# Modelling High-Frequency Data in Continuous Time

Nour Meddahi,* Eric Renault† and Bas J.M. Werker‡

February 1998

Keywords: Continuous time models, High-frequency data, Micro-structure noise.

**Abstract**

Nowadays, many models in finance are specified in continuous time. It seems natural to use all observations available in estimating these models. In this way one arrives quickly at using high-frequency data. However, an important component of high-frequency data are several microstructure effects. In estimating the underlying structural model it is important to recognize these effects. It is the goal of the current paper to present a methodology in this direction, i.e. to identify a continuous time structural model from high-frequency observations contaminated by microstructure effects. Previous econometric work in this area contains the work of Engle and Russell (1995), Engle (1996), and Ghysels and Jasiak (1996).

The idea of our paper is as follows. We start with the modelisation of an underlying structural model in continuous time. Then, we explicitely take into account the microstructure effects. There are three important aspects. First is the uncertainty in the transaction times. Secondly, we allow for possible jumps in the underlying continuous time process for the spot price. Finally, we allow for additional microstructure noise that implies that observed returns are not necessarily equal to those in the structural model. In this way we get a discrete time model implied by the jump-diffusion observed at random frequencies. We are able to calculate explicitely conditional moments in some relevant cases. These moments are then used to construct estimators for the parameters of interest in the model. We plan to implement the methods presented using high-frequency data.

*Université de Montréal, dépt de sciences économiques, CRDE, CIRANO, CEPR, CP 6128, succursale centre-ville, Montréal (QC) H3C 3J7, E-mail: nour.meddahi@umontreal.ca

†CREST-ENSAI. E-mail: erenault@ensai.fr

‡Université Libre de Bruxelles, Institut de Statistique, CP210, B-1050, Bruxelles, Belgique. E-mail: BWerker@ulb.ac.be

# 1 Introduction

R. Engle introduced his seminal paper about ARCH modelling (Engle (1982)) via the one-period forecast of a random variable $y_t$: "the variance of this one-period forecast is given by $\mathrm{Var}(y_t|y_{t-1})$. Such an expression recognizes that the conditional forecast variance depends upon past information and may therefore be a random variable". But, while Engle (1982) was concerned with forecasting the rate of inflation considered as a univariate Markov process (of order 1), this simple ARCH model, although widely used nowadays to capture the temporal dynamics of asset return volatility, is not actually sufficiently rich to make optimal forecasts using all the multivariate information in transaction data. Moreover, if the econometrician is not able to characterize efficient use of such public information by the economic agents, then he will not be able to identify the asymmetric information models of the market micro-structure literature. For all these reasons, there is a need to define a more sophisticated statistical model that takes into account at least two types of additional information.

On the one hand, as noticed by Engle (1996) himself, the price of a financial contract is almost never the only relevant information which is available about this contract. Other potentially important information is "the volume of the contract (...), the posted bid and asked prices at the time (...), the counter-parties to the trade, the posted bid and asked prices for other stocks, the order mechanism, and many other features of a trade which are of interest in studying market microstructure". On the other hand, high-frequency financial time series which are transaction-based (and therefore irregularly spaced) are becoming more widely available and used in recent years. This provides the econometrician with a new type of data which contains the time of the transaction itself.

Making an adjustment of ARCH models to capture such tick-by-tick volatility dynamics would lead to a severe aggregation of available information about the environmental variables of the return process itself and the time stamps of the realized value of this return as well. There is naturally a loss of information in such aggregates and in turn a lack of efficiency in associated forecasts. The interest of this paper is to develop a general methodology which is well-suited to avoid such aggregation. In order to do this, we refer to two different strands of the recent litterature.

The first strand discusses temporal aggregation properties of GARCH models (Drost and Nijman (1993), Drost and Werker (1996), and Meddahi and Renault (1996)). In this paper we will follow Meddahi and Renault (1996) that proposes to reconcile the GARCH and stochastic volatility (SV) paradigms, by claiming that the traditional debate about observable past information (in the GARCH spirit) versus unobservable state-space conditioning information (in the SV spirit) should be avoided since the econometrician has no idea about something like a structural level of disaggregation. The main point of that paper is that a well-specified volatility model should be such that one is always allowed to reduce the information set without invalidating the model. They introduce a square-root autoregressive stochastic volatility (SR-SARV) model which remains true to the GARCH paradigm of ARMA dynamics for squared innovations, but weakens the GARCH structure (in a spirit close to the weak GARCH modelling in Drost and Nijman (1993)). This provides one with the required robustness properties with respect to various kinds of aggregation, including temporal aggregation. A by product of their modelling is that, since it can always be viewed as resulting from discrete time sampling in an underlying continuous time SV model (compare Drost and Werker (1996)), it provides a versatile tool to characterize stochastic volatility dynamics when data on asset prices are irregularly spaced. The second ingredient of our recipe is the recent literature about modelling of transaction arrival times, including Engle and Russell (1995), Engle (1996), Ghysels and Jasiak (1996), Bauwens and Giot (1997), and Ghysels, Gourieroux and Jasiak (1997). The central idea of this literature is an autoregressive conditional duration (ACD) modelling, which will be discussed in more detail in Section 3.

The contribution of the present paper is threefold. First, in Section 2, we extend Drost and Werker (1996) and Meddahi and Renault (1996) to define a stochastic volatility model for unequally spaced observations. In short, volatility over durations in a stochastic volatility setting is affected by two things. First, of course, longer durations generally imply higher total volatility. Secondly, the irregularly spaced datapoints induce (by definition) a temporal aggregation effect, as is modelled in Ghysels and Jasiak (1996) in a GARCH framework. Intuitively, the influence of previous volatility depends on how long ago previous volatility refers to, i.e. on the previous duration. We, using a stochastic volatility framework, identify both effects by starting from a continuous time model.

