CORE DISCUSSION PAPER
2003/103

DYNAMIC LATENT FACTOR MODELS FOR INTENSITY PROCESSES

L. Bauwens[1] and N. Hautsch[2]

December 22, 2003

Abstract

This paper introduces a new framework for the dynamic modelling of univariate and multivariate point processes. The so-called latent factor intensity (LFI) model is based on the assumption that the intensity function consists of univariate or multivariate observation driven dynamic components and a univariate dynamic latent factor. In this sense, the model corresponds to a dynamic extension of a doubly stochastic Poisson process. We illustrate alternative parameterizations of the observation driven component based on autoregressive conditional intensity (ACI) specifications, as well as Hawkes types models. Based on simulation studies, it is shown that the proposed model provides a flexible tool to capture the joint dynamics of multivariate point processes. Since the latent component has to be integrated out, the model is estimated by simulated maximum likelihood based upon efficient importance sampling techniques. Applications of univariate and bivariate LFI models to transaction data extracted from the German XETRA trading system provide evidence for an improvement of the econometric specification when observable as well as unobservable dynamic components are taken into account.

Keywords: Multivariate point process, latent factor, transaction durations, efficient importance sampling.

JEL Classification: C22, C32, C41.

[1]CORE and Department of Economics, Université catholique de Louvain.
[2]Center of Finance and Econometrics, University of Konstanz.

Former versions of this paper have been presented at the 2002 EC^2 meeting in Bologna, and in 2003, at the workshop of the EC RTN "Microstructure of Financial Markets in Europe" in Louvain-la-Neuve, 54th ISI conference in Berlin, ESEM conference in Stockholm, 7th METU Conference in Ankara, conference on Time Series Econometrics in Linz, and the annual meeting of the German Economic Association in Zurich. For helpful comments and discussions we would like to thank András Heinen, Winfried Pohlmeier, David Veredas, as well as the participants of seminars at Toulouse, Marseille (GREQAM), CORE, and Erasmus University Rotterdam. We also thank particularly Jean-Franois Richard and Roman Liesenfeld for providing information and their GAUSS code for the estimation of stochastic volatility models. The conception of the paper started when N. Hautsch was visiting CORE. In this context he would like to thank CORE, as well as the University of Konstanz, for financial support.
This paper presents results of the Research and Training Network "Microstructure of Financial Markets in Europe" funded by the European Commission, and of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

Corresponding author. Email: Nikolaus.Hautsch@uni-konstanz.de. Adress: Box D124, University of Konstanz, Universitaetsstr. 10, D-78457 Konstanz, Germany, tel: ++ 49 7531 88 2204, fax: ++ 49 7531 88 4450.

1 Introduction

Point processes are particularly relevant in the analysis of economic behavior whenever one is interested in the timing of the activity of economic agents, and the timing is itself an aspect of agents' decisions. Relevant examples are individual actions on a particular market, like a trade on a financial market, the buy of a specific good (which might be recorded by scanner data), the visit of a particular website, and the decisions of a public authority. One may think of many other examples. Suffice it to recall that in general equilibrium theory, goods may be characterized also by the time at which they are produced or consumed.[1] Especially in financial econometrics point processes are of importance in order to model market activity on a trade-by-trade level.

Even though there exist alternative statistical methods to model point processes, the most common way is a direct modelling of the waiting time between successive events. Since the introduction of the autoregressive conditional duration (ACD) model proposed by Engle and Russell (1998) and Engle (2000), much research has been devoted to the specification and application of discrete time autoregressive duration models (see Bauwens and Giot, 2000, Grammig and Maurer, 2000, Dufour and Engle, 2000, Fernandes and Grammig, 2001, and Hautsch, 2002, among others). An obvious reason for the focus on duration models is that the inclusion of dynamic structures, which is essential for modelling financial point processes, is quite straightforward. Nonetheless, the major drawback of autoregressive duration models is their discrete time nature. Typically, the conditioning set of the process is only updated at the individual event arrival times but not during a duration spell. This makes it extremely difficult to account for changes of the information set during a spell, which is particularly important in the context of multivariate data or time-varying covariates. In particular, the problem which has to be resolved is that in a multivariate context, the particular processes occur asynchronously, and thus, there exist no joint points that can be used to couple the processes. For this reason, it is very difficult to estimate contemporaneous correlations between the single autoregressive processes.[2] One possibility to circumvent the asynchronity problem is to model a point processes in a counting framework. The strength of an autoregressive count data model is that these models are based (per assumption) on equidistant time intervals. For this reason, they are relatively easily extended to a multivariate framework (see Davis, Rydberg, Shephard, and Streett, 2001 or Heinen and Rengifo, 2003). Nevertheless, the application of count data approaches requires to aggregate transaction data to equi-distant time intervals which induces a loss of information. These effects become particularly relevant when the event activity within such an interval is very high, as in the case of transaction data of heavily traded stocks.

A valuable alternative to a discrete time modelling of point processes is to switch to a continuous time framework. A convenient and powerful way to describe a point process in continuous time is to

[1] Another potential field of application is spatial economics, where rather than the timing, the location is a crucial dimension of analysis, such as the decision to locate a firm in the space, which may be viewed as a realization of a spatial point process. See e.g. Snyder and Miller (1991).

[2] Because of this difficulty, multivariate extensions of ACD models typically result in competing risks models (see e.g. Bauwens and Giot, 2003 or Engle and Lunde, 2003). In these approaches, one models the time until the occurrence of one of the individual processes and treats all other (non-observed) processes as right-censored. However, such an approach implies some information loss when successive points of one process occur without intervening points of the other processes. For this reason, such an approach is not appropriate for a complete modelling of multivariate point processes.

specify the *intensity* function. The intensity function is a central concept in the theory of point processes and is defined as the instantaneous rate of occurrence given the process history and observable factors. The major advantage of the modelling of the intensity function is that it allows us to account for events that occur in *any* point in time. This property is an important requirement for an extension to the multivariate case since here, the conditioning set can be updated at every point of the pooled process. Moreover, since the intensity function defines the instantaneous event arrival rate, it serves as a natural and nicely interpretable continuous time measure for economic activity.

In this paper, we introduce a very flexible class of dynamic intensity models which can be interpreted as a dynamic extension of a doubly stochastic Poisson process (see, e.g., Grandell, 1976 or Cox and Isham, 1980) and is called "latent factor intensity" (LFI) model. We propose a fully parametric model for the intensity function, assuming that the intensity function is driven not only by the observable history of the process but also by a dynamic latent component. Thus, the major idea of the LFI model is to assume that the conditional intensity function given the observable history of the process is itself stochastic and follows an autoregressive process. The intensity is parameterized in terms of two components, a univariate latent one, and an "observable" one which is driven by the (observable) history of the process and can be specified univariately or multivariately. In the latter case, the latent factor corresponds to a common component that captures the impact of a general factor that influences all individual process components. In the context of financial markets, the latent factor may be economically interpreted as a variable representing the information flow that cannot be observed directly but influences the general activity of the markets (and hence the intensity of the process). In this sense, the LFI approach combines the idea of latent factor models arising from the mixture-of-distribution hypothesis (see Clark, 1973) with the concept of dynamic intensity processes. It can be seen as the counterpart of the stochastic volatility (SV) model (Taylor, 1982) or the stochastic conditional duration (SCD) model (Bauwens and Veredas, 1999). However, while in the SV or SCD model, the process dynamics are completely driven by the dynamics of the latent component, the LFI model is based on observation driven *and* latent dynamics. Hence, using the terminology of Cox (1981), the LFI model combines the idea of observation driven models and parameter driven models. In this sense, the latent component models unobserved dynamic heterogeneity which is not captured by the observation driven part of the model. Then, two limit cases emerge naturally: one when the latent factor is irrelevant and the intensity is completely described by observation driven dynamics, and the other when the observable components are not relevant and the latent factor completely dominates. Hence, econometrically, the existence of a latent factor can be interpreted as an indication that the observation driven part of the model is not able to completely capture the dynamics in the data.

While the latent component is assumed to follow an AR dynamic that is updated by latent innovations, the observation driven component may be specified in different ways. In this paper, we discuss parameterizations of the observation driven component based on autoregressive conditional intensity (ACI) processes proposed by Russell (1999), as well as self-exciting processes as introduced by Hawkes (1971). In the so-called latent factor autoregressive conditional intensity (LF-ACI) model, the conditional intensity given the observable history of the process follows itself a dynamic process which is updated by observable innovations. Correspondingly, in the latent factor Hawkes (LF-Hawkes) model, the conditional intensity is driven by an exponentially weighted function of the backward recurrence time to all previous points.

Based on simulation experiments, we analyze the dynamic properties of the LFI specification under different settings of the model. It turns out that the LFI model is a flexible tool able to capture the joint dynamics of multivariate point processes. Estimation of the LFI model by maximum likelihood (ML) is not easy since the dynamic latent factor is not observable and must be integrated out. We use the efficient importance sampling (EIS) technique introduced by Richard (1998) and illustrate the application of this approach to the proposed model. Then, the LFI model is applied to the bivariate trading process of the Allianz and BASF stock in the German electronic trading system XETRA. We find evidence for the existence of a joint unobservable component which influences the individual processes. Particularly, in the ACI framework, the inclusion of a latent dynamic component strongly improves the goodness-of-fit of the model. However, in the Hawkes model, only slight improvements of the specification can be observed revealing some evidence for an overall better fit of a Hawkes type specification compared to an ACI approach.

The remainder of the paper is organized in the following way. In Section 2 we present the basic structure of the LFI model. In the Sections 3 and 4 we discuss different parameterizations of the observation driven component leading to the LF-ACI model and the LF-Hawkes model. Dynamic properties of both types of models are illustrated in Section 5 while estimation and diagnostic issues are exposed in Section 6. The empirical illustration is provided in Section 7 and Section 8 concludes.

2 The latent factor intensity model

2.1 Definitions

Let t denote the physical (calendar) time. Furthermore, let $\{t_i^s\}_{i\in\{1,2,\dots\}}$, $s = 1,\dots S$, be S sequences of nonnegative random variables on some probability space $(\Omega, \mathcal{F}, \mathcal{P})$ associated with random arrival times $0 \leq t_i^s \leq t_{i+1}^s$. Then, the sequences $\{t_i^s\}$, $s = 1,\dots S$, represent a S-dimensional point process on $[0,\infty)$.

Accordingly, $N^s(t) := \sum_{i\geq 1} \mathbb{1}_{\{t_i^s \leq t\}}$ and $\breve{N}^s(t) := \sum_{i\geq 1} \mathbb{1}_{\{t_i < t\}}$ denote the right-continuous and, respectively, left-continuous counting functions associated with s-type events. Correspondingly, $N(t)$ and $\breve{N}(t)$ are the right-continuous and, respectively, left-continuous counting functions of the *pooled* process, which pools and orders the arrival times of *all* single processes. Let n denote the number of points in the pooled process and n^s the number of s-type elements in the sample. Furthermore, define y_i^s as an indicator variable that takes the value 1 if the i-th point of the pooled process is of type s. Moreover, $x_i^s := t_i^s - t_{i-1}^s$, $i = 1,\dots n^s$, and $x^s(t) := t - t^s_{\breve{N}^s(t)}$ denote the inter-event durations and backward recurrence times associated with the s-th process. Furthermore, let $\{z_i\}_{i\in\{1,2,\dots\}}$ be a sequence of random vectors corresponding to the characteristics associated with the arrival times $\{t_i\}_{i\in\{1,2,\dots\}}$.

In the following, we assume that for small positive Δ,

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

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

A central concept in the theory of point processes is the intensity function[3] that is defined as following:

Definition 1 *Let $N^s(t)$ be the s-type component of a S-dimensional point process on $[0, \infty)$ that is adapted to some history $\mathcal{F}_t$ and assume that $\lambda^s(t|\mathcal{F}_t)$ is a positive process with sample paths that are left-continuous and have right-hand limits. Then, the process*

$$\lambda^s(t|\mathcal{F}_t) := \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \mathrm{E}\left[N^s(t+\Delta) - N^s(t)|\,\mathcal{F}_t\right], \qquad \lambda^s(t|\mathcal{F}_t) > 0,\ \forall\, t, \tag{2}$$

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

Under assumption eq. (1), the intensity function is alternatively written as

$$\lambda^s(t|\mathcal{F}_t) = \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \Pr\left[(N^s(t+\Delta) - N^s(t)) > 0 \,|\mathcal{F}_t\right], \tag{3}$$

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

An important result in stochastics literature is that under fairly weak regularity conditions[4], the integrated intensity function

$$\Lambda^s(t^s_{i-1}, t^s_i) := \int_{t^s_{i-1}}^{t^s_i} \lambda^s(u|\mathcal{F}_u) du \tag{4}$$

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

$$\Lambda^s(t^s_{i-1}, t^s_i) \sim \text{ i.i.d. } Exp(1). \tag{5}$$

This property also holds for the integrated intensity of the pooled process. Since the $\mathcal{F}_t$-intensity of the pooled process is $\lambda(t|\mathcal{F}_t) = \sum_{s=1}^{S} \lambda^s(t|\mathcal{F}_t)$, it is easy to show that

$$\Lambda(t_{i-1}, t_i) := \int_{t_{i-1}}^{t_i} \sum_{s=1}^{S} \lambda^s(t|\mathcal{F}_t) dt := \epsilon_i \sim \text{i.i.d. } Exp(1). \tag{6}$$

Therefore, $\Lambda(t_{i-1}, t_i)$ can be interpreted as a generalized error (e.g. in the spirit of Cox and Snell, 1968) that establishes the link between the intensity function and the duration until the occurrence of the next point. For example, in the simple case of an intensity function that does not depend on the

[3]Note that the notation is not consistent in all papers or textbooks. For example, Brémaud (1981) and Karr (1991) denote it as "stochastic intensity function". Here, we follow the notation of, for example, Aalen (1978) and Snyder and Miller (1991).

[4]See e.g. Theorem T16 in Brémaud (1981) or Bowsher (2002).

backward recurrence time since the last point, the duration x_i is given by $\epsilon_i/\lambda(t_i|\mathcal{F}_{t_i})$. Moreover, the generalized error indicates whether the intensity function under predicts ($\Lambda(t_{i-1}, t_i) > 1$) or over predicts ($\Lambda(t_{i-1}, t_i) < 1$) the number of events between t_{i-1} and t_i.

For ease of exposition, we refrain from accounting explicitly for time-varying covariates. However, this extension is straightforward in the given setting (see, for example, Hall and Hautsch, 2003).

2.2 The basic latent factor intensity model

Note that the $\mathcal{F}_t$-intensity function $\lambda(t|\mathcal{F}_t)$ completely characterizes the evolution of the point process in dependence of the process history $\mathcal{F}_t$. Hence, under the assumption that $\mathcal{F}_t$ consists of the complete (observable) history of the pooled process, i.e., for example, $\mathcal{F}_t = \sigma(t_{N(t)}, z_{N(t)}, t_{N(t)-1}, z_{N(t)-1}, \ldots, t_1, z_1)$, the intensity $\lambda(t|\mathcal{F}_t)$ is conditionally deterministic given $\mathcal{F}_t$. However, the assumption that $\lambda(t|\mathcal{F}_t)$ is completely described by the *observable* process history is rather unrealistic. For this reason, in the classical duration literature, the inclusion of unobserved heterogeneity effects plays an important role to obtain well specified econometric models. Correspondingly, in the framework of point processes, the consideration of unobserved factors leads to the class of doubly stochastic Poisson processes. The standard doubly stochastic Poisson process (see, e.g., Grandell, 1976, or Cox and Isham, 1980), is characterized by the intensity $\lambda(t|\mathcal{F}_t^*)$, where $\mathcal{F}_t^*$ denotes the history of some unobserved process up to t.

