Causality Effects in Return Volatility Measures with Random Times

Eric Renault* and Bas J.M. Werker†‡

University of North Carolina and *Tilburg University*

First version: February 2002
This version: November 2005

Abstract

We provide a structural approach to identify instantaneous causality effects between quote-to-quote durations and stock price volatility. So far, in the literature, instantaneous causality effects have either been excluded or cannot be identified separately from Granger type causality effects. By giving explicit moment conditions for observed returns over (random) duration intervals, we are able to identify the instantaneous causality effect where news events drive both surprises in durations and surprises in volatilities. We conclude that instantaneous volatility forecasts for, e.g., IBM stock returns must be decreased by as much as 40% when not having seen the next quote before its (conditionally) median time. For less liquidly traded stocks at NYSE this effect is even stronger. Also, instantaneous volatilities are found to be much higher than indicated by standard volatility assessment procedures. Finally, the documented causality effect has significant impact on statistical inference for tick-by-tick data.

Keywords: Continuous time models, Granger causality, Instantaneous causality, Quote-to-quote durations, Realized variance, Ultra-high frequency data, Volatility per trade.

---

*University of North Carolina at Chapel Hill, CIRANO, and CIREQ.

†Finance and Econometrics Group, CentER, Tilburg University, Tilburg, The Netherlands.

‡The authors would like to thank Jeff Russell for his very useful remarks as a discussant at the 2003 winter meetings in Washington D.C., Laura Spierdijk for kindly providing (well-organized) data, and two referees, Rob Engle, Nour Meddahi, and Harald Uhlig for several suggestions that improved the paper significantly. The second author gratefully acknowledges support from the EU within the MICFINMA network. This paper circulated previously under the title "Stochastic volatility models with transaction time risk".

1 Introduction

Engle (2000) defines "ultra-high frequency" data as data where all transactions of/quotes for an asset in a financial market and their characteristics are available. These data are typically recorded at random times of transactions or quote revisions. This paper proposes a way to assess return volatility from this tick-by-tick data, while preserving the continuous-time paradigm on the underlying prices. The continuous time framework has well-known advantages, both for addressing statistical issues, like temporal aggregation, and asset pricing. At the same time, we point out a precise assessment of the instantaneous causality relationship between the volatility of returns and the duration between quote revisions.

The identification of this latter effect has important repercussions for risk measurement and statistical inference. To start with the former, note that two ways to infer return volatility from tick-by-tick data have been recently put forward in the literature. First, starting from the seminal papers of Andersen, Bollerslev, and co-authors on realized variance computed from high-frequency data (see, e.g., Andersen, Bollerslev, and Diebold, 2004, and references therein), a bunch of papers including Barndorff-Nielsen and Shephard (2002), Andreou and Ghysels (2002), Bandi and Russell (2004), and Hansen and Lunde (2005) discuss several bias and variance issues connected to the use of realized variance as an estimator of integrated volatility. While, due to bias problems, most papers do not use tick-by-tick data and, typically, sample at moderate frequencies, Zhang, Mykland, and Aït-Sahalia (2004) explicitly recognizes the tick-by-tick grid as a benchmark. Actually, the conversion of tick-by-tick data into fixed-time interval sampling has several drawbacks including not only information loss about returns but also loss of the informational content of durations (see Engle and Sun, 2005, for a discussion). We show that realized variance calculated directly on the basis of observed returns between quote revisions exhibits a fundamental causality bias, next to the well documented market microstructure noise bias.

Moreover, Engle and Sun (2005) points out an alternative approach to assess tick-by-tick return volatility that amounts to directly calculating the conditional volatility per trade. Their empirical results lead to the conclusion that contemporaneous durations have little effect on the conditional volatility per trade, which is in line with the intuition of an inverse relationship between spot volatility and durations. We argue in this paper that a precise measurement of instantaneous causality between spot volatility and contemporaneous duration is needed to support such a claim. Actually, we are going to show that even a qualitative claim like time invariance of volatility per trade "means that per second volatility is inversely related to the duration of returns" (Engle and Sun, 2005) implicitly assumes that the causality effect we document in this paper is time-invariant.

In line with the above, the causality measure put forward in the present paper actually features a simple economic interpretation in terms of the relative update in instantaneous volatility predictions depending upon the hypothetical knowledge that a subsequent quote revision has occurred or not by a given time. To be more precise, let us assume that a quote revision has been observed at some date t_i and let t_{i+1} denote the next time. Denote by $\mathrm{E}_{t_i}\{\sigma^2_{t_i+u}\}$ the conditional prediction, at time t_i, of the future instantaneous variance $\sigma^2_{t_i+u}$. Furthermore, let $\mathrm{E}_{t_i}\{\sigma^2_{t_i+u}|t_{i+1} > t_i + u\}$ be the forecast associated to the hypothetical knowledge that no new quote or transaction is observed between time t_i and $t_i + u$. From the general point of view (see, e.g., Easley and O'Hara, 1992) that longer durations will be associated with lower volatility, one would expect that the latter forecast is below the former. The empirical specification chosen in this paper is such that, when u corresponds to the median of the conditional (given the information at time t_i) distribution of the random duration, the relative update in the instantaneous variance due to not having seen a new transaction or quote at time $t_i + u$, is a constant depending on –using the notations maintained throughout– a parameter β^*. According to our estimations for several randomly picked stocks traded on NYSE, for the period January 2005-March 2005, the coefficient β^* is always significantly negative, both from an economic and a statistical point of view. The main contribution of the present paper is to show that the knowledge of this causality factor β^* is precisely what is needed to correctly infer integrated volatility from realized variance (corrected for microstructure noise) on tick-by-tick data or to infer spot volatility from conditional volatility per trade or quote revision. Omitting this factor amounts to a severe underestimation of volatility since we show that realized variance and conditional variance per duration are only $1 + \beta^* < 1$ times the integrated variance or variance per expected duration, respectively. Our empirical model points out that, neglecting the causality effect, may lead to volatility assessments which underestimate the actual one by, for IBM, 41% (in relative terms). It is important to note that, while we show that the causality effect disappears when durations converge to zero, this latter number is based on IBM, one of the most frequently traded stocks. Such an order of magnitude clearly makes the so-called causality factor important, next to the microstructure noise that is the main focus of interest of the current literature.

Our instantaneous causality characterization also sheds some light on the way parametric likelihood functions for ultra-high frequency data should be specified. Typically, by contrast, for instance, with the Aït-Sahalia and Mykland (2003), Grammig and Wellner (2002), or Meddahi, Renault, and Werker (2003) approaches to irregularly spaced data, our estimation results lead us to claim that a parametric model should explicitly accommodate a causal role of current durations in the dynamics of return volatility. In this respect, the approach closest to ours is Engle (2000) where current durations explicitly show up in the right-hand side of the ultra-high frequency GARCH equation.

However, we argue additionally that the instantaneous causality effect can only be identified without ambiguity in a continuous time stochastic volatility framework. The important contribution of Duffie and Glynn (2004) must be acknowledged in this respect. For reasons that are going to become apparent in the rest of the paper (Section 3.1), it turns out that a duration model is better suited for addressing the issue of our main interest than a random intensity model as in Duffie and Glynn (2004).

The paper is organized as follows. In Section 2, we present a general framework for incorporating random durations in a continuous time stochastic volatility model. The main result is a decomposition of the conditional volatility over observed durations into two components: (a) the time to build, which reflects the simple idea that the variance of returns is longer over longer intervals and (b) the additional effect of instantaneous causality between durations and volatility. We put forward, in Section 3, a duration based model for the econometric specification of this decomposition. We stress that, following Engle and Russell (1998), this ACD kind of duration model is well suited for the identification of the causality effects that are of prime interest to us. In Section 4, we derive explicit moment conditions that identify our parameters of interest and take into account some well-known phenomena in ultra-high frequency data. Section 5 presents the empirical analysis that shows convincingly that the effect we document is significant, both from a statistical and a financial point of view. In Section 6 we elaborate on the statistical consequences of the causality effects. Section 7 concludes and the proof of our main result is presented in the appendix.

2 A general framework for modeling times and volatilities

We introduce our framework for the analysis of continuous time price processes observed at random times. Our framework allows us to identify separately the marginal price volatility process, the marginal process for the times at which prices are recorded, and the interaction between both.

In the literature, see, e.g., Engle (2000), one often models the marginal distribution of times and, subsequently, the conditional distribution of prices given the times. This, clearly, requires a priori information on the form of the conditional distribution of returns given (future) times. We feel that it is more natural to model the marginal process for prices, as the majority of the empirical finance literature so far deals with this marginal price processes. We show that, given the (marginal) distributions of times and prices, we can model possible causality relations between both using a simple (conditional) regression coefficient. This regression coefficient is sufficient to derive observable moment conditions. In Section 6 we use these results to identify the noncausality assumptions made in previous papers. We want to stress that not all previous papers assume absence of causality from times to prices (e.g., Engle, 2000, and Duffie and Glynn, 2004). However, we think that the present paper is the first to explicitly address the question of (non)causality in a structural way and does not rely on ad hoc reduced form specifications.

The basis of our model is the filtration that generates the information accumulation in the market. Following the majority of the literature, we suppose that this information structure is exogenously given and that it satisfies the so-called 'usual conditions' with respect to the underlying probability space (see, e.g., Protter, 2003, p. 3).

Assumption A The information flow in the market is described by the filtration $(\mathcal{F}_t)_{t\geq 0}$ that is supposed to satisfy the usual conditions.

All stochastic processes that appear in the sequel are assumed to be adapted to the filtration $(\mathcal{F}_t)$, unless explicitly stated otherwise. Note that the filtration $(\mathcal{F}_t)$ is generally not completely observed by the econometrician.

Consider a financial asset with time t price given by S_t. The evolution of the price S_t is supposed to be given by $S_0 = 1$ and

$$\mathrm{d}\log S_t = \sigma_t \mathrm{d}L_t, \;\; t \geq 0. \tag{2.1}$$

In our specification, (σ_t) is a predictable process and (L_t) is some Lévy process. In particular, we do not assume that the volatility process (σ_t) is continuous or Markovian. For ease of exposition, we momentarily ignore a possible drift term. This is also in line with, e.g., Engle (2000). We will allow for a possibly non-zero drift in the empirical analysis of Section 5. Specifications like (2.1) have been used in Carr and Wu (2004) and Eberlein and Papapantoleon (2005).

In order to derive moment conditions, we impose some further conditions.

Assumption B The innovation process (L_t) is assumed to be a zero-mean Lévy process with unit variance, i.e., $\mathrm{Var}\{L_t\} = t$. The volatility process (σ_t) is assumed to be predictable with respect to the filtration $(\mathcal{F}_t)$ and square-integrable in the sense $\mathrm{E}\{\int_0^T \sigma_t^2 \mathrm{d}[L, L]_t\} < \infty$, for all $T > 0$. For any stopping time T, with respect to the filtration $(\mathcal{F}_t)$, we write E_T for the conditional expectation operator given the σ-field $\mathcal{F}_T$ (Protter, 2003, p. 5). Moreover, we define

$$\xi_T(u) = \mathrm{E}_T\left\{\sigma_{T+u}^2\right\}. \tag{2.2}$$

We denote by Ξ_T the integral of ξ_T, with the normalization that $\Xi_T(0) = 0$, i.e., $\Xi_T(u) = \int_{v=0}^{u} \xi_T(v)\mathrm{d}v$.

Assumption B implies that (S_t) is a semimartingale adapted to the filtration $(\mathcal{F}_t)$. In fact, this provides a desirable price model since it is well-known that ruling out arbitrage possibilities in continuous time (in the appropriate way) implies that the price processes are semimartingales (Back, 1991, and Delbaen and Schachermayer, 1999). The unit variance assumption on the Lévy driving process L identifies σ_t as the volatility process. Assuming that L is continuous would, by Lévy's characterization theorem (Protter, 2003, p. 86), imply that L is a Brownian motion. A Brownian motion for L is the only way to exclude jumps in S. In that case the integrability condition on the volatility process becomes simply $\mathrm{E}\{\int_0^T \sigma_t^2 \mathrm{d}t\} < \infty$. Alternatively, for, say, a bounded volatility process, Assumption B is clearly satisfied. Alternative Lévy processes that could be considered are sums of Brownian motions and zero-mean compound Poisson processes with finite variance. The integral Ξ_T of the conditional variance predictor ξ_T will appear in the moment condition that we derive below for returns observed over random durations.