Our second contribution is to extend the stochastic volatility point of view to duration modelling. In the spirit of Meddahi and Renault (1996), we define (Section 3) a dynamic model of transaction dates which extends the ACD model in the same way that the SARV model extends traditional GARCH models. Related work is Ghysels, Gourieroux and Jasiak (1997). A by product of the general framework developped in Sections 2 and 3 is that it allows one to perform inference about a SV model specified in continuous time from high-frequency (irregularly spaced) data.

However, it is well-known that the use of tick-by-tick data to identify underlying continuous time structures (although probably optimal in the sense of statistical efficiency) poses some serious challenges because observations are contaminated (with respect to a continuous diffusion model which should reflect asset prices in equilibrium) by microstructure effects. The third contribution of this paper (Section 4) is to propose an integrated framework to estimate continuous time models from high-frequency data. There are several important aspects. First is the uncertainty in the transaction times. Secondly, we allow for possible jumps in the underlying continuous time process for the spot price. Finally, we allow for additional microstructure noise that implies that observed returns are not necessarily equal to those in the structural model. In this way we get a discrete time model implied by the jump-diffusion observed at random frequencies. We are able to calculate explicitely conditional moments in some relevant cases. By the way, a general methodology of building moments well-suited for inference in this type of data has been recently proposed in Duffie and Glynn (1997). These moments are then used to construct estimators for the parameters of interest in the model in traditional ways.

# 2 Stochastic volatility and unequally spaced observations

In this section we consider the unequally spaced transaction dates to be deterministically fixed. As mentioned in the introduction this is an unrealistic assumption but the results should be interpreted as a modellization of the transaction prices conditionally on the observed dates. Starting from a model specified in continuous time, we will derive (following the ideas put forward in Meddahi and Renault (1997) for the equally spaced case) first and second order moment conditions that easily allow for estimation. For convenience, we will restrict attention in this section to modellization of the processes within a single day. The pecularities that arise in the closure-to-open period of the markets will be discussed shortly in Section 4. In that section we will also consider additional noise on the observed prices that are a consequence of market microstructure effects. The present section ignores this as well, thus assuming that the observed high-frequency returns come from the specified price dynamics.

Let $t_1, t_2, \ldots, t_i, \ldots, t_n$ denote the time stamps of n subsequent transactions[1]. As mentioned above, these will be considered to be fixed, but the durations $x_i := t_i - t_{i-1}$ will not be assumed to be constant. In contrast to related literature we start from a continuous time specification of the underlying price process $S$. It is precisely the irregularity in the transaction dates that makes a continuous time model a natural choice. We will assume that the underlying prices $S$ can be described by

$$\mathrm{d}\log S_t = \mu \mathrm{d}t + \sigma_{t-}\mathrm{d}L_t, \tag{2.1}$$

$$\mathrm{d}\sigma_t^2 = \kappa(\theta - \sigma_{t-}^2)\mathrm{d}t + \mathrm{d}K_t, \tag{2.2}$$

where $\mu \in \mathbb{R}$, $\kappa > 0$, and $\theta > 0$ are constants, $L$ is a Lévy process, and $K$ is an arbitrary local martingale. We allow for possible jumps in both the price process as well as the volatility. The driving processes $L$ and $K$ will largely remain unspecified. In that sense we take a semiparametric point of view, and our interest lies in the mean return $\mu$ and the mean-reversion parameters of the volatility. We only make some regularity assumptions on $L$ and $K$, e.g., we will assume that $K$ is such that (2.2) allows for a strictly positive solution. Moreover, we will assume sufficient integrability properties for $L$ and $K$ such that the expectations to be derived below exist. We will

[1]The reader will easily be convinced that the main ideas of the paper also apply when modelling quotes or other irregular spaced financial data.

not make these assumptions explicit as they can be found in any standard textbook on stochastic integration (like, e.g., Protter (1995)) and are not very restrictive. We will denote by $\mathcal{F}_t := \sigma(L_u, K_u : \ u \leq t)$ the filtration generated by the driving processes $L$ and $K$.

We assume that the prices $S$ are observed at the dates $t_1, \ldots, t_n$. This means that we observe the returns

$$y_{t_i} := \log\left(S_{t_i}/S_{t_{i-1}}\right) = \mu x_i + \int_{t_{i-1}}^{t_i} \sigma_u \mathrm{d}L_u. \tag{2.3}$$

The volatility of these returns is described by (2.2). The important assumption that we make here, is that the volatility process exhibits linear mean reversion. As is well known, (2.2) implies

$$\mathrm{d}\exp(\kappa t)(\sigma_t^2 - \theta) = \exp(\kappa t)\mathrm{d}K_t, \tag{2.4}$$

and hence, under sufficient integrability conditions on $K$,

$$\mathrm{E}\left(\sigma_{t+h}^2 \middle| \mathcal{F}_t\right) = \theta + \exp(-\kappa h)\left(\sigma_t^2 - \theta\right). \tag{2.5}$$

As a consequence, the innovation in the returns $y_{t_i}$, i.e.

$$\varepsilon_{t_i} := \int_{t_{i-1}}^{t_i} \sigma_u \mathrm{d}L_u, \tag{2.6}$$

satisfy, using that the compensated quadratic variation of $L$ equals time,

$$\begin{aligned} \mathrm{Var}\left(\varepsilon_{t_i} \middle| \mathcal{F}_{t_{i-1}}\right) &= \int_{t_{i-1}}^{t_i} \mathrm{E}\left(\sigma_u^2 \middle| \mathcal{F}_{t_{i-1}}\right) \mathrm{d}u \\ &= \theta x_i + c(\kappa x_i)(\sigma_{t_{i-1}}^2 - \theta) x_i, \end{aligned} \tag{2.7}$$

where $c(v) := (1 - \exp(-v))/v$. Note that $c(v) \approx 1$ for small $v$, which will be very useful in interpreting the following results. For instance, it shows that the variance of the innovations is approximately equal to $\sigma_{t_{i-1}}^2 x_i$, which is intuitively appealing. This approximate linearity leads us, following Engle (1996), to study the discrete time behaviour of the *volatility per time-unit*, i.e.

$$f_{t_{i-1}} := \mathrm{Var}\left(\varepsilon_{t_i} \middle| \mathcal{F}_{t_{i-1}}\right)/x_i. \tag{2.8}$$

Note that, somewhat in contrast with the GARCH literature, the volatility per time-unit over the interval $(t_{i-1}, t_i]$ is denoted by $f_{t_{i-1}}$ and not by $f_{t_i}$. We will follow the convention that variables with index $t_{i-1}$ are always $\mathcal{F}_{t_{i-1}}$ measurable and usually not $\mathcal{F}_{t_i}$-measurable.

