.372***	0.504***	0.494***	0.316***	1.079***
ω_2	−0.089***	−1.267***	−0.784***	−1.420***	−1.831***	−1.307***	−1.403***
ω_3	−0.036***	−0.192***	−0.200***	−0.517***	−0.373***	−0.160***	−0.305***
α_0^{12}	0.077***	0.821***	0.792***	0.704***	0.596***	0.128***	0.140***
α_0^{13}	0.074***	0.712***	0.746***	0.182**	0.585***	0.099***	0.429***
α_0^{23}	−0.080***	−0.394***	−0.783***	−1.355***	−0.768***	−0.866***	−0.789***
α_1^{11}	0.024***	0.203***	0.183***		0.070***	0.068***	0.067***
α_1^{12}	0.003***	0.044***	0.025***			0.018**	0.015**
α_1^{13}	0.000	0.003	0.000			−0.001	0.002
α_1^{21}	−0.005***	−0.015***	−0.052***			−0.061***	−0.064***
α_1^{22}	0.002***	0.019***	0.021***		0.016***	0.012***	0.013***
α_1^{23}	0.000***	0.001***	0.001***			−0.001**	0.001**
α_1^{31}	−0.012***	−0.138***	−0.136***			−0.086**	0.015
α_1^{32}	0.031***	−0.061***	0.344***			0.110**	0.136***
α_1^{33}	0.016***	0.188***	0.191***		0.153***	0.182***	0.151***
α_2^{11}		0.230***	0.131***			0.086***	0.048**
α_2^{22}		0.014***	−0.004***			0.001	0.001
α_2^{33}		−0.110***	−0.101***			−0.105***	−0.087***
β_1^{11}	0.063***	0.340***	0.580***		−0.228***	0.471***	0.596***
β_1^{12}	0.007***		0.048***				−0.050
β_1^{13}	0.000		0.001				−0.007**
β_1^{21}	−0.006**		−0.108***				0.477***
β_1^{22}	0.081***	0.071**	0.866***		0.176***	0.400***	0.436***
β_1^{23}	0.001		0.000				−0.059***
β_1^{31}	−0.030***		−0.347***				0.114
β_1^{32}	0.068***		0.701***				0.105
β_1^{33}	0.091***	1.289***	1.211***		0.900***	1.272***	1.122***
β_2^{11}		0.133***	0.044			0.496***	0.414***
β_2^{22}		0.642***	−0.009			−0.091**	−0.115**
β_2^{33}		−0.329***	−0.252***			−0.313***	−0.229**
p_2	0.083***	0.759***	0.847***	1.168***	0.965***	1.190***	1.115***
m_2	7.271***	7.778***	6.936***	4.466***	6.492***	4.953***	5.470***
p_3	2.225***	2.222***	2.303***	2.294***	2.156***	2.278***	2.172***
m_3	4.373***	4.424***	4.131***	3.499***	4.642***	4.255***	4.620***
Latent component							
a				0.942***	0.967***	0.940***	0.944***
δ_1				0.154***	0.141***	0.263***	0.256***
δ_2				0.146***	0.087***	0.133***	0.123***
δ_3				0.044***	0.004***	0.012***	0.003**
Diagnostics							
LL	−15389	−15798	−15338	−16959	−15349	−15106	−15086
BIC	−15515	−15924	−15490	−17021	−15439	−15250	−15257
MLB	324.925***	686.107***	240.460***	19077.193***	833.269***	95.978***	76.669*
Diagnostics for the return process							
Mean	−0.009	−0.008	−0.009	−0.009	−0.006	0.000	0.000
S.D.	1.000	0.999	1.000	1.014	1.009	1.026	1.028
LB	18.485	17.055***	17.720	17.105	19.767	23.732	24.735
LB2	175.325***	282.111***	154.973***	622.605***	313.360***	18.042	16.570

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

Table 10 (*continued*)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Diagnostics for the volume process							
Mean	1.000	1.000	0.999	1.003	1.001	1.005	1.004
S.D.	0.485	0.512	0.484	0.512	0.487	0.481	0.483
LB	50.970***	444.839***	47.007***	257.354***	54.532***	56.413***	54.164***
Diagnostics for the trading intensity process							
Mean	0.999	0.999	0.999	0.999	0.999	1.000	1.000
S.D.	0.217	0.216	0.216	0.246	0.217	0.216	0.216
LB	83.299***	9.630	12.195	6935.761 ***	84.642***	11.464	11.405