We assume that S_t is only observed at some particular (random) times $t_1, t_2, \ldots$.

Assumption C The times $t_1, t_2, \ldots$ form an increasing sequence of bounded stopping times with respect to the filtration $(\mathcal{F}_t)$. We denote durations by $\Delta t_{i+1} = t_{i+1} - t_i$. Moreover, F_{t_i} denotes the distribution function of the conditional distribution of Δt_{i+1} given $\mathcal{F}_{t_i}$, i.e.,

$$F_{t_i}(u) = \mathbb{P}\left\{ \Delta t_{i+1} \leq u | \, \mathcal{F}_{t_i} \right\}. \tag{2.3}$$

In this paper, t_i will refer to times at which the midquote changes, but other application can be imagined. The stopping time assumption merely states that, at time t_i, all previous midquote changes have been observed by both the investor and the econometrician. For notational convenience we define $t_0 = 0$. Under (2.1), returns on the asset S as they are observed over the interval $(t_i, t_{i+1}]$, are given by

$$R_{t_i:t_{i+1}} = \log \frac{S_{t_{i+1}}}{S_{t_i}} = \int_{t_i}^{t_{i+1}} \sigma_t \mathrm{d}L_t, \;\; i = 0, 1, 2, \ldots. \tag{2.4}$$

Note that, under the assumptions stated, $R_{t_i:t_{i+1}}$ is the increment of a martingale stopped at time Δt_{i+1}, so that Doob's optional sampling theorem (Protter, 2003, p. 9) implies

$$\mathrm{E}_{t_i}\left\{ R_{t_i:t_{i+1}} \right\} = 0, \;\; i = 0, 1, 2, \ldots. \tag{2.5}$$

The following proposition relates the conditional variance of observed returns $R_{t_i:t_{i+1}}$ to the variance predictor Ξ_{t_i}, to the distribution function of the durations F_{t_i}, and to some regression coefficient that we denote $\beta_{t_i}(\cdot)$ and formally define below.

Proposition 2.1 *Under Assumptions A-C we have the following moment condition:*

$$\begin{aligned}\mathrm{Var}_{t_i}\{R_{t_i:t_{i+1}}\} &= \mathrm{E}_{t_i}\{R^2_{t_i:t_{i+1}}\} \\ &= \int_0^\infty \Xi_{t_i}(u)dF_{t_i}(u) + \int_0^\infty \beta_{t_i}(u)F_{t_i}(u)(1-F_{t_i}(u))du, \end{aligned} \tag{2.6}$$

where $\beta_{t_i}(\cdot)$ *is the (conditional) regression coefficient (given* $\mathcal{F}_{t_i}$*)*

$$\beta_{t_i}(u) = \frac{\mathrm{Cov}_{t_i}\{\sigma^2_{t_i+u}, I_{(0,\Delta t_{i+1}]}(u)\}}{\mathrm{Var}_{t_i}\{I_{(0,\Delta t_{i+1}]}(u)\}}, \tag{2.7}$$

and where $I_{(0,\Delta t_{i+1}]}$ *denotes the indicator function of the (random) interval* $(0, \Delta t_{i+1}]$.

Note that, since the conditional expectation given an indicator coincides with affine regression, we can write

$$\mathrm{E}_{t_i}\left\{\sigma^2_{t_i+u}\middle| I_{(0,\Delta t_{i+1}]}(u)\right\} - \mathrm{E}_{t_i}\left\{\sigma^2_{t_i+u}\right\} = \beta_{t_i}(u)\left[I_{(0,\Delta t_{i+1}]}(u) - (1-F_{t_i}(u))\right]. \tag{2.8}$$

From (2.8), we see that the β function characterizes by how much an instantaneous variance assessment is influenced by the information that no quote revision occurred for some time. It is then not surprising that this information matters as well for measuring the volatility of returns between two consecutive quote revisions as in (2.6). Generally speaking, when returns are considered over random time intervals $(t_i, t_{i+1}]$, the duration Δt_{i+1} between two consecutive stopping times may convey (through a non-zero coefficient β) some relevant information about the risk borne at time t_i over the horizon Δt_{i+1}.

For sake of exposition, let us call, following Engle and Sun (2005), the quantity $\mathrm{Var}_{t_i}\{R_{t_i:t_{i+1}}\}$ the "conditional volatility per trade". It may, as in our empirical analysis, actual be the volatility per quote revision. The main message of Proposition 2.1 is that this volatility has two components. First, there is a time-to-build part which simply reflects the fact that the variance of returns has to be accumulated over the relevant duration, before computing an expectation with respect to the conditional distribution F_{t_i} of the next duration. This time-to-build effect can be seen as an expected integrated volatility *imposing* independence between times and prices. More precisely,

$$\begin{aligned} TB_{t_i} &= \int_0^\infty \Xi_{t_i}(\Delta)\mathrm{d}F_{t_i}(\Delta) \\ &= E^{\otimes}_{t_i}\left\{\int_0^{\Delta t_{i+1}} \sigma^2_{t_i+u}\mathrm{d}u\right\} \qquad (2.9) \\ &= \int_0^\infty (1-F_{t_i}(u))\,\mathrm{E}_{t_i}\sigma^2_{t_i+u}\mathrm{d}u. \qquad (2.10) \end{aligned}$$

The second equality comes from Fubini's theorem where $\otimes$ indicates that the expectation is taken with respect to the product measure of the marginal (yet conditional on $\mathcal{F}_{t_i}$) distributions of Δt_{i+1}

and $\left(\sigma^2_{t_i+u}:\ u \geq 0\right)$. The third equality follows by integration by parts. This product measure ignores any possible instantaneous causality effects. The conditional expectation $\Xi_{t_i}(\Delta)$ of integrated volatility for deterministic durations Δ is studied in detail in Bollerslev and Zhou (2002). The moment conditions they derive can be directly translated in terms of the time-to-build effect, just by performing the appropriate averaging using the duration distribution.

The second component in (2.6) is the additional effect of causality between quote revision times and volatilities. Let us call this the instantaneous causality effect:

$$\begin{aligned} IC_{t_i} &= \int_0^\infty \beta_{t_i}(u) F_{t_i}(u)\left(1 - F_{t_i}(u)\right) \mathrm{d}u \\ &= \int_0^\infty \mathrm{Cov}_{t_i}\left\{\sigma^2_{t_i+u}, I_{(0,\Delta t_{i+1}]}(u)\right\} \mathrm{d}u. \end{aligned} \tag{2.11}$$

It is worth noticing that this effect would be negligible if the durations were infinitely small. However, as we find empirically later, even for the most frequently traded stocks the effect is negative and far from negligible.

Relation (2.8) gives the informational content in contemporaneous durations for volatility predictions at the time t_i in case a quote revision or transaction has (or has not) occurred. It is also interesting to have such predictions at times in between quote revisions or transactions. Relation (2.8) can be used to that extent as well, if we consider the situation where only quote revision information on the stock at hand is available. In such case, the information available to the econometrician at time $t \in [t_i, t_{i+1})$ would be $\mathcal{F}_{t_i}$ and the fact that $t_{i+1} > t$. Denote the expectation given this information by $\tilde{\mathrm{E}}_t$. In that case the prediction of σ^2_{t+u}, given the fact that the next quote revision will take still at least $u > 0$ seconds, is given by

$$\begin{aligned} \tilde{\mathrm{E}}_t\left\{\sigma^2_{t+u}\middle| t_{i+1} \geq t+u\right\} &= \mathrm{E}_{t_i}\left\{\sigma^2_{t_i+(t-t_i)+u}\middle| \Delta t_{i+1} \geq t - t_i + u\right\} \\ &= \mathrm{E}_{t_i}\left\{\sigma^2_{t+u}\right\} + \beta_{t_i}(t-t_i+u) F_{t_i}(t-t_i+u). \end{aligned} \tag{2.12}$$

To see this, just apply (2.8) with u replaced by $i - t_i + u$.

Moreover, (2.8) gives the update in the prediction of $\sigma^2_{t_i+u}$ given the information $\Delta t_{i+1} > u$. One may be equally interested in the update in the prediction of $\sigma^2_{t_i+v}$, with $v > u$ given this information. Proposition 2.1 can be used for this problem as well. To see this, not that we obviously have, for $v > u$ and on the set $\Delta t_{i+1} \geq u$,

$$\begin{aligned} \mathrm{E}_{t_i}\left\{\sigma^2_{t_i+v}\middle| t_{i+1} \geq t_i+u\right\} &= \mathrm{E}_{t_i}\left\{\sigma^2_{t_i+u}\middle| t_{i+1} \geq t_i+u\right\} \\ &\quad + \mathrm{E}_{t_i}\left\{\sigma^2_{t_i+v} - \sigma^2_{t_i+u}\middle| t_{i+1} \geq t_i+u\right\} \\ &= \mathrm{E}_{t_i}\left\{\sigma^2_{t_i+u}\middle| t_{i+1} \geq t_i+u\right\} \\ &\quad + \mathrm{E}_{t_i}\left\{\mathrm{E}_{t_i+u}\left\{\sigma^2_{t_i+v} - \sigma^2_{t_i+u}\right\}\middle| t_{i+1} \geq t_i+u\right\}. \end{aligned} \tag{2.13}$$

The first conditional expectation is the one derived in (2.8). The second one involves the prediction of $\sigma^2_{t_i+v} - \sigma^2_{t_i+u}$ given the information available at time $t_i + u$. In the empirical section, we will impose a martingale assumption for the very high frequency instantaneous variance process. Under such an assumption, the second term in (2.13) vanishes. Under other specifications that lead to linear functions of $\sigma^2_{t_i+u}$, like linear mean-reversion, (2.8) could be used once more to derive the appropriate expression.

The analysis in this section is fairly general and we introduce, and motivate, a duration based model that we will essentially use in the empirical analysis in the next section.

3 A duration based model for conditional volatility by trade

In this section we present (and motivate) the use of a duration based model to identify the instantaneous causality effect between durations and volatilities.

3.1 ACD specification and proportional updates of volatility predictions

Engle and Russell (1998) proposes the general framework of Autoregressive Conditional Duration models, which are characterized by the fact that durations Δt_{i+1} divided by their conditional expectations are serially independent and identically distributed. The following assumption makes this precise in terms of the distribution functions F_{t_i} in Assumption C.

Assumption D Let $\psi_{t_i} = \mathrm{E}_{t_i}\Delta t_{i+1}$ denote the conditionally expected next duration at time t_i. We have $F_{t_i}(u) = F(u/\psi_{t_i})$ where F is a probability distribution function (on the positive part of the real line) with unit expectation.

Note that ψ_{t_i} may be a function not only of past durations, but also of past returns in case of Granger causality effects from returns towards durations. However, we are primarily interested in instantaneous causality as characterized by the function β_{t_i} in Proposition 2.1. Using Assumption D, we may rewrite (2.8) in terms of the rescaled forecasting horizon $v = u/\psi_{t_i}$. This yields

$$\mathrm{E}_{t_i}\left\{\sigma^2_{t_i+v\psi_{t_i}} \,\middle|\, \Delta t_{i+1} \geq v\psi_{t_i}\right\} - \mathrm{E}_{t_i}\sigma^2_{t_i+v\psi_{t_i}} = \beta_{t_i}\left(v\psi_{t_i}\right)F\left(v\right). \tag{3.1}$$

For the purpose of econometric specification, we need to specify how the function β_{t_i} depends on the conditioning information $\mathcal{F}_{t_i}$. Both Engle (2000) and Manganelli (2004) estimate a discrete time model of the conditional variance of $R_{t_i:t_{i+1}}$ given not only $\mathcal{F}_{t_i}$ but also the current duration Δt_{i+1}. Whatever the difference in approach, their empirical results gives us some guidelines about the way the forecast at time t_i of $\sigma^2_{t_i+v\psi_{t_i}}$ should be modified by the additional knowledge that

$\Delta t_{i+1} \geq v\psi_{t_i}$. Under the working hypothesis that linear approximations give a correct account of these relations, it seems natural to relate the shape of $\beta_{t_i}(v\psi_{t_i})$ to the (unconditional) volatility predictions at the corresponding horizon. This is formalized in the next assumption.