Following Meddahi and Renault (1996), observe that (2.4) implies

$$\exp(\kappa t_i)\left(\sigma_{t_i}^2 - \theta\right) = \exp(\kappa t_{i-1})\left(\sigma_{t_{i-1}}^2 - \theta\right) + \int_{t_{i-1}}^{t_i} \exp(\kappa u)\mathrm{d}K_u.$$

The linear relationship between $f_{t_{i-1}}$ and $\sigma^2_{t_{i-1}}$, i.e.

$$f_{t_{i-1}} = \theta + c(\kappa x_i)(\sigma^2_{t_{i-1}} - \theta),$$

implies that

$$\begin{aligned} f_{t_{i-1}} &= \theta + \exp(-\kappa x_{i-1})\frac{c(\kappa x_i)}{c(\kappa x_{i-1})}(f_{t_{i-2}} - \theta) + c(\kappa x_i)\int_{t_{i-2}}^{t_{i-1}} \exp(-\kappa(t_{i-1} - u))\mathrm{d}K_u \\ &= \omega_i + \gamma_i f_{t_{i-2}} + \nu_{t_{i-1}}, \end{aligned} \tag{2.9}$$

where

$$\omega_i = \theta\left(1 - \exp(-\kappa x_i)\frac{c(\kappa x_i)}{c(\kappa x_{i-1})}\right), \tag{2.10}$$

$$\gamma_i = \exp(-\kappa x_{i-1})\frac{c(\kappa x_i)}{c(\kappa x_{i-1})}, \text{ and} \tag{2.11}$$

$$\nu_{t_{i-1}} = c(\kappa x_i)\int_{t_{i-2}}^{t_{i-1}} \exp(-\kappa(t_{i-1} - u))\mathrm{d}K_u. \tag{2.12}$$

Under sufficient integrability conditions on $K$, we have that $\mathrm{E}\left(\nu_{t_i}\middle|\, \mathcal{F}_{t_{i-1}}\right) = 0$, so that $f_{t_i}$ follows an autoregressive structure. Hence, by definition, the innovations $\varepsilon_{t_i}$ follow a SARV (Stochastic AutoRegressive Volatility) process. This SARV structure can be used to derive moment restrictions, which in turn can be used to derive estimators for the volatility parameters. This will be discussed next.

The problem in deriving these moment conditions are the unobserved volatilities per time-unit $f_{t_i}$. The linear autoregressive structure for $f_{t_i}$, however, implies that also the conditional variance of $\varepsilon_{t_i}$ given $\mathcal{F}_{t_{i-2}}$ is linear in $f_{t_{i-2}}$. Hence, both $\varepsilon_{t_i}$ and $\varepsilon_{t_{i-1}}$ have volatilities conditional on $\mathcal{F}_{t_{i-2}}$ that are linear in $f_{t_{i-2}}$. Taking an appropriate linear combination of $\varepsilon_{t_i}$ and $\varepsilon_{t_{i-1}}$ then yields a conditional expectation zero quantity on $\mathcal{F}_{t_{i-2}}$. This is of course a method that can be used in general with linear moment restrictions in unobserved factors. A similar interpretation is along the following lines (compare Meddahi and Renault (1996)). Note that

$$\begin{aligned} \mathrm{E}\left(\varepsilon^2_{t_i}/x_i\middle|\, \mathcal{F}_{t_{i-2}}\right) &= \mathrm{E}\left(f_{t_{i-1}}\middle|\, \mathcal{F}_{t_{i-2}}\right) \\ &= \omega_i + \gamma_i f_{t_{i-2}}, \text{ and} \\ \mathrm{E}\left(\varepsilon^2_{t_{i-1}}/x_{i-1}\middle|\, \mathcal{F}_{t_{i-2}}\right) &= f_{t_{i-2}}. \end{aligned}$$

This yields the conditional moment restriction

$$\mathrm{E}\left(\varepsilon^2_{t_i}/x_i - \omega_i - \gamma_i\varepsilon^2_{t_{i-1}}/x_{i-1}\middle|\, \mathcal{F}_{t_{i-2}}\right) = 0. \tag{2.13}$$

This equation can be used in a straightforward way to derive GMM estimators for the volatility parameters $\theta$ and $\kappa$. In Section 4 we will use equation (2.13) to derive a similar moment condition in case prices are observed with an additional microstructure noise.

