Supplementary information: The excess volatility puzzle explained by financial noise amplification from endogenous feedbacks

Alexander Wehrli, Didier Sornette

emails: alexander.wehrli@snb.ch, dsornette@ethz.ch

1 Inference scheme for the epidemic model

This section provides an overview of the estimation schemes applied for the different model types used in the main text.

1.1 Description of the estimation algorithm

To estimate the epidemic model, we use an Expectation Maximization (EM) algorithm. Generally, the EM can be used to iteratively build an MLE using the complete data likelihood function \( L^c(\theta; \omega, Z) \). Here, \( \omega = \{t_i\}_{i=1}^{N(T)} \) denotes the observation times of the events we want to model, in our specific setup these are the times when the observed price changed. \( N(T) \) counts the number of events observed on the sample interval of length \( T \). \( \theta \) is a vector of model parameters and \( Z \) introduces a set of latent variables which help decompose the estimation into manageable sub-problems. In our specific setup, these latent variables are taken as the branching structure of the epidemic model. Formally, this is encoded as a matrix of indicator variables \( Z_{N(T) \times N(T)} \in \{0, 1\}^{N(T) \times N(T)} \), where \( Z_{ij} = 1 \) if an event at \( t_i \) in the cascade of events is an offspring of an event at \( t_j \). If an event at \( t_i \) was exogenously triggered (was an “immigrant” to the system), \( Z_{ij} = 1 \). Using the branching structure as the latent information in the EM achieves two important goals. First, it reduces estimation of the epidemic model iteratively build an MLE using the complete data likelihood function \( L^c(\theta; \omega, Z) \).

The modification consists in dropping any assumption regarding a particular model for the fertility in the inference scheme for the epidemic model

E-step

This can be obtained using random thinning\(^8\) as

\[
\pi_{i,j} = \frac{\mu(t_i) \hat{\theta}^{(l-1)}}{\lambda(t_i) \hat{\theta}^{(l-1)}} \quad \text{and} \quad \pi_{i,j} = \frac{R^1 h(t_i - t_j) \hat{\theta}^{(l-1)}}{\lambda(t_i) \hat{\theta}^{(l-1)}} \quad \text{for} \quad i > j.
\]
3. **M-step:** In the maximization step, new values of the parameters \( \hat{\theta}^{(l)} \) are obtained by maximizing the expected complete data log-likelihood, \( \ell^c(\theta) \), defined as

\[
\ell^c(\theta) = \sum_{i=1}^{N(T)} \pi_{i,j} \log \mu(t_i) - \int_0^T \mu(t) dt + \sum_{j=1}^{N(T)} \sum_{t_{i,j} - t_i} \pi_{i,j} \log (R^T h(t_i - t_j)) - \sum_{j=1}^{N(T)} R^T H(T - t_j),
\]

where \( H(t) = \int_0^t h(s) ds \) is the cumulative distribution corresponding to the offspring density \( h \). Assuming that \( \mu \) and \( h \) do not share any parameters, the two parts \( \ell^c \) and \( \ell^h \) can be maximized separately. Furthermore, we obtain estimates for the \( R^T_j \)'s from the branching structure as

\[
\hat{R}^T_j = \frac{\sum_{i>j} \pi_{i,j}}{H(T - t_j[\hat{\theta}^{(l)}]).}
\]

4. **Check convergence:** Stop estimation if \( ||\hat{\theta}^{(l)} - \hat{\theta}^{(l-1)}||_\infty \) is below some small threshold, or go back to the E-step for another iteration \( l = l + 1 \) otherwise.

An estimate of the branching ratio of the epidemic process can finally obtained as

\[
\bar{\eta} = \frac{1}{N(T)} \sum_{j=1}^{N(T)} \hat{R}^T_j.
\]

In our empirical calibrations, we stop the EM algorithm if the convergence criterion is satisfied at a threshold of \( 10^{-4} \) for a few consecutive iterations, or if a maximum number of 500 iterations was performed.