Following this strand of the literature, we assume that the intensity function depends not only on the observable process history, but also on some *unobservable* (dynamic) factor. In this context, we define the information set $\mathcal{F}_t$ more explicitly as $\mathcal{F}_t := \sigma(\mathcal{F}_t^o \cup \mathcal{F}_t^*)$, consisting of an observable conditioning set $\mathcal{F}_t^o$ that includes the complete observable history of the pooled process and possible marks up to t, as well as an unobservable history $\mathcal{F}_t^*$ of some latent factor $\lambda^*(t)$.

The basic latent factor intensity (LFI) model combines features of observation driven models and parameter driven models. Hence, it is based on the assumption that the intensity function is driven by observation driven dynamics, as well as latent dynamics. In other words, we assume that the conditional intensity function given the observable process history is not deterministic, but stochastic and follows itself a dynamic process. Then, the basic LFI model for the s-type process is given by

$$\lambda^s(t|\mathcal{F}_t) = \lambda^{o,s}(t) \left(\lambda^*_{\breve{N}(t)+1}\right)^{\sigma_s^*}, \tag{7}$$

where $\lambda^{o,s}(t) := \lambda^{o,s}(t|\mathcal{F}_t^o)$ denotes a conditionally deterministic univariate or multivariate intensity component given the *observable* history up to t, $\mathcal{F}_t^o$. Likewise, $\lambda^*_{\breve{N}(t)+1} := \lambda^*(t_{\breve{N}(t)+1})$ stands for the latent intensity component that depends on its past own history $\mathcal{F}_t^*$ and is updated at each point of the (pooled) process. Furthermore, σ_s^* denotes a process-specific parameter that allows the common latent factor to have different weights in each intensity component.

As a starting point, we assume that $\lambda^{o,s}(t)$ is constant and discuss the specification of the latent component. Obviously, the latter is only identifiable at the particular observed points t_i. To guarantee its positivity, we specify it as conditionally i.i.d. lognormally distributed, i.e.

$$\ln \lambda_i^* | \mathcal{F}_{t_i}^* \sim \text{ i.i.d. } N(m_i^*, 1). \tag{8}$$

Notice that the latent factor is indexed by the left-continuous counting function, i.e. it does not change between t_{i-1} and t_i. More precisely, it is assumed that λ_i^* has left-continuous sample paths with right-hand limits which means that it is updated instantaneously after the occurrence of t_{i-1} and remains constant until t_i (inclusive). The choice of a normal distribution allows us to parameterize the conditional mean and the conditional variance of the latent component separately. For simplicity, we consider an AR(1) specification of the latent factor process, even if the specification can be extended to more general processes, like, for instance, ARMA processes of higher order. Thus,

$$\ln \lambda_i^* = a^* \ln \lambda_{i-1}^* + u_i^*, \qquad u_i^* \sim \text{i.i.d. } N(0,1). \tag{9}$$

In order to obtain a valid intensity process, it is assumed that the latent innovations u_i^* are independent from the series of the integrated intensities $\epsilon_i := \Lambda(t_{i-1}, t_i)$ which are i.i.d. standard exponentially distributed. An important prerequisite for weak stationarity of the LFI model is weak stationarity of the latent component which is fulfilled for $|a^*| < 1$. Notice that we omit a constant term since we include it in the observable component $\lambda^{o,s}(t)$.

By defining $\lambda_i^{s*} := \sigma_s^* \ln \lambda_i^*$ as the latent factor that influences the s-type component, it is easy to see that

$$\lambda_i^{s*} = a^* \lambda_{i-1}^{s*} + \sigma_s^* u_i^*.$$

Hence, σ_s^* scales the standard deviation of the latent factor influencing the $s-$type intensity process.[5] This flexibility ensures that the impact of a latent shock u_i^* on the individual processes can differ and is driven by the parameter σ_s^*.[6] Hence, the major idea behind this specification is to assume a *joint* latent dynamic which influences the single components of the multivariate process differently.

This approach can be extended by specifying σ_s^* time-varying. Then, the process-specific scaling factors can change over time in order to allow for conditional heteroscedasticity. An important source of heteroscedasticity could be intradaily seasonality associated with deterministic fluctuations of the overall information and activity flow. For example, it could be due to institutional settings, like the opening of other related markets. Hence, a reasonable specification could be to index σ_s^* itself by the counting function and parameterize it in terms of a linear spline function:

$$\sigma_{s,i}^* = \sigma_0 \left(1 + \sum_{k=1}^{K} \nu_{0,k}^* 1_{\{\tau(t_i) \geq \bar{\tau}_k\}} (\tau(t_i) - \bar{\tau}_k) \right), \tag{10}$$

where $\tau(t)$ denotes the calendar time at t, $\bar{\tau}_k$, $k = 1, \ldots, K-1$ denote the exogenously given (calendar) time points and $\nu_{0,k}^*$ the corresponding coefficients of the spline function.

[5]For this reason, we assume $\text{Var}[u_i^*] = 1$.

[6]Note that σ_s^* can be even negative. Hence, theoretically it is possible that the latent component simultaneously increases one component while decreasing the other component.

A further valuable generalization of the LFI model is to allow for regime switching latent dynamics. Thus, a more flexible LFI model is obtained by specifying the autoregressive parameter in dependence of the length of the previous spell. Such a specification is in line with a threshold model (see e.g. Tong, 1990, or Zhang, Russell, and Tsay, 2001) and is obtained by

$$\ln \lambda_i^* = a_r^* \mathbb{1}_{\{\bar{x}_{r-1} < x_{\breve{N}(t)} \leq \bar{x}_r\}} \ln \lambda_{i-1}^* + u_i^*, \qquad r = 1, \ldots, R-1, \tag{11}$$

where $\bar{x}_r$ denotes the exogenously given thresholds (with $\bar{x}_0$:=0), and a_r^* are the regime dependent latent autoregressive parameters.

Moreover, the latent parameters a^* and σ^{s*} might be specified in dependence of the type of the most recent observed point. Thus, in this context, the LFI model can be expressed in terms of $\lambda_i^{sr*} := \sigma_{s,r}^* \ln \lambda_i^*$, with

$$\lambda_i^{sr*} = a_r^* \mathbb{1}_{\{y^r_{\breve{N}(t)}=1\}} \lambda_{i-1}^{sr*} + \sigma_{s,r}^* \mathbb{1}_{\{y^r_{\breve{N}(t)}=1\}} u_i^*, \tag{12}$$

where $\sigma_{s,r}^*$ denotes the s-type scaling factor if the most recent event was of type r.

The observation driven component $\lambda^{o,s}(t)$ can also be specified as a dynamic process. Two possible choices are an autoregressive conditional intensity (ACI) specification as proposed by Russell (1999) and a Hawkes process (Hawkes, 1971), giving rise to the latent factor ACI (LF-ACI) and the latent factor Hawkes (LF-Hawkes) models. Both types of models are presented in the following sections.

3 The latent factor autoregressive conditional intensity model

The so-called latent factor autoregressive conditional intensity (LF-ACI) model is obtained by specifying $\lambda^{o,s}(t)$ according to an autoregressive conditional intensity (ACI) approach. Then, each observation driven component is parameterized as

$$\lambda^{o,s}(t) = \Psi^s_{\breve{N}(t)} \lambda_0^s(t) s^s(t), \quad s = 1, \ldots S, \tag{13}$$

where $\Psi_i^s := \Psi^s(t_i)$ captures the dynamic structure and possible exogenous variables, $\lambda_0^s(t)$ is a baseline intensity component (also called backward recurrence time function), and $s^s(t)$ is a deterministic function (e.g. to capture seasonality).

As proposed by Russell (1999), a simple way to specify the baseline intensity functions is to assume that $\lambda_0^s(t)$ is constant, but depends on the type of the event that occurred most recently, i.e.

$$\lambda_0^s(t) = \exp(\omega_r^s)\, y^r_{\breve{N}(t)}, \quad r = 1, \ldots, S, \quad s = 1, \ldots, S. \tag{14}$$

Alternatively, the baseline intensity function may be specified in terms of a (multivariate) Weibull parameterization, i.e.,

$$\lambda_0^s(t) = \exp(\omega^s) \prod_{r=1}^{S} x^r(t)^{p_r^s - 1}, \; (p_r^s > 0). \tag{15}$$

Then, a special case occurs when the s-th process depends only on its own backward recurrence time, which corresponds to $p_r^s = 1 \; \forall \, r \neq s$.

Russell (1999) proposes to specify the vector of specific elements $\tilde{\Psi}_i = (\tilde{\Psi}_i^1 \, \tilde{\Psi}_i^2 \ldots \tilde{\Psi}_i^S)'$, where

$$\tilde{\Psi}_i^s := \ln \Psi_i^s - z_{i-1}' \gamma^s, \tag{16}$$

with γ^s process-specific parameters associated with marks, as a vectorial ARMA process, which is parameterized by[7]

$$\tilde{\Psi}_i = \left(A^s \breve{\epsilon}_{i-1} + B \tilde{\Psi}_{i-1} \right) y_{i-1}^s, \tag{17}$$

where $A^s = \{\alpha_j^s\}$ is a $S \times 1$ coefficient vector and $B = \{\beta_{ij}\}$ denotes a $S \times S$ coefficient matrix.

Like the latent component, Ψ_i is a left-continuous function that does not change between t_{i-1} and t_i, i.e., $\Psi_i = \Psi(t_i)$ for $t_{i-1} < t \leq t_i$, meaning that Ψ_i is known instantaneously after the occurrence of t_{i-1} and does not change until t_i. Then, $\lambda^{o,s}(t)$ changes between t_{i-1} and t_i only as a deterministic function of time (according to $\lambda_0^s(t)$ and $s^s(t)$).

The innovation term is specified by

$$\breve{\epsilon}_i = \breve{\epsilon}_i^s y_{i-1}^s, \tag{18}$$

where

$$\breve{\epsilon}_i^s := 1 - \Lambda^{o,s}(t_{i-1}^s, t_i^s) \tag{19}$$

or (in logarithmic terms)

$$\breve{\epsilon}_i^s := -0.5772 - \ln \Lambda^{o,s}(t_{i-1}^s, t_i^s) \tag{20}$$

with

$$\Lambda^{o,s}(t_{i-1}^s, t_i^s) := \sum_j \int_{t_j}^{t_{j+1}} \lambda^{o,s}(u) du = \sum_j \left(\lambda_j^* \right)^{-\sigma_s^*} \int_{t_j}^{t_{j+1}} \lambda^s(u|\mathcal{F}_u) du \tag{21}$$

with $j : t_{i-1}^s \leq t_j \leq t_i^s$ and $\sum_j \int_{t_j}^{t_{j+1}} \lambda^s(u|\mathcal{F}_u) du = \varepsilon_i \sim$ i.i.d. $Exp(1)$.

Hence, $\Lambda^{o,s}(t_{i-1}^s, t_i^s)$ corresponds to an i.i.d. exponential variate that is piecewise standardized by a strictly stationary lognormal random variable. However, since these pieces (i.e., the duration between consecutive points of the pooled process) are determined endogenously and depend themselves on the (multivariate) intensity, it is not possible to establish the distributional and dynamical properties of $\Lambda^{o,s}(t_{i-1}^s, t_i^s)$ analytically. A further difficulty arises by the fact that the resulting innovation $\breve{\epsilon}_i$ is itself a random mixture of the particular processes $\breve{\epsilon}_i^s$, $s = 1, \ldots, S$, which renders the derivation of the distributional properties of $\breve{\epsilon}_i$ even in the benchmark case of a pure ACI model (for $\lambda_i^* = 1$) rather difficult[8].

[7]For simplicity, we restrict our presentation to a lag order of one. The extension to higher order specifications is straightforward.

[8]See also the discussion in Bowsher (2002).

Nevertheless, own simulation studies show that in that specific case, $\mathrm{E}[\breve{\epsilon}_i] = \mathrm{E}[\epsilon_i^{s*}]$. Thus, centering the innovation terms as presented in eq. (19) and (20) makes them comparable to zero mean innovations in the benchmark case of a pure ACI model. Furthermore, notice that eq. (17) implies a regime-switching structure since A^s is a vector of coefficients reflecting the impact of the innovation term on the intensity of the S processes when the previous point $(t_{\breve{N}(t)})$ was of type s. Since we assume that the process is orderly, and thus only one type of point can occur at each instant, the innovation $\breve{\epsilon}_i$ is a scalar random variable and is associated with the most recently observed process.

The major advantage of this specification is that the innovations can be computed directly based on the *observable* history of the process, leading to a clear-cut separation between the observation driven and the parameter driven components of the model. Nevertheless, notice that $\Lambda^{o,s}(t_{i-1}, t_i)$, and thus the LFI-ACI innovation $\breve{\varepsilon}_i$ depends on lags of λ_i^*. Therefore, the components $\lambda^{o,s}(t)$ themselves are directly affected by lags of the latent factor. Hence, λ_i^* influences the intensity $\lambda^s(t|\mathcal{F}_t)$ not only contemporaneously (according to eq. (7)), but also through its lags. For this reason, the latent factor causes cross-autocorrelations not only between $\lambda^s(t|\mathcal{F}_t)$, but also between the particular intensity components $\lambda^{o,s}(t)$, $s = 1, \ldots, S$. This will be illustrated in more detail in Section 5.

Necessary conditions for the weak stationarity of the LF-ACI model are the weak stationarity of λ_i^* ($|a^*| < 1$), as well as of $\tilde{\Psi}_i$ given the history of the latent factor $\Lambda_i^* := \{\lambda_j^*\}_{j=1}^i$. Given Λ_i^*, it is obvious that the innovations $\breve{\epsilon}_i$ are i.i.d., and thus, $\tilde{\Psi}_i|\Lambda_i^*$ is weakly stationary whenever the eigenvalues of B lie inside the unit circle. However, because of the strong interaction of both dynamic components λ_i^* and $\tilde{\Psi}_i$, it is not possible to derive the dynamical properties of the LF-ACI model analytically (see also the discussion in Section 5). Nevertheless, own simulation experiments as outlined in Section 5 strongly support the hypothesis that the aforementioned restrictions are not only necessary but also sufficient for the model to be weakly stationary.

In order to provide more insights into the dynamic properties of the LF-ACI model, we rewrite a bivariate version of the model in logarithmic terms. Due to the log-linear structure of both components $\lambda^{o,s}(t)$ and λ_i^*, the model dynamics are characterized by a direct interaction between the latent dynamic factor, which is updated by latent innovations u_i^* and the observation driven dynamic component, which is updated by the innovations $\breve{\epsilon}_i$. By excluding covariates and assuming that $\lambda_0^s(t) = \exp(\omega^s)$ and $s^s(t) = 1$ for $s = 1, 2$, the model can be written in logarithm as

$$\begin{aligned}\ln \lambda^1(t|\mathcal{F}_t) - \omega^1 = \Big\{ & \alpha_1^s \breve{\epsilon}_{\breve{N}(t)} + (\sigma_1^* a^* - \sigma_1^* \beta_{11} - \sigma_2^* \beta_{12}) \ln \lambda^*_{\breve{N}(t)} + \sigma_1^* u^*_{\breve{N}(t)} \\ & + \beta_{11} [\ln \lambda^1(t_{\breve{N}(t)}|\mathcal{F}_{t_{\breve{N}(t)}}) - \omega^1] + \beta_{12} [\ln \lambda^2(t_{\breve{N}(t)}|\mathcal{F}_{t_{\breve{N}(t)}}) - \omega^2] \Big\} \, y^s_{\breve{N}(t)}\end{aligned} \tag{22}$$

$$\begin{aligned}\ln \lambda^2(t|\mathcal{F}_t) - \omega^2 = \Big\{ & \alpha_2^s \breve{\epsilon}_{\breve{N}(t)} + (\sigma_2^* a^* - \sigma_1^* \beta_{21} - \sigma_2^* \beta_{22}) \ln \lambda^*_{\breve{N}(t)} + \sigma_2^* u^*_{\breve{N}(t)} \\ & + \beta_{21} [\ln \lambda^1(t_{\breve{N}(t)}|\mathcal{F}_{t_{\breve{N}(t)}}) - \omega^1] + \beta_{22} [\ln \lambda^2(t_{\breve{N}(t)}|\mathcal{F}_{t_{\breve{N}(t)}}) - \omega^2] \Big\} \, y^s_{\breve{N}(t)}.\end{aligned} \tag{23}$$

This is a switching VARMA(1,1) model, augmented by the latent component, where the operating regime is determined by the occurrence of one of the individual processes. Notice that *each* particular process is updated at every occurrence of a new event.