It is possible to derive conditions under which the above SARV structure, simplifies to a semi-strong GARCH model. Allthough we will not impose these additional conditions in the remaining sections, it is insightful to consider this case in order to compare our results with other results in the literature. Suppose that

$$\begin{aligned} f_{t_{i-1}} &= \omega_i + \gamma_i f_{t_{i-2}} + \nu_{t_{i-1}} \\ &= \omega_i + \alpha_i \varepsilon^2_{t_{i-1}}/x_{i-1} + \beta_i f_{t_{i-2}}, \end{aligned} \tag{2.14}$$

i.e. suppose that we are in the semi-strong GARCH framework. Taking conditional expectations with respect to $\mathcal{F}_{t_{i-2}}$, we find that $\gamma_i = \alpha_i + \beta_i$ and hence

$$\nu_{t_{i-1}} = \alpha_i(\varepsilon^2_{t_{i-1}}/x_{i-1} - f_{t_{i-2}}).$$

This shows that the semi-strong GARCH case occurs if and only if in the $MA(1)$ decomposition of the (here assumedly covariance stationary) process

$$\begin{aligned} \varepsilon^2_{t_i}/x_i - \omega_i - \gamma_i \varepsilon^2_{t_{i-1}}/x_{i-1} &= (\varepsilon^2_{t_i}/x_i - f_{t_{i-1}}) - \gamma_i(\varepsilon^2_{t_{i-1}}/x_{i-1} - f_{t_{i-2}}) + \nu_{t_{i-1}} \\ &= (\varepsilon^2_{t_i}/x_i - f_{t_{i-1}}) - \beta_i(\varepsilon^2_{t_{i-1}}/x_{i-1} - f_{t_{i-2}}), \end{aligned}$$

in terms of $\xi_{t_i} - \beta_i \xi_{t_{i-1}}$ is such that $\xi_{t_i}$ forms a martingale difference sequence (and hence $\xi_{t_i} = \varepsilon^2_{t_i}/x_i - f_{t_{i-1}}$).

It is essentially the GARCH specification (2.14) that is used in Engle (1996) and Ghysels and Jasiak (1996). However, both papers are different from ours in significant ways. Engle (1996) focusses indeed on the modellization of *volatility per time-unit*, but his specification is (2.14) where the parameters $\omega_i$, $\alpha_i$, and $\beta_i$ are constant over $i$. Thus, this paper does not take into account the effects of temporal aggregation on the model parameters. From (2.10) and (2.11) we see that this effect is quite important. It is clear that it plays a role in other specifications as well. For example, for long durations $x_i$, the conditional volatility per time-unit over the next duration should be close to the unconditional one. As we see from (2.9), $f_{t_{i-1}} \to \theta$ for $x_i \to \infty$. Specifying time-invariant parameters in (2.14) assumes away this intuitively clear effect. On the other hand, Ghysels and Jasiak (1996) does take into account the temporal aggregation effects, but models the *total variance* over the next duration. It is clear that total variances are foremost influenced by the associated duration. For example, in their model a high variance for the current duration induces a high volatility for the following duration, even if both durations are of unequal length.

In short, volatility over durations in a stochastic volatility setting is affected by two things. First, of course, longer duration generally imply higher total volatility over the duration as a result of cumulating uncertainty. Following Engle (1996), we incorporate this by considering volatility per time-unit. Secondly, the irregularly spaced datapoints induce (by definition) a temporal aggregation effect, as is modelled in Ghysels and Jasiak (1996) in a GARCH framework. Intuitively, the influence of previous volatility on current volatility depends on how long ago previous volatility refers to, i.e. on the previous duration. We, using a stochastic volatility framework, identify both effects by starting from a continuous time model.

# 3 Modelling transaction dates

The seminal paper Engle and Russell (1995) proposes to subsume the time dependence in the durations between transactions in their conditional expectations via

$$\psi_{i-1} = \mathrm{E}(x_i | \mathcal{F}_{t_{i-1}}), \tag{3.1}$$

where $\mathcal{F}_{t_{i-1}}$ denotes now the information in the processes $L$ and $K$ up to time $t_{i-1}$ and the information in all time stamps $t_1, t_2, ..., t_j,\ j < i$. Their so-called ACD (Autoregressive Conditional Duration) model specifies the observed duration as mixing two components.

On the one hand, "standardized durations"

$$\tilde{x}_i = \frac{x_i}{\psi_{i-1}}, \tag{3.2}$$

are assumed to be i.i.d. and to follow an exponential distribution of unit intensity, i.e.

$$\mathrm{Prob}(\tilde{x}_i \leq u) = 1 - \exp(-u). \tag{3.3}$$

It is worthwhile to notice that Engle and Russell (1995) proposes more generally to introduce an additional parameter $\lambda$ such that $\tilde{x}_i^\lambda$ is conformable to (3.3). $\tilde{x}_i$ then follows a Weibull probability distribution. $\lambda < 1$ (resp. $\lambda > 1$) allows one to capture decreasing (resp. increasing) hazard functions (see also Engle (1996) for evidence of a decreasing hazard function with $\lambda \approx 0.8$). For sake of notational simplicity, this additional degree of freedom is not considered in this theoretical section; the shape of the hazard function will be discussed in the empirical section. On the other hand, by a natural adaptation of the GARCH(1,1) idea, Engle and Russell (1995) proposes the following dynamics for the expected conditional duration

$$\psi_{i-1} = d + a x_{i-1} + b \psi_{i-2}. \tag{3.4}$$

Note that, in contrast with the GARCH and ACD literature, but according to our notations for the stochastic volatility process in Section 2 above, the expected conditional duration at time $t_{i-1}$ is denoted by $\psi_{i-1}$ and not by $\psi_i$. Ghysels, Gourieroux, and Jasiak (1998) (GGJ hereafter) have, among others, recently stressed that, concerning the above model, the "most obvious and serious drawback (...) is the assumption that the dynamics of the conditional distribution is entirely channelled through the single factor" $\psi_{i-1}$, that is through the conditional expectation. They propose the concept of SVD (Stochastic Volatility Duration) model as a solution to this drawback. Indeed, it is worthwhile to systematize the GGJ criticism and to enrich the ACD model in five stages which are described below.