Assumption E The regression function β_{t_i} introduced in Proposition 2.1 satisfies

$$\beta_{t_i}(u) = \beta\left(\frac{u}{\psi_{t_i}}\right) \mathrm{E}_{t_i}\left\{\sigma^2_{t_i+u}\right\}, \tag{3.2}$$

for a given function β defined on the support of the distribution function F such that, for all v, $\beta(v)F(v) \geq -1$.

Note that, following the common wisdom that long durations are associated with high levels of volatility, we expect negative values for the function β. Assumption E implies that, for a given level of the rescaled forecasting horizon v, the update in variances predictions is constant in relative terms:

$$\frac{\mathrm{E}_{t_i}\left\{\sigma^2_{t_i+v\psi_{t_i}} \,\middle|\, \Delta t_{i+1} \geq v\psi_{t_i}\right\}}{\mathrm{E}_{t_i}\sigma^2_{t_i+v\psi_{t_i}}} - 1 = \beta(v)\, F(v)\,, \tag{3.3}$$

and, similarly,

$$\frac{\mathrm{E}_{t_i}\left\{\sigma^2_{t_i+v\psi_{t_i}} \,\middle|\, \Delta t_{i+1} < v\psi_{t_i}\right\}}{\mathrm{E}_{t_i}\sigma^2_{t_i+v\psi_{t_i}}} - 1 = -\beta(v)\,(1 - F(v))\,. \tag{3.4}$$

The condition $\beta(v)F(v) \geq -1$ in Assumption E assures that, even when volatility forecasts are updated downward in (3.3), they never become negative. It is worth stressing that we document empirically that the prediction updates (3.3) and (3.4) are not negligible. To get a compelling assessment of their economic significance, let us consider the simplest model where the function β is constant and F corresponds to the exponential distribution: $F(v) = 1 - \exp(-v)$. Even though these assumptions are never maintained in the rest of the paper, we can use them to get a visual appraisal of the orders of magnitude in the volatility updates (3.3) and (3.4). Using the GMM-based estimated parameters for IBM (and their standard errors) as they are presented in Section 5, Figure 3.1 and 3.2 present the updates according to Equations (3.3) and (3.4), respectively. Figure 3.1 shows, for instance, that a present time prediction made for the instantaneous volatility 1.5 seconds from now (the median duration), conditional on not having seen a quote revision by that time, is about 40% less than the unconditional prediction. Note, however, that since Figure 3.1 has been built under the working assumption that the function β is constant, we are likely to exaggerate volatility updates for large durations. Similarly, Figure 3.2 gives the volatility update, in present time predictions, conditional on having seen a quote revision within a given period. At the median duration of 1.5 seconds, the instantaneous volatility prediction now has to be increased by about 28% if we now that a new quote is available. In other words, the instantaneous causality effect is clearly economically

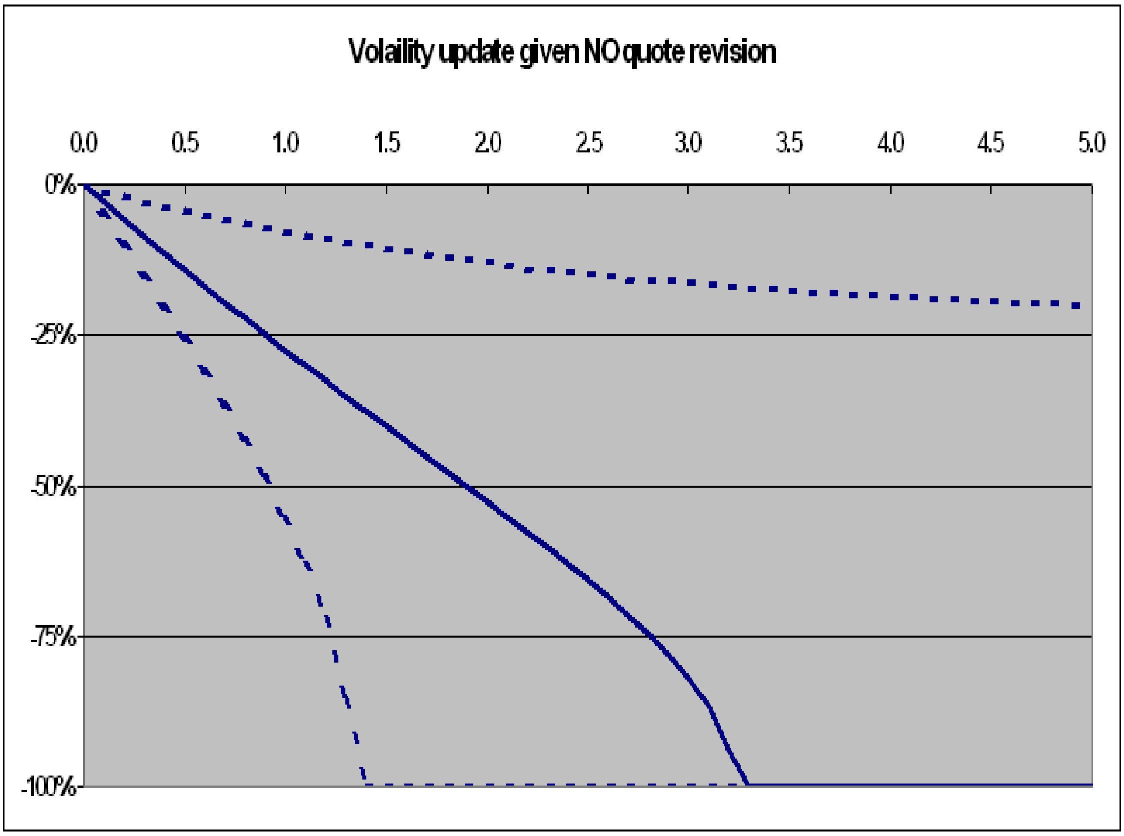

Figure 3.1: Relative update in instantaneous volatility prediction due to *not* having seen a new quote by the time (in seconds) indicated on the horizontal axis. The graph is based on the estimated parameters for IBM (Section 5) and the additional hypotheses of a constant regression function β and exponentially distributed durations. The solid line gives the point estimate and the dotted lines give 95% confidence intervals.

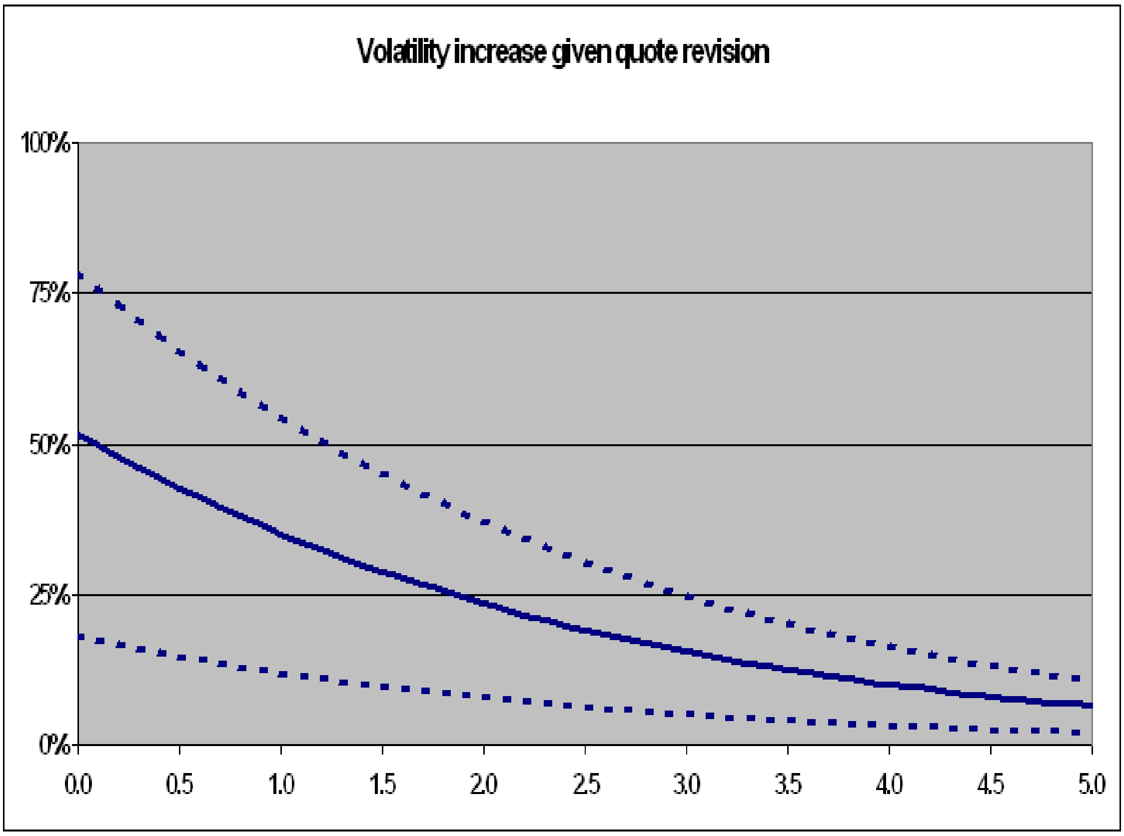

Figure 3.2: Relative update in instantaneous volatility prediction due to having seen a new quote by the time (in seconds) indicated on the horizontal axis. The graph is based on the estimated parameters for IBM (Section 5) and the additional hypotheses of a constant regression function β and exponentially distributed durations. The solid line gives the point estimate and the dotted lines give 95% confidence intervals.

significant. A similar effect would also show up with durations between trades as shown in a previous version of this paper. Note that the nice feature of proportional updates of volatility predictions as displayed in Figures 3.1 and 3.2 is a direct consequence of the ACD type Assumptions D and E. The relative adjustment given the hypothetical information that the next duration exceeds (or is below) its conditional median, its conditional first quartile, or any given conditional quantile, is always the same, irrespective of the other available forecasting information. This nice, albeit simple, updating rule is quite useful for economic reasoning. This is the reason why the direct specification of a duration model, while remaining nonparametric about the distribution F of rescaled durations and the causality function β, is more convenient for our purposes than a more general model about the point process of quote revisions.

3.2 Conditional volatility per trade

Following Engle and Sun (2005), we keep the terminology "conditional volatility per trade" to designate $\mathrm{Var}_{t_i}\{R_{t_i:t_{i+1}}\}$ even though, given the empirical illustration in this paper, we should rather say "conditional volatility per quote revision". Under the Assumptions D and E, we can rewrite the volatility decomposition of Proposition 2.1 through a change of variables $u = v\psi_{t_i}$. We find

$$\mathrm{Var}_{t_i}\{R_{t_i:t_{i+1}}\} = \psi_{t_i} \int_0^\infty (1+\beta(v)F(v))(1-F(v))\,\mathrm{E}_{t_i}\left\{\sigma^2_{t_i+v\psi_{t_i}}\right\} \mathrm{d}v. \tag{3.5}$$

It is then worth introducing a modified density for the distribution of durations:

$$g(v) = \frac{(1+\beta(v)F(v))(1-F(v))}{1+\beta^*}, \tag{3.6}$$

with

$$\beta^* = \int_0^\infty \beta(v)F(v)(1-F(v))\,\mathrm{d}v. \tag{3.7}$$

Its easy to see that g defines indeed a probability density function over the positive real line. If we denote by G the corresponding cumulative distribution function, we find the following result.