Data extracted from the 2001 TAQ database. The sample period was from 02.01.01 to 31.05.01. Overnight observations are excluded. The models are re-initialized on each trading day. Standard errors are computed based on the inverse of the estimated Hessian. The ML-EIS estimates are computed using $R = 50$ Monte Carlo replications based on 5 EIS iterations.
Diagnostics: Log likelihood function (LL), Bayes Information Criterion (BIC), mean, standard deviation and Ljung–Box statistics of the filtered residuals (LB) and squared filtered residuals (LB2, only for the return process) as well as multivariate Ljung–Box statistic (MLB). The Ljung–Box statistics are computed based on 20 lags.

demand associated with a high trading intensity naturally consumes market depth and induces revisions in the ask and bid quotes, which results in increases in the mid-quote return volatility.

In contrast, the parameter α_0^{23} is significantly negative and mainly unaffected by the inclusion of the latent component. This indicates that the (negative) contemporaneous relation between trade size and trading intensity is *not* driven by a latent common component. Instead, it is explained by the typical finding that high trading volumes absorb a non-trivial part of the offered liquidity supply. This induces high spreads increasing transaction costs and thus reduces traders' incentives for market order trading (see, e.g., Foucault, 1999). Another explanation is that high volumes are split over time leading to higher trading frequencies but also to smaller trade sizes. Our results indicate that these effects are obviously not primarily driven by an underlying common process.

Nevertheless, the negative relation between trading intensities and sizes is overcompensated in periods of high information intensity. Then, the latent factor increases both volumes and the speed of trading. In contrast, if λ_i is small, the opposite effect dominates. However, as shown by the descriptive statistics, unconditionally we see a negative relationship between trade size and trading intensity.

(iv) The estimations omitting a common latent component (panels (1)–(3)) clearly reveal significant evidence for cross-dependencies between volatilities, volumes and trading intensities. In particular, as indicated by a mostly positive parameter $\hat{\alpha}_1^{12}$, we observe a positive relationship between innovations in the lagged trade size and the current volatility. Hence, higher than expected volumes imply significant quote revisions and consequently increase the subsequent volatility. As revealed by $\hat{\alpha}_1^{32} > 0$, this effect is also accompanied by an increase in the trading intensity. In contrast, unexpected increases of the volatility reduce both the subsequent trade size and the trading intensity ($\hat{\alpha}_1^{21} < 0$ and $\hat{\alpha}_1^{31} < 0$). This result is very much in line with theory (see, e.g., Foucault, 1999) suggesting that a higher transitory volatility increases spreads and transaction costs, which in turn reduces trade sizes and trading intensities.

However, as shown in panels (5)–(7), the inclusion of λ_i clearly reduces the magnitude of the aforementioned cross-effects. In most cases, the latter become close to zero and/or insignificant. Similar effects are also observed for the non-diagonal elements in B_1. This finding indicates that the latent common component indeed captures a substantial part of the cross-dependencies. Hence, most of the observed causalities between the individual variables are mainly due to the existence of a subordinated common (information) process jointly directing the individual components.[18] This suggests the usefulness of more parsimonious parameterizations of the observation-driven dynamics, which might be mainly reduced to a diagonal specification of the autoregressive parameter matrices.

Moreover, the inclusion of the latent factor reduces the impact of the process-specific innovations ($\hat{\alpha}_1^{ii}$ and $\hat{\alpha}_2^{ii}$ for $i = 1, 2, 3$) and increases the persistence in the observation-driven dynamics. Hence, in accordance with the results for the univariate models, we find evidence for the effect that news enters the model primarily through the latent factor, whereas the impact of the process-specific innovations declines. Nonetheless, given the subordinated process, volatilities, trade sizes and trading intensities are still positively autocorrelated. As suggested by asymmetric information-based market microstructure theory, such effects reflect that high volatilities, trade sizes and trading intensities serve as signals allowing (non-informed) market participants to infer the existence of information. In this sense, the idiosyncratic dynamics reflect how common information shocks (induced by λ_i) are processed in the multivariate trading process.