First, we will show that the ACD specification (3.4) is too restrictive with respect to a dynamic factor duration modelling in a "Stochastic Volatility" style, in the same way that Meddahi and Renault (1996) have shown that a convenient weakening of the GARCH concept (in the spirit of weak GARCH processes à la Drost and Nijman (1993)) leads to a SARV structure. A similar reasoning is along the following lines (compare Section 2 above). If one starts from the maintained assumption of an autoregressive conditional duration

$$\psi_{i-1} = \mathrm{E}(x_i|\mathcal{F}_{t_{i-1}}) = d + c\psi_{i-2} + \eta_{i-1}, \tag{3.5}$$

where

$$\mathrm{E}(\eta_{i-1}|\mathcal{F}_{t_{i-2}}) = 0, \tag{3.6}$$

one remains true to the genuine meaning of the term ACD but relaxes significantly (3.4). To see this, note that

$$\begin{aligned} \mathrm{E}(x_i|\mathcal{F}_{t_{i-2}}) &= \mathrm{E}(\mathrm{E}[x_i|\mathcal{F}_{t_{i-1}}]|\mathcal{F}_{t_{i-2}}) = \mathrm{E}(\psi_{i-1}|\mathcal{F}_{t_{i-2}}) \\ &= d + c\psi_{i-2} = d + c\mathrm{E}(x_{i-1}|\mathcal{F}_{t_{i-2}}). \end{aligned}$$

This yields the following conditional moment restriction for conditional durations (which is the exact equivalent of (2.13) for volatility dynamics)

$$\mathrm{E}(x_i - d - cx_{i-1}|\mathcal{F}_{t_{i-2}}) = 0. \tag{3.7}$$

This equation can be used in a straightforward way to derive GMM estimators for the conditional duration parameters $c$ and $d$. Indeed, (3.7) is a stochastic volatility type conditional duration model while it is easy to check that the ACD case (3.4) occurs if and only if, in addition to the $MA(1)$ decomposition of the (here assumedly covariance stationary) process $x_i - d - cx_{i-1}$ (implied by (3.7))

$$x_i - d - cx_{i-1} = \xi_i - b\xi_{i-1}, \tag{3.8}$$

the innovation process $\xi_i$ is a martingale difference sequence (and the coefficient $b$ in (3.8) coincides with $b$ in (3.4)). Indeed, $\xi_i$ is a martingale difference sequence if and only if $\xi_i = x_i - \psi_{i-1}$ and thus

$$\begin{aligned} x_i - d - cx_{i-1} &= (x_i - \psi_{i-1}) - c(x_{i-1} - \psi_{i-2}) + \eta_{i-1} \\ &= (x_i - \psi_{i-1}) - b(x_{i-1} - \psi_{i-2}), \end{aligned}$$

which implies

$$\eta_{i-1} = a(x_{i-1} - \psi_{i-2}), \tag{3.9}$$

where $c = a + b$. In turn, (3.9) jointly with (3.5) provides (3.4).

Equation (3.5) can be thought of as being generated by a continuous time diffusion, much in spirit of the time-deformation literature. In order to see this, let $Z$ follow a non-negative continuous time diffusion process with linear mean-reversion, like

$$\mathrm{d}Z_t = \kappa_1(\theta_1 - Z_t)\mathrm{d}t + \delta Z_s^{\lambda}\mathrm{d}W_t^{(Z)}, \tag{3.10}$$

with appropriate conditions on $\kappa_1$, $\theta_1$, $\delta$, and $\lambda$ to ensure non-negativity of the solution. If we define the transaction times $t_i$ by $t_0 = 0$ and

$$t_i = \int_0^i Z_s \mathrm{d}s,$$

we obtain

$$x_i = t_i - t_{i-1} = \int_{i-1}^i Z_s \mathrm{d}s.$$

By the same calculations as in Section 2 one finds

$$Z_{s+h} = \theta_1(1-\exp(-\kappa_1 h)) + \exp(-\kappa_1 h) Z_s + \exp(-\kappa_1 h) \int_s^{s+h} \exp(\kappa_1(u-s)) \delta Z_u^\lambda \mathrm{d}W_u^{(Z)},$$

so that expected durations follow

$$\begin{aligned} \psi_{i-1} &= \mathrm{E}(x_i|\mathcal{F}_{t_{i-1}}) = c(\kappa_1) Z_{i-1} + \theta_1(1 - c(\kappa_1)) \\ \psi_i &= \theta_1(1-\exp(-\kappa_1)) + \exp(-\kappa_1)\psi_{i-1} + v_i, \end{aligned}$$

where, again, $c(v) = (1 - \exp(-v))/v$ and $v_i$ is a martingale difference sequence of innovations. Note that the specification (3.10) is only essential with respect to the mean-reversion part. Otherwise, along the same lines, heteroskedasticity with and without leverage effects, etc. can be included without much problems.

Even though the more restrictive ACD model has been enriched by reducing the specification of the dynamics to the basic $AR(1)$ structure (3.5), it remains true that this linear dynamic specification is somewhat arbitrary for a nonlinear variable as a duration. Therefore, in the framework (3.1)-(3.3), one should preferably replace (3.5) by

$$F_{i-1} = d + cF_{i-2} + \eta_{i-1}, \tag{3.11}$$

where the $\eta_i$ are i.i.d. and normally distributed with zero mean and variance $\sigma^2$, and

$$\begin{aligned} \psi_{i-1} &= -\log[1 - \phi(F_{i-1})], \\ \phi(u) &= \exp\left(-u^2/2\right)/\sqrt{2\pi}. \end{aligned} \tag{3.12}$$

In other words, $\psi_{i-1}$ is directly related to a Gaussian factor $F_{i-1}$ (which is conformable to linear $AR(1)$ dynamics) through a natural monotone mapping between the real line and its positive part. The duration clustering will be captured by the persistence parameter $c$ of the AR process while for a given $c$, $d$ and $\sigma^2$ are one-to-one related to the unconditional mean and variance of the conditional duration process $\psi_i$. Such an approach can be compared with what GGJ (1997) have termed a Stochastic Volatility Duration Model. They propose to define the volatility dynamics by

$$x = \lambda \log[1 - \phi(F_i)], \tag{3.13}$$

where $F_i$ is a Gaussian $AR(1)$ factor whose unconditional distribution is standard normal and $\lambda$ is given real number[2]. The basic motivation of (3.13) is that, since by inversion of (3.3), $H(u) = -\log(1-u)$ is the quantile function of the exponential distribution with unit intensity, $x_i = -\lambda H(\phi(F_i))$ will follow an exponential probability distribution with parameter $(-1/\lambda)$. This latter distribution is unconditional which means that one has to translate the dynamics of $F_i$ in terms of the conditional distribution of $x_i$ given $\mathcal{F}_{t_{i-1}}$. This is the reason why we prefer to remain true to a genuine stochastic volatility type framework like (3.1)–(3.3) completed with (3.11)–(3.12) in order to take into account the intrinsic nonlinearities of duration data.