Proposition 3.1 *Under Asumptions A-E and with the distribution G defined via (3.6), we have*

$$\mathrm{Var}_{t_i}\left\{R_{t_i:t_{i+1}}\right\} = (1+\beta^*)\,\psi_{t_i}\mathrm{E}_G\left[\mathrm{E}_{t_i}\left\{\sigma^2_{t_i+V\psi_{t_i}}\right\}\right], \tag{3.8}$$

where E_G *denotes the expectation operator when the variable V is endowed with the distribution G.*

In other words, volatility per trade involves an average of volatility predictions, when the average is computed with some modified distribution G for rescaled durations and a causality factor $1+\beta^*$. To get some intuition about the modified distribution, several remarks are in order. First, G

coincides with F when F is the exponential distribution and there is no causality effect ($\beta(v) = 0$ for all v). More generally, in case there is no causality effect, $\mathrm{E}_G(V)$ may be slightly larger than one when the distribution function F features some overdispersion. More precisely, if we write

$$\varphi = 1/\mathrm{E}_F\{V^2\} = \left(\int_{v=0}^{\infty} v^2 \mathrm{d}F(v)\right)^{-1}, \tag{3.9}$$

we have $\mathrm{E}_G\{V\} = (2\varphi)^{-1}$. Hence, there is overdispersion in F if and only if $\varphi < 1/2$, and, in that case, $\mathrm{E}_G\{V\} > 1$. A more general result can be formulated as well.

Proposition 3.2 *When the β function is constant and negative ($\beta(v) = \beta \leq 0$ for all v), the distribution of V, under G, is decreasing (in the sense of first-order stochastic dominance) in the absolute value of β. Its maximum value, reached for $\beta = 0$, is the distribution with density $1 - F(v)$ and expectation $(2\varphi)^{-1}$.*

We can conclude that horizons of volatility predictions involved in the expectation E_G in Proposition 3.1 should not much exceed, on average, the actual conditionally expected duration ψ_{t_i}, that is typically in the order of magnitude of a few seconds (see Table 1 in Section 5). Since on very short term intervals volatilities are known to be highly persistent, it means that for all practical purposes we can see the volatility predictions $\mathrm{E}_{t_i}\left\{\sigma^2_{t_i+V\psi_{t_i}}\right\}$ in Proposition 3.1 as almost identical to the present spot volatility level $\sigma^2_{t_i}$. Concerning this approximation, the orders of magnitude are extensively discussed in Fouque, Papanicolaou, and Sircar (2000). They report, for instance, a time of mean-reversion for S&P-500 volatility of 1.5 days. Therefore, there is no significant mean-reversion effect over the horizons of a few seconds that we study. In other words, Proposition 3.1 implies

$$\mathrm{Var}_{t_i}\left\{R_{t_i:t_{i+1}}\right\} = (1 + \beta^*)\,\psi_{t_i}\sigma^2_{t_i;G}, \tag{3.10}$$

where $\sigma^2_{t_i;G}$ is an instantaneous volatility level, depending on the distribution G, but very close to $\sigma^2_{t_i}$. In particular, irrespective of the distribution G, we have exactly

$$\mathrm{Var}_{t_i}\left\{R_{t_i:t_{i+1}}\right\} = (1 + \beta^*)\,\psi_{t_i}\sigma^2_{t_i}, \tag{3.11}$$

in case the volatility process is a martingale like for IGARCH or unit root volatility models (Hansen, 1995). Note that using (3.11) as an accurate approximation of (3.10) is crucial for a semiparametric identification of the conditional volatility per trade. To see this, just note that any Ornstein-Uhlenbeck like model of volatility mean reversion (see, e.g., Drost and Werker,1996, Barndorff-Nielsen and Shephard, 2002, or Meddahi and Renault, 2004) would imply

$$\mathrm{E}_{t_i}\left\{\sigma^2_{t_i+V\psi_{t_i}}\right\} = a\left(1 - \exp\left(-\kappa V\psi_{t_i}\right)\right) + \exp\left(-\kappa V\psi_{t_i}\right)\sigma^2_{t_i},$$

so that calculation of $\sigma^2_{t_i;G}$ for an unknown mean reversion parameter κ would involve the complete specification of the Laplace transform of the probability distribution G. In other words, there would be no way to avoid a parametric model for durations and for the causality function β.

Finally, note that (3.11) provides a very convenient way to correct the common rule of thumb that the current value of the spot volatility process can be inferred by just dividing the conditional volatility per trade by the expected duration. Such a rule of thumb is one way to understand the Engle and Sun (2005) observation that volatility is inversely related to expected durations as the conditional volatility per trade appears to be nearly independent of durations. Actually, this latter observation is consistent with (3.11), although the rule of thumb is incorrect. More precisely, there is an additional instantaneous causality factor (β^*) that the above rule of thumb overlooks but, insofar as we assume it constant over time, we still obtain the aforementioned inverse proportionality relationship, albeit with a very different level for the instantaneous volatility. To illustrate the consequences of neglecting the causality factor $1+\beta^*$, that is, to compute the instantaneous variance by unit of time as $\mathrm{Var}_{t_i}\left\{R_{t_i:t_{i+1}}\right\}/\psi_{t_i}$, just remember that the estimated β^* is -65% for IBM and even more negative for some other stocks. As a result, the rule of thumb underestimated the actual instantaneous volatility by $\sqrt{1+\beta^*} = 59\%$ (in relative terms). Clearly, this may have important repercussions for risk management. The next subsection shows that the same kind of underestimation appears in realized volatility assessments based on ultra-high frequency data when neglecting the causality factor $1+\beta^*$.

3.3 Causality effect in realized variance

Consider the situation that one is interested in the realized variance on the stock S. Relation (3.10) can be used to assess the effect of the causality effect we document. More precisely, rewrite (3.10) as

$$\begin{aligned} R^2_{t_i:t_{i+1}} &= (1+\beta^*)\psi_{t_i}\sigma^2_{t_i;G} + \tilde{e}_{t_i:t_{i+1}} \\ &= (1+\beta^*)\,(\Delta t_{i+1})\,\sigma^2_{t_i;G} + e_{t_i:t_{i+1}}, \end{aligned}$$

where both $\tilde{e}_{t_i:t_{i+1}}$ and $e_{t_i:t_{i+1}}$ are conditionally zero-mean errors. In particular, we may write $e_{t_i:t_{i+1}} = M(t_{i+1}) - M(t_i)$ for some martingale M.

Let us now assume that one wants to compute daily realized variances from transaction data or from all the observed quote revisions, i.e., from all returns observed over the intervals $(t_i, t_{i+1}]$. Such a realized variance for day h would be computed as

$$V(h) = \sum_{i=i_{min}(h)}^{i_{max}(h)} R^2_{t_i:t_{i+1}} = (1+\beta^*) \sum_{i=i_{min}(h)}^{i_{max}(h)} \left\{\sigma^2_{t_i;G}\Delta t_{i+1}\right\} + M(t_{i_{max}(h)}) - M(t_{i_{min}(h)}), \quad (3.12)$$

where $i_{max}(h)$ and $i_{min}(h)$ denote the index i of the last and first price observation on day h, respectively. Starting with independent and concurrent papers by Comte and Renault (1998), Andersen et al. (2001), and Barndorff-Nielsen and Shephard (2002), there is now a well established tradition to resort to the theory of quadratic variation of continuous time martingales to study the limit behavior of realized variance when the sampling frequency goes to infinity. With ultra-high frequency data (all the relevant data are recorded) we can a fortiori approximate the above sum by its Riemann integral counterpart to obtain

$$V(h) \approx (1+\beta^*) \int_{t=t_{i_{min}(h)}}^{t_{i_{max}(h)}} \left\{\sigma_{t;G}^2 \mathrm{d}t\right\} + M(t_{i_{max}(h)}) - M(t_{i_{min}(h)}). \tag{3.13}$$

Assuming a stationary volatility process, a natural thing to do would be to assess, for instance, mean daily volatility from a sample of H days as

$$V = \frac{1}{H} \sum_{h=1}^{H} V(h) \approx (1+\beta^*) \, \mathrm{E}\left\{\sigma_{t;G}^2\right\} L, \tag{3.14}$$

where the martingale M is assumed to vanish by a law of large numbers argument and L denotes the length of the trading day. In order to discuss the practical consequences of (3.14), several remarks are important. First, the above argument neglects the microstructure noise which is known to have a huge effect with high frequency data. An empirically sound procedure which would take into account both the microstructure noise and the causality effect in realized variances is beyond the scope of this paper. While our empirical study in Section 5 focuses on conditional volatility per trade, and accommodates microstructure noise, we just want to isolate here the likely bias in realized variance due to the causality effect. Secondly, using the approximation (3.11) of (3.10), formula (3.14) basically tells us that, on average, daily realized variances are going to give us an assessment of $(1+\beta^*)$ times the (daily) integrated volatility and not, as it is commonly believed, the integrated volatility itself. Actually, the relative order of magnitude of underestimation of volatility that is implied by overlooking the causality effect is exactly the same as the one documented in the previous section. The intuition behind this intriguing result is the following. When computing realized variances with ultra-high frequency data, the period of highest volatility are precisely the shortest ones and thus, the corresponding squared returns are multiplied by relatively short durations. It is then not surprising that we end up with an underestimation of quadratic variation while an exogenous sampling technique would have provided a closer approximation of it.

It is worth stressing, however, that genuine exogenous sampling is unfeasible in practice precisely because durations between quote revisions are never infinitely small. In other words, we claim that the causality bias is likely to be much more significant than the bias due to approximation of quadratic variation by realized variance on small deterministic intervals. A simple example may illustrate this point even further. Let $\left(\sigma_i^2, \Delta t_i\right)_{i=1}^n$ be i.i.d. from some distribution over the positive quadrant and assume that the return R_i, conditional on the volatility σ_i and duration Δt_i, has variance $\sigma_i^2 \Delta t_i$. In this setup volatility is i.i.d. and, in particular, the expectation of integrated volatility over a deterministic interval of length L is $L \mathrm{E} \sigma_i^2$. Note that $\mathrm{E}\sigma_i^2$ is not the unconditional expectation (per unit of time) of a squared return over a time interval between quotes since the latter is deduced from:

$$\mathrm{E}\left\{R_i^2\right\} = \mathrm{E}\left\{\sigma_i^2 \Delta t_i\right\} = \mathrm{E}\left\{\sigma_i^2\right\} \mathrm{E}\left\{\Delta t_i\right\} + \mathrm{Cov}\left\{\sigma_i^2, \Delta t_i\right\}. \tag{3.15}$$

The causality effects documented in this paper typically leads us to expect a negative covariance term in (3.15), such that the unconditional expectation of squared returns per unit of time underestimates the average (squared) volatility $\mathrm{E}\sigma_i^2$. It is important to realize that computing the expected realized variance for n subsequent returns will have no impact on this underestimation, irrespective of the size of n. Even though this discrete time setting prevents us from a clear disentangling of Granger causality and instantaneous causality effects, this negative bias of integrated volatility computed on tick by tick data is tightly related to the one put forward in (3.14). Consider what happens if realized variance is calculated over long, say five minute, deterministic intervals. Effectively, this leads to

$$\mathrm{E}\left\{\sum_{t=1}^{T}\left(\sum_{i \in I_t} R_i\right)^2\right\},$$

where I_t indicates the set of returns that fall in the t-th deterministic interval and T is the number of deterministic intervals within the day. It's easy to see that, in our example of independent returns and using the law of iterated expectations, this expected realized variance equals

$$\sum_{t=1}^{T} \mathrm{E}\left\{\sum_{i \in I_t} \sigma_i^2 \Delta t_i\right\}. \tag{3.16}$$

Note that T is not random and thus the outer sum and the expectation in (3.16) may be interchanged. However, due to the randomness of the set I_t, one cannot conclude that (3.16) corresponds to n times (3.15) where n is the total number of observed returns, i.e., the sum of all the numbers of elements of the sets I_t for $t = 1, \ldots, T$, as n itself is random. However, for T sufficiently large, one may expect that this makes a little difference.