[18] A notable exception is the negative relationship between past volatility innovations and current average trade sizes as reflected by $\hat{\alpha}^{21} < 0$. This relationship is obviously not information-driven and becomes even more pronounced when the latent factor is taken into account. This finding is not easily explained in the given setting and requires further investigation.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

(v) From a statistical viewpoint, the specifications without a latent factor (columns (1)–(3)) are not able to completely capture the dynamics of the system as indicated by highly significant Ljung–Box statistics for the residuals. Inmost cases, the inclusion of the latent component improves the dynamic properties of the model. This is particularly true for the volatility and the volume component, whereas in some cases the dynamics in the trading intensity are still not completely captured by the model. The latter results are not surprising given that the latent factor impact on the trading intensity is only very weak. Moreover, the inclusion of the latent factor leads to a reduction of the multivariate Ljung–Box statistic indicating that the common component does a good job in capturing the multivariate dynamics and interdependencies between the individual processes. Furthermore, as revealed by the Bayes information criterion (BIC), the SMEM yields a better goodness-of-fit compared to MEMs without a latent factor.

(vi) The worst performance is observed for specification (4), where any observation-driven dynamics are omitted and only a parameter-driven dynamic is included. Hence, a single common autoregressive component is not sufficient to completely capture the dynamics of the multivariate system which confirms the findings by Andersen (1996) or Liesenfeld (1998). Therefore, as in the univariate models we can neither reject the parameter-driven dynamic nor the observation-driven dynamic. Actually, the best performance is revealed by specifications that include both types of dynamics, which confirms the basic idea of the proposed model.

Our results are widely robust over the cross-section of stocks. A notable exception are the findings for the Boeing stock, which deviate in several respects from those of the other stocks. Here, the latent factor is less persistent and does not capture the dynamics of the processes very well.

5.4. Impulse response dynamics and graphical illustrations

In order to analyze the impact of shocks on the SMEM process, we rely on the concept of the generalized impulse response function (GIRF) introduced by Koop et al. (1996) which is given by

$$GIRF_{X_i}(s,\delta,\mathscr{F}_{i-1}) = \mathrm{E}[X_{i+s}|\varpi_i=\delta,\mathscr{F}_{i-1}] - \mathrm{E}[X_{i+s}|\mathscr{F}_{i-1}], \tag{26}$$

where $X_i \in \{\lambda_i, Y_i^2, V_i, \rho_i\}$, $\varpi_i \in \{v_i, \eta_i, u_i, \varepsilon_i\}$, δ is the magnitude of the shock, and s denotes the number of periods over which the GIRF is computed. As shown in this representation, the GIRF is conditioned on the shock and on the history of the process whereas innovations occurring in intermediate time periods are averaged out. Then, the GIRF can be interpreted as a random variable in terms of the history $\mathscr{F}_{i-1}$. In nonlinear models, analytical expressions of the conditional expectations used in (26) are often not available and thus, Monte Carlo simulation techniques must be performed. Figs. 11–14 show the generalized impulse response functions for a shock in the latent innovation v_i with a magnitude of one standard deviation. The GIRF is computed by conditioning on the unconditional means $\mathrm{E}[X_i]$ and $\mathrm{E}[\varpi_i]$ and is estimated by

$$\widehat{GIRF}_{X_i}(s,\delta,\mathscr{F}_{i-1}) = \hat{\mathrm{E}}[X_{i+s}|\varepsilon_i=1,\mathscr{F}_{i-1}] - \hat{\mathrm{E}}[X_{i+s}|\mathscr{F}_{i-1}],$$

where the conditional expectations are estimated by sample averages based on 5,000 simulated paths of $X_i, X_{i+1}, \ldots, X_{i+h}$, given the corresponding conditioning information and using the parameter estimates of specification (8) in Tables 7–10. For all processes, we observe a positive, persistent response of ξ_i^2, V_i and ρ_i due to a shock in the latent component. In most cases, the impulse response function declines monotonically and approaches zero after 30–40 lags corresponding to 150–200 trading minutes. Hence, common (information) shocks remain present in the trading process for up to about 3 h. In accordance with the parameter estimates, these effects are mostly pronounced for the volatility and trade size component but, not surprisingly, are quite weak for the trading intensity.