On the other hand, GGJ stress that "empirical results based, for example, on the intratrade times on the stock markets suggest on the contrary, existence of distinct dynamic patterns of the conditional mean and dispersion as well as different degrees of temporal dependence (...) To accomodate this complex duration pattern clearly at least two time varying parameters are required. Each parameter could represent one conditional moment, i.e. the location (or the mean) and the dispersion (or the variance)". But, while GGJ are led to introduce this second time-varying parameter through a gamma heterogeneity factor (i.e. a time-varying $\lambda$ in (3.13)), our setting (3.11)–(3.12) provides an alternative way to introduce not only dynamics in the conditional mean ($\mathrm{E}(F_{i-1}|\mathcal{F}_{t_{i-2}}) = d + cF_{i-2}$) but also dynamics in the conditional variance. In order to do this, we have just to equip the innovation process $\eta_i$ of $F_i$ with a $ARCH$ or $SV$ structure. For instance, the simple $ARCH(1)$ approach will consist of replacing the i.i.d. assumption on $\eta_i$ by

$$\eta_i|\mathcal{F}_{t_{i-1}} \sim N(0, A + B(F_{i-1} - d - cF_{i-2})^2). \tag{3.14}$$

With such a model one may capture the following two stylized facts: On the one hand a high expected duration $\psi_{i-1}$ at time $t_{i-1}$ corresponds to a high underlying factor $F_{i-1}$ (see (3.12)) and in turn (if $0 < c < 1$) will lead to forecast high $F_i$ and $\psi_i$, that is high expected duration for the consecutive transaction period. On the other hand, there is a volatility clustering effect in consecutive durations as in the underlying $AR(1) - ARCH(1)$ factor.

Our fourth point is about temporal aggregation issues in the duration modelling. One may wonder why we have neglected this temporal aggregation effect that we stressed in Section 2 above. While we noticed in Section 2 (see (2.11)) that the persistence coefficient $\gamma_i$ concerning dates $t_{i-2}$ and $t_{i-1}$ was (since $c(v) \approx 1$) exponentially decreasing with the duration $t_{i-1} - t_{i-2}$, this effect has not yet been taken into account here for the persistence in the mean (fixed coefficient $c$) and in the variance (fixed coefficient $B$). This effect can easily be taken into account but one should note that there are two kinds of persistence coefficients with two different interpretations. Let us for instance consider the following enrichment of the model (3.11)

$$F_i = d + c(\bar{c})^{x_i} F_{i-1} + \eta_i. \tag{3.15}$$

[2]An additional gamma heterogeneity is introduced later; see below.

The two persistence coefficients $c$ and $\bar{c}$ have then the following interpretations. On the one hand, large $c$ means, as above, that a large expected duration $\mathrm{E}(x_i|\mathcal{F}_{t_{i-1}})$ at time $t_{i-1}$ leads one to forecast at this time a large expected duration for the period $(t_i, t_{i+1}]$. This is the well-documented duration clustering effect. But, on the other hand, if one has effectively observed at time $t_i$ a large realized value of $x_i$, we will consider that what was to be expected at time $t_{i-1}$ (far in the past) should be to a large extent forgotten and the expected duration $F_i$ for the next period $(t_i, t_{i+1}]$ is then unpredictable. By the product $c(\bar{c})^{x_i}$, these two antagonist effects are both taken into account in an identifiable way. In particular one may test $c = 1$ or $\bar{c} = 1$ to check if only one effect among the two is present. The same type of trade-off between two kinds of persistence may also be imagined for the conditional variance process defined by (3.14).

An implicit assumption which is maintained in all the interpretations above is the stochastic independence between the various innovation processes. In practice, this independence may be questioned and some interpretations have to be modified slightly in case of rejection. First, leverage effect as stressed by Black (1976), that is instantaneous (negative) correlation between the innovation processes $K$ and $L$ of the price process and the volatility process, may matter for the volatility dynamics. Meddahi and Renault (1996) have shown how leverage in the underlying continuous time SV model produces leverage effect in the discrete time model. Their work is easily extended to the case of irregularly spaced data. Secondly, since the transaction dates model of Section 3 is analogous to a stochastic volatility model, some kind of leverage effect might also be considered here. More precisely, a non zero correlation between $\tilde{x}_i$ (in (3.2)) and $\eta_i$ (in (3.11)) may explain that abnormally long or short observed durations have an asymetric effect on future conditionally expected durations. But the most important challenge in transaction data modelling is the evidence of causality between asset returns and durations, which imply that the strong exogeneity of these durations has to be questioned. In that case, the model of Section 2 which was interpreted as a modellization of the transaction prices conditionally on the observed dates does no longer provide a sequential factorization of the likelihood, since future durations matter in the conditioning information of returns. This lack of exogeneity poses some serious challenges for statistical inference, particularly for indirect inference à la Gourieroux, Monfort, and Renault (1993). Indeed, this is the only case where our model is not easy to simulate and could not be easily estimated.