Exact calculations are possible if we assume that Δt_i is exponentially distributed with mean $1/\lambda$ seconds and $\mathrm{E}\{\sigma_i^2 | \Delta t_i\} = a - b\left(\Delta t_i - 1/\lambda\right)$. Clearly, such linear specification allows for the theoretical possibility of negative variances in the empirically relevant case that $b > 0$, but the reader will easily be convinced that this does not affect the essential part of the argument that follows. Consider the first interval I_1 and denote the number of times that belong to this interval by N. For the sake of argument, add a $N+1$-st time $t_{N+1}=L$ so that we have exactly $\sum_{i=1}^{N+1}\Delta t_i = L$, where L denotes the length of the interval in seconds, e.g., $L = 300$ for five minute intervals. Clearly, N is Poisson distributed with mean λL. We now find, for each term in (3.16),

$$\begin{aligned}
\mathrm{E}\left\{\sum_{i\in I_t}\sigma_i^2\Delta t_i\right\} &= \mathrm{E}\left\{\sum_{i=1}^{N+1}\sigma_i^2\Delta t_i\right\} \\
&= \mathrm{E}\left\{\sum_{i=1}^{N+1}\left(a-b\left(\Delta t_i-\frac{1}{\lambda}\right)\right)\Delta t_i\right\} \\
&= \left(a+\frac{b}{\lambda}\right)L - b\mathrm{E}\left\{\sum_{i=1}^{N+1}(\Delta t_i)^2\right\}. \qquad (3.17)
\end{aligned}$$

It is well-known that, conditionally on the value of N, the times t_i are uniformly distributed over the interval $[0, L]$. The durations Δt_i are thus the spacings between uniform order statistics over this interval. The expectation of the sum of the squared spacings in (3.17), still conditionally on N, equals $2L^2/(N+2)$. This latter result follows directly from the observation (attributed by Greenwood, 1946, to Whitworth, 1898), that, even for all positive integers m, the expectation of each of the terms in the expansion of $\left(\sum_{i=1}^{N+1}\Delta t_i\right)^m$ is the same. Since, for $m = 2$, there are $(N+1)(N+2)/2$ such terms, the result follows. Consequently, we have

$$\mathrm{E}\left\{\sum_{i=1}^{N+1}\sigma_i^2\Delta t_i\,\middle|\, N\right\} = \left(a+\frac{b}{\lambda}\right)L - 2b\frac{L^2}{N+2}.$$

Calculating the expectation of $1/(N+2)$, we find

$$\mathrm{E}\left\{\sum_{i=1}^{N+1}\sigma_i^2\Delta t_i\right\} = aL - b\left(\frac{L}{\lambda}+\frac{2}{\lambda^2}\left(\exp(-\lambda L)-1\right)\right) = aL - b\frac{L}{\lambda} + o(L), \qquad (3.18)$$

for large L. The bias induced by the causality parameter b is thus of the same order of magnitude as the first term in (3.18) or (3.15). In particular, this bias does not disappear when using longer intervals L. In other words, even seemingly deterministic intervals do not get rid of the causality bias. The point is that since we are always dependent on dates of quotes for observing prices, in periods of high volatility, durations tend to be small and thus the number of quotations observed within a deterministic interval is large. Actually, by using these seemingly deterministic intervals, we have just reduced the amount of underestimation by the term $2b\left(1-\exp(-\lambda L)\right)/\lambda^2$ but, for given finite λ, this is of little importance since for L large, the bias reduction $2b/\lambda^2$ will be negligible compared to the inherent bias bL/λ.

Of course, if durations converge to zero, i.e., $\lambda \to \infty$ in this simple example, all the bias terms disappear. At the limit, if it were really the case that $\max \Delta t_{i+1}$ converges to zero, we would be back to the standard asymptotic theory of quadratic variation, irrespective of the kind of sampling:

deterministic or random. This is in line with recent results of Barndorff-Nielsen et al. (2005). However, our empirical results imply that the causality effect should not be ignored, even for the most liquidly traded stocks.

4 Explicit moment conditions

The previous section showed that, for our purposes, an ACD type duration model is convenient (Assumptions D and E). In this section we specialize further to the actual model we use in the empirical analysis of Section 5. In particular, we show how we take into account Granger type causality effects from durations to volatilities, market microstructure noise, possibly non-zero expected returns, and intraday seasonality.

4.1 Granger causality effects

Condition (3.11) paves the way for feasible GMM inference insofar as the product $\sigma^2_{t_i}\psi_{t_i}$ can be related to a conditional expectation of a known function of observables, that is returns and durations. We choose to avoid any explicit specification of both $\sigma^2_{t_i}$ and ψ_{t_i} in terms of observed past durations and quotes, but instead impose merely

$$\mathrm{E}\{\sigma^2_{t_i} \,|\, \psi_{t_i}; \Delta t_i, \Delta t_{i-1}, \Delta t_{i-2}, \ldots\} = \alpha_0 + \alpha_1 \psi_{t_i}. \tag{4.1}$$

In this specification, α_1 may be interpreted as a regression coefficient measuring the sensitivity of instantaneous volatility with respect to expected duration. Given that larger volatility usually goes together with more expected quote revisions, i.e., smaller expected durations, we expect α_1 to be negative. However, note that this volatility-expected duration relationship has nothing to do with the instantaneous causality effect between volatility and durations as measured by β^*. While the former will generate a kind of Granger causality effect from past durations to current volatility (see, e.g., Dufour and Engle, 2000), the latter relates instantaneously surprises in durations to surprises in volatility. Relation (4.1) could be extended to a quadratic or even higher order polynomial specification. Our empirical investigations show that a linear specification suffices, at least for the stocks we study over our sample period. Obviously, in any such specification, care has to be taken that the right-hand side of (4.1) remains positive over the relevant domain of ψ_{t_i}.

Note that specifying a relation between $\sigma^2_{t_i}$ and ψ_{t_i} directly, as in (4.1), avoids the need to specify how each of them individually depends on past observables. This is an additional benefit of our approach. As a result, our conclusions are not driven by any possible misspecification that could occur when writing explicitly a parametric model for $\sigma^2_{t_i}$ and/or ψ_{t_i}. In particular, Granger causality from returns to durations, making ψ_{t_i} dependent on past returns, is not ruled out in our analysis.

From (4.1) we deduce

$$\begin{aligned}
\mathrm{E}\{\,\sigma_{t_i}^2\psi_{t_i}\big|\,\Delta t_i,\Delta t_{i-1},\Delta t_{i-2},\ldots\} &= \mathrm{E}\{\,\alpha_0\psi_{t_i}+\alpha_1\psi_{t_i}^2\big|\,\Delta t_i,\Delta t_{i-1},\Delta t_{i-2},\ldots\}\\
&= \mathrm{E}\{\,\alpha_0\Delta t_{i+1}+\alpha_1\varphi\,(\Delta t_{i+1})^2\Big|\,\Delta t_i,\Delta t_{i-1},\Delta t_{i-2},\ldots\},
\end{aligned}$$

where φ is the dispersion parameter previously introduced in (3.9).

4.2 Market microstructure noise

There exists currently a large body of literature documenting that prices observed at high frequencies are contaminated with market microstructure noise. We follow Bandi and Russell (2004) and Zhang, Mykland, and Aït-Sahalia (2004) and impose an independent market microstructure noise which effectively adds $2\sigma_{mms}^2$ to the variance of observed returns.

As will be explained later (Section 4.5), our empirical analysis uses both returns over random durations and over (long) deterministic durations. In order to allow for possible correlation in the microstructure noise for subsequently observed midquotes, we will actually estimate two separate variances of market microstructure noise. The first ($\sigma_{mms,1}^2$) will, like in the current literature, capture the noise in midquotes that are far apart in time. The second, $\sigma_{mms,2}^2$ will be used for subsequent midquotes and thus, implicitly, allows for correlation in the corresponding market microstructure noises.

4.3 Expected returns

So far we have ignored a possible drift in the price process. If a general semi-martingale model for the price process were considered, returns over the interval $(t_i, t_{i+1}]$ would be given by

$$\begin{aligned}
R_{t_i:t_{i+1}} &= \log\frac{S_{t_{i+1}}}{S_{t_i}}\\
&= \int_0^{\Delta t_{i+1}}\mu_{t_i+u}\mathrm{d}u+\int_0^{\Delta t_{i+1}}\sigma_{t_i+u}\mathrm{d}L_{t_i+u},
\end{aligned}$$

where μ_{t_i+u} denotes the drift of the log-price process. Of course, any source of randomness in the drift term possibly introduces other causality relationships with the times t_i. For instance, a risk premium related to $\sigma_{t_i}^2$ would introduce causality in higher order moments. For simplicity, we restrict attention to the case of a constant drift, i.e.,

$$R_{t_i:t_{i+1}}=\mu\Delta t_{i+1}+\int_0^{\Delta t_{i+1}}\sigma_{t_i+u}\mathrm{d}L_{t_i+u}. \tag{4.2}$$

Given the assumptions made, this representation leads immediately to the moment condition

$$\mathrm{E}_{t_i}\left\{R_{t_i:t_{i+1}} - \mu\Delta t_{i+1}\right\} = 0. \tag{4.3}$$

Concerning the conditional variance of returns between quote revisions, the occurrence of a non-zero drift may complicate our fundamental variance decomposition in Proposition 2.1. Additional causality effects between revision times and the Lévy process L may come into play. These effects can, for instance, be excluded by assuming that L remains a Lévy process with respect to the extended filtration $\mathcal{F}_t^* = \mathcal{F}_t \vee \sigma\left(\Delta t_{n_t+1}\right)$, where n_t denotes the index of the first quote revision after time t. Under such condition, both terms in (4.2) are uncorrelated and we obtain

$$\mathrm{Var}_{t_i}\left\{R_{t_i:t_{i+1}}\right\} = \mu^2 \mathrm{Var}_{t_i}\left\{\Delta t_{i+1}\right\} + \mathrm{Var}_{t_i}\left\{\int_0^{\Delta t_{i+1}} \sigma_{t_i+u} \mathrm{d}L_{t_i+u}\right\},$$

which, using (3.11), immediately leads to a moment condition. Note, that these latter considerations become void in case a drift would be absent as is often assumed in the empirical market microstructure literature.

4.4 Seasonality

There is abundant evidence that tick-by-tick market data exhibits strong seasonality patterns over the day. As a result, it is unlikely that the moment conditions derived above hold throughout the day with identical parameter values. We address this problem by estimating the parameters for each stock and for each 15 minute interval over the day separately. Given the large number of observations available, we still have at least $10,000$ observations in each single GMM estimation. In the empirical section, we report precision weighted (using the estimated standard errors) average parameter estimates for conciseness. In line with the literature, we do not use the first 15 minute interval of the day (9:30am-9:45am) in this average.

4.5 Summary of moment conditions

Summarizing, we use the following moment conditions in Section 5:

$$\mathrm{E}\left\{ R_{t_i:t_{i+1}} - \mu\Delta t_{i+1} \middle| \Delta t_i, \Delta t_{i-1}, \ldots\right\} = 0, \tag{4.4}$$

$$\mathrm{E}\left\{ R^2_{t_i:t_{i+1}} - \mu^2\varphi\left(\Delta t_{i+1}\right)^2 - \left(1+\beta^*\right)\left[\alpha_0\Delta t_{i+1} + \alpha_1\varphi\left(\Delta t_{i+1}\right)^2\right] - 2\sigma^2_{mms,2} \middle| \Delta t_i, \Delta t_{i-1}, \ldots\right\} = 0. \tag{4.5}$$

Not surprisingly, the two causality parameters α_1 and β^* cannot be identified separately from (4.4) and (4.5). The two sensitivity factors, α_1 (for Granger causality) and $1 + \beta^*$ (for instantaneous causality) play multiplicative roles in the moment condition (4.5). But this problem can be easily resolved by adding extra identifying moment restrictions based on *deterministic* duration intervals. By definition, the instantaneous causality effect is no longer at stake when observing returns over fixed time intervals of length h. Proposition 2.1 applied to deterministic durations of length h leads straightforwardly to the additional moment conditions

$$\mathrm{E}\left\{ R_{t_i:t_i+h} - \mu h \middle| \Delta t_i, \Delta t_{i-1}, \ldots \right\} = 0, \tag{4.6}$$

$$\mathrm{E}\left\{ R^2_{t_i:t_i+h} - \mu^2 h^2 - \alpha_0 h - \alpha_1 h \Delta t_{i+1} - 2\sigma^2_{mms,1} \middle| \Delta t_i, \Delta t_{i-1}, \ldots \right\} = 0. \tag{4.7}$$

In the empirical section we actually use two deterministic duration intervals. More precisely, we use intervals which are, respectively, 25 and 50 times the average observed duration for a stock. Once more, the use of different parameters for the variance of market microstructure noise in (4.5) and (4.7) allows for the possibility that this noise, at the quote-revision frequency, admits some serial correlation. Such correlation would affect the total variance of the microstructure noise for observed returns.