Figs. 15 and 16 show the time series plots of the (filtered) estimates of $\exp(\lambda_i)$, Y_i^2, V_i and ρ_i for the AOL and IBM stock, respectively.[19] The latent factor captures common shocks in the volatility and volume component, whereas the trading intensity is widely unaffected by co-movements. We observe that there are certain periods, where volatilities, volumes and the common factor move in lock-steps, whereas during other periods, the individual processes seem to be clearly disentangled. These results are confirmed by the correlations between the individual components as given by $\mathrm{Corr}[e^{\lambda_i}, \xi_i^2] = 0.18$ (= 0.22), $\mathrm{Corr}[e^{\lambda_i}, v_i] = 0.52$ (= 0.55) and $\mathrm{Corr}[e^{\lambda_i}, \rho_i] = -0.05$ (= 0.01) for the AOL (IBM) stock.[20] Hence, we observe significant commonalities between the latent factor and return volatility as well as the trade size but not necessarily between the latent component and the trading intensity.

6. Conclusions

In this paper, we have proposed a new type of multivariate multiplicative error model for intra-day trading processes. The basic idea of the so-called multivariate stochastic multiplicative error model (SMEM) is to combine a multivariate

[19] For the sake of brevity, the corresponding plots for the two other stocks are not shown, but are available upon request from the author.

[20] For the other stocks the correlation structures look quite similar.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

observation-driven (multiplicative error) dynamic with an underlying univariate parameter-driven factor which jointly affects all individual components of the system. Whereas the observation-driven dynamic is updated by process-specific innovations, which are completely observable given the process history, the parameter-driven component follows an autoregressive process that is updated by unobservable innovations independent from the idiosyncratic errors. We propose the model as a tool for identifying a common component in multivariate systems while still allowing for idiosyncratic dynamics given the subordinated process. It is a computationally more tractable alternative to multiple latent factor models. Moreover, if there is a common component serving as a major source for cross-dependencies between the individual processes, then its explicit consideration should result in a more parsimonious specification of the multivariate process. This becomes even more important when the dimension of the process is very high.

The SMEM was designed to allow for the possibility that intra-day return volatility, the trade size and trading intensity are driven not only by their own history but also by a joint dynamic latent factor capturing the (unobservable) information process. It is economically motivated by the idea that trading activity is driven by (i) an underlying information component and (ii) idiosyncratic, process-specific dynamics. This structure can be considered to be a reduced form representation of trading processes arising from asymmetric information-based market microstructure theory (see, e.g., Easley and O'Hara, 1992; Easley et al., 1996, among others). The latter assumes that the trading process is driven by the interactions between informed market participants who can observe the underlying information process and uninformed agents who infer the true value of the traded asset by observing the trading history. Under this assumption, the common parameter-driven component serves as a proxy for the unobserved information process, whereas the process-specific observation-driven dynamics result from the fact that common (observable) shocks are processed in different ways in volatility, trade size and trading intensity, causing genuine trading effects.

Applying the model to 5-min data of four blue chip stocks traded at the NYSE yields the following conclusions: (i) There is significant evidence for the existence of a common unobservable component following persistent dynamics and jointly driving volatilities, trade sizes and trading intensities in the same direction. This finding confirms the notion that underlying common dynamics are not only identifiable based on a daily level but also on an intradaily level providing evidence for a 'micro-foundation' of the well-known volume–volatility relationship. (ii) We find that the common subordinated (information) process mainly affects volatility and trade size but has only a comparably weak impact on the trading intensity. This confirms Blume et al. (1994), Xu and Wu (1999), Chan and Fong (2000) and Huang and Masulis (2003), who reported that the trade size captures non-trivial information. Based on our results, we conclude that the trade size seems to be an important proxy for common information shocks. (iii) In most cases, the conditional cross-dependencies between the individual variables given the latent factor are significantly lower than the corresponding unconditional ones. This documents that a substantial part of the causal relationship between volatility, trade size and trading intensity are indeed driven by a common subordinated process. This holds particularly for the positive relationship between volatility and trade size, which is nearly completely captured by the subordinated common process. (iv) We also find evidence for causal effects that are not explained by the presence of a common component. This is particularly true for a positive relationship between volatility and trading intensity that does not stem from an underlying common (information) process but rather reflects trade-specific effects. This result provides deeper insights of the interplay between volatility and trading intensity and complements the findings of Jones et al. (1994) and Ané and Geman (2000). Moreover, given the latent process, high trading volumes enter the market at a lower speed. Nevertheless, this effect can be overcompensated in periods of high informational intensity where both trading components are driven upwards. (v) Model diagnostics indicate that it is necessary to allow for both parameter-driven as well as observation-driven dynamics. Our findings suggest that a single latent component is not sufficient to capture the dynamics of the multivariate system and that common (information) shocks are processed individually in the particular trading components. (vi) In univariate specifications of the individual trading components, SMEMs significantly outperform models without a latent factor. This finding strongly suggests the need for flexible two-factor models and complements the findings of Ghysels et al. (2004).