The choice of the parameterization of A^s and B determines the dynamic and interdependence structure of the particular processes. For example, if for each s, α_j^s is different from 0 only if $s = j$ and $\beta_{ij} = 0$ except for $i = j$, the intensity of process j is updated only when a point of type j has occurred, and only the lagged value of the intensity of process j contributes to the current intensity of process j.

4 The latent factor Hawkes model

A valuable alternative to an autoregressive specification of the observable intensity component $\lambda^{o,s}(t)$ is a self-exciting process. The basic linear self-exciting process for the s-type component is given by

$$\lambda^{o,s}(t) = \exp(\omega^s) + \int_{-\infty}^{t} w(t-u)dN^s(u) = \exp(\omega^s) + \sum_{i\geq 1} \mathbb{1}_{\{t_i^s \leq t\}} w(t - t_i^s), \tag{24}$$

where ω^s is a constant, and $w(\cdot)$ is a non-negative weighting function of the backward recurrence time to all previous points (see, for example, Hawkes, 1971 or Cox and Isham, 1980). Hawkes (1971) introduces a class of self-exciting processes given by

$$\begin{aligned} \lambda^{o,s}(t) &= \mu^s(t) + \int_{(0,t)} \sum_{r=1}^{S} \sum_{j=1}^{P} \alpha_j^{sr} \exp(-\beta_j^{sr}(t-u))dN^r(u) \\ &= \mu^s(t) + \sum_{r=1}^{S} \sum_{j=1}^{P} \sum_{k=1}^{\breve{N}^r(t)} \alpha_j^{sr} \exp(-\beta_j^{sr}(t - t_k^r)), \end{aligned} \tag{25}$$

where $\alpha_j^{sr} \geq 0$, $\beta_j^{sr} \geq 0$ $(j = 1, \ldots, P)$ are parameters, and $\mu^s(t)$ is a deterministic function of time. If $P = 1$, the impact of a previous s-type event on the s-type intensity function at t decays exponentially, from the level α_1^{ss}, at a rate driven by β_1^{ss}. For $P > 1$, there is a superposition of exponentially decaying weighted sums of the backward recurrence times to all previous points. Moreover, for $S > 1$, $\lambda^{o,s}(t)$ depends not only on the backward recurrence time to all s-type points, but also on the backward recurrence time to all other points of the pooled process. Note that the marginal contribution of a previous event to the current intensity is independent from the number of intervening events.

Obviously, alternative functional forms for the decay function are possible.[9] However, the choice of an exponential decay simplifies the derivation of the theoretical properties of the model. More details can be found in Hawkes (1971), Hawkes and Oakes (1974) or Ogata and Akaike (1982) among others. Hawkes (1971) provides parameter restrictions for α^{sr} and β^{sr} under which $\lambda^{o,s}(t)|\Lambda_i^*$ is weakly stationary. However, in general, these conditions cannot be explicitly expressed in closed form and have to be verified by numerical methods.[10] Nevertheless, in the univariate case $(S = 1)$, weak stationarity of $\lambda^{o,s}(t)|\Lambda_i^*$ is ensured by the restriction

$$\sum_{j=1}^{P} \alpha_j / \beta_j < 1.$$

[9] For example, Kawasaki (2002) uses Laguerre polynomials.
[10] For more details, see Hawkes (1971).

The function $\mu^s(t)$ can be specified in terms of covariates $z_{\breve{N}(t)}$ and a seasonality function $s^s(t)$, thus

$$\mu^s(t) = \exp(\omega^s) + s^s(t) + z'_{\breve{N}(t)}\gamma^s. \tag{26}$$

Alternatively, one might assume a multiplicative impact of the seasonality function.

Hawkes-type intensity models serve as epidemic models since the occurrence of a number of events increases the probability of further events. In natural sciences, they play an important role in modelling the emission of particles from a radiating body or in forecasting seismic events (see, for example, Vere-Jones, 1970). As far as we know, with the exception of the studies by Bowsher (2002) and Kawasaki (2002), they have never been applied to financial data.[11] As a bivariate LF-ACI model given by eq. (22) and (23), the bivariate LF-Hawkes model, with $P = 1$ and $\mu^s(t) = \exp(\omega^s)$ has 13 parameters. If $\alpha_{12} = \alpha_{21} = 0$ (in which case β_{12} and β_{21} may be set to 0), the two Hawkes components of the intensities are unrelated: each intensity is updated only by its own past points.

5 Dynamic properties of LFI models

The conditional moments of the intensity function given the information set $\mathcal{F}_{t_i}$ are easily obtained by using the moments of the lognormal distribution, but they are of limited interest since the conditioning information is partly unobservable. The computation of conditional moments of $\lambda(t|\mathcal{F}_t)$, given the *observable* information set, is difficult since the latent variable has to be integrated out. In the following, we illustrate the computation of conditional moments given $\mathcal{F}^o_{t_{i-1}}$ in a general way for arbitrary parameterizations of $\lambda^o(t)$. Denote by w_i a row of the matrix of observable variables W and correspondingly define $W_i = \{w_j\}_{j=1}^i$. Furthermore, let $f(W_i, \Lambda_i^*|\theta)$ denote the joint density function of W_i and Λ_i^*, and $p(\lambda_i^*|W_i, \Lambda_i^*, \theta)$ the conditional density of λ_i^* given W_i and Λ_i^*, where θ denotes the parameter vector of the model. Then, the conditional expectation of an arbitrary function $\vartheta(\lambda_i^*)$ given the observable information set up to t_{i-1} can be computed as

$$\mathrm{E}\left[\vartheta(\lambda_i^*)\left|\mathcal{F}^o_{t_{i-1}}\right.\right] = \frac{\int \vartheta(\lambda_i^*)p(\lambda_i^*|W_{i-1}, \Lambda_{i-1}^*, \theta) f(W_{i-1}, \Lambda_{i-1}^*|\theta)d\Lambda_i^*}{\int f(W_{i-1}, \Lambda_{i-1}^*|\theta)d\Lambda_{i-1}^*} \tag{27}$$

(the expectation depends on θ but we drop it in the notations). In general, the integrals in this ratio cannot be computed analytically, but they can be computed numerically, e.g. by efficient importance sampling (see Subsection 6.3).

Hence, the computations of the conditional expectations $\mathrm{E}\left[t_i|\,\mathcal{F}^o_{t_{i-1}}\right]$ and $\mathrm{E}\left[\lambda(t_i)|\,\mathcal{F}^o_{t_{i-1}}\right]$ are conducted by exploiting the exponential distribution of the integrated intensity $\epsilon_i := \Lambda(t_{i-1}, t_i)$. The calculations are performed by conditioning on predetermined values of ϵ_i and

[11] Bowsher (2002) proposes a generalization (the so-called generalized Hawkes model) that nests the basic model and allows for spillover effects between particular trading days. Kawasaki (2002) focuses on the specification of the seasonal intradaily component through a flexible Fourier form (i.e. a sum of sinusoids).

λ^*_{i-1} and applying the law of iterated expectations. For example, $\mathrm{E}\left[t_i | \mathcal{F}^o_{t_{i-1}}\right]$ is computed as

$$\mathrm{E}\left[t_i | \mathcal{F}^o_{t_{i-1}}\right] = \mathrm{E}\left[g_1(\cdot) | \mathcal{F}^o_{t_{i-1}}\right] \tag{28}$$

$$g_1(\cdot) = \mathrm{E}\left[t_i | \lambda^*_{i-1}, \mathcal{F}^o_{t_{i-1}}\right] = \mathrm{E}\left[g_2(\cdot) | \lambda^*_{i-1}, \mathcal{F}^o_{t_{i-1}}\right], \tag{29}$$

where $t_i = g_2(\epsilon_i; \mathcal{F}^o_{t_{i-1}}, \lambda^*_{i-1})$ is determined by solving the equation $\Lambda(t_i, t_{i-1}) = \epsilon_i$ for given values of ϵ_i, λ^*_{i-1} and $\mathcal{F}^o_{t_{i-1}}$. Clearly, the complexity of the expression for t_i depends on the parameterization of $\lambda^o(t)$. However, a closed form solution for $g_2(\cdot)$ generally does not exist. After computing the conditional expectation of t_i given λ^*_{i-1} (eq.(29)), the next step is to integrate over the latent variable (eq. (28)) based on the integrals in eq. (27). The computation of $\mathrm{E}\left[\lambda(t_i) | \mathcal{F}_{t_{i-1}}\right]$ is done similarly.

In any event, the computation of fully unconditional moments has to be performed by simulation since no closed form analytical expression can be derived. We simulated several processes and computed the ACF of the intensity components and of the implied duration series (see Appendix). Figures 1 through 5 show the (cross-)autocorrelation functions of the individual intensity components and the corresponding duration processes based on simulated bivariate LF-ACI(1,1) processes using logarithmic innovations (eq. (20)). All simulations are based on constant baseline intensity functions, and a diagonal specification of B. Figures 1 and 2 are based on LF-ACI specifications implying no interactions between the observation driven components $\lambda^{o,s}(t)$ and strong serial dependencies in the latent factor. Not surprisingly, it turns out that the impact of the latent factor strongly depends on the magnitude of the latent variances. In particular, the simulation in Figure 1 is based on $\sigma^*_1 = \sigma^*_2 = 0.01$. Here, only a very weak cross-autocorrelation between $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$ can be observed. In contrast, in Figure 2, the impact of the latent factor is clearly stronger. Here, it causes also slight contemporaneous correlations between the observable components $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. As already discussed above, this is due to the fact that λ^*_i influences the intensity components not only contemporaneously but also through the lagged innovations $\breve{\epsilon}_i$. One sees that the latent dynamics dominate the dynamics of the particular processes $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$, as well as of x^1_i and x^2_i, leading to quite similar autocorrelation functions. Moreover, a clear increase of the autocorrelations in the duration processes is observed in comparison with Figure 1.

Figure 3 shows the corresponding plots of symmetrically interdependent ACI components ($\alpha^1_2 = \alpha^2_1 = 0.05$). Here, the latent variable enforces the contemporaneous correlation between the two processes and drives the autocorrelation functions of the individual intensity components towards higher similarity. Moreover, while the CACF of $\lambda^{o,1}(t_i)$ vs. $\lambda^{o,2}(t_i)$ shows an asymmetric shape (due to the difference between $\beta_{11} = 0.97$ and $\beta_{22} = 0.7$), the joint latent factor causes a more symmetric shape of the CACF between $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$.

The DGP associated with Figure 4 is based on individual weak dynamics in the particular intensity components $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$ while the impact of the latent factor is quite strong. Here, λ^*_i completely dominates the dynamics of the joint system. It causes strong, similar autocorrelations in the components $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$, as well as in $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Moreover, its impact on the CACF is clearly stronger than in the cases discussed above. In particular, the contemporaneous correlation between $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$ is nearly equal to one. Nevertheless, the CACF dies out quite quickly which is due to the latent AR(1) structure.

The parameterization underlying Figure 5 resembles the specification in Figure 3, except that the parameter σ_1^* is fixed at 0.1 and σ_2^* at -0.1. One sees clearly that the latent component influences $\lambda^1(t_i|\mathcal{F}_{t_i})$ positively while influencing $\lambda^2(t_i|\mathcal{F}_{t_i})$ negatively which causes negative cross-autocorrelations between $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$ as well as a flattening of the CACF between the observation driven components compared to Figure 3.

Figures 6 through 10 are based on simulated LF-Hawkes(1) processes (i.e. with the order $P = 1$). The plots in Figure 6 are based on *independent* Hawkes(1) processes for the observable component that interact with a relatively weak latent factor ($\sigma_1^* = \sigma_2^* = 0.05$). One sees that the dynamics of the individual intensity components remain nearly unaffected, thus, the latent factor has only a weak influence. Nevertheless, slight contemporaneous correlations between $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$ as well as $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$ are observed. The asymmetric dip in the CACF between $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$ is obviously caused by the different persistence of the underlying Hawkes processes.

Figure 7 is based on the same parameterization as Figure 6, except that the standard deviations of the latent factor are increased to $\sigma_1^* = \sigma_2^* = 0.5$. The graphs show that in that case, the latent factor dominates the dynamics of the complete system causing significant contemporaneous correlations and cross-autocorrelations between $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$, as well as driving the dynamics of the individual processes to a higher degree of similarity. The fact that even the dynamics of the observation driven components strongly resemble each other (compared to Figure 6) illustrate the strong impact of λ_i^* on the innovations (in this context past backward recurrence times) of the model.

Figure 8 is based on interdependent Hawkes(1) processes ($\alpha^{12} = \alpha^{21} = 0.1$). Here, the latent factor clearly amplifies the contemporaneous correlation structure, but weakens the persistence in the cross-autocorrelations. Figure 9 is the counterpart to Figure 4 revealing only weak dynamics in the particular components $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$ combined with a strong latent dynamic. Again, the resulting autocorrelation functions of the intensities, as well as of the duration processes are clearly dominated by the dynamics of the latent factor. Finally, in Figure 10, we assume different persistent independent Hawkes processes for the observation driven components which are oppositely influenced by the latent factor ($\sigma_1^* = 0.1$, $\sigma_2^* = -0.1$). This parameterization leads to significant negative cross-autocorrelations not only between $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$ but also between $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. However, in that case, the latent factor obviously does not drive the dynamics of the individual components towards higher similarity.

6 Inference

6.1 Construction of the likelihood function

In the following we discuss the construction and computation of the likelihood function for the multivariate case. Let W denote the $n \times (1 + M + S)$ data matrix, consisting of the n points t_i, the vector z_i of covariates, and the particular indicator variables y_i^s. Thus $w_i := (t_i \; z_i' \; y_i^1 \; \dots \; y_i^S)$ is a row of this matrix and, correspondingly, we define $W_i := \{w_j\}_{j=1}^i$. Let θ denote the vector of parameters of the LFI model, and introduce the notation $l_i := \ln \lambda_i^*$.

If the latent variables $L_n^* := \{l_i^*\}_{i=1}^n$ were observable, the likelihood function would be given by

$$\mathcal{L}(W, L_n^*; \theta) = \prod_{i=1}^{n} \prod_{s=1}^{S} \exp(-\Lambda^s(t_{i-1}, t_i)) \left[\lambda^{o,s}(t_i)\lambda_i^{*\sigma_s^*}\right]^{y_i^s}, \tag{30}$$

where

$$\Lambda^s(t_{i-1}, t_i) = \lambda_i^{*\sigma_s^*} \Lambda^{o,s}(t_{i-1}, t_i). \tag{31}$$

For any i, the process to which t_i corresponds (assume it is process j so that $y_i^j = 1$ and $y_i^s = 0 \; \forall \; s \neq j$) contributes to the likelihood function by its probability density function, given by the product of the intensity function $\lambda^j(t_i|\mathcal{F}_{t_i})$ and the survivor function $\exp(-\Lambda^j(t_{i-1}, t_i))$. All other processes contribute only by their survivor function.

However, the main difficulty in this context is that the latent process is not observable, hence the conditional likelihood function must be integrated with respect to L_n^* using the assumed normal distributions $N(m_i^*, 1)$ for each $l_i := \ln \lambda_i$. Hence the unconditional or integrated likelihood function is given by

$$\mathcal{L}(W; \theta) = \int \prod_{i=1}^{n} \prod_{s=1}^{S} \lambda_i^{*\sigma_s^* y_i^s} \exp(-\lambda_i^{*\sigma_s^*} \Lambda^{o,s}(t_{i-1}, t_i)) \, [\lambda^{o,s}(t_i)]^{y_i^s} \tag{32}$$

$$\times \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(l_i^* - m_i^*)^2}{2}\right] dL_n^*.$$

$$= \int \prod_{i=1}^{n} f(w_i, l_i^* | W_{i-1}, L_{i-1}^*, \theta) dL_n^*, \tag{33}$$

where the last equality clearly defines the function $f(\cdot)$.