### 1.2 Adaptive estimation of \( \mu(t) \)

To facilitate a flexible and systematic estimation of the exogenous driving process \( \mu \), we use flexible logspline\(^{10} \) estimators to control for trends and shocks in the data in an adaptive fashion\(^{5} \). The flexibility and smoothness provided by these modern, non-parametric estimators has been found to be crucial in controlling for the significant exogenous heterogeneity in high-frequency financial data\(^{6} \). To select models, we vary the degrees of freedom allowed for \( \mu(t) \) in \( p \in \{0, 3, 6, 10, 15, 20, 25, 30, 40, 50, 60\} \), where \( p = 0 \) denotes a model with constant exogenous driving \( \mu(t) \equiv \mu > 0 \), compute log-likelihoods

\[
\ell(\hat{\theta}^{(l)}) = \sum_{i=1}^{N(T)} \log \lambda(\eta_i[\hat{\theta}^{(l)}] - \int_0^T \lambda(s[\hat{\theta}^{(l)}] ds
\]

and select the best model for \( \mu(t) \) according to the Bayesian Information Criterion (BIC) \(-2\ell(\hat{\theta}^{(l)}) + \log(N(T)) p\).

### 1.3 Simulation study of the EM estimator

We test the ability of the proposed EM procedure to characterize the distribution of the \( R^T_j \) in a simulation study. For this purpose, we simulate realizations of Hawkes processes with constant immigration \( \mu = 0.1 \), a branching ratio of \( \bar{\eta} = 0.6 \) and three different fertility distributions \( f(\eta) \), namely:

1. Exponentially distributed fertilities \( f_\lambda(\eta) = \lambda e^{-\lambda \eta} \) with \( \lambda = 1/\bar{\eta} \).
2. Weibull distributed fertilities \( f_{\lambda, \beta}(\eta) = \beta \lambda (\lambda \eta)^{\beta-1} e^{-(\lambda \eta)^\beta} \) with \( \lambda = \bar{\eta}/2 \) and \( \beta = 0.5 \).
3. Power law distributed fertilities \( f_\gamma(\eta) = \gamma \eta^{\gamma} \) with \( \eta_0 = 0.3 \) and \( \gamma = 1.5 \).

We use a simple exponential offspring density with ground truth \( h(t) = 2e^{-2t} \) in all three cases. Fitting is then done using the EM on 100 realizations of length \( T = 864000 \) seconds from each of these three processes. The results from this simulation study are summarized in Figure S1. First, we notice that the branching ratios \( \bar{\eta} \) recovered by the EM are quite consistent, with estimates \( \approx 0.62 \) for all three fertility distributions considered. Regarding the exact shape of the distributions, we find that the head of the distribution of \( R^T \) appears difficult to estimate – not surprising given that the true distribution is discrete, whereas the estimator is real-valued. Regarding the tail, we see that even though there is clearly some bias, the shape of the tail is approximated quite well. In particular, the tail of the distribution recovered by the EM exhibits approximate power.
law dynamics in the case where the underlying true fertility distribution also follows a power law – the heaviness of the tail is however underestimated. In the case of exponential fertility, the EM estimates also clearly drop off exponentially in the tail. Therefore, we conclude that, while inferring the exact distribution of the fertilities directly from the EM estimates of the number of first generation offspring is difficult, testing for certain properties of the underlying distribution (e.g. computing its mean or conditioning it on some observable, testing for power law tail, asserting the share of zero-offspring points) works well.

### 1.4 Extension of the estimation algorithm in the presence of exogenous stochastic clustering

Here, we give a short practical description of the extended algorithm required to estimate the model in the presence of stochastic exogenous clustering for the special case where \( \gamma(s) \equiv \tilde{\gamma} \) and \( \eta(s) \equiv \tilde{\eta} \), i.e. with conditional intensity

\[
\lambda(t) = \mu(t) + \int_0^t \tilde{\gamma}(t-s)dN^\mu(s) + \int_0^t \tilde{\eta}(t-s)dN(s).
\]

The algorithm synthesizes the principles described above for the generalized Hawkes model, with the MCEM algorithm as derived in \(^{14}\) for the inclusion of the additional shot-noise effects.