In summary, our results provide evidence that a substantial part of the cross-dependencies between the individual processes can be captured by a common dynamic component. This should open up the possibility of specifying high-dimensional trading processes in a more parsimonious way. Future research is devoted to more extensive applications of the model. On one hand, it might be interesting to analyze the performance of the model when even more dimensions, such as bid-ask spreads or market depths are added. We expect that the importance of the common component in such a setting will become even stronger. Important applications of such a model would be the prediction of liquidity and trading costs over short intra-day time horizons. Here, expected trading intensities, market depths, bid-ask spreads and return volatilities will be important determinants of expected trading costs and associated risks. The latter are important inputs for automated trading algorithms as more and more heavily used in the financial industry. Moreover, we also plan to confront the model with observable news announcements as an additional component. This should shed some light on the question of whether missing information is captured by the latent factor and how observable and unobservable information interact and jointly drive the trading process. This should lead to a deeper understanding of how information is processed and how it depends on the state of the market and the institutional environment of the market. Such information might be helpful to optimize trading structures as well as corresponding trading models.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

Acknowledgments

Earlier versions of this paper have been presented at the EC^2 conference 2003 in London, at the International Conference on Finance in Copenhagen in 2005, the 2005 Arne Ryde Workshop in Financial Economics in Lund, the International Conference on High Frequency Finance in Konstanz in 2006, the 2006 meeting of the European Econometric Society in Vienna as well as the 2006 meeting of the German Economic Association in Bayreuth. For valuable comments we would like to thank two anonymous referees, Torben G. Andersen, Luc Bauwens, Tim Bollerslev, Robert F. Engle, Timo Teräsvirta, Winfried Pohlmeier and the seminar participants at the Stockholm School of Economics and the Université Libre de Bruxelles. This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

Appendix A. EIS algorithm for the estimation of SMEM processes

To define the importance sampler let $k(\Lambda_i, \phi_i)$ denote a density kernel for $m(\lambda_i|\Lambda_{i-1}, \phi_i)$, given by

$$k(\Lambda_i, \phi_i) = m(\lambda_i|\Lambda_{i-1}, \phi_i)\chi(\Lambda_{i-1}, \phi_i), \tag{27}$$

where

$$\chi(\Lambda_{i-1}, \phi_i) = \int k(\Lambda_i, \phi_i)\, \mathrm{d}\lambda_i \tag{28}$$

denotes the integrating constant. The implementation of EIS requires the selection a class of density kernels $k(\cdot)$ for the auxiliary sampler $m(\cdot)$, which provides a good approximation to the product $f(\cdot)\chi(\cdot)$. As discussed by Richard and Zhang (2007), a convenient and efficient possibility is to use a parametric extension of the direct samplers, which stem from Gaussian distributions in this context. Since the function $g(\cdot)$ appearing in (18) is essentially a product of different exponential functions, we propose to approximate it using a normal density kernel

$$\zeta(\lambda_i, \phi) = \exp\left(\phi_{1,i}\lambda_i + \phi_{2,i}\lambda_i^2\right), \tag{29}$$

which is itself an exponential function in terms of λ_i based on the auxiliary parameters $\phi_i = (\phi_{1,i}, \phi_{2,i})$. Exploiting the property that the product of normal densities is itself a normal density, we parameterize $k(\cdot)$ as

$$k(\Lambda_i, \phi_i) = p(\lambda_i|\Lambda_{i-1}; \theta)\zeta(\lambda_i, \phi_i)$$