6.2 Numerical computation of the likelihood function

The computation of the n-dimensional integral in eq. (32) must be done numerically and represents a challenge. We use the efficient importance sampling (EIS) method of Richard (1998) because this method has proven to be quite efficient for the computation of the likelihood function of stochastic volatility models (see also Liesenfeld and Richard 2002), to which the LFI model resembles in this respect.

In order to implement the EIS algorithm, the integral (33) is rewritten as

$$\mathcal{L}(W; \theta) = \int \prod_{i=1}^{n} \frac{f(w_i, l_i^* | W_{i-1}, L_{i-1}^*, \theta)}{m(l_i^* | L_{i-1}^*, \phi_i)} \prod_{i=1}^{n} m(l_i^* | L_{i-1}^*, \phi_i) dL_n^*, \tag{34}$$

where $\{m(l_i^*|L_{i-1}^*, \phi_i)\}_{i=1}^n$ denotes a sequence of auxiliary importance samplers for $\{l_i^*\}_{i=1}^n$ indexed by the parameters $\{\phi_i\}_{i=1}^n$. EIS relies on a choice of the auxiliary parameters $\{\phi_i\}_{i=1}^n$ that provide a good match between $\{m(l_i^*|L_{i-1}^*, \phi_i)\}_{i=1}^n$ and $\{f(w_i, l_i^*|W_{i-1}, L_{i-1}^*, \theta)\}_{i=1}^n$, so that the integral (33) can be computed reliably by importance sampling, i.e.

$$\mathcal{L}(W; \theta) \approx \hat{\mathcal{L}}_R(W; \theta) = \frac{1}{R} \sum_{r=1}^{R} \prod_{i=1}^{n} \frac{f(w_i, l_i^{*(r)} | W_{i-1}, L_{i-1}^{*(r)}, \theta)}{m(l_i^{*(r)} | L_{i-1}^{*(r)}, \phi_i)}, \tag{35}$$

where $\{l_i^{*(r)}\}_{i=1}^n$ denotes a trajectory of random draws from the sequence of auxiliary importance samplers $\{m(l_i^*|L_{i-1}^*, \phi_i)\}_{i=1}^n$, and R such trajectories are generated. More precisely, the "good match" criterion is the minimization of the sampling variance of the Monte Carlo importance sampling estimator $\hat{\mathcal{L}}_R(W;\theta)$, by choice of the auxiliary parameters $\{\phi_i\}_{i=1}^n$. Before describing precisely how this is implemented, we explain how the importance samplers themselves are defined.

Let $k(L_i^*, \phi_i)$ denote a density kernel for $m(l_i^*|L_{i-1}^*, \phi_i)$, defined by

$$k(L_i^*, \phi_i) = m(l_i^*|L_{i-1}^*, \phi_i)\chi(L_{i-1}^*, \phi_i), \tag{36}$$

where

$$\chi(L_{i-1}^*, \phi_i) = \int k(L_i^*, \phi_i) dl_i^*. \tag{37}$$

The implementation of EIS requires the selection of a class of density kernels $k(\cdot)$ for the auxiliary sampler $m(\cdot)$ that provide a good approximation to the product $f(w_i, l_i^*|W_{i-1}, L_{i-1}^*, \theta)\ \chi(L_{i-1}^*, \phi_i)$. A seemingly obvious choice of auxiliary samplers is the sequence of log-normal densities $N(m_i^*, 1)$ that is a component of the sequence of functions $f(w_i, l_i^*|W_{i-1}, L_{i-1}^*, \theta)$. However, this choice is not good enough because these direct samplers do not take the observed data into account.

As argued convincingly by Richard (1998), a convenient and efficient possibility is to use a parametric extension of the direct samplers, normal distributions in our context. We can approximate the function

$$\prod_{s=1}^{S} \lambda_i^{*\sigma_s^* y_i^s} \exp(-\lambda_i^{*\sigma_s^*}\ \Lambda^{o,s}(t_{i-1}, t_i)), \tag{38}$$

that appears in eq. (32) by the normal density kernel

$$\zeta(l_i^*) = \exp(\phi_{1,i} l_i^* + \phi_{2,i} l_i^{*2}). \tag{39}$$

We can also include the $N(m_i^*, 1)$ density function in the importance sampler $m(l_i^*|L_{i-1}^*, \phi_i)$. Using the property that the product of normal densities is itself a normal density, we can write a density kernel of $m(\cdot)$ as

$$k(L_i^*, \phi_i) \propto \exp\left((\phi_{1,i} + m_i^*)\, l_i^* + \left(\phi_{2,i} - \frac{1}{2}\right) l_i^{*2} \right) \tag{40}$$

$$= \exp\left(-\frac{1}{2\pi_i^2} (l_i^* - \mu_i)^2 \right) \exp\left(\frac{\mu_i^2}{2\pi_i^2} \right), \tag{41}$$

where

$$\pi_i^2 = (1 - 2\phi_{2,i})^{-1} \tag{42}$$

$$\mu_i = (\phi_{1,i} + m_i^*)\, \pi_i^2. \tag{43}$$

Hence, the integrating constant (37) is given by

$$\chi(L_{i-1}^*, \phi_i) = \exp\left(\frac{\mu_i^2}{2\pi_i^2} - \frac{m_i^{*2}}{2} \right) \tag{44}$$

(neglecting the factor $\pi_i\sqrt{2\pi}$ since it depends neither on L^*_{i-1} nor on ϕ_i).

As mentioned above, the choice of the auxiliary parameters has to be optimized in order to minimize the MC variance of $\hat{\mathcal{L}}_R(W;\theta)$. This can be split into n minimization problems of the form (see Richard, 1998, for a detailed explanation):

$$\min_{\phi_{i,0},\,\phi_i} \sum_{r=1}^{R} \left\{ \ln\left[f(w_i, l_i^{*(r)}|W_{i-1}, L_{i-1}^{*(r)}, \theta)\chi(L_i^{*(r)}, \phi_{i+1})\right] - \phi_{0,i} - \ln k(L_i^{*(r)}, \phi_i)\right\}^2, \tag{45}$$

where $\phi_{0,i}$ are auxiliary proportionality constants. These problems must be solved sequentially starting at $i = n$, under the initial condition $\chi(L_n^*, \phi_{n+1}) = 1$ and ending at $i = 1$. In eq. (45), $\{l_i^{*(r)}\}_{i=1}^n$ denotes a trajectory of random draws from the sequence of direct samplers $\{N(m_i^*, 1)\}_{i=1}^n$. Given that $k(L_i^*, \phi_i)$ is a normal density kernel (see (41)), the minimization problem (45) is solved by computing an ordinary least squares estimate of $\phi_{i,0}$ and $\phi_i = (\phi_{1,i}, \phi_{2,i})$.

To summarize, the EIS algorithm requires the following steps:

Step 1: Generate R trajectories $\{l_i^{*(r)}\}_{i=1}^n$ using the direct samplers $\{N(m_i^*, 1)\}_{i=1}^n$.

Step 2: For each i (from n to 1), estimate by OLS the regression (with R observations) implicit in (45), which takes precisely the following form:

$$\begin{aligned}\sigma_s^* y_i^s \ln\lambda_i^{*(r)} - \left(\lambda_i^{*(r)}\right)^{\sigma_s^*} \sum_{s=1}^{S} \Lambda^{o,s}(t_{i-1}, t_i) + \sum_{s=1}^{S} y_i^s \ln\lambda^{o,s}(t_i) + \ln\left(\chi(L_i^{*(r)}, \phi_{i+1})\right) \\ = \phi_{0,i} + \phi_{1,i}\ln\lambda_i^{*(r)} + \phi_{2,i}(\ln\lambda_i^{*(r)})^2 + u^{(r)},\ r = 1, \ldots, R,\end{aligned} \tag{46}$$

(where $u^{(r)}$ is an error term), using $\chi(L_n^{*(r)}, \phi_{n+1}) = 1$ as initial condition, and then (44).

Step 3: Generate R trajectories $\{l_i^{*(r)}\}_{i=1}^n$ using the EIS samplers $\{N(\mu_i, \pi_i^2)\}_{i=1}^n$ (see eq. (42) and eq. (43)) to compute $\hat{\mathcal{L}}_R(W;\theta)$ as defined in eq. (35). Per assumption the computation of the terms $\Lambda^{o,s}(t_{i-1}, t_i)$ and $\lambda^{o,s}(t_i|\mathcal{F}^o_{t_i})$ can be done separately and is done in a step (before the EIS algorithm).

As recommended by Liesenfeld and Richard (2003), steps 1 to 3 should be iterated about five times to improve the efficiency of the approximations. This is done by replacing the direct sampler in step 1 by the importance samplers built in the previous iteration. It is also possible to start step 1 (of the first iteration) with another sampler than the direct one. This is achieved by immediately multiplying the direct sampler by a normal approximation to $\sigma_s^*\ln\lambda_i^* - \lambda_i^{*\sigma_s^*}\sum_{s=1}^S \Lambda^{o,s}(t_{i-1}, t_i)$, using a second order Taylor expansion (TSE) of the argument of the exponential function around $l_i^* = 0$. This yields

$$\sigma_s^* y_i^s \ln\lambda_i^* - \lambda_i^{*\sigma_s^*} \sum_{s=1}^{S} \Lambda^{o,s}(t_{i-1}, t_i) \approx \text{constant} + \ln\lambda_i^* - (\ln\lambda_i^*)^2 \sum_{s=1}^{S} \Lambda^{o,s}(t_{i-1}, t_i), \tag{47}$$

which implies that $\phi_{1,i} = 1$ and $\phi_{2,i} = -\sum_{s=1}^S \Lambda^{o,s}(t_{i-1}, t_i)$ must be inserted into eq. (42) and eq. (43) to obtain the moments of the TSE normal importance sampler. In this way, the initial importance sampler takes into account the data. This enables one to reduce to three (instead of five) the number of iterations over the three steps.

6.3 Testing the Latent Factor Intensity Model

In order to provide diagnostics with respect to the latent process, it is necessary to produce sequences of filtered estimates of functions of the latent variable λ_i^*, which take the form of the ratio of integrals in eq. (27).[12] Liesenfeld and Richard (2003) propose evaluating the integrals in the denominator and numerator by MC integration using the EIS algorithm where θ is set equal to its ML estimate. The integral in the denominator corresponds to $\mathcal{L}(W_{i-1};\theta)$, the marginal likelihood function of the first $i-1$ observations, and is evaluated on the basis of the sequence of auxiliary samplers $\{m(l_j^*|L_{j-1}^*, \hat{\phi}_j^{i-1})\}_{j=1}^{i-1}$ where $\{\hat{\phi}_j^{i-1}\}$ denotes the value of the EIS auxiliary parameters associated with the computation of $\mathcal{L}(W_{i-1};\theta)$. Furthermore, the numerator of eq. (27) is approximated by

$$\frac{1}{R}\sum_{r=1}^{R}\left\{\vartheta\left(l_i^{*(r)}(\theta)\right)\prod_{j=1}^{i-1}\left[\frac{f\left(w_j, l_j^{*(r)}(\hat{\phi}_j^{i-1})|W_{j-1}, L_{j-1}^{*(r)}(\hat{\phi}_{j-1}^{i-1}), \theta\right)}{m\left(l_j^{*(r)}(\hat{\phi}_j^{i-1})|L_{j-1}^{*(r)}(\hat{\phi}_{j-1}^{i-1}), \hat{\phi}_j^{i-1}\right)}\right]\right\}, \tag{48}$$

where $\{l_j^{*(r)}(\hat{\phi}_j^{i-1})\}_{j=1}^{i-1}$ denotes a trajectory drawn from the sequence of importance samplers associated with $\mathcal{L}(W_{i-1};\theta)$, and $l_i^{*(r)}(\theta)$ is a random draw from the conditional density $p(l_i^*|W_{i-1},\ L_{i-1}^{*(r)}(\hat{\phi}_{i-1}^{i-1}), \theta)$. The computation of the sequence of filtered estimates requires rerunning the complete EIS algorithm for every function $\vartheta(l_i^*)$ and for every i (=1 to n).

The sequence of filtered estimates can be computed to evaluate the series of the latent variable, to compute conditional moments (see Section 5) and forecasts.

The LFI residuals of the s-th process are computed on the basis of the trajectories drawn from the sequence of auxiliary samplers associated with $\mathcal{L}(W;\theta)$, i.e.,

$$\begin{aligned}\hat{\epsilon}_i^{s,(r)} &:= \int_{t_{i-1}^s}^{t_i^s}\left[\lambda_i^{*(r)}(\hat{\phi}_i^n)\right]^{\sigma_s^*}\hat{\lambda}^{o,s}(u)du \\ &= \sum_{j=N(t_{i-1}^s)}^{N(t_i^s)}\left[\lambda_i^{*(r)}(\hat{\phi}_i^n)\right]^{\sigma_s^*}\int_{t_j}^{t_{j+1}}\hat{\lambda}^{o,s}(u)du.\end{aligned} \tag{49}$$

The residual diagnostics are computed for each of the R sequences separately.[13] Under correct specification, the residuals $\hat{\epsilon}_i^{s,(r)}$ should be i.i.d. $Exp(1)$. Hence, model evaluations can be done by testing the dynamical and distributional properties of the residual series using e.g. Ljung-Box statistics, and by testing the distribution properties, e.g. against overdispersion. Engle and Russell (1998) propose a test against excess dispersion based on the asymptotically normal test statistic $\sqrt{n^s/8}\,\hat{\sigma}^2_{e^{s,(r)}}$, where n^s denotes the number of points associated with the s series and $\hat{\sigma}^2_{e^{s,(r)}}$ the empirical variance of the residual series.

[12] Although the integrals in that ratio are written using the latent variables λ_i^*, we write them in this section using $l_i := \ln\lambda_i^*$.

[13] Note that it is not reasonable to evaluate the model based on the average trajectory since dispersion and dynamic effects would eliminate each other then and would lead to misleading results.

7 Applications of LFI Models

The LFI model is applied to financial duration series from German XETRA trading of the Allianz and BASF stock. The sample period covers two trading weeks from 11/15/99 to 11/26/99, corresponding to $6,035$ observations for the Allianz stock and $5,710$ observations for the BASF stock. Deterministic time-of-the-day effects are modelled using a linear spline function with nodes for every hour. Each original series is divided by its average trade duration mean.[14] Note that this is only a matter of scaling and does not change the order of the processes. Overnight spells, as well as all trades before 9:00 and after 17:00 are removed. Descriptive statistics for the original series are given in Table 1. They show that the series are overdispersed and strongly autocorrelated.

We estimate univariate and bivariate (LF-)ACI and (LF-)Hawkes models. The ACI and Hawkes models are estimated by standard ML, while the estimation of the LFI models requires using the EIS technique discussed in Section 6. In the ML-EIS procedure, we use $R = 50$ Monte Carlo replications, while the efficiency steps in the algorithm are repeated 5 times. In the first iteration, we start with the TSE normal importance sampler. It turns out that the ML-EIS procedure performs very well and leads to a proper convergence of the estimation algorithm. Note that in the LF-ACI model, the computation of the log likelihood requires the numerical integration of the intensity components $\lambda^{o,s}(t)$ since the backward recurrence function is specified according to (15). In this case we use the Gauss-Legendre quadrature algorithm implemented in GAUSS. The standard errors are computed based on the inverse of the estimated Hessian. The lag order P and Q is chosen on the basis of the Bayesian (Schwarz) information criterion BIC.