For the estimation of a model like \((8)\), or \((7)\) of the main text, the missing data has to be extended to \((Z,C)\), where \(C\) denotes the immigrant process induced by \(N^\mu\), and the complete data log-likelihood thus takes the form

\[
Q(\theta | \omega, \hat{\theta}^{(i)}) = \mathbb{E}_{Z,C|\omega,\hat{\theta}^{(i)}}[\log L^c(\theta | \omega, Z, C)],
\]

which can be approximated by averaging over a sample \(c^{(k)}, k = 1, ..., K\) of immigrant realizations from a density \(\propto p(c | \omega, \hat{\theta}^{(i)})\) as

\[
Q(\theta | \omega, \hat{\theta}^{(i)}) \approx \frac{1}{K} \sum_{k=1}^K \mathbb{E}_{Z,\omega,\hat{\theta}^{(i)},c^{(k)}}[\log L^c(\theta | \omega, Z, c^{(k)})].
\]

To obtain a conditional MCMC sample from the immigrant process, a Metropolis-Hastings algorithm is employed. Given some state \(c^{(k)}\) of the Markov chain, each Metropolis-Hastings iteration starts with randomly proposing the birth or death of an immigrant. In case we choose birth, we uniformly select some state \(c^* \in \omega \setminus c^{(k)}\). The Metropolis-Hastings birth ratio to decide whether \(c^*\) is accepted, given some immigrant vector \(c\), is obtained from

\[
r_b(c,c^*) = \frac{N(T) - |c|}{|c| + 1} \mu(c^*) e^{-\tilde{\gamma}G(T-c^*)} \times \prod_{t_i \in \omega^{(c,c^*)}} \left( 1 + \frac{\tilde{\gamma}g(t_i-c^*)}{\lambda(t_i|c) - \mu(t_i)} \right) / (\lambda(c^*|c) - \mu(c^*)).
\]
and we accept \( c^{(k+1)} = c^{(k)} \cup c^{*} \) with acceptance probability \( \min\{ r_{b}(c^{(k)}, c^{*}), 1 \} \). In case we chose death, \( c^{*} \in c^{(k)} \) is removed from the current state of the Markov chain with acceptance probability \( \min\{ 1 / r_{b}(c^{(k)} \setminus c^{*}, c^{*}), 1 \} \). The Metropolis-Hastings iterations can then be repeated until a sufficiently long chain is available, and the desired \( K \) samples be drawn from this chain.

Then, given some \( c^{(k)} = (c_{1}, \ldots, c_{n_{k}}) \) and due to the fact that the process maps to a branching process, the inner expectation in (10) decouples analogously to (4), using branching probabilities which can be evaluated due to random thinning\(^9\) in a similar fashion like (3) as

\[
\pi_{i,j}^{k} = \begin{cases} 
\frac{\gamma g(t_{i} - c_{i})}{\lambda(t_{i}, \hat{\theta}^{(k)}) - \mu(t_{i}, \hat{\theta}^{(k)})} & \text{for } t_{i} \in \omega \setminus c^{(k)}, t_{j} \in \omega \\
0 & \text{otherwise}
\end{cases} \quad (12)
\]

\[
\pi_{i,j}^{h} = \frac{\eta h(t_{i} - c_{i})}{\lambda(t_{i}, \hat{\theta}^{(k)}) - \mu(t_{i}, \hat{\theta}^{(k)})} \quad (13)
\]

In the M-step of the algorithm, the objective (10) is maximized, which – as for estimation without shot-noise effects – has the structure of decoupled density estimation. First, \( \mu(t) \) is estimated on the MCMC sample of immigration times \( \{c^{(1)}, \ldots, c^{(K)}\} \).

Next, the offspring density \( h(\cdot) \) is estimated with objective \( \ell^{b} \) from (4), with the difference that the offspring probabilities used in the objective function need to be averaged over the MCMC samples as

\[
\bar{\pi}_{i,j}^{h} = \frac{1}{K} \sum_{k=1}^{K} \pi_{i,j}^{h(k)}.
\]