and can show that

$$\begin{aligned} k(\Lambda_i, \phi_i) &\propto \exp\left((\phi_{1,i} + \mu_{0,i})\lambda_i + \left(\phi_{2,i} - \frac{1}{2}\right)\lambda_i^2\right) \\ &= \exp\left(-\frac{1}{2\pi_i^2}(\lambda_i - \mu_i)^2\right)\exp\left(\frac{\mu_i^2}{2\pi_i^2}\right), \end{aligned} \tag{30}$$

where

$$\pi_i^2 = (1 - 2\phi_{2,i})^{-1}, \tag{31}$$

$$\mu_i = (\phi_{1,i} + \mu_{0,i})\pi_i^2. \tag{32}$$

Hence, the auxiliary sampler $m(\cdot)$ is a normal distribution with conditional mean μ_i and conditional variance π_i^2. By omitting irrelevant multiplicative factors, we obtain the integrating constant as

$$\chi(\Lambda_{i-1}, \phi_i) = \exp\left(\frac{\mu_i^2}{2\pi_i^2} - \frac{\mu_{0,i}^2}{2}\right). \tag{33}$$

As shown by Richard and Zhang (2007), the Monte Carlo variance of $\hat{\mathscr{L}}_R(W;\theta)$ can be minimized by splitting the minimization problem into n minimization problems of the form

$$\min_{\phi_{i,0},\phi_i} \sum_{r=1}^{R} \{\ln f[(w_i, \lambda_i^{(r)}(\theta)|W_{i-1}, \Lambda_{i-1}(r)(\theta), \theta) \cdot \chi(\Lambda_i^{(r)}(\theta), \phi_{i+1}(\theta))] - \phi_{0,i} - \ln k(\Lambda_i^{(r)}(\theta), \phi_i(\theta))\}^2, \tag{34}$$

where $\phi_{0,i}$ is a constant and $\{\lambda_i^{(r)}(\theta)\}_{i=1}^n$ with $\lambda_i^{(r)}(\theta) := \lambda_i^{(r)}(\phi_i(\theta))$ denotes a trajectory of random draws from the sampler m with auxiliary parameters $\phi_i(\theta)$ which themselves depend on the model parameters θ.

Then, in practice, the implementation of the ML-EIS estimator requires the following steps:

(i) Draw R trajectories of the latent factor $\{\lambda_i^{(r)}(\phi_i)\}_{i=1}^n$ using the direct sampler $p(\cdot)$.
(ii) For $i : n \to 1$ solve the least squares problem characterized by the (auxiliary) linear regression

$$D_{1,i}^{(r)} + D_{2,i}^{(r)} + D_{3,i}^{(r)} + D_{4,i}^{(r)} = \phi_{0,i} + \phi_{1,i}\lambda_i^{(r)}(\theta) + \phi_{2,i}[\lambda_i^{(r)}(\theta)]^2 + \varepsilon_i^{(r)}, \quad r = 1, \ldots, R,$$

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

where

$$D_{1,i}^{(r)} = -\frac{1}{2}(\ln h_i + \delta_1 \lambda_i^{(r)}(\theta) + \ln s_{h,i}) - \frac{\xi_i^2}{2h_i s_{h,i}(e^{\delta_1 \lambda_i^{(r)}(\theta)})},$$

$$D_{2,i}^{(r)} = (p_2 m_2 - 1)\ln V_i - (p_2 m_2 \ln \Phi_i + \delta_2 p_2 m_2 \lambda_i^{(r)}(\theta) + \ln s_{V,i}) - \left(\frac{V_i}{\Phi_i s_{V,i}\, e^{\delta_2 \lambda_i^{(r)}(\theta)}}\right)^{p_2},$$

$$D_{3,i}^{(r)} = (p_3 m_3 - 1)\ln \rho_i - (p_3 m_3 \ln \Psi_i + \delta_3 p_3 m_3 \lambda_i^{(r)}(\theta) + \ln s_{\rho,i}) - \left(\frac{\rho_i}{\Psi_i s_{\rho_i}\, e^{\delta_3 \lambda_i^{(r)}(\theta)}}\right)^{p_3},$$

$$D_{4,i}^{(r)} = \ln \chi(\Lambda_i^{(r)}(\theta), \phi_{i+1}(\theta)),$$

and $\varepsilon_i^{(r)}$ denotes the regression error term. These problems are solved sequentially starting at $i = n$, under the initial condition $\chi(\Lambda_n, \phi_{n+1}) = 1$ and ending at $i = 1$. Liesenfeld and Richard (2003) recommend three to five iterations of this procedure to improve the efficiency of the approximations.