# 4 Modelling microstructure noise

The previous sections propose a modellization of the transaction times and, conditional on these dates, a modellization of the observed returns. Uncertainty in the transaction times is one of the important aspects of transaction data. A second important effect is microstructure noise. Before we continue to study the effects of this on the derived moment conditions, let us discuss in some detail what we mean by this.

In our opinion, modelisation of asset prices is significantly different for high frequency (i.e. intraday returns) and low frequency (i.e. daily or lower returns). For instance, the simple GARCH(1,1) model is known to provide a good description of day-to-day or lower returns. It is also known that this model does not describe well intraday returns. In that sense, there seem to be quite different dynamics for intra- and extra-day returns. Our goal is the estimation of the continuous time model (2.1)-(2.2). This model is based on low-frequency considerations for which it is known to perform quite well. On the other hand, it seems reasonable to estimate continuous time models using observations at the highest frequency available, i.e. using all transactions. But, as mentioned above, it would be heroic to assume that (2.1)-(2.2) would also accurately describe intraday returns. We therefore view (2.1)-(2.2) as a description of the long term evolution of asset prices, and assume that the observed transaction prices $\tilde{S}$ equal the long term ones $S$ plus an additional noise $M$. Hence, $M$ describes short term intraday effects and will be assumed to be independent over distinct days.

In order to distinguish the different days, we will adopt a slightly adapted notation in this section. We will write $S_t(u)$ for $S_{t+u}$, where the integer $t$ denotes the number of the day and $u \in [0,1)$ the within day time. Let $N_t$ denote the number of transactions on day $t$[3]. We will assume that these transactions take place at (intraday) times $\tau_{ti}$, $i = 1, \ldots, N_t$. The microstructure noise will be modelled by a sequence of Gaussian processes $M_t = (M_t(u) : \ 0 \le u < 1)$. As mentioned before, the processes $M_t(\cdot)$ are assumed to be independent for distinct days $t$. Moreover, we will assume that the processes $M_t(\cdot)$ are Gaussian with mean function $m : [0,1) \to \mathbb{R}$ and covariance $c : [0,1) \times [0,1) \to \mathbb{R}$. Finally, we assume that the microstructure noise is independent of $L$ and $K$.

The observed transaction price for the $i$-th transaction on day $t$ is then given by

$$\tilde{S}_t(\tau_{ti}) = S_t(\tau_{ti}) \exp(M_t(\tau_{ti})). \tag{4.1}$$

The observed return for $t + \tau_{t,i-1}$ to $t + \tau_{ti}$ hence equals

$$\begin{aligned} \tilde{y}_t(i) &= \log\left(\tilde{S}_t(\tau_{ti}) / \tilde{S}_t(\tau_{t,i-1})\right) \\ &= \mu x_t(i) + \varepsilon_t(i) + M_t(\tau_{ti}) - M_t(\tau_{t,i-1}), \end{aligned} \tag{4.2}$$

where $x_t(i) = \tau_{ti} - \tau_{t,i-1}$ and

$$\varepsilon_t(i) = \int_{\tau_{t,i-1}}^{\tau_{ti}} \sigma_{t+u} \mathrm{d}L_{t+u}. \tag{4.3}$$

[3]It will be clear that, while we use a day a unit of time this is completely unessential for the results in the paper.

Note that this implies that $m$ is only identified up to affine terms in $u$. The constant is not identifiable because we the observed returns contain differences of the microstructure noise. Linear terms cannot be distinguished from $\mu u$. Therefore, we need the identifying constraints

$$\int_0^1 m(u)\mathrm{d}u = \int_0^1 u m(u)\mathrm{d}u = 0. \tag{4.4}$$

For the moment we will consider the mean function $m$ and the covariance kernel $c$ to be completely unspecified. It is clear, that in that case $m$ and $c$ can only be consistently estimated if the number of days for which all returns are observed converges to infinity. If one is ready to make parametric assumptions about $m$ and $c$, this is no longer necessary. To estimate the parameters in (2.1)-(2.2) and $m$ and $c$, we will derive moment conditions based on the observed returns $\tilde{y}_t(i)$ and the transaction dates $\tau_{ti}$. Let $\tilde{\mathcal{F}}_t(i)$ be the information in the processes $L$ and $K$ up to time $t+\tau_{ti}$, in $M_1(u), \ldots, M_{t-1}(u)$ for $0 \le u < 1$, and in all times $\tau_{sj}$, $j = 1, \ldots, N_s$, $s \le t$. From (4.2) one then obtains

$$\mathrm{E}\left(\tilde{y}_t(i) \middle| \tilde{\mathcal{F}}_t(i-1)\right) = \mu x_t(i) + m(\tau_{ti}) - m(\tau_{t,i-1}). \tag{4.5}$$

This gives a moment condition that allows immediately for estimation of the mean parameters $\mu$ and $m$. Note that only previous day prices are allowed as instruments. If $m$ is specified parametrically, a standard GMM estimator would do. For nonparametric $m$, one would have to resort to nonparametric GMM estimators, e.g. the kernel M-estimators introduced in Gourieroux, Monfort, and Tenreiro (1995). Equation (4.5) can also be used to obtain a simple nonparametric estimator for $\mu$ and $m$. Let $\tilde{z}_t(u)$ be the total return on day $t$ up to time $u$. That is, $\tilde{z}_t(u) = \sum_i \tilde{y}_t(i)$ for those $i$ for which $\tau_{ti} \le u$. Let $\bar{z}(u)$ be the average of $\tilde{z}_t(u)$ over all trading days, i.e. $\bar{z}(u) = \sum_t \tilde{z}_t(u)/T$. The reader is easily convinced that $\bar{z}(u)$ is consistent for $z(u) = \mu u + m(u) - m(0)$. Estimates for $\mu$, $m(0)$, and $m(\cdot)$ can be obtained from projecting the function $\bar{z}$ on the space of affine function. This gives the following estimates for $m(0)$ and $\mu$

$$\begin{bmatrix} \hat{m}(0) \\ \hat{\mu} \end{bmatrix} = \begin{bmatrix} \int_0^1 \mathrm{d}u & \int_0^1 u\mathrm{d}u \\ \int_0^1 u\mathrm{d}u & \int_0^1 u^2\mathrm{d}u \end{bmatrix}^{-1} \begin{bmatrix} \int_0^1 \bar{z}(u)\mathrm{d}u \\ \int_0^1 u\bar{z}(u)\mathrm{d}u \end{bmatrix} = \begin{bmatrix} 4 & -6 \\ -6 & 12 \end{bmatrix} \begin{bmatrix} \int_0^1 \bar{z}(u)\mathrm{d}u \\ \int_0^1 u\bar{z}(u)\mathrm{d}u \end{bmatrix}.$$