Estimation of \( g(\cdot) \) finally requires to pool inter-event times and weights over all MCMC samples \( \pi^{k} := \{(\bar{\pi}_{i,j}^{k}, t_{i} - c_{i}^{k})\}_{k=1}^{K} \) and maximize

\[
\ell_{k}^{c}(\hat{\theta} | \bar{\pi}) = \frac{1}{K} \sum_{k=1}^{K} \sum_{i=1}^{N(T)} \sum_{j=1}^{n_{k}} \bar{\pi}_{i,j}^{k} \log \left( \frac{\gamma g(t_{i} - c_{i})}{\lambda(t_{i}, \hat{\theta}^{(k)}) - \mu(t_{i}, \hat{\theta}^{(k)})} \right) - \frac{\eta}{K} \sum_{k=1}^{K} \sum_{i=1}^{n_{k}} G(T - c_{i}^{k}),
\]

where \( G(t) = \int_{0}^{t} g(s)ds \) is the cumulative distribution corresponding to the exogenous offspring density \( g(t) \). Estimating nested models can be done by enforcing the restrictions \( \bar{\gamma} = 0 \) and/or \( \bar{\eta} = 0 \). For the applications presented in the main text, the MCEM estimation is done with \( K = 50 \) samples taken uniformly from the second half of a chain of length 100000, generated by the Metropolis-Hastings algorithm presented here.

### 1.5 Model selection in the presence of exogenous stochastic clustering

In the absence of stochastic exogenous clustering (\( \gamma = 0 \)), model selection is possible consistently using information criteria (see section 1.2), because the unconditional likelihood is accessible. If however \( \gamma > 0 \), then conditional likelihoods must be computed using MCMC as discussed in\(^{13}\). To fix notation, denote by \( df(\cdot) \) the degrees of freedom of a function, thus we have the exogenous flexibility as \( p = df(\mu) \) and the endogenous flexibility of the model as \( r = df(h) + df(g) \). Then \( \hat{\theta}_{p,r} \) denotes the parameter estimates of a particular model with overall \( p + r \) degrees of freedom.

To compute conditional likelihoods, we create an MCMC sample \( c^{(1)}, c^{(2)}, \ldots, c^{(k)}, \ldots, c^{(K)} \) of immigrant realizations with density \( \propto p(\cdot | \omega, \hat{\theta}_{p,r}) \) – just like in the E-step of the MCEM algorithm in the previous section – being the density of an immigrant realization \( c \), conditional on the parameter estimates \( \hat{\theta}_{p,r} \) and the data \( \omega \). Each \( c^{(k)} \) defines a conditional log-likelihood

\[
\ell(\hat{\theta}_{p,r} | \omega, c^{(k)}) = \sum_{i=1}^{N(T)} \log \lambda(t_{i} | c^{(k)}, \hat{\theta}_{p,r}) - \int_{0}^{T} \mu(s | \hat{\theta}_{p,r})ds - \sum_{i=1}^{N(T)} \eta H(T - t_{i} | \hat{\theta}_{p,r}) - \sum_{i=1}^{n_{k}} \gamma G(T - c_{i}^{k} | \hat{\theta}_{p,r})
\]

and a conditional BIC

\[
\text{BIC}(\hat{\theta}_{p,r} | \omega, c^{(k)}) = -2\ell(\hat{\theta}_{p,r} | \omega, c^{(k)}) + \log(N(T))(p + r).
\]

In order to select models, we apply a two-step procedure:

1. For each model type – represented with a given \( r \) – we first select the optimal degrees of freedom \( p^{\ast} \) for the immigration intensity \( \mu \). For this purpose, for \( b = 1, \ldots, B \) and \( p = 0, \ldots, P \), we sample \( j_{p}^{b} \in [1, K] \) to obtain a bootstrap sample

\[
\mathcal{B}_{b} = \left\{ \text{BIC}(\hat{\theta}_{p,r} | \omega, c^{(j_{p}^{b})}) \right\}_{p \in \{0, \ldots, P\}, b \in \{1, \ldots, B\}}
\]
We can then compute conditionally (on the particularly sampled immigrant realizations) optimal degrees of freedom $p_b^r = \arg \min_p BIC_p$, being the optimal degrees of freedom per bootstrap sample, according to the randomly chosen conditional BICs. The final choice across bootstrap samples is then $p^r = \text{mode}(\{p_b^r\}_{b=1,\ldots,B})$.

2. To compare two different models, say $r'$ and $r''$, we sample pairs $(i_b, j_b)$ from $[1, K] \times [1, K]$, again $b = 1, \ldots, B$ times, and then select $r'$ over $r''$ on a given bootstrap sample $b$ if

$$BIC(\hat{\theta}_{p^r'}, r'\mid \omega, c^{(i_b)}) < BIC(\hat{\theta}_{p^r''}, r''\mid \omega, c^{(j_b)}).$$

(18)

We can then finally compute selection probabilities $P(r', r'') = B^{-1} \sum_{b=1}^B I\{BIC(\hat{\theta}_{p^r'}, r'\mid \omega, c^{(i_b)}) < BIC(\hat{\theta}_{p^r''}, r''\mid \omega, c^{(j_b)})\}$, giving an estimate of the likelihood that a certain model type $r'$ is to be preferred over another model type $r''$.