To conclude our discussion of the estimation procedure we follow, we note that, using standard GMM practice, the conditional moment conditions derived above are transformed into unconditional ones using instruments. Given (4.1), valid instruments are past durations and functions thereof. As it is well-known that durations are strongly autocorrelated, past durations can be expected to be informative for the parameters of interest. Therefore, besides the constant, we use Δt_i and $(\Delta t_i)^2$ as instruments. We use the standard optimal weighting matrix for weighting the unconditional moment conditions. The use of both returns over random durations and deterministic intervals of length h induces a overlapping samples problem, since clearly $R_{t_i:t_{i+1}} = R_{t_i:t_i+\Delta t_{i+1}}$ and $R_{t_i:t_i+h}$ are correlated. To resolve this problem, we estimate the variance of the unconditional moment conditions using a Newey-West estimator with a fixed number of lags. The number of lags is fixed at 100, which, given the fact that h is either 25 or 50 times the average duration, is more than enough.

5 Empirical evidence for duration/volatility causality

The theoretical results derived in the previous sections are equally valid whether the times refer to transactions or quote revisions. In the empirical analysis of this paper we, however, only deal with the latter. In order to assess the economic and statistical relevance of possible instantaneous causality between quote durations and volatility, we estimate the causality parameter β^* as introduced in Section 3, for ten liquidly traded stocks at the NYSE. We first discuss, in Section 5.1, the ten stocks

	DDS	FD	IBM	JCP	MAT	MAY	MCD	SKS	SLB	WMT
Observations	328167	354205	657906	413551	405697	442760	588747	291770	521279	676793
Average dur.	4.3	4.0	2.2	3.4	3.5	3.2	2.4	4.9	2.7	2.1
Stand.dev. dur.	6.9	6.8	2.5	5.0	4.7	4.6	2.9	7.4	4.5	2.5
Average ret.	-0.0	0.0	-0.0	0.0	0.0	-0.0	0.0	0.0	0.0	-0.0
Stand.dev. ret.	2.8	1.5	0.8	1.5	2.1	2.0	1.2	3.1	1.2	0.9

Table 1: Summary statistics for durations and returns for ten stocks from the TAQ database January 3, 2005, until March 31, 2005. The rows of the table present, from top to bottom, the number of observations, the average duration between quote revisions, the standard deviation of durations, the average return between quote revisions, and the standard deviation. All durations are measured in seconds (*sec*) and returns in basis points (*bp*).

that we analyse and, subsequently, in Section 5.2 we show that, at least for these stocks and the time period we study, instantaneous causality effects from durations to volatilities are statistically and economically significant.

5.1 Data description

We consider ten randomly selected liquidly traded stocks at NYSE. We use data on quotes from the TAQ dataset for 61 days from January 3, 2005, until March 31, 2005. The times we consider are those where the midquote for a stock changes. We define the midquote as the geometric average of the bid and ask price. The ten stocks we use, with ticker symbol in parentheses, are Dillard's (DDS), Federated (FD), IBM (IBM), JCPenney (JCP), Mattel (MAT), May (MAY), McDonald's (MCD), Saks (SKS), Schlumberger (SLB), and Walmart (WMT). We remove zero durations, i.e., durations where the midquote does not change from one observation to the next. Moreover, we replace returns above 100 basis points (in absolute value) by the average return. The latter only affected three out of the ten stocks for in total 41 observations. It is important to note that we performed no other data cleaning. In particular we did not seasonally adjust the data in any way. The reason for this is that it is not clear how such an adjustment would interfere with the causality effects we are interested in. As discussed in Section 4.4 we estimate our model on 15 minute intervals to take possible seasonality effects into account. Throughout, durations are measured in seconds (*sec*) and returns in basis points (*bp*). Summary statistics are in Table 1.

The first row in Table 1 gives, for each of the ten stocks, the number of observations that are available in the estimation. For a fairly illiquid stock like Saks (SKS), we still have almost $300,000$

observations available. For the most liquid stocks (IBM and WMT), we have twice as many. The difference in liquidity also follows from the second row, that gives the average duration (in seconds) between subsequent quote revisions for each stock, ranging from 2.1 seconds for WMT to 4.9 seconds for SKS. The standard deviation of durations is always above the average, which shows unconditional excess dispersion with respect to the exponential distribution. Finally, we present the average and standard deviation of returns. Note that there is a clear positive relationship between average durations and standard deviations of returns, due to, in particular, the time-to-build effect.

5.2 Empirical results

We present estimation results on the causality effects of interest in this paper using the moment conditions detailed in Section 4 and the ten stocks described above. The estimation results are in Table 2. Recall that, in order to account for possible intraday seasonality, the estimates presented are averages over GMM estimates obtained for subsequent 15 minute intervals. The parameter α_0 determines the level of the instantaneous variance. Given the average durations in Table 1 and the estimated values for α_1, we can easily derive the average level of the instantaneous variance for each of the ten stocks, ignoring market microstructure noise. Focusing on IBM, we would find $0.18 - 0.0152 \times 2.2 = 0.15$ bp^2/sec. The parameter α_1 is estimated significantly[1] negative in all cases. Recall that α_1 measures the relation between instantaneous volatility and expected durations. Consequently, a higher instantaneous volatility indeed goes together with smaller expected durations. Note, moreover, that the estimates are such that expected variances in (4.1) remain positive over the relevant domain of expected durations for all stocks.

However, the key interest in the present paper is instantaneous causality between future volatilities and surprises in durations as measured by β^*. This parameter is estimated significantly negative for all stocks in our sample. Figures 3.1 and 3.2 in Section 3 are based on the estimates for IBM ($\beta^* = -0.65$), considering, for illustrative purposes only, the typical case of exponentially distributed durations[2] and a constant function β. Using (3.7) and since for the exponential distribution $\int F(v)[1 - F(v)]\mathrm{d}v = 1/2$, we find that the function β is equal in size to about $2\beta^*$ Now consider the event that, after waiting the (conditional) median duration, we have not seen the next quote yet. Then, according to (3.3), we should update our current instantaneous variance prediction with $\beta/2 = \beta^* = -0.65$, i.e., a 65% decrease in variances and a corresponding $1 - \sqrt{1 + \beta^*} = 41\%$ decrease in volatility. For each of the individual stocks, the row "underestimation" in Table 2 gives

[1] All statements about statistical significance in this paper are at a level of 1%.

[2] Formally, the exponential distribution for durations is ruled out in our setup as the support is not bounded. We ignore this in the present empirical section.

	DDS		FD		IBM		JCP		MAT	
Parameter	est.	t-val	est.	t-val	est.	t-val	est.	t-val	est.	t-val
μ (%)	-0.01	-0.22	-0.05	-1.03	-0.01	-0.18	0.00	0.04	-0.00	-0.01
α_0	0.76	26.07	0.58	29.82	0.18	32.42	0.51	27.09	0.49	25.45
α_1 (%)	-3.82	-11.39	-3.19	-11.47	-1.52	-9.65	-2.78	-7.77	-2.67	-8.17
β^*	-0.94	-5.75	-0.93	-9.77	-0.65	-2.90	-0.88	-7.13	-0.88	-4.74
φ	1.06	1.98	0.71	2.99	0.46	1.96	0.32	2.00	0.68	4.69
$\sigma^2_{mms,1}$	1.21	1.56	0.99	1.67	1.00	10.80	0.16	0.37	3.43	6.90
$\sigma^2_{mms,2}$	1.73	9.90	0.60	7.66	0.19	5.63	0.65	7.25	1.21	8.63
p-value	0.38		0.32		0.70		0.44		0.40	
underestimation	0.24		0.26		0.59		0.35		0.35	
	MAY		MCD		SKS		SLB		WMT	
μ (%)	-0.01	-0.22	-0.05	-1.03	-0.01	-0.18	0.00	0.04	-0.00	-0.01
α_0	0.73	26.22	0.40	33.42	0.74	22.84	0.58	34.66	0.24	32.83
α_1 (%)	-7.76	-15.25	-2.69	-9.19	-3.50	-11.58	-1.59	-4.12	-1.63	-6.71
β^*	-0.88	-7.22	-0.54	-2.59	-0.92	-5.19	-0.86	-4.59	-0.63	-2.63
φ	0.56	5.38	0.77	6.00	0.65	6.04	0.49	0.80	0.45	2.43
$\sigma^2_{mms,1}$	1.54	2.92	0.90	4.75	2.92	2.93	0.51	1.39	0.84	8.02
$\sigma^2_{mms,2}$	1.17	8.99	0.43	6.32	3.01	8.78	0.37	3.31	0.22	5.06
p-value	0.52		0.65		0.37		0.44		0.73	
underestimation	0.34		0.68		0.28		0.37		0.61	

Table 2: Point estimates and t-values for the expected return (μ), the relation between instantaneous volatility and expected durations (α_0 and α_1), the instantaneous causality parameter (β^*), the duration dispersion parameter (φ), and the variances of market microstructure noise ($\sigma^2_{mms;1}$ referring to deterministic time intervals and $\sigma^2_{mms;2}$ referring to subsequent quote revisions). All estimates presented are precision weighted averages over 25 independent 15 minute intervals per trading day. The last two lines in each panel present, respectively, the average p-values of the GMM J-test for overidentifying restrictions and the relative volatility underestimation due to not taking into account the instantaneous causality ($\sqrt{1+\beta^*}$). See main text for details.

the relative underestimation in instantaneous volatility due to ignoring the documented causality effect as discussed in Section 3.2. In all cases we find an economically significant effect, with some variation for the individual stocks.

Observe that, in line with the intuition that the causality effect disappears at higher frequencies, there is a strong positive (rank)correlation between the estimated β^* for each stock and the liquidity, as measured, e.g., by the average duration. This observation confirms the aforementioned intuition. However, we stress that even for the currently most liquidly traded stocks (IBM and WMT) the causality effect is far from negligible. These empirical results are consistent with those in Engle and Sun (2005). They specify the conditional variance per trade proportional, given $\mathcal{F}_{t_i}$, to $(\Delta t_{i+1})^\delta$ and find empirically $\delta < 1$ and even smaller for the least liquid stocks. In other words, the variance per unit of time is decreasing with the corresponding duration (as duration to the power $\delta - 1$) and the causality effect is even stronger for the least liquid stocks.

The rows σ^2_{mms} in Table 2 provide estimates of the variance of market microstructure noise. As mentioned before, we allow for the possibility that market microstructure noise for consecutively observed midquotes is correlated. As a result, we present two variance estimates. $\sigma^2_{mms,1}$ refers to the variance of market microstructure noise for returns measured over long intervals, i.e., the deterministic intervals we use in the estimation. The parameter $\sigma^2_{mms,2}$ refers to market microstructure noise in quote-to-quote prices. Observe that this estimate is for some stocks smaller than the estimate for long duration returns. In these cases, the results indicate a negative correlation in quote-to-quote microstructure noise. These differences are however not statistically significant.