Table 2 shows the estimation results of univariate ACI(1,1) and LF-ACI(1,1) models for both stocks. Here, we restrict our consideration to the basic LFI specification without a time-varying latent parameter. This case will be considered in a bivariate framework below. The estimates of $\hat{\alpha}$ and $\hat{\beta}$ are typical for trade duration series. For $\hat{\alpha}$, we find a relatively low value, while $\hat{\beta}$ is very close to one, which indicates a strong persistence in the data. Moreover, we find values for $\hat{p}$ lower than one indicating a monotonous decline of the conditional intensity in the absence of a new event. Such a behavior is associated with a decreasing hazard function which is a typical result for stock market trade durations. For the seasonality variables, we find insignificant estimates in nearly all cases. However, it turns out that the coefficients are jointly significant. The resulting seasonality function (see Figure 11, Appendix) indicates a high trading intensity after the opening of the market, followed by the well known lunch time dip around noon. After 14:00 GMT we observe a strong increase of the trading intensity which is probably caused by the opening of the American markets.

We observe highly significant coefficients (a^* and σ^*) associated with the latent process, thus, evidence for the existence of a latent autoregressive process is provided. Furthermore, we notice a clear increase of the likelihood and the BIC values. Hence, the inclusion of a latent factor improves the goodness-of-fit of the model. It turns out that the autoregressive parameter of the observation

[14] In the bivariate models, we use the average duration with respect to the pooled process.

driven component $\lambda^o(t)$ moves towards more persistence while the impact of the (observable) innovations decreases when the latent factor is taken into account. Thus, the latent innovations displace the observable ones and drive the observable component towards a more constant process. The residuals are computed based on eq. (49) for the LF-ACI model and based on $\hat{\epsilon}_i = \hat{\Lambda}(t_{i-1}, t_i)$ for the ACI model. Note that in the LF-ACI framework, the latter is obviously not anymore i.i.d. exponentially distributed. Nevertheless, we also report statistics for these 'residuals' in order to illustrate the properties of the particular components of the model. The LF-ACI residuals are computed based on each single trajectory of random draws of $\lambda_i^{*(r)}$, $r = 1, \ldots, R$. In the tables, we state the *average* values of the particular statistics over all trajectories and the corresponding p-values.[15] We find that the inclusion of the latent factor leads to a clear reduction of the Ljung-Box-statistics. In particular, the null hypothesis of no serial correlation in the residual series cannot be rejected for the LF-ACI model. Thus, it turns out that the latent factor model captures the duration dynamics in a better way. Based on the test against excess dispersion, the ACI specifications are clearly rejected at the 5 per cent level. In contrast, the p-values of the test statistics exceed 5 per cent for the LF-ACI models, indicating an improvement of the specification.

Table 3 presents the estimation results of Hawkes(2) and LF-Hawkes(2) processes. The use of a lag order of $P = 2$ is absolutely necessary to obtain a satisfactory specification. In particular, Hawkes(1) specifications (which are not reported) are not able to capture the dynamics in the data, leading to distinctly higher Ljung-Box statistics and lower BIC values. With respect to the latent factor, we only find mixed evidence. For the Allianz series, highly significant estimates of a^* and σ^* are observed coming along with a value of $\hat{a}^*$ close to one. For this series, the latent factor seems to capture remaining persistence in the data. However, based on the BIC, the inclusion of latent dynamics does not increase the goodness-of-fit of the model. This result is confirmed by the individual residual diagnostics that indicate no improvements of the LF-Hawkes specifications compared to the pure Hawkes model. For the BASF series, the result is even more clear-cut. Here, the latent variance σ^* tends towards zero, thus, the existence of a latent factor is clearly rejected. For this reason, the autoregressive parameter a^* is not identified, and thus estimable anymore. These results provide evidence against the need for a latent factor in a Hawkes model for the analyzed data.

Table 4 contains the estimation results of bivariate Weibull (LF-)ACI models. We estimate one common seasonality function for both series. Columns (1) and (2) in Table 4 show the results based on an ACI(1,1) and a basic LF-ACI(1,1) specification. As a starting point, we assume identical latent standard deviations for both processes, i.e. $\sigma_1^* = \sigma_2^*$. We find clear interdependencies between the two processes, thus, the backward recurrence time of one process influences not only its own intensity but also the intensity of the other process. Accordingly, the innovation components reveal significant interdependencies as well. Moreover, as indicated by the estimates of B, we find again a high persistence in the processes. Interestingly, the latent autoregressive process is more dominant than in the univariate case. The latent parameters are highly significant and lead to a strong increase of the BIC. This result indicates that the latent factor captures unobserved information that *jointly* influences both transaction

[15] For the excess dispersion test, we take the average value of the test statistic.

processes. As in the univariate models, the introduction of a latent factor increases the persistence of the observable components, while the magnitude of the ACI innovations is reduced. The diagnostic checks show that the LF-ACI model captures the dynamics in a better way than the pure ACI model leading to an important reduction of the Ljung-Box statistics. However, in contrast to the univariate case, the introduction of the latent factor does not improve the distributional properties of the model. In particular, it leads to still significant excess-dispersion in the residuals of both stocks.

In specification (3), we estimate an extended LF-ACI model with process-specific latent variances and a regime switching latent autoregressive parameter. The parameter a^* is specified according to eq. (11) based on $R = 2$ regimes with an exogenously given threshold value $\bar{x} = 1$. In fact, we find a significantly higher estimate for the autoregressive parameter associated with small (previous) durations (a_1^*). Hence, the serial dependence of the latent factor obviously decreases over the length of the previous spell, and thus, the importance of latent information declines the longer it dates back. The estimates of σ_1^* and σ_2^* are not significantly different, hence, it turns out that the latent factor influences both components in a similar way. We find a clear increase of the BIC and marginal improvements in the residual diagnostics indicating a better overall goodness-of-fit of this specification. Specification (4) is based on a LF-ACI model where the latent parameters depend on the type of the most current point (eq. (12)). However, as indicated by the estimates, this higher flexibility is not supported by the data. In particular, it turns out that the individual regime dependent parameter are nearly identical in both cases. Finally, column (5) gives the estimation result for an LF-ACI specification without any dynamics in the observable component, thus $\lambda^o(t) = 1$. In this case, the joint latent component captures the dynamics of both processes, which is associated with an increase of a_1^*, while the latent variances are nearly unaffected. However, it turns out that, according to the BIC, specification (5) outperforms the pure ACI model (specification (1)) which illustrates the usefulness of modelling the system dynamics by a *joint* latent component. However, we observe that the dynamic specification of the latent factor is not sufficient to completely capture the dynamics of the model. In terms of the residual diagnostics, specification (5) underperforms a fully parameterized LF-ACI specification. This result is not surprising since a simple AR(1) dynamic in the latent factor is probably not appropriate for modelling the persistence in the process.[16]

Table 5 displays the estimation results of bivariate (LF-)Hawkes(2) processes. As in the univariate case, the choice of a lag order $P = 2$ is absolutely necessary to obtain a satisfactory specification of the model. However, according to the BIC, we find a restricted version[17] of a (LF-)Hawkes(2) process as the best and most parsimonious specification. Column (1) and (2) show the results based on the basic Hawkes(2) model, as well as the basic LF-Hawkes(2) specification. As for the LF-ACI estimates, specification (2) is based on the restriction $\sigma_1^* = \sigma_2^*$. The dynamic parameters α^{sr} and β^{sr} are highly significant in most cases, yielding evidence for clear interdependencies between both processes. For the

[16] We also estimated a LF-ACI model based on heteroscedastic latent variances that are specified according to eq. (11) based on one hour nodes. It turns out that the seasonality effects are not individually or collectively significant, thus, we find no evidence for seasonality effects driving the variance of the latent processes.

[17] The second order cross-terms are set to zero, i.e. $\alpha_2^{12} = \alpha_2^{21} = \beta_2^{12} = \beta_2^{21} = 0$.

latent factor, we obtain highly significant estimates with a value of $\hat{a}^*$ close to one and a quite low standard deviation. Hence, as in the univariate case, the latent factor seems to capture remaining persistence in the data. Nonetheless, the latent factor's influence is clearly weaker than in the LF-ACI model. This is confirmed by the reported BIC value and the residual diagnostics that do not indicate improvements of the goodness-of-fit of the LF-Hawkes model compared to the pure Hawkes model. In the columns (3) and (4), we present generalized LF-Hawkes(2) specifications based on process-specific latent variances and a regime-switching latent dynamic parameter. Again, we observe clear empirical evidence for a significantly lower serial dependence in the latent component when the previous spell was long. The fact that the highest BIC value and thus, the best goodness-of-fit is found for specification (3), indicates that this higher flexibility in the LF-Hawkes specification is supported by the data. This finding is confirmed by the residual diagnostics that report slight improvements of the dynamical and distributional properties at least for the BASF series. Nevertheless, as in the LF-ACI model, the use of of process-specific latent variances leads to no improvement of the model specification. In particular, the influences of the latent factor on the individual intensity components do not differ significantly.

Summarizing, our results provide evidence for the existence of an unobservable autoregressive component that is common to both duration series. In the ACI model, the introduction of a latent factor leads to a better model performance, as witnessed by a clear increase of the maximized likelihood and the BIC value, as well as improved residuals diagnostics. In contrast, in the Hawkes model, the inclusion of a latent factor appears to be less important. In this framework, only slight improvements of the model's goodness-of-fit can be observed. Obviously, Hawkes processes provide an overall better goodness-of-fit compared to the ACI model since no strong evidence for omitted (dynamic) factors can be found with our data.

8 Conclusions

In this paper, we introduce a new type of intensity model for dynamic point processes. The main idea is to specify the intensity function based on two components, an observation driven autoregressive process which is updated by observable innovations, and a dynamic latent factor which is updated by unobservable innovations. Hence, the so-called latent factor intensity (LFI) model combines features of observation driven models and parameter driven models and can be interpreted as a dynamic extension of a doubly stochastic Poisson process. The latent factor is specified as a log-linear model based on a Gaussian AR(1) process. In this sense, the model resembles a stochastic volatility model and enriches an intensity function with a latent factor process. The observation driven component can be specified univariately and multivariately. In the latter case, the latent factor serves as an unobservable component that jointly influences the individual intensity component and drives the joint dynamics of the multivariate system. In the spirit of the mixture-of-distribution hypotheses, such a factor can be economically interpreted as a variable that captures the unobservable information flow that drives the overall trading activity on the market. We discuss different ways to specify the observation driven component based on autoregressive conditional intensity (ACI) parameterizations (Russell, 1999) and Hawkes type models (Hawkes, 1971). Since conditional and unconditional moments typically cannot be computed analytically, we provide simulation results to provide insights into the model dynamics, as well as of those of the resulting duration processes.

Since the latent dynamic process is not observable, the model cannot be estimated by standard maximum likelihood techniques. We suggest the efficient importance sampling (EIS) method proposed by Richard (1998), which has proven to be highly efficient for very accurate Monte Carlo evaluations of the likelihood depending upon high-dimensional interpedendent integrals. Based on this technique, it is possible to compute simulated residuals which are the basis of diagnostic tests.

We apply the proposed model to intertrade duration series of the Allianz and BASF stocks traded in the German XETRA system. Based on estimations of univariate and bivariate LF-ACI and LF-Hawkes models, we provide evidence for the existence of an unobservable autoregressive component which is common to both duration series. It turns out that the introduction of a parameter driven component in the ACI model leads to a better model performance, witnessed by a clear increase of the maximized likelihood and the BIC criterion, while diagnostic checks indicate a better specification. However, in a Hawkes framework, only slight improvements of the model can be observed. Nevertheless, our results illustrate that the LFI model is a useful specification to model financial point processes in a very flexible way and to get deeper insights into the joint market dynamics.

References

Aalen, O. (1978): "Nonparametric Inference for a Family of Counting Processes," *The Annals of Statistics*, 6(4), 701–726.

Bauwens, L., and P. Giot (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.

——— (2003): "Asymetric ACD Models: Introducing Price Information in ACD Models with a Two State Transition Model," *Empirical Economics*, 28 (4), 1–23.

Bauwens, L., and D. Veredas (1999): "The Stochastic Conditional Duration Model: A Latent Factor Model for the Analysis of Financial Durations," Discussion Paper 9958, CORE, Université Catholique de Louvain, forthcoming *Journal of Econometrics*, 2003.

Bowsher, C. G. (2002): "Modelling Security Markets in Continuous Time: Intensity based, Multivariate Point Process Models," Discussion Paper 2002-W22, Nuffield College, Oxford.

Brémaud, P. (1981): *Point Processes and Queues, Martingale Dynamics.* Springer, New York.

Clark, P. (1973): "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica*, 41, 135–155.

Cox, D. (1981): "Statistical Analysis of Time Series: Some Recent Developments," *Scandinavian Journal of Statistics*, 8, 93–115.

Cox, D. R., and V. Isham (1980): *Point Processes*. Chapman and Hall, London.

Cox, D. R., and E. J. Snell (1968): "A General Definition of Residuals," *Journal of the Royal Statistical Society*, B30, 248–265.

Davis, R. A., T. H. Rydberg, N. Shephard, and S. B. Streett (2001): "The CBin Model for Counts: Testing for Common Features in the Speed of Trading, Quote Changes, Limit and Market Order Arrivals," Discussion paper.

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

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

Engle, R. F., and A. Lunde (2003): "Trades and Quotes: A Bivariate Point Process," *Journal of Financial Econometrics*, 1(2), 159–188.

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

Fernandes, M., and J. Grammig (2001): "A Family of Autoregressive Conditional Duration Models," Discussion Paper 2001/31, CORE, Université Catholique de Louvain.

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

Grandell, J. (1976): *Doubly Stochastic Poisson Processes*. Springer, Berlin.

Hall, A. D., and N. Hautsch (2003): "Estimating the Intensity of Buy and Sell Arrivals in a Limit Order Book Market," Discussion paper, University of Technology, Sydney.

Hautsch, N. (2002): "Testing the Conditional Mean Function of Autoregressive Conditional Duration Models," Discussion paper, University of Konstanz.

Hawkes, A. G. (1971): "Spectra of Some Self-Exciting and Mutually Exciting Point Processes," *Biometrika*, 58, 83–90.

Hawkes, A. G., and D. Oakes (1974): "A Cluster Process Representation of a Self-Exciting Process," *Journal of Applied Probability*, 11, 493–503.

Heinen, A., and E. Rengifo (2003): "Multivariate Autoregressive Modelling of Time Series Count Data Using Copulas," Discussion Paper 2003/25, CORE, Université Catholique de Louvain.

Karr, A. F. (1991): *Point Processes and their Statistical Inference*. Dekker, New York.

KAWASAKI, Y. (2002): "Modeling Periodicity in High Frequent Financial Data," Discussion paper, Institute of Statistical Mathematics, Tokyo.

LIESENFELD, R., AND J.-F. RICHARD (2003): "Univariate and Multivariate Stochastic Volatility Models: Estimation and Diagnostics," *Journal of Empirical Finance*, 10, 505–531.

OGATA, Y., AND H. AKAIKE (1982): "On Linear Intensity Models for Mixed Doubly Stochastic Poisson and Self-exciting Point Processes," *Journal of the Royal Statistical Society, Series B*, 44, 102–107.

RICHARD, J.-F. (1998): "Efficient High-dimensional Monte Carlo Importance Sampling," Discussion paper, University of Pittsburgh.

RUSSELL, J. R. (1999): "Econometric Modeling of Multivariate Irregularly-Spaced High-Frequency Data," Discussion paper, University of Chicago.

SNYDER, D. L., AND M. I. MILLER (1991): *Random Point Processes and Time and Space*. Springer, New York, 2 edn.

TAYLOR, S. J. (1982): "Financial Returns Modelled by the Product of Two Stochastic Processes - a Study of Daily Sugar Prices," in *Time Series Analysis: Theory and Practice*, ed. by O. D. Anderson. North-Holland, Amsterdam.

TONG, H. (1990): *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford.

VERE-JONES, D. (1970): "Stochastic Models for Earthquake Occurrence," *Journal of the Royal Statistical Society, Series B*, 32, 1–62.

ZHANG, M. Y., J. RUSSELL, AND R. S. TSAY (2001): "A Nonlinear Autoregressive Conditional Duration Model with Applications to Financial Transaction Data," *Journal of Econometrics*, 104, 179–207.