The function $m$ is then estimated by the residual of this projection, i.e. $\hat{m}(u) = \bar{z}(u) - \hat{\mu}u + \hat{m}(0)$.

To obtain estimates for the variance parameters $\theta$, $\kappa$, and $c(\cdot,\cdot)$, we need a conditional variance condition for the observed returns $\tilde{y}_t(i)$. Observe that (4.2) and (2.8) imply that (with the obvious notation $f_t(i)$)

$$\mathrm{Var}\left(\tilde{y}_t(i) \middle| \tilde{\mathcal{F}}_t(i-1)\right) = f_t(i-1)x_t(i) + c(\tau_{ti}, \tau_{ti}) + c(\tau_{t,i-1}, \tau_{t,i-1}) - 2c(\tau_{ti}, \tau_{t,i-1}), \tag{4.6}$$

so that

$$\mathrm{E}\left(\tilde{y}_t(i)^2\middle|\,\tilde{\mathcal{F}}_t(i-1)\right) = f_t(i-1)x_t(i) + a_t(i)x_t(i), \tag{4.7}$$

where

$$a_t(i)x_t(i) = c(\tau_{ti},\tau_{ti}) + c(\tau_{t,i-1},\tau_{t,i-1}) - 2c(\tau_{ti},\tau_{t,i-1}) + \left[\mu x_t(i) + m(\tau_{ti}) - m(\tau_{t,i-1})\right]^2. \tag{4.8}$$

Essentially we obtain the same structure as in Section 2, but for the additional and observable term $a_t(i)$. Using the same idea, we observe

$$\mathrm{E}\left(\tilde{y}_t(i)^2\middle|\,\tilde{\mathcal{F}}_t(i-2)\right) = [\omega_i + \gamma_i f_t(i-2)]x_t(i) + a_t(i)x_t(i). \tag{4.9}$$

Therefrom, we obtain the second order moment condition

$$\mathrm{E}\left(\tilde{y}_t(i)^2/x_t(i) - \gamma_i\tilde{y}_t(i-1)^2/x_t(i-1)\middle|\,\tilde{\mathcal{F}}_t(i-2)\right) = \omega_i + a_t(i) - \gamma_i a_t(i-1). \tag{4.10}$$

Note that, for the estimation of the structural parameters $\theta$ and $\kappa$, the moment condition (4.10) can be used, with instruments from the previous day. At first sight, it might seem inefficient not to be able to use returns of the prevailing day. On the other hand, the structural parameters are exactly defined to describe the low frequency (i.e. day-to-day) returns. In that respect, it is only reasonable to use previous day returns as instruments.

In this section we derived first and second order moment conditions that easily allow for GMM-type estimators for the parameters of interest. These estimators are based on intraday returns. As mentioned above, in the fully nonparametric setting for the microstructure noise, it is necessary that the number of observation days converges to infinity in order to be able to construct consistent estimators. In a parametric setting such a condition is not necessary. The above discussion did not take into account the problems arising when considering non-trading hours. During market closure, many interesting effects occur. However, in order to keep the empirical analysis tractable, we chose not to model these effects seperately. In the empirical section, overnight returns are therefore simply excluded from the moment conditions.

## References


BAUWENS, L. and P. GIOT (1997), The logarithmic ACD model: an application to market microstructure and NASDAQ, *CORE Discussion paper*.

BLACK, F. (1976), Studies in stock price volatility changes, *Proceedings of the 1976 Business Meeting of the Business and Statistics Section, American Statistical Association*, 177–181.

DROST, F.C. and TH.E. NIJMAN (1993), Temporal aggregation of GARCH processes, *Econometrica* **61**, 909–927.

DROST, F.C. and B.J.M. WERKER (1996), Closing the GARCH gap: continuous time GARCH modelling, *Journal of Econometrics* **74**, 31–57.

DUFFIE, D. and P. GLYNN (1997), Estimation of Continuous-Time Markov Processes Sampled at Random Time Intervals, mimeo, Stanford University.

ENGLE, R.F. (1982), The autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica* **45**, 987–1007.

ENGLE, R.F. (1996), The econometrics of ultra-high frequency data, Fisher-Schultz Lecture, ESEM, Turkey.

ENGLE, R.F. and J.R. RUSSELL (1995), Forecasting transaction rates: The autoregressive conditional duration model, UCSD discussion paper.

GHYSELS, E., C. GOURIEROUX, and J.JASIAK (1997), Stochastic volatility duration models, preprint.

GHYSELS, E. and J. JASIAK (1996), GARCH for irregularly spaced financial data: The ACD-GARCH model, preprint.

GOURIEROUX, C., A. MONFORT, and E. RENAULT (1993), Indirect inference, *Journal of Applied Econometrics* **8**, 85–118.

GOURIEROUX, C., A. MONFORT, and TENREIRO (1995), Kernel M-estimators, preprint.

MEDDAHI, N. and E. RENAULT (1996), Aggregations and marginalization of GARCH and Stochastic Volatility Models, preprint.

PROTTER, P. (1995), Stochastic Integration and Differential Equations: A New Approach, Springer-Verlag, Berlin.