(iii) Compute the EIS sampler $\{m(\lambda_i|\Lambda_{i-1}, \hat{\phi}(\hat{\theta})\}_{i=1}^n$ on the basis of the conditional mean and variance as given by

$$\pi_i^2 = (1 - 2\phi_{2,i})^{-1}, \tag{35}$$

$$\mu_i = (\phi_{1,i} + \mu_{0,i})\pi_i^2. \tag{36}$$

in order to draw R trajectories $\{\lambda_i^{(r)}(\hat{\phi}_i(\hat{\theta}))\}_{i=1}^n$. These trajectories are used to calculate the likelihood according to (20). Then, as suggested by Richard and Zhang (2007), the variance–covariance matrix of the estimated parameters is estimated in a straightforward manner based on the inverted Hessian.

References

Admati, A., Pfleiderer, P., 1988. A theory of intraday patterns: volume and price variability. Review of Financial Studies 1, 3–40.
Andersen, T.G., 1996. Return volatility and trading volume: an information flow interpretation of stochastic volatility. Journal of Finance 51, 169–204.
Ané, T., Geman, H., 2000. Order flow, transaction clock, and normality of asset returns. Journal of Finance 55, 2259–2284.
Bauwens, L., Giot, P., 2000. The logarithmic ACD model: an application to the bid/ask quote process of two NYSE stocks. Annales d'Economie et de Statistique 60, 117–149.
Bauwens, L., Giot, P., 2001. Econometric Modelling of Stock Market Intraday Activity. Kluwer Academic Publishers, Boston, Dordrecht, London.
Bauwens, L., Hautsch, N., 2006. Stochastic conditional intensity processes. Journal of Financial Econometrics 4, 450–493.
Bauwens, L., Veredas, D., 2004. The stochastic conditional duration model: a latent factor model for the analysis of financial durations. Journal of Econometrics 119, 381–412.
Bauwens, L., Galli, F., Giot, P., 2008. Quantitative and qualitative analysis in social sciences, forthcoming.
Blazsek, S., Escribano, A., 2005. Dynamic latent factor intensity models of knowledge spillovers: evidence based on patent analysis. Working Paper, Universidad Carlos III de Madrid.
Blume, L., Easley, D., O'Hara, M., 1994. Market statistics and technical analysis. Journal of Finance 49 (1), 153–181.
Bollerslev, T., Jubinski, D., 1999. Equity trading volume and volatility: latent information arrivals and common long-run dependencies. Journal of Business and Economic Statistics 17, 9–21.
Bowsher, C.G., 2007. Modelling security markets in continuous time: intensity based, multivariate point process models. Journal of Econometrics 141, 876–912.
Chan, K., Fong, W., 2000. Trade size, order imbalance, and the volatility–volume relation. Journal of Financial Economics 57, 247–273.
Cipollini, F., Engle, R.F., Gallo, G.M., 2007. Vector multiplicative error models: representation and inference. Mimeo, University of Firenze.
Clark, P., 1973. A subordinated stochastic process model with finite variance for speculative prices. Econometrica 41, 135–155.
Dufour, A., Engle, R.F., 2000. Time and the impact of a trade. Journal of Finance 55, 2467–2498.
Easley, D., O'Hara, M., 1992. Time and process of security price adjustment. The Journal of Finance 47, 577–605.
Easley, D., Kiefer, N.M., O'Hara, M., Paperman, J.B., 1996. Liquidity, information and infrequently traded stocks. Journal of Finance 51 (4), 1405–1436.
Engle, R.F., 2000. The econometrics of ultra-high-frequency data. Econometrica 68 (1), 1–22.
Engle, R.F., 2002. New frontiers for ARCH models. Journal of Applied Econometrics 17, 425–446.
Engle, R.F., Gallo, G.M., 2006. A multiple indicators model for volatility using intra-daily data. Journal of Econometrics 131, 3–27.
Engle, R.F., Ng, V.K., 1993. Measuring and testing the impact of news on volatility. Journal of Finance 48, 1749–1778.
Epps, T.W., Epps, M.L., 1976. The stochastic dependence of security price changes and transaction volumes: implications for the mixture-of-distributions hypothesis. Econometrica 44, 305–321.
Fernandes, M., Grammig, J., 2006. A family of autoregressive conditional duration models. Journal of Econometrics 130, 1–23.
Foucault, T., 1999. Order flow composition and trading costs in a dynamic limit order market. Journal of Financial Markets 2, 99–134.
Ghysels, E., Gouriéroux, C., Jasiak, J., 2004. Stochastic volatility duration models. Journal of Econometrics 119, 413–433.
Glosten, L.R., Milgrom, P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71–100.
Grammig, J., Wellner, M., 2002. Modeling the interdependence of volatility and inter-transaction duration process. Journal of Econometrics 106, 369–400.
Hasbrouck, J., 1991. Measuring the information content of stock trades. Journal of Finance 46, 179–207.
Hautsch, N., 2004. Modelling irregularly spaced financial data. Vol. 539 of Lecture Notes in Economics and Mathematical Systems. Springer, Berlin.
Hautsch, N., 2006. Testing the conditional mean function of autoregressive conditional duration models. Discussion Paper 2006-06, Finance Research Unit, Department of Economics, University of Copenhagen.
Hentschel, L., 1995. All in the family: nesting symmetric and asymmetric GARCH models. Journal of Financial Economics 39, 71–104.
Hosking, J.R.M., 1980. The multivariate portmanteau statistic. Journal of the American Statistical Association 75, 602–608.
Huang, R.D., Masulis, R.W., 2003. Trading activity and stock price volatility: evidence from the London Stock Exchange. Journal of Empirical Finance 10, 249–269.
Jones, C.M., Kaul, G., Lipson, M.L., 1994. Information, trading, and volatility. Journal of Financial Economics 36, 127–154.
Koop, G., Pesaran, M.H., Potter, S.M., 1996. Impulse response analysis in nonlinear multivariate models. Journal of Econometrics 74, 119–174.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009