Appendix

Figures

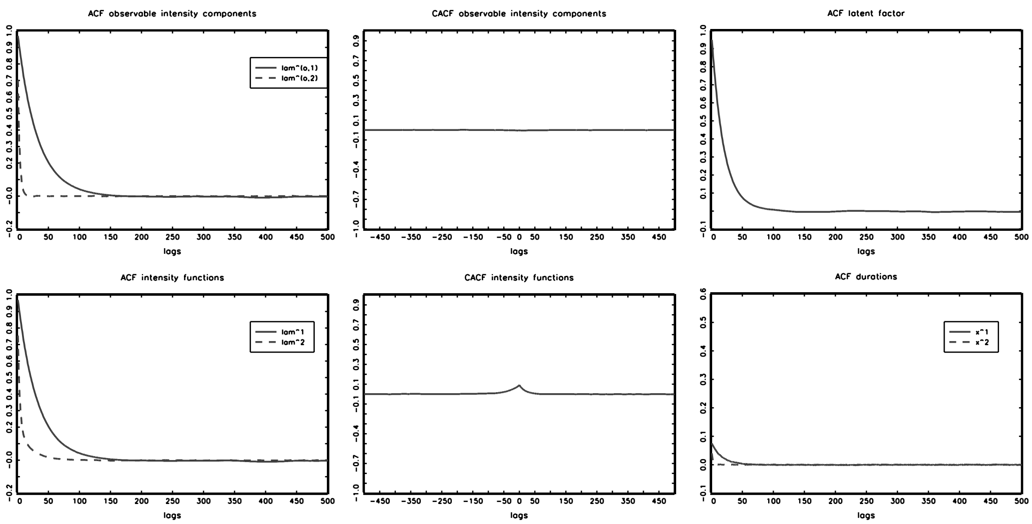

Figure 1: Bivariate LF-ACI(1,1) processes. Upper left: ACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper middle: CACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper right: ACF of λ_i^*. Lower left: ACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower middle: CACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower right: ACF of x_i^1 and x_i^2. Simulations based on $5,000,000$ drawings. $\omega^1 = \omega^2 = 0$, $\alpha_1^1 = \alpha_2^2 = 0.05$, $\alpha_2^1 = \alpha_1^2 = 0$, $\beta_{11} = 0.97$, $\beta_{12} = \beta_{21} = 0$, $\beta_{22} = 0.7$, $a^* = 0.95$, $\sigma_1^* = \sigma_2^* = 0.01$.

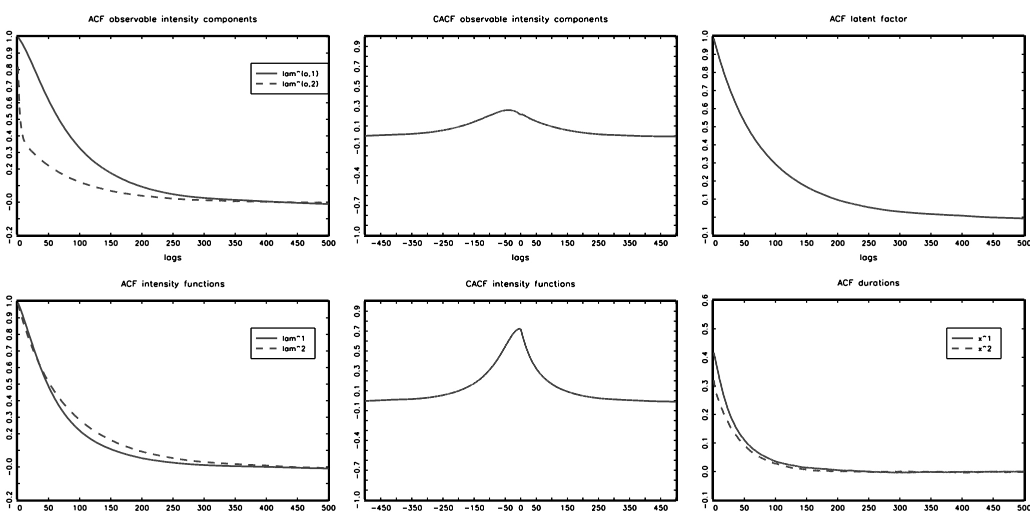

Figure 2: Bivariate LF-ACI(1,1) processes. Upper left: ACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper middle: CACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper right: ACF of λ_i^*. Lower left: ACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower middle: CACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower right: ACF of x_i^1 and x_i^2. Simulations based on $5,000,000$ drawings. $\omega^1 = \omega^2 = 0$, $\alpha_1^1 = \alpha_2^2 = 0.05$, $\alpha_2^1 = \alpha_1^2 = 0$, $\beta_{11} = 0.97$, $\beta_{12} = \beta_{21} = 0$, $\beta_{22} = 0.7$, $a^* = 0.99$, $\sigma_1^* = \sigma_2^* = 0.1$.

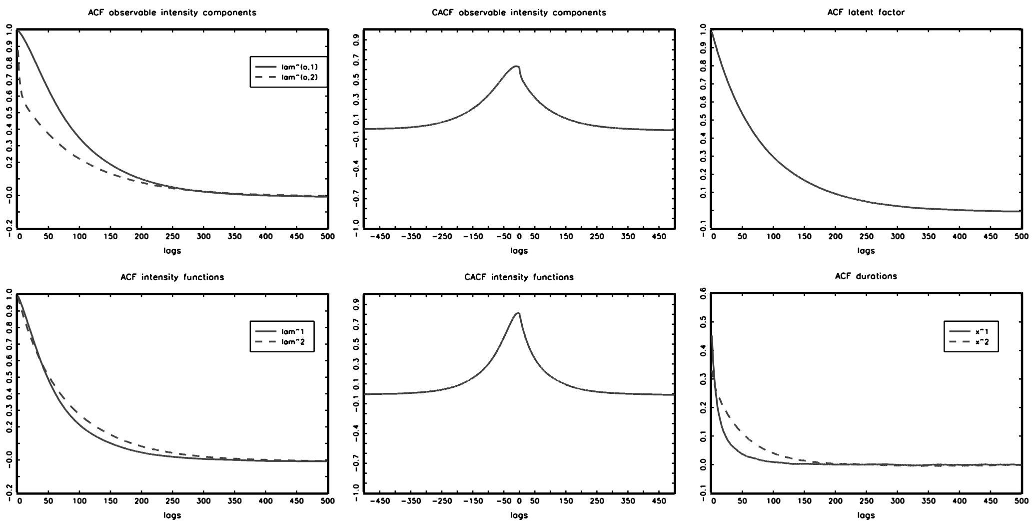

Figure 3: Bivariate LF-ACI(1,1) processes. Upper left: ACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper middle: CACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper right: ACF of λ_i^*. Lower left: ACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower middle: CACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower right: ACF of x_i^1 and x_i^2. Simulations based on $5,000,000$ drawings. $\omega^1 = \omega^2 = 0$, $\alpha_1^1 = \alpha_2^1 = \alpha_1^2 = \alpha_2^2 = 0.05$, $\beta_{11} = 0.97$, $\beta_{12} = \beta_{21} = 0$, $\beta_{22} = 0.7$, $a^* = 0.99$, $\sigma_1^* = \sigma_2^* = 0.1$.

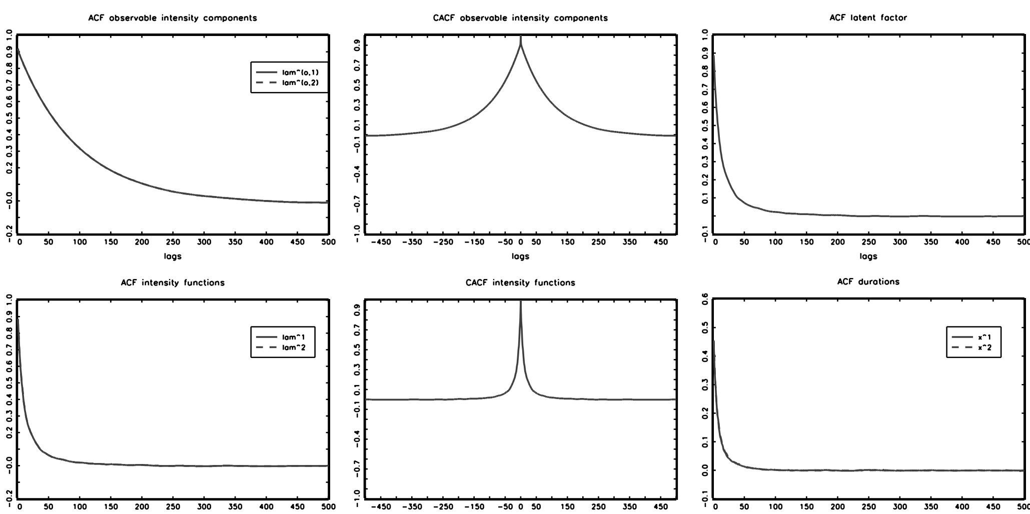

Figure 4: Bivariate LF-ACI(1,1) processes. Upper left: ACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper middle: CACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper right: ACF of λ_i^*. Lower left: ACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower middle: CACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower right: ACF of x_i^1 and x_i^2. Simulations based on $5,000,000$ drawings. $\omega^1 = \omega^2 = 0$, $\alpha_1^1 = \alpha_2^1 = \alpha_1^2 = \alpha_2^2 = 0.05$, $\beta_{11} = \beta_{22} = 0.2$, $\beta_{12} = \beta_{21} = 0$, $a^* = 0.95$, $\sigma_1^* = \sigma_2^* = 0.5$.

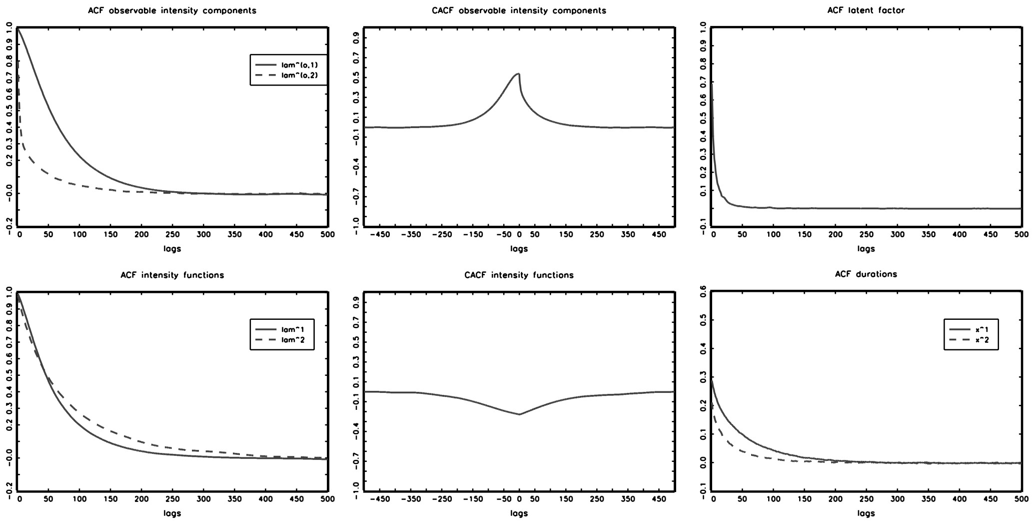

Figure 5: Bivariate LF-ACI(1,1) processes. Upper left: ACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper middle: CACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper right: ACF of $\lambda^*(t_i)$. Lower left: ACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower middle: CACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower right: ACF of x_i^1 and x_i^2. Simulations based on $5,000,000$ drawings. $\omega^1 = \omega^2 = 0$, $\alpha_1^1 = \alpha_1^2 = \alpha_2^1 = \alpha_2^2 = 0.05$, $\beta_{11} = 0.97$, $\beta_{12} = \beta_{21} = 0$, $\beta_{22} = 0.7$, $a^* = 0.99$, $\sigma_1^* = 0.1$, $\sigma_2^* = -0.1$.

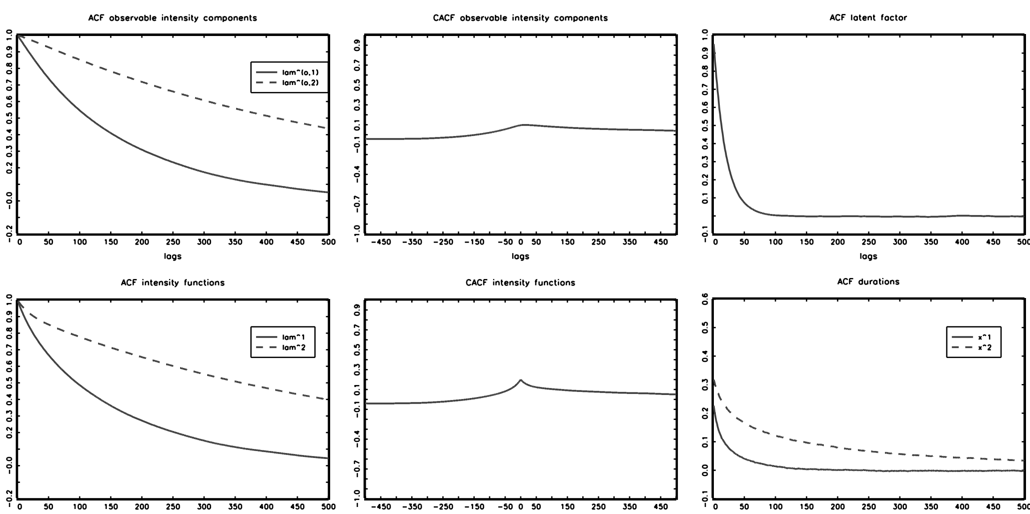

Figure 6: Bivariate LF-Hawkes(1) processes. Upper left: ACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper middle: CACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper right: ACF of λ_i^*. Lower left: ACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower middle: CACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower right: ACF of x_i^1 and x_i^2. Simulations based on $5,000,000$ drawings. $\omega^1 = \ln(0.4)$, $\omega^2 = \ln(0.3)$, $\alpha^{11} = \alpha^{22} = 0.2$, $\beta^{11} = 0.25$, $\beta^{22} = 0.21$, $\alpha^{12} = \alpha^{21} = 0$, $a^* = 0.95$, $\sigma_1^* = \sigma_2^* = 0.05$.

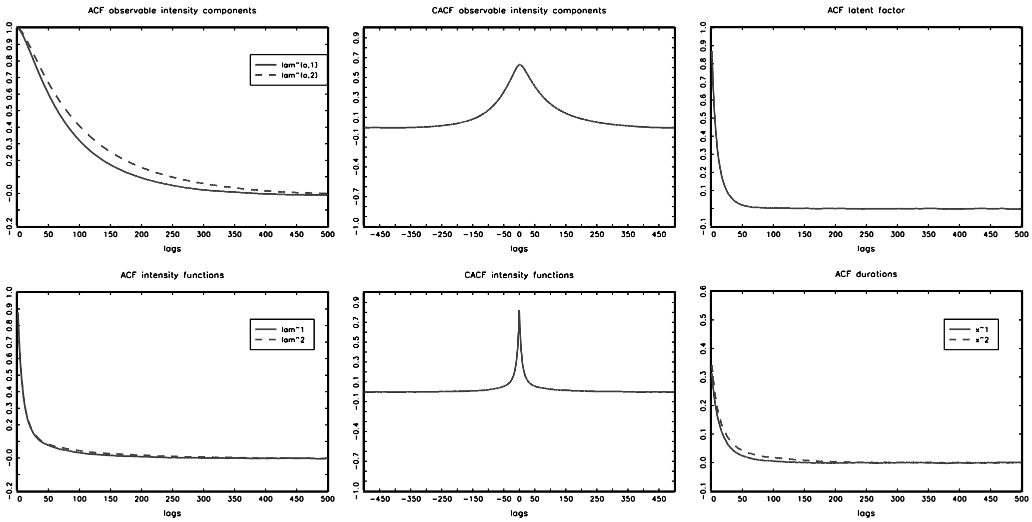

Figure 7: Bivariate LF-Hawkes(1) processes. Upper left: ACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper middle: CACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper right: ACF of λ_i^*. Lower left: ACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower middle: CACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower right: ACF of x_i^1 and x_i^2. Simulations based on $5,000,000$ drawings. $\omega^1 = \ln(0.4)$, $\omega^2 = \ln(0.3)$, $\alpha^{11} = \alpha^{22} = 0.2$, $\beta^{11} = 0.25$, $\beta^{22} = 0.21$, $\alpha^{12} = \alpha^{21} = 0$, $a^* = 0.95$, $\sigma_1^* = \sigma_2^* = 0.5$.