For all applications in the main text, we choose $B = 1000$ and perform the test on an MCMC sample of $K = 100$ draws from a Markov chain of length $300'000$.

2 Endogenous vs. exogenous clustering

In order to test which of the clustering mechanisms are dominant in practice, we conduct the following test. For some randomly selected days of data we calibrate models with constant branching coefficients $\gamma(s) \equiv \bar{\gamma}, \eta(s) \equiv \bar{\eta}$ and the adaptive estimator using logsplines from section 1.2 for $\mu^c(t)$, which also allows for deterministic clustering through time-variation in $\mu^c(t)$. We then estimate increasingly complex models, starting from no stochastic clustering (Deterministic Exo: $\gamma = 0, \bar{\eta} = 0$), to include only exogenous clustering (Stochastic Exo: $\bar{\eta} = 0$), or deterministic exogenous and stochastic endogenous clustering (Stochastic Endo: $\gamma = 0$), or finally a mixture of all three clustering mechanisms (Stochastic Endo-Exo). Additionally, we allow $h$ to either exhibit short- (exponential) or heavy-tailed (approximate power law) memory.

In Table S1, we illustrate how frequently (across samples and thresholds $\epsilon$) a certain clustering mechanism is selected according to the procedure described in section 1.5. We observe first that the most frequently selected model is a model with deterministic exogenous clustering and stochastic endogenous clustering (mechanisms (A) and (C) from the main text). Moreover, an endogenous offspring distribution with heavy tail is preferred over short-tailed memory. A more detailed summary of the estimates for the stochastic endo model with heavy-tailed memory, being the most frequently selected model, is additionally given in Figure S2.

### Table S1

| EUR/USD | Deterministic Exo | Stochastic Exo (ST) | Stochastic Endo (ST) | Stochastic Endo-Exo (ST) |
|---------|-------------------|---------------------|---------------------|------------------------|
| Stochastic Exo (ST) | 0% | 100% | 60% | 100% |
| Stochastic Endo (ST) | 100% | 100% | 100% | 100% |
| Stochastic Endo-Exo (ST) | 60% | 100% | 0% | 100% |

| E-mini | Deterministic Exo | Stochastic Exo (ST) | Stochastic Endo (ST) | Stochastic Endo-Exo (ST) |
|--------|-------------------|---------------------|---------------------|------------------------|
| Stochastic Exo (ST) | 17% | 97% | 97% | 97% |
| Stochastic Endo (ST) | % | 80% | 80% | 65% |

Table S1. Shows a summary of the model selection procedure described in section 1.5, applied to three randomly chosen days of data (EUR/USD: Aug 9, 2016; Sep 30, 2016, Oct 3, 2016; E-mini: Feb 8, 2019, Feb 26, 2019, Mar 7, 2019). The $\epsilon$ are chosen as the (5%, 10%, ..., 95%)-quantiles of the distribution of squared returns. The offspring densities $g(t)$ and $h(t)$ are modelled using an exponential distribution for short-tailed (ST) models, and using an approximate Pareto distribution for heavy-tailed (HT) models. The branching coefficients are taken as constants. The table shows the selection frequency of the respective row-model over the corresponding column-model, pooled across all thresholds $\epsilon$ and samples.

with deterministic exogenous clustering and stochastic endogenous clustering (mechanisms (A) and (C) from the main text). Moreover, an endogenous offspring distribution with heavy tail is preferred over short-tailed memory. A more detailed summary of the estimates for the stochastic endo model with heavy-tailed memory, being the most frequently selected model, is additionally given in Figure S2.