Koopman, S.J., Lucas, A., Monteiro, A., 2008. The multi-state latent factor intensity model for credit rating transitions. Journal of Econometrics 142, 399–424.
Lamoureux, C.G., Lastrapes, W.D., 1990. Heteroskedasticity in stock return data: volume versus GARCH effects. Journal of Finance 45, 221–229.
Liesenfeld, R., 1998. Dynamic bivariate mixture models: modeling the behavior of prices and trading volume. Journal of Business and Economic Statistics 16, 101–109.
Liesenfeld, R., 2001. A generalized bivariate mixture model for stock price volatility and trading volume. Journal of Econometrics 104, 141–178.
Liesenfeld, R., Richard, J.-F., 2003. Univariate and multivariate stochastic volatility models: estimation and diagnostics. Journal of Empirical Finance 10, 505–531.
Ljung, G.M., Box, G.E.P., 1978. On a measure of lack of fit in time series models. Biometrika 65, 297–303.
Manganelli, S., 2005. Duration, volume and volatility impact of trades. Journal of Financial Markets 8, 377–399.
Meddahi, N., Renault, E., Werker, B., 2006. GARCH and irregularly spaced data. Economic Letters 90, 200–204.
Nelson, D., 1991. Conditional heteroskedasticity in asset returns: a new approach. Journal of Econometrics 43, 227–251.
Renault, E., Werker, B.J., 2003. Stochastic volatility models with transaction risk. Discussion Paper, Tilburg Univeristy.
Richard, J.-F., 1998. Efficient high-dimensional Monte Carlo importance sampling. Working Paper, University of Pittsburgh.
Richard, J.-F., Zhang, W., 2007. Efficient high-dimensional importance sampling. Journal of Econometrics 141, 1385–1411.
Tauchen, G.E., Pitts, M., 1983. The price variability–volume relationship on speculative markets. Econometrica 51, 485–505.
Taylor, S.J., 1986. Modelling Financial Time Series. Wiley, New York.
Tran, D.T., 2006. Relationship between persistent and erratic volatility factors and trading activity. Working Paper, Duke University.
Xu, X.E., Wu, C., 1999. The intraday relation between return volatility, transactions, and volume. International Review of Economics and Finance 8, 375–397.

Please cite this article as: Hautsch, N., Capturing common components in high-frequency financial time series: A multivariate stochastic.... Journal of Economic Dynamics and Control (2008), doi:10.1016/j.jedc.2008.01.009