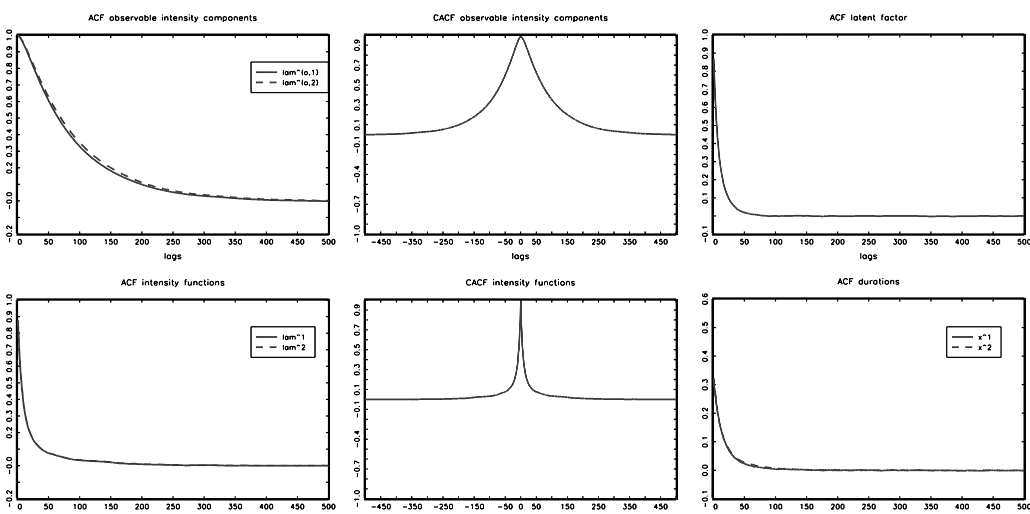

Figure 8: Bivariate LF-Hawkes(1) processes. Upper left: ACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper middle: CACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper right: ACF of λ_i^*. Lower left: ACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower middle: CACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower right: ACF of x_i^1 and x_i^2. Simulations based on $5,000,000$ drawings. $\omega^1 = \ln(0.2)$, $\omega^2 = \ln(0.1)$, $\alpha^{11} = \alpha^{22} = 0.2$, $\beta^{11} = 0.4$, $\beta^{22} = 0.25$, $\alpha^{12} = \alpha^{21} = 0.1$, $\beta^{12} = 0.3$, $\beta^{21} = 0.6$, $a^* = 0.95$, $\sigma_1^* = \sigma_2^* = 0.5$.

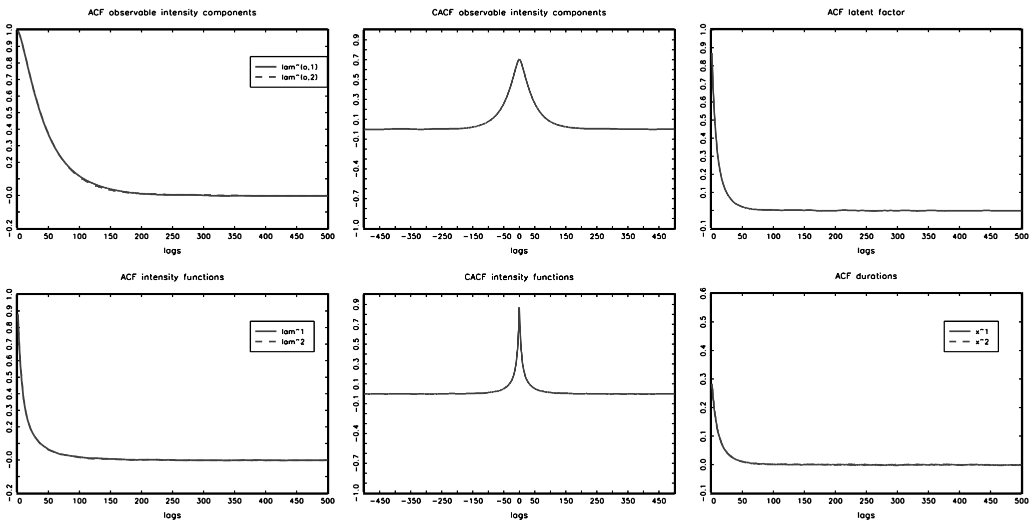

Figure 9: Bivariate LF-Hawkes(1) processes. Upper left: ACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper middle: CACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper right: ACF of λ_i^*. Lower left: ACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower middle: CACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower right: ACF of x_i^1 and x_i^2. Simulations based on $5,000,000$ drawings. $\omega^1 = \omega^2 = \ln(0.6)$, $\alpha^{11} = \alpha^{22} = 0.2$, $\beta^{11} = \beta^{22} = 0.5$, $\alpha^{12} = \alpha^{21} = 0$, $a^* = 0.95$, $\sigma_1^* = \sigma_2^* = 0.5$.

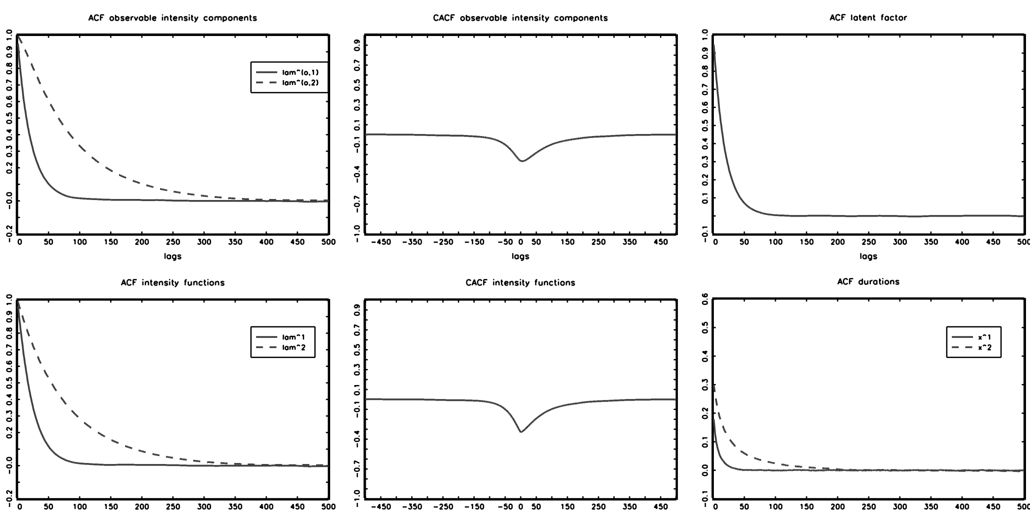

Figure 10: Bivariate LF-Hawkes(1) processes. Upper left: ACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper middle: CACF of $\lambda^{o,1}(t_i)$ and $\lambda^{o,2}(t_i)$. Upper right: ACF of λ_i^*. Lower left: ACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower middle: CACF of $\lambda^1(t_i|\mathcal{F}_{t_i})$ and $\lambda^2(t_i|\mathcal{F}_{t_i})$. Lower right: ACF of x_i^1 and x_i^2. Simulations based on $5,000,000$ drawings. $\omega^1 = \ln(0.2)$, $\omega^2 = \ln(0.1)$, $\alpha^{11} = \alpha^{22} = 0.2$, $\beta^{11} = 0.4$, $\beta^{22} = 0.25$, $\alpha^{12} = \alpha^{21} = 0$, $a^* = 0.95$, $\sigma_1^* = 0.1$, $\sigma_2^* = -0.1$.

Empirical results

Table 1: Descriptive statistics and Ljung-Box statistics of intertrade durations for the Allianz and BASF stocks traded in the German XETRA system. Sample period from 11/15/99 to 11/26/99.

	Allianz	BASF
Number of observations	6035	5710
Mean	50.90	53.78
Standard deviation	68.18	69.11
Ljung-Box of order 20	405.4	644.0

Descriptive statistics in seconds.

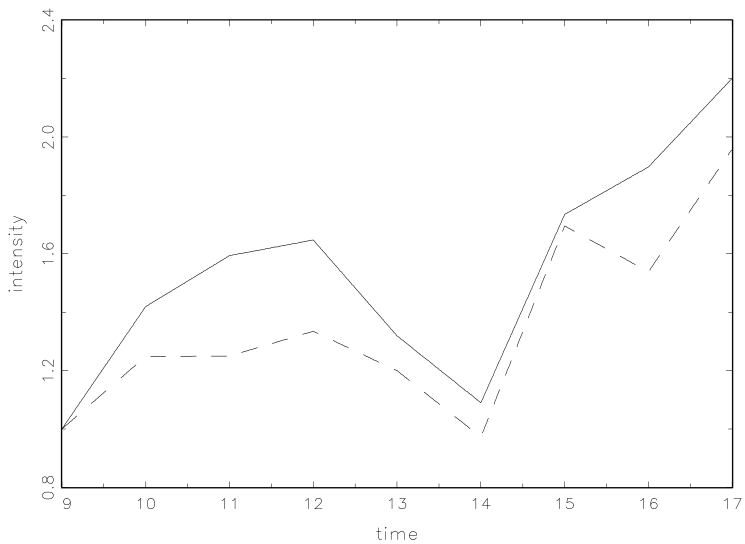

Figure 11: Seasonality function of trading intensities for the Allianz (solid line) and the BASF stock traded in the German XETRA system. Sample period from 11/15/99 to 11/26/99. Estimation based on LF-ACI model.

Table 2: ML(-EIS) estimates of univariate Weibull ACI(1,1) and LF-ACI(1,1) models for Allianz and BASF trade durations. Data based on German XETRA trading, sample period from 11/15/99 to 11/26/99. Standard errors based on the inverse of the estimated Hessian.

	Allianz				BASF			
	ACI(1,1)		LF-ACI(1,1)		ACI(1,1)		LF-ACI(1,1)	
	est.	*S.E.*	*est.*	*S.E.*	*est.*	*S.E.*	*est.*	*S.E.*
ω	-0.54	0.13	-0.21	0.18	-0.36	0.12	-0.13	0.15
p	0.81	0.01	0.99	0.03	0.85	0.01	0.99	0.02
α	0.05	0.01	0.03	0.01	0.05	0.01	0.04	0.01
β	0.96	0.01	0.98	0.01	0.98	0.01	0.98	0.01
s_1	1.31	0.76	1.17	1.03	0.97	0.64	0.69	0.71
s_2	-0.72	1.19	-0.68	1.60	-1.00	1.03	-0.69	1.06
s_3	-0.83	1.13	-0.34	1.39	0.17	0.97	0.23	0.92
s_4	-0.39	1.07	-1.06	1.34	-0.44	0.96	-0.61	1.07
s_5	0.53	1.02	0.28	1.17	-0.55	0.92	-0.26	1.06
s_6	1.40	1.09	2.43	1.37	2.88	0.98	2.64	1.13
s_7	-0.99	1.21	-1.34	1.53	-2.32	1.11	-2.44	1.35
s_8	0.39	1.09	0.39	1.46	1.26	1.02	1.60	1.26
a^*			0.39	0.06			0.39	0.08
σ^*			0.63	0.06			0.53	0.06
Obs	6035		6035		5710		5710	
LL	-5539		-5484		-5291		-5254	
BIC	-5591		-5545		-5343		-5315	
	Diagnostics of ACI and LF-ACI residuals							
Mean of $\hat{\Lambda}^o(t_{i-1},t_i)$	1.00		1.24		1.00		1.17	
S.D. of $\hat{\Lambda}^o(t_{i-1},t_i)$	1.05		1.61		1.04		1.42	
LB(20) of $\hat{\Lambda}^o_1(t_{i-1},t_i)$	34.40	0.02	41.01	0.00	35.05	0.02	42.76	0.00
Mean of $\hat{\epsilon}_i^{s,(r)}$			1.01				1.01	
S.D. of $\hat{\epsilon}_i^{s,(r)}$			1.03				1.02	
LB(20) of $\hat{\epsilon}_i^{s,(r)}$			18.63	0.54			20.99	0.40
Exc. disp.	2.55	0.01	1.88	0.06	2.12	0.04	1.18	0.24

Diagnostics: Log Likelihood (LL), Bayes Information Criterion (BIC) and diagnostics (mean, standard deviation and Ljung-Box statistic (inclusive p-values)) of ACI residuals $\hat{\Lambda}(t_{i-1},t_i)$, average diagnostics (mean, standard deviation, Ljung-Box statistic, as well as excess dispersion test (the latter two statistics inclusive p-values)) over all trajectories of LFI residuals $\hat{\epsilon}_i^{s,(r)}$.

Estimated specifications:
ACI model specified according to eq. (13), (15), (16) and (17) with $S=1$, $p=q=1$ and $\lambda_i^*=1$. Innovation term specified according to eq. (18) and (20) with $\lambda_i^*=1$.
LF-ACI models specified according to eq. (7), (13), (15), (16) and (17) with $S=1$ and $p=q=1$. Innovation term specified according to eq. (18) and (20). Latent factor specified according to eq. (9).
Seasonality functions computed as

$$s(t) = 1 + \sum_{k=1}^{K} s_k 1_{\{\tau(t_i)\leq\bar{\tau}_k\}}(\tau(t_i)-\bar{\tau}_k),$$

where $\tau(t)$ denotes the calendar time at t, $\bar{\tau}_k$, $k=1,\ldots,K-1$ denote the exogenously given (calendar) time nodes and s_k the corresponding coefficients of the spline function. The nodes are chosen as 10:00, 11:00,. . ., 17:00.

Table 3: ML(-EIS) estimates of univariate Hawkes(2) and LF-Hawkes(2) models for Allianz and BASF trade durations. Data based on German XETRA trading, sample period from 11/15/99 to 11/26/99. Standard errors based on the inverse of the estimated Hessian.

	Allianz				BASF			
	Hawkes		LF-Hawkes		Hawkes		LF-Hawkes	
	est.	*S.E.*	*est.*	*S.E.*	*est.*	*S.E.*	*est.*	*S.E.*
ω	0.38	0.04	0.48	0.06	0.40	0.04	0.40	0.04
α_1	0.02	0.00	0.01	0.00	0.63	0.08	0.03	0.01
β_1	0.03	0.01	0.01	0.00	2.33	0.31	0.06	0.02
α_2	0.93	0.12	0.94	0.15	0.03	0.01	0.62	0.08
β_2	2.67	0.25	2.76	0.25	0.06	0.02	2.33	0.31
s_1	-0.68	0.28	-0.37	0.48	-0.58	0.27	-0.57	0.28
s_2	0.80	0.39	0.22	0.80	0.63	0.40	0.61	0.47
s_3	-0.19	0.33	0.03	0.71	-0.04	0.34	-0.03	0.43
s_4	-0.09	0.31	-0.21	0.50	-0.14	0.33	-0.15	0.36
s_5	0.07	0.31	0.14	0.38	0.09	0.34	0.09	0.35
s_6	0.50	0.33	0.71	0.41	0.53	0.36	0.53	0.37
s_7	-0.47	0.31	-0.53	0.41	-0.69	0.34	-0.69	0.35
s_8	0.15	0.26	0.13	0.34	0.42	0.28	0.42	0.28
a^*			0.96	0.02			0.13	6.17
σ^*			0.05	0.02			0.00	0.03
Obs	6035		6035		5710		5710	
LL	-5398		-5393		-5196		-5196	
BIC	-5454		-5458		-5252		-5261	
Diagnostics of Hawkes and LF-Hawkes residuals								
Mean of $\hat{\Lambda}^o(t_{i-1}, t_i)$	1.00		1.02		1.00		1.00	
S.D. of $\hat{\Lambda}^o(t_{i-1}, t_i)$	1.01		1.05		0.99		0.99	
LB(20) of $\hat{\Lambda}^o_1(t_{i-1}, t_i)$	17.87	0.60	53.65	0.00	21.92	0.34	21.89	0.35
Mean of $\hat{\epsilon}_i^{s,(r)}$			1.00				1.00	
S.D. of $\hat{\epsilon}_i^{s,(r)}$			1.01				0.99	
LB(20) of $\hat{\epsilon}^{s(r)}$			18.26	0.55			21.89	0.35
Exc. disp.	0.32	0.75	0.71	0.48	0.36	0.72	0.36	0.72

Diagnostics: Log Likelihood (LL), Bayes Information Criterion (BIC) and diagnostics (mean, standard deviation and Ljung-Box statistic (inclusive p-values)) of ACI residuals $\hat{\Lambda}(t_{i-1}, t_i)$, average diagnostics (mean, standard deviation, Ljung-Box statistic, as well as excess dispersion test (the latter two statistics inclusive p-values)) over all trajectories of LFI residuals $\hat{\epsilon}_i^{s,(r)}$.