We also calculated p-values for the standard GMM J-test for overidentifying restrictions. Given the six moments we use, the three instruments we have, and the seven parameters we estimate, for each 15 minute interval, the test distribution has eleven degrees of freedom. We present the average p-values in the table and remark that for all the individual test, only a single rejections occurs in the whole sample of all ten stocks. Moreover, assuming that the individual 15 minute interval estimates are independent, such that the individual J-tests can be combined to a single one, our specification is rejected for none of the ten stocks under consideration. Clearly, this J-test has a large number of degrees of freedom and is, therefore, not reported.

Finally, let us consider the parameter φ which measures the dispersion of the rescaled (by their conditional expectation) durations. For the exponential distribution, we have $\varphi = 1/2$. The results for the ten stocks we study vary in this respect, leading to the conclusion that some stocks exhibit some overdispersion and others might exhibit underdispersion for the conditional duration distribution. These effects are, however, never statistically significant.

To the best of our knowledge, the present paper is the first one that specifically addresses empirically the origin of observed dependencies between durations and volatility. Reduced form VAR-models do not allow for disentangling dependencies between expected durations and current instantaneous volatility on the one hand, and surprises in durations and in future instantaneous volatility on the other hand. As mentioned before, the approach of Grammig and Wellner (2002) implicitly imposes that all dependence takes place through the relation between expected durations and instantaneous volatility. We confirm this effect, but find in addition that exogenous news events apparently drive both durations and volatility.

6 Implications of causality effects for modeling and estimating price processes

Following Engle (2000), the general statistical issue we have to address is inference about a marked point process. The so-called marks describe the actual event that occurs at time t_i and consist of a k-vector y_i at this time. Engle (2000) states that "the relevant economic questions can all be determined" from the densities:

$$p\left(y_{i+1}, \Delta t_{i+1} | \mathcal{G}_{t_i}\right) = p\left(y_{i+1} | \Delta t_{i+1}, \mathcal{G}_{t_i}\right) p\left(\Delta t_{i+1} | \mathcal{G}_{t_i}\right), \tag{6.1}$$

which decomposes the joint conditional density of $(y_{i+1}, \Delta t_{i+1})$ given the natural past in discrete time, i.e., given $\mathcal{G}_{t_i} = \sigma(y_j, \Delta t_j : j \leq i)$.

The focus of interest in the present paper has been the economic interpretation of the occurrence of the current duration Δt_{i+1} in the function $p(y_{i+1}|\Delta t_{i+1}, \mathcal{G}_{t_i})$. We have considered exclusively the effect of durations on prices, i.e., $y_i = S_{t_i}$ is the price at time t_i. In particular, we have focused on the effect of durations on volatilities. However, the results of this paper could be extended to other marks, e.g., volume traded at time t_i or cross-stock effects.

A consequence of our analysis is that the influence of durations on prices, i.e., the occurrence of Δt_{i+1} in $p(S_{t_{i+1}}|\Delta t_{i+1}, \mathcal{G}_{t_i})$, is twofold and should be split, in an identifiable way, into a temporal aggregation effect and an informational effect. Since both effects have different repercussions for risk measurement and management, this separate identification has important consequences.

We have shown in a previous paper (Meddahi, Renault, and Werker, 2003), that, even if the time sequence Δt_i , $i = 1, \ldots, n$, were purely deterministic or strongly exogenous, the current duration Δt_i would explicitly appear in the model $p(S_{t_{i+1}}|\Delta t_{i+1}, \mathcal{G}_{t_i})$ of the price dynamics, simply through a "time-to-build" effect in volatility fluctuations. This dependence is caused by two effects. On the one hand, the application of a standard discrete time volatility model in itself must consider the "volatility per unit of time", as in Engle (2000) in the context of GARCH modeling. On the other hand, the volatility clustering effect is likely to be erased by longer durations and, therefore, the model of volatility persistence must be conformable to temporal aggregation formulas (see, e.g., Drost and Werker, 1996, Ghysels and Jasiak, 1998, or Grammig and Wellner, 2002, for proposals to apply the Drost and Nijman, 1993, formulas of temporal aggregation of weak GARCH processes). The exact formulas taking both into account are rigorously derived in Meddahi, Renault, and Werker (2003) using the Meddahi and Renault (2004) formulas for temporal aggregation of continuous time linear autoregressive volatility dynamics. Without the continuous time paradigm, the application of temporal aggregation formulas with random times has to be justified by resorting to something like a latent "normal duration GARCH process" (Grammig and Wellner, 2002) whose structural foundations are not clear.

But in addition to these deterministic effects of irregular time sampling, an even more interesting issue is to see the time between trades as a measure of trading activity which could affect price behavior. This is the reason why the economic interpretation of the informational content of times, in models of price and trade dynamics, is better founded by identifying a structural continuous time model. Actually, only such a continuous time model will be able to disentangle what we have called the time-to-build effect from the genuine information effect. Typically, this structural model specifies the joint probability distribution of the price process S_t over some reference period $[0, T]$ as well as a sequence of stopping times t_i, $i = 1, \ldots, n$, over the same period. The marginal probability distribution of the price process provides, for any (fixed and deterministic) time interval h, the density function $p_h(S_{t_i+h}|\mathcal{G}_{t_i})$ of the conditional distribution of S_{t_i+h} given the natural past $\mathcal{G}_{t_i}$. Then, the economic issue of interest is the validity of the condition:

$$p_{\Delta_{t_{i+1}}}\left(S_{t_{i+1}}|\mathcal{G}_{t_i}\right) = p\left(S_{t_{i+1}}|\Delta t_{i+1}, \mathcal{G}_{t_i}\right). \tag{6.2}$$

When this equality is fulfilled, and under the additional assumption that the marginal process describing the relevant times does not contain information about the structural parameters in the price dynamics, the times contain no genuine information regarding these asset price dynamics and there is no cost when these times are considered to be deterministic, still taking into account that they are irregularly spaced. Aït-Sahalia and Mykland (2003) studies the full information maximum likelihood under the maintained assumption (6.2). They also document the fact that there is, of course, an efficiency loss when one decides to integrate out the likelihood with respect to the random durations and, even worse, a misspecification bias if one incorrectly supposes that durations are fixed (i.e., $\Delta t_{i+1} = \bar{\Delta}$ for all i).

But if, on the contrary, some instantaneous causality relationship between durations and asset prices leads to a violation of (6.2), the incremental information content of Δt_{i+1} about $S_{t_{i+1}}$ given the past $\mathcal{G}_{t_i}$ is crucial for statistical inference. Typically, when the observed values S_{t_i} are plugged into a likelihood function based on the densities $p_{\Delta t_{i+1}}\left(S_{t_{i+1}}|\mathcal{G}_{t_i}\right)$ as if the times t_i were deterministic, one would introduce some kind of selection bias which may be significant. For the purpose of statistical inference about the continuous time price processes, the contribution of this paper is to provide a semiparametric specification test to decide whether the noncausality assumption (6.2) is satisfied. The answer, as we have seen, is negative.

The incremental information of the current duration Δt_{i+1} in the function $p\left(S_{t_{i+1}}|\Delta t_{i+1}, \mathcal{G}_{t_i}\right)$, in excess of the deterministic time-to-build effect, is typically neglected in the current literature. The ACD-GARCH model as proposed by Ghysels and Jasiak (1998) or Grammig and Wellner (2002) uses the temporal aggregation formulas for weak GARCH processes as derived by Drost and Nijman (1993) with time-varying aggregation period (expected duration). This setup does not allow for a parameter taking into account instantaneous causality between durations and prices. For example, the volatility equation of Grammig and Wellner (2002), which just takes into account the temporal aggregation effect in a "normal duration GARCH process", implicitly assumes that this "normal" regime is not influenced by unexpected durations. In spite of its name ("interdependent duration-volatility model") the model of Grammig and Wellner (2002) cannot capture any instantaneous causality relationship between volatility and duration since both the volatility equation and the duration equation are only about conditionally expected squared returns and expected durations given the past.

This is the reason why the only discrete time model which can be compared with the moment restrictions that we derive from our continuous time structural model is the one of Engle (2000). In this model, according to (6.1), the conditional expectation of squared returns is computed given not only the past but also given the current duration. While volatility depends on past durations through the reciprocal of the past conditional expectation of the current duration, the dependence on the current duration goes not only through the reciprocal of the current duration but also through "surprises in durations", as measured by the relative difference between the current duration and its past conditional expectation. While the first of the three duration/volatility causality effects is typically a Granger one, the two others, and especially the last one, are more focused on instantaneous causality relationships. The general conclusion is that longer (shorter) durations lead to lower (higher) volatility. However, it is important to note that the instantaneous causality and the Granger causality relationships may play in opposite directions. We find that our continuous time structural model is useful for disentangling precisely the two causality effects, making tests of various microstructure models possible (e.g., those of Easley and O'Hara, 1992, or that of Admati and Pfeiderer, 1988). Actually, it allows to test without ambiguity the significance and the sign of an instantaneous causality relationship between duration and volatility, in the presence of, but separate from, possible Granger causality.

In addition it may also be argued that the GARCH assessment of causality between duration and volatility may be biased by a kind of filtering effect, due to latent stochastic volatility. Since our model is a stochastic volatility one, the information $\mathcal{F}_{t_i}$ that defines the conditioning in the risk measurement $\mathrm{Var}_{t_i}\{R_{t_i:t_{i+1}}\}$ does contain the current latent value σ_{t_i} of the spot volatility process. Then, if one wants to specify a GARCH type model that characterizes the dynamics of the conditional variance given the smaller information set defined only from the past observations of the asset price ($\mathcal{G}_{t_i}$), one has to reproject the above conditional variance on this smaller information set. If the current value Δt_{i+1} of the duration is added, as, e.g., in Engle (2000), to this smaller information set, it may have an informational content, just as way to better filter the past values of the volatility process. This informational content may occur even when the regression coefficient β is zero. This would be akin to some indirect Granger causality effect from durations to prices through volatility (see, e.g., Renault, Sekkat, and Szafarz, 1998) and does not correspond to the instantaneous causality relationship between duration and volatility. Of course, the empirical evidence documented by Engle (2000) is fairly convincing. The functional forms (39) and (40) in that paper are sufficiently specific to make it difficult to imagine that the significant role of the duration (Δt_{i+1}) is just a filtering effect. However, we do consider that, to fully disentangle the filtering effect from the instantaneous causality effect of interest, the stochastic volatility framework in continuous time is better suited.