3 First-order properties and relation to autoregressive time series

A string of recent literature has developed first- and second-order statistical properties of self-excited / Hawkes point processes in the context of financial applications, which were originally formulated and their behavior...
we note that ε (cf. also the unconditionally expected activity profiles in Figure 1 of the main text), across various thresholds h corresponding to the (5%, 10%, ..., 95%)-quantiles of the distribution of squared returns. The smaller plots below the heatmaps show, also as a function of ε, estimates of the branching ratio \( \bar{\eta} \), the characteristic time scale \( \tau_0 \) and the tail index \( \alpha \) of the memory kernel \( h(t) \).

characterized carefully and in great detail even before the recent spur of interest from the quantitative finance literature\textsuperscript{21–26}. Here we build on these results to derive the time-dependent properties of first-order characteristics of the epidemic process underlying the study in the main text. In a few extensive comments, we furthermore highlight some important parallels to the analysis of autoregressive time series.

In the following, we consider the conditional intensity \( \lambda(t) = \mu(t) + \sum_{t_i<t} \bar{\eta} h(t-t_i) = \mu(t) + \int_t^\infty \bar{\eta} h(t-s) dN(s) \) after having averaged over the distribution of fertilities. With some abuse of notation, we denote this averaged intensity also as \( \lambda(t) \). Taking the expectation of \( \lambda(t) \), we find the expected intensity as the solution to the equation

\[
\mathbb{E}[\lambda(t)] = \mu(t) + \int_0^\infty \phi(t-s) \mathbb{E}[\lambda(s)] ds,
\]

(19)

for \( \phi(t) = \bar{\eta} h(t) \) and \( \bar{\eta} = \mathbb{E}[\eta] \). Equation (19) has the form of a classical renewal equation and can be expressed as the convolution

\[
e(t) = \mu(t) + \phi * e(t)
\]

(20)

for \( e(t) = \mathbb{E}[\lambda(t)] \). Denoting the Laplace transform of a function \( x(t) \) as \( \tilde{x}(s) = \int_0^\infty e^{-st} x(t) dt \), the solution of (20) is given in the Laplace domain as

\[
\tilde{e}(z) = \frac{\tilde{\mu}(z)}{1 - \tilde{\phi}(z)}.
\]

(21)

If we define the auto-convolution of \( \phi \) for \( n \geq 1 \) as \( \phi^{*n+1}(t) = \int_0^t \phi(t-s) \phi^{*n}(s) ds = \phi * \phi^{*n}(t) \) with \( \phi^{*1} = \phi \) and \( \phi^{*0} = 1 \), we note that

\[
\phi^{*n}(t) = \bar{\eta}^n h^{*n}(t).
\]

(22)

Since \( h \) is a probability density function, \( h^{*n} \) itself is the cumulative distribution of a sum of \( n \) independent random variables, each with density \( h \). By explicitly calculating the \( L^1 \)-norm of \( h^{*n} \) and remarking that \( \int_0^\infty dt \int_0^\infty ds = \int_0^\infty ds \int_0^\infty dr \), it can be shown that the \( L^1 \)-norm of \( h^{*n}(t) \) is equal to the \( L^1 \)-norm of \( h^{*(n-1)}(t) \). By recurrence, it is also equal to the norm of \( h(t) \) and consequently \( \phi^{*n}(t) \) has \( L^1 \)-norm equals \( \bar{\eta}^n \) for all \( n \). Thus the auxiliary function \( \psi^\phi \), defined as

\[
\psi^\phi(t) = \sum_{n=0}^\infty \phi^{*n}(t)
\]

(23)
We note that since \( \psi \) we can invert the polynomial, we get the equivalent moving-average form

\[ y(t) = \sum_{n=0}^{\infty} \hat{\phi}(z)^n = \frac{1}{1 - \hat{\phi}(z)}. \]

(24)

To finally find the solution of (20), we see that \( \psi^0(t) \) has Laplace transform

\[ \hat{\psi}^0(z) = \sum_{n=0}^{\infty} (\hat{\phi}(z))^n = \frac{1}{1 - \hat{\phi}(z)}. \]

(25)