Estimated specifications:
Hawkes model specified according to eq. (7) and (25) with $S = 1$, $P = 2$ and $\lambda_i^* = 1$.
LF-Hawkes models specified according to eq. (7) and (25) with $S = 1$ and $P = 2$. Latent factor specified according to eq. (9).
Seasonality functions computed as

$$s(t) = 1 + \sum_{k=1}^{K} s_k 1_{\{\tau(t_i) \le \bar{\tau}_k\}} (\tau(t_i) - \bar{\tau}_k),$$

where $\tau(t)$ denotes the calendar time at t, $\bar{\tau}_k$, $k = 1, \ldots, K-1$ denote the exogenously given (calendar) time nodes and s_k the corresponding coefficients of the spline function. The nodes are chosen as 10:00, 11:00,…, 17:00.

Table 4: ML(-EIS) estimates of bivariate Weibull (LF-)ACI(1,1) models for Allianz and BASF trade durations. Data based on German XETRA trading, sample period from 11/15/99 to 11/26/99. Standard errors based on the inverse of the estimated Hessian.

	(1)		(2)		(3)		(4)		(5)	
	ACI(1,1)		LF-ACI(1,1)		LF-ACI(1,1)		LF-ACI(1,1)		LF-ACI(1,1)	
	est.	*S.E.*	*est.*	*S.E.*	*est.*	*S.E.*	*est.*	*S.E.*	*est.*	*S.E.*
					Allianz					
ω^1	−0.90	0.10	−0.82	0.18	−1.01	0.13	−0.84	0.17	−0.82	0.129
p_1^1	0.83	0.01	0.98	0.01	1.00	0.02	1.00	0.02	0.99	0.018
p_2^1	0.88	0.01	1.02	0.01	1.04	0.02	1.03	0.02	1.06	0.016
α_1^1	0.08	0.02	0.02	0.01	0.02	0.00	0.02	0.01		
α_2^1	0.06	0.01	0.01	0.01	0.01	0.00	0.02	0.01		
β_{11}	0.91	0.04	0.99	0.00	0.99	0.00	0.99	0.00		
β_{12}	−0.06	0.03	−0.01	0.00	−0.01	0.00	−0.01	0.00		
					BASF					
ω^2	−0.93	0.10	−0.80	0.18	−0.96	0.13	−0.78	0.17	−0.87	0.13
p_1^2	0.92	0.01	1.07	0.02	1.09	0.02	1.07	0.02	1.11	0.02
p_2^2	0.86	0.01	1.01	0.02	1.03	0.02	1.00	0.02	1.03	0.02
α_1^2	0.04	0.01	0.00	0.01	0.00	0.00	0.01	0.00		
α_2^2	0.07	0.01	0.03	0.01	0.03	0.00	3.01	0.01		
β_{21}	−0.03	0.01	−0.00	0.00	−0.00	0.00	−0.00	0.00		
β_{22}	0.97	0.01	0.99	0.00	0.99	0.00	0.10	0.00		
					Seasonalities					
s_1	0.83	0.49	1.13	0.99	0.43	0.53	1.28	1.10	0.36	0.57
s_2	−0.89	0.75	−1.10	1.45	−0.28	0.63	−1.30	1.10	−0.33	0.86
s_3	0.13	0.60	0.21	1.09	0.02	0.10	0.30	1.59	0.27	0.58
s_4	−0.31	0.59	−0.69	1.36	−0.68	0.49	−0.74	1.10	−0.86	0.55
s_5	−0.30	0.59	−0.34	1.38	−0.11	0.55	−0.43	1.57	0.05	0.91
s_6	1.76	0.62	2.68	1.37	2.36	0.73	2.86	1.61	2.07	1.06
s_7	−1.36	0.63	−2.28	1.37	−2.08	0.93	−2.38	1.44	−1.65	0.85
s_8	0.76	0.54	1.53	1.15	1.39	0.87	1.61	1.20	1.12	0.70
					Latent Factor					
a_1^*			0.58	0.03	0.74	0.02	0.55	0.04	0.79	0.02
a_2^*					0.25	0.06	0.62	0.05	0.42	0.03
σ_1^*			0.57	0.03	0.59	0.04			0.56	0.04
σ_2^*					0.57	0.04			0.58	0.04
$\sigma_{1,1}^*$							0.61	0.05		
$\sigma_{1,2}^*$							0.61	0.05		
$\sigma_{2,1}^*$							0.58	0.05		
$\sigma_{2,2}^*$							0.50	0.05		
Obs	11745		11745		11745		11745		11745	
LL	-18838		-18709		-18666		-18707		-18800	
BIC	-18941		-18821		-18788		-18838		-18884	
					Diagnostics of ACI and LF-ACI residuals for the Allianz series					
Mean $\hat{\Lambda}^o(\cdot)$	1.00		1.23		1.06		1.25		1.10	
S.D. $\hat{\Lambda}^o(\cdot)$	1.05		1.57		1.38		1.62		1.45	
LB $\hat{\Lambda}_1^o \cdot ()$	31.23	0.05	51.65	0.00	60.52	0.00	49.13	0.00	292.88	0.00
Mean $\hat{\epsilon}_i^{s,(r)}$			1.01		1.01		1.01		1.02	
S.D. $\hat{\epsilon}_i^{s,(r)}$			1.05		1.05		1.05		1.05	
LB $\hat{\epsilon}_i^{s,(r)}$			26.32	0.16	24.35	0.23	25.49	0.18	87.75	0.00
Exc. disp.	2.90	0.00	2.67	0.01	2.80	0.01	2.66	0.01	2.94	0.00
					Diagnostics of ACI and LF-ACI residuals for the BASF series					
Mean $\hat{\Lambda}^o(\cdot)$	1.00		1.27		1.10		1.27		1.13	
S.D. $\hat{\Lambda}^o(\cdot)$	1.04		1.61		1.41		1.59		1.50	
LB $\hat{\Lambda}_1^o(\cdot)$	38.59	0.01	51.60	0.00	57.96	0.00	50.91	0.00	481.19	0.00
Mean $\hat{\epsilon}_i^{s,(r)}$			1.01		1.01		1.01		1.02	
S.D. $\hat{\epsilon}_i^{s,(r)}$			1.05		1.05		1.04		1.06	
LB $\hat{\epsilon}_i^{s,(r)}$			26.64	0.15	24.67	0.21	27.25	0.13	134.19	0.00
Exc. disp.	2.35	0.02	2.68	0.01	2.61	0.01	2.39	0.02	3.12	0.00

Diagnostics: Log Likelihood (LL), Bayes Information Criterion (BIC) and diagnostics (mean, standard deviation and Ljung-Box statistic (inclusive p-values)) of ACI residuals $\hat{\Lambda}(t_{i-1}, t_i)$, average diagnostics (mean, standard deviation, Ljung-Box statistic, as well as excess dispersion test (the latter two statistics inclusive p-values)) over all trajectories of LFI residuals $\hat{\epsilon}_i^{s,(r)}$.

Estimated specifications:

ACI model (column (1)) specified according to eq. (13), (15), (16) and (17) with $S = 2$ and $p = q = 1$. Innovation term specified according to eq. (18) and (20). $\lambda_i^* = 1$.

LF-ACI models (columns (2) through (5)) specified according to eq. (7), (13), (15), (16) and (17) with $S = 2$ and $p = q = 1$. Innovation term specified according to eq. (18) and (20). In column (5), $\Psi_i^1 = \Psi_i^2$ is set to 1.

Specification of the latent factor: In column (2) specified according to eq. (9), where $\sigma_1^* = \sigma_2^*$. In column (3) specified according to eq. (11), where $R = 1$ and $\bar{x}_1 = 1$. In column (4) specified according to eq. (12). In column (5) specified as in column (3).

Seasonality functions computed as $s^1(t) = s^2(t) = 1 + \sum_{k=1}^{K} s_k 1_{\{\tau(t_i) \leq \bar{\tau}_k\}}(\tau(t_i) - \bar{\tau}_k)$, where $\tau(t)$ denotes the calendar time at t, $\bar{\tau}_k$, $k = 1, \ldots, K-1$ denote the exogenously given (calendar) time nodes and s_k the corresponding coefficients of the spline function. The nodes are chosen as 10:00, 11:00,..., 17:00.

Table 5: ML(-EIS) estimates of bivariate (LF-)Hawkes(2) models for Allianz and BASF trade durations. Data based on German XETRA trading, sample period from 11/15/99 to 11/26/99. Standard errors based on the inverse of the estimated Hessian.

	(1)		(2)		(3)		(4)	
	LF-Hawkes		LF-Hawkes		LF-Hawkes		LF-Hawkes	
	est.	*S.E.*	*est.*	*S.E.*	*est.*	*S.E.*	*est.*	*S.E.*
				Allianz				
ω^1	0.14	0.02	0.16	0.02	0.16	0.02	0.15	0.02
α_1^{11}	0.44	0.04	0.46	0.05	0.44	0.05	0.45	0.05
β_1^{11}	1.44	0.14	1.50	0.15	1.55	0.15	1.52	0.15
α_1^{12}	0.29	0.04	0.30	0.04	0.29	0.04	0.29	0.04
β_1^{12}	1.77	0.26	1.92	0.30	1.20	0.31	1.99	0.31
α_2^{11}	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
β_2^{11}	0.02	0.00	0.01	0.00	0.02	0.00	0.02	0.00
				BASF				
ω^2	0.16	0.02	0.18	0.02	0.18	0.02	0.18	0.02
α_1^{21}	0.02	0.00	0.01	0.00	0.01	0.00	0.01	0.00
β_1^{21}	0.03	0.01	0.02	0.01	0.02	0.01	0.02	0.01
α_1^{22}	0.18	0.03	0.18	0.03	0.17	0.03	0.17	0.03
β_1^{22}	1.70	0.29	1.86	0.33	1.90	0.34	1.91	0.35
α_2^{22}	0.30	0.04	0.31	0.04	0.30	0.04	0.29	0.04
β_2^{22}	1.17	0.17	1.19	0.17	1.26	0.18	1.32	0.20
				Seasonalities				
s_1	−0.65	0.18	−0.73	0.21	−0.70	0.19	−0.71	0.19
s_2	0.74	0.28	0.77	0.32	0.76	0.28	0.78	0.28
s_3	−0.12	0.23	−0.06	0.28	−0.12	0.24	−0.14	0.23
s_4	−0.08	0.23	−0.11	0.26	−0.12	0.23	−0.09	0.23
s_5	0.07	0.23	0.06	0.26	0.18	0.24	0.16	0.24
s_6	0.42	0.24	0.45	0.27	0.40	0.25	0.40	0.25
s_7	−0.47	0.23	−0.51	0.26	−0.53	0.24	−0.53	0.24
s_8	0.21	0.18	0.22	0.21	0.24	0.20	0.23	0.19
				Latent Factor				
a_1^*			0.93	0.05	0.99	0.02	0.98	0.02
a_2^*					0.51	0.21	0.45	0.24
σ_1^*			0.05	0.03	0.06	0.02	0.05	0.02
σ_2^*							0.07	0.03
Obs	11745		11745		11745		11745	
LL	-18578		-18576		-18564		-18564	
BIC	-18682		-18688		-18681		-18685	
			Diagnostics of Hawkes and LF-Hawkes residuals for the Allianz series					
Mean $\hat{\Lambda}^o(\cdot)$	1.00		1.01		0.98		0.99	
S.D. $\hat{\Lambda}^o(\cdot)$	1.01		1.03		1.01		1.01	
LB $\hat{\Lambda}_1^o(\cdot)$	20.58	0.42	23.59	0.26	23.09	0.28	22.25	0.33
Mean $\hat{\epsilon}_i^{s,(r)}$			1.00		1.00		1.00	
S.D. $\hat{\epsilon}_i^{s,(r)}$			1.01		1.01		1.01	
LB $\hat{\epsilon}_i^{s,(r)}$			20.11	0.45	20.47	0.43	20.59	0.42
Exc. disp.	0.49	0.63	0.47	0.64	0.52	0.60	0.49	0.63
			Diagnostics of Hawkes and LF-Hawkes residuals for the BASF series					
Mean $\hat{\Lambda}^o(\cdot)$	1.00		1.01		0.98		0.98	
S.D. $\hat{\Lambda}^o(\cdot)$	0.99		1.01		0.99		1.00	
LB $\hat{\Lambda}_1^o(\cdot)$	24.64	0.22	28.13	0.11	26.93	0.14	28.00	0.11
Mean $\hat{\epsilon}_i^{s,(r)}$			1.00		1.00		1.00	
S.D. $\hat{\epsilon}_i^{s,(r)}$			0.99		0.99		0.99	
LB $\hat{\epsilon}_i^{s,(r)}$			25.18	0.19	21.77	0.35	21.11	0.39
Exc. disp.	0.44	0.66	0.44	0.66	0.39	0.70	0.35	0.72

Diagnostics: Log Likelihood (LL), Bayes Information Criterion (BIC) and diagnostics (mean, standard deviation and Ljung-Box statistic (inclusive p-values)) of ACI residuals $\hat{\Lambda}(t_{i-1}, t_i)$, average diagnostics (mean, standard deviation, Ljung-Box statistic, as well as excess dispersion test (the latter two statistics inclusive p-values)) over all trajectories of LFI residuals $\hat{\epsilon}_i^{s,(r)}$.

Estimated specifications:

Hawkes model (column (1)) specified according to eq. (7) and (25) with $S = 2$ and $P = 2$, where $\alpha_2^{12} = \alpha_2^{21} = \beta_2^{12} = \beta_2^{21} = 0$ and $\lambda_i^* = 1$.

LF-Hawkes models (columns (2) through (4)) specified according to eq. (7) and (25) with $S = 2$ and $P = 2$, where $\alpha_2^{12} = \alpha_2^{21} = \beta_2^{12} = \beta_2^{21} = 0$.

Specification of the latent factor: In column (2) specified according to eq. (9) with $\sigma_1^* = \sigma_2^*$. In column (3) specified according to eq. (11), where $R = 1$, $\bar{x}_1 = 1$ and $\sigma_1^* = \sigma_2^*$. In column (4) specified according to eq. (11).

Seasonality functions computed as $s^1(t) = s^2(t) = 1 + \sum_{k=1}^{K} s_k 1_{\{\tau(t_i) \leq \bar{\tau}_k\}}(\tau(t_i) - \bar{\tau}_k)$, where $\tau(t)$ denotes the calendar time at t, $\bar{\tau}_k$, $k = 1, \ldots, K-1$ denote the exogenously given (calendar) time nodes and s_k the corresponding coefficients of the spline function. The nodes are chosen as 10:00, 11:00,..., 17:00.