Finally, a few remarks are in order about the specific way we characterized causality relationships between volatility and durations. This way was well-suited for designing a semiparametric test of noncausality but, of course, more would be needed for a parametric specification of causality within a maximum likelihood framework. To see this, note that our focus of interest has only been the causality property which makes β non-zero, that is

$$\mathrm{E}_{t_i}\left\{\sigma^2_{t_i+u}\middle|\,\Delta t_{i+1} > u\right\} \neq \mathrm{E}_{t_i}\left\{\sigma^2_{t_i+u}\right\}. \tag{6.3}$$

The equality

$$\mathrm{E}_{t_i}\left\{\sigma^2_{t_i+u}\middle|\,\Delta t_{i+1} > u\right\} = \mathrm{E}_{t_i}\left\{\sigma^2_{t_i+u}\right\}, \tag{6.4}$$

is actually a testable implication of the noncausality property:

$$\mathrm{E}_{t_i}\left\{\sigma^2_{t_i+u}\middle|\,\Delta t_{i+1}\right\} = \mathrm{E}_{t_i}\left\{\sigma^2_{t_i+u}\right\}. \tag{6.5}$$

Following the Florens and Fougère (1996) terminology (more precisely, their Definition 2.1, p. 1197), (6.5) means that the filtration $\mathcal{F}^*_t = \mathcal{F}_t \vee \sigma(\Delta t_{n_t+1})$ does not weakly globally cause the volatility process, given $\mathcal{F}_t$, where $n_t = \max(i: \; t_i \leq t)$ denotes the number of quote revisions up to time t. In more intuitive terms, the next time to come does not weakly (i.e., in expectation) cause the spot volatility process. Note that, given the absence of a drift function, (6.5) would imply also that $(\mathcal{F}_t^*)$ does not weakly instantaneously cause the price process given $(\mathcal{F}_t)$ in the Granger sense (Florens and Fougère, 1996, Definition 3.1., p. 1202), insofar as it does not cause the innovation process L in (2.1). Then, the price process remains a martingale with respect to the augmented filtration $(\mathcal{F}_t^*)$. If we knew more generally that the Doob-Meyer decomposition would not change for any $(\mathcal{R}_t)$-adapted special semimartingale, where $(\mathcal{R}_t)$ denotes the filtration generated by prices only, we would say (Florens and Fougère, 1996, Definition 3.2., p. 1203) that $(\mathcal{F}_t^*)$ does not strongly instantaneously cause the price process given $(\mathcal{F}_t)$ in the Granger sense. In this case, for any function of the price process, the Doob-Meyer decomposition is not modified by the knowledge of the next time. This strong instantaneous noncausality property in the Granger sense is obviously implied by the strong global noncausality property (Florens and Fougère, 1996, Definition 2.2., p. 1197):

$$\mathcal{F}_t^* \text{ and } \mathcal{R}_{t+h} \text{ are conditionally independent given } \mathcal{F}_t, \text{ for all } h > 0. \tag{6.6}$$

The converse is less clear. Theorem 3.1, p. 1203, in Florens and Fougère (1996) states that "strong global noncausality" and "strong instantaneous noncausality in the Granger sense" are equivalent when $\mathcal{F}_t = \mathcal{R}_t$, that is typically not our case since a stochastic volatility process has been added to the filtration $(\mathcal{R}_t)$ of past returns to define the filtration $(\mathcal{F}_t)$. The additional instantaneous causality effects in continuous time to consider to get strong global noncausality in the context of stochastic volatility are sketched in Comte and Renault (1996). The reason why strong global noncausality of times towards the price process is not guaranteed, even when strong instantaneous noncausality is, is that the Doob-Meyer decomposition of the volatility process itself might also be modified by the knowledge of the relevant times. Testing for this later causality effect is beyond the scope of the present paper.

7 Concluding remarks

The present paper considers a structural continuous time model for the analysis of instantaneous causality relations between price volatility and durations, in addition to possible Granger causality. We argue that these instantaneous causality effects are significant and that failure to take them into account may lead to severely biased volatility estimates and, consequently, possibly inadequate risk management.

We identify the instantaneous causality effects using appropriate moment conditions. These conditions (see Proposition 2.1) are sufficiently general to be applicable for a wide range of model specifications. The analysis does not yet take into account other relevant microstructure variables, like volume or information in other assets. Since our results for the variance of observed returns is based on a specification of volatility predictions given all current information (the function ξ_T in Assumption B), these could easily be included. Also, while we focus on an interpretation of t_i as quote revision times, this is not required in our main Proposition 2.1. As such, interesting empirical applications could include situations where transaction times are studied or cross-causality effects where surprises in durations for one stock, may cause instantaneous volatility in another stock.

Appendix: Proofs

Proof of Proposition 2.1: All references in this proof are to Protter (2003). We consider the conditional expectation of squared observed returns. Note that, under Assumption B, L is a square-integrable martingale and so is $\int_0^t \sigma_{t_i+u} \mathrm{d}L_{t_i+u}$ by applying the lemma on Page 171. Using Corollary 3 on Page 73 and Theorem II.29, we find

$$\mathrm{E}_{t_i} \left(\int_{u=0}^{t} \sigma_{t_i+u} \mathrm{d}L_{t_i+u} \right)^2 = \mathrm{E}_{t_i} \int_{u=0}^{t} \sigma_{t_i+u}^2 \mathrm{d}[L, L]_{t_i+u}.$$

The quadratic variation $[L, L]$ is obviously increasing and, thus, of integrable variation since $\mathrm{E}[L, L]_t = t < \infty$. Moreover, the compensator of this quadratic variation is time itself and, hence, Theorem III.16 implies

$$\mathrm{E}_{t_i} \int_{u=0}^{t} \sigma_{t_i+u}^2 \mathrm{d}[L, L]_{t_i+u} = \mathrm{E}_{t_i} \int_{u=0}^{t} \sigma_{t_i+u}^2 \mathrm{d}u.$$

Since Δt_{i+1} is a bounded stopping time for the filtration $(\mathcal{F}_{t_i+u} : u \geq 0)$, the optional sampling theorem (Theorem I.16) shows that the above arguments remain valid if we stop the martingales at $t = t_{i+1}$. This leads to

$$\mathrm{E}_{t_i} R_{t_i:t_{i+1}}^2 = \mathrm{E}_{t_i} \int_0^{\Delta t_{i+1}} \sigma_{t_i+u}^2 \mathrm{d}u.$$

Consequently,

$$\begin{aligned}
\mathrm{E}_{t_i} R_{t_i:t_{i+1}}^2 &= \mathrm{E}_{t_i} \int_0^\infty I_{(0,\Delta t_{i+1}]}(u) \sigma_{t_i+u}^2 \mathrm{d}u \\
&= \int_0^\infty \mathbb{P}_{t_i}\{\Delta t_{i+1} \geq u\} \xi_{t_i}(u) \mathrm{d}u + \int_0^\infty \mathrm{Cov}_{t_i}\{I_{(0,\Delta t_{i+1}]}(u), \sigma_{t_i+u}^2\} \mathrm{d}u \\
&= \int_0^\infty \Xi_{t_i}(u) \mathrm{d}F_{t_i}(u) + \int_0^\infty \beta_{t_i}(u) F_{t_i}(u)(1 - F_{t_i}(u)) \mathrm{d}u,
\end{aligned}$$

where the Fubini exchange in the second equality is allowed as the integrand is nonnegative and the expectation of the product is written as the product of the expectations and the covariance.

Proof of Proposition 3.2: Consider a bounded increasing function u. Observe that $\mathrm{E}_G\{u(V)\}$ is the ratio of two linear functions in β, i.e.,

$$\mathrm{E}_G\{u(V)\} = \frac{a + b\beta}{c + d\beta}.$$

The constants a, b, c, and d are given by

$$\begin{aligned}
a &= \int_{v=0}^{\infty} u(v)\,(1 - F(v))\,\mathrm{d}v, \\
b &= \int_{v=0}^{\infty} u(v)F(v)\,(1 - F(v))\,\mathrm{d}v, \\
c &= \int_{v=0}^{\infty} (1 - F(v))\,\mathrm{d}v = 1, \\
d &= \int_{v=0}^{\infty} F(v)\,(1 - F(v))\,\mathrm{d}v.
\end{aligned}$$

Note that $c + d\beta$ is positive for $\beta > -1$. Calculating the derivative with respect to β of $\mathrm{E}_G\{u(V)\}$, we find that its sign is determined by $bc - ad$. Moreover, since $c = 1$, the function $1 - F$ defines a density on the positive real line. Denoting expectations under this density by $\tilde{\mathrm{E}}$, we find that $bc - ad$ is actually a covariance:

$$bc - ad = \tilde{\mathrm{E}}\{u(V)F(V)\} - \tilde{\mathrm{E}}\{u(V)\}\tilde{\mathrm{E}}\{F(V)\}.$$

The result now follows from the fact that a covariance between two increasing functions $u(V)$ and $F(V)$ is always positive.

References

Admati, A., and P. Pfleiderer (1988), A theory of intraday patterns: volume and price variability, *The Review of Financial Studies*, **1**, 3–40.

Aït-Sahalia, Y. and P. Mykland (2003), The effects of random and discrete sampling when estimating continuous-time diffusions, *Econometrica*, **71**, 483–549.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2001), The distribution of exchange rate volatility, *Journal of the American Statistical Association*, **96**, 42–55. Correction published in 2003, **98**, Page 501.

Andersen, T.G., T. Bollerslev, and F.X. Diebold (2004), Parametric and non-parametric volatility measurement, *Handbook of Financial Econometrics*, Y. Aït-Sahalia and L.P. Hansen (eds), forthcoming.

Andreou, E., and E. Ghysels (2002), Rolling volatility estimators: Some new theoretical, simulation, and empirical results, *Journal of Business and Economic Statistics*, **20**, 363–376.

Back, K. (1991), Asset prices for general processes, it Journal of Mathematical Economics, **20**, 317–395.

Bandi, F.M. and J.R. Russell (2004), Separating market microstructure noise from volatility, *Journal of Financial Economics*, forthcoming.

Barndorff-Nielsen, O. and N. Shephard (2002), Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society, Series B*, **64**, 253–280.

Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard (2005), Robust variation estimation using kernels in financial econometrics, 2005, working paper.

Bollerslev, T., and H. Zhou (2002), Estimating stochastic volatility diffusion using conditional moments of integrated volatility, *Journal of Econometrics*, **109**, 33–65.

Carr, P. and L. Wu (2004), Time-changed Lévy processes and option pricing, *Journal of Financial Economics*, **71**, 113–141.

Comte, F. and E. Renault (1996), Noncausality in continuous time models, *Econometric Theory*, **12**, 215–256.

Comte, F. and E. Renault (1998), Long-memory in continuous time stochastic volatility models, *Mathematical Finance*, **8**, 291–323.

Delbaen, F., and W. Schachermayer (1999), Non-Arbitrage and the Fundamental Theorem of Asset Pricing: Summary of Main Results, "Proceedings of Symposia in Applied Mathematics" of the AMS, **57**, 49–58.

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 (2004), Estimation of continuous-time Markov processes sampled at random times, *Econometrica*, 1773–1808.

Dufour, A., and R.F. Engle (2000), Time and the price impact of a trade, *The Journal of Finance*, **55**, 2467–2498.

Easley, D., and M. O'Hara (1992), Time and the process of security price adjustment, *The Journal of Finance*, **47**, 577–605.

Eberlein, E. and A. Papapantoleon (2005), Equivalence of floating and fixed strike Asian and lookback options, *Stochastic Processes and their Applications*, **115**, 31–40.

Engle, R.F. (2000),The econometrics of ultra-high frequency data, *Econometrica*, **68**, 1–22.

Engle, R.F., and J.R. Russell (1998), Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica*, **66**, 1127–1162.

Engle, R.F., and Z. Sun (2005), Forecasting volatility using tick by tick data, working paper.

Florens, J.-P. and D. Fougère (1996), Noncausality in continuous time, *Econometrica*, **64**, 1195–1212.

Fouque, J.-P., G. Papanicolaou, and R. Sircar (2000). Derivatives in Financial Markets with Stochastic Volatility. Cambridge University Press.

Ghysels, E., and J. Jasiak (1998), GARCH for irregularly spaced financial data: The ACD-GARCH model, *Studies in Nonlinear Dynamics and Econometrics*, **2**, 133–149.

Grammig, J. and M. Wellner (2002), Modelling the interdependence of volatility and inter-transaction duration processes, *Journal of Econometrics*, **106**, 369–400.

Greenwood, M. (1946), The statistical study of infectious diseases, *Journal of the Royal Statistical Society*, **109**, 85–110.

Hansen, B.E. (1995), Regression with nonstationary volatility, *Econometrica*, **63**, 1113–1132.

Hansen, P.R. and A. Lunde (2005), Realized variance and market microstructure noise, Invited lecture with discussion, *Journal of Business and Economic Statistics*, forthcoming.

Manganelli, S. (2004), Duration, volume, and volatility impact of trades, working paper.

Meddahi, N. and E. Renault (2004), Temporal aggregation of volatility models, *Journal of Econometrics*, **119**, 355–379.

Meddahi, N., E. Renault, and B.J.M. Werker (2003), GARCH and irregularly spaced data, *Economics Letters*, forthcoming.

Protter, P. (2003), Stochastic Integration and Differential Equations, 2nd edition, Springer-Verlag, Berlin.

Renault, E., K. Sekkat, and A. Szafarz (1998), Testing for spurious causality in exchange rates, *Journal of Empirical Finance*, **5**, 47–66.

Zhang, L., P.A. Mykland, and Y. Aït-Sahalia (2004), A tale of two time scales: Determining integrated volatility with noisy high-frequency data, *Journal of the American Statistical Association*, forthcoming.