Note that the infinite sum in (25) resembles closely the characteristic polynomial of an MA(\(\infty\)) process. To see this, define the lag operator \( L \) as \( L \gamma_t = \gamma_{t-1} \) and consider the AR(\(p\)) process \( \gamma_t = -\sum_{j=1}^{p} \phi_j L \gamma_{t-j} + \epsilon_t \) with innovations \( \epsilon_t \sim IID(0, \sigma^2) \). Then the characteristic polynomial \( \Phi(\eta) = \sum_{n=0}^{\infty} \phi_n \eta^n \) – defined on \( |\eta| < 1 \) – expresses the innovations as \( \epsilon_t = \Phi(\eta) \gamma_t \) and assuming we can invert the polynomial, we get the equivalent moving-average form \( \gamma_t = \Phi^{-1}(\eta) \epsilon_t = \Psi(\eta) \epsilon_t \), where \( \Phi(\eta)\Psi(\eta) = 1 \). In the specific case of an AR(1) process, the characteristic polynomial is \( \Phi(\eta) = 1 - \eta L \) and thus the characteristic polynomial of the moving-average representation in the time series case has an obvious conceptual correspondence in the point process case as

\[ \Psi(L) = \sum_{n=0}^{\infty} \eta^n L^n \iff \hat{\psi}^0(z) = \sum_{n=0}^{\infty} \eta^n (\hat{h}(z))^n. \]

(26)

In the Hawkes process case, we have a similar equivalence regarding the “characteristic polynomials” as \( (1 - \hat{\phi}(z))(1 + \hat{\psi}(z)) = 1 \). Furthermore, the indicated comparison to the Laplace transform of \( \psi^0 \) can be interpreted such that – as intuitively reasonable – the “lag operator” in this moving average representation of the Hawkes process is given by the Laplace transform of the offspring density \( \hat{h}(z) \), which precisely defines the probability of lagged event occurrences given some mother event. Furthermore, the integrability condition in (24) is equivalent to the condition in MA(\(\infty\)) models where the absolute MA-coefficients must be summable for \( \gamma_t \) to take finite values and for the process not to explode. The similarities to time series analysis are conceptually striking: in order to learn about the probability structure of the Hawkes process (the autoregressive point process), we have ended up looking for an equivalent moving-average representation. But when working with data, we prefer the simple structure of the Hawkes. In time series analysis, when dealing with data, autoregression as a linear method is often very convenient and computationally fast. But when evaluating the structural properties of a process, they can often more easily be obtained via the moving-average representation – one example being the autocovariance function, which can be obtained from the convolution of the moving-average coefficients. Continuing the derivation, the expression for \( \hat{\psi}^0(z) \) from (25) can be inserted into (21) to find the solution of the renewal equation in the time domain by applying the inverse Laplace transform \( \mathcal{L}^{-1} \) as

\[
\begin{align*}
e(t) &= \mathcal{L}^{-1} \left[ \hat{\mu}(z) \sum_{n=0}^{\infty} (\hat{\phi}(z))^n \right] = \mathcal{L}^{-1} \left[ \sum_{n=0}^{\infty} (\hat{\phi}(z))^n \right] * \mu(t) \\
&= \left( \sum_{n=0}^{\infty} \mathcal{L}^{-1} \left[ (\hat{\phi}(z))^n \right] \right) * \mu(t) = \left( \sum_{n=0}^{\infty} \phi^{*n}(t) \right) * \mu(t) \\
&= \left( 1 + \sum_{n=1}^{\infty} \phi^{*n}(t) \right) * \mu(t) = \mu(t) + \psi * \mu(t)
\end{align*}
\]

(27)

where \( \psi(t) = \sum_{n \geq 1} \phi^{*n}(t) \), which is in turn defined by its Laplace transform

\[ \hat{\psi}(z) = \frac{\hat{\phi}(z)}{1 - \hat{\phi}(z)}. \]

(28)

We note that since \( \phi(t) \) is positive and supported by \( \mathbb{R}^+ \), and \( \psi(t) \) is simply defined as the sum of convolutions of \( \phi(t) \), \( \psi(t) \) is also positive and supported by \( \mathbb{R}^+ \).

The mapping of the Hawkes process to a Galton-Watson branching process\(^{27}\) gives valuable intuition on the derivations at this point. We recall that the number of direct offspring of any point in the genealogy of the mapped branching process has Poisson distribution with mean \( \bar{\eta} \). The expected size of the \( n \)-th generation of the equivalent branching process is thus \( \bar{\eta}^n \) (cf.
Further, the total number of individuals in the entire cluster formed by the branching process, including the immigrant (0-th generation), has a Borel distribution, having mean \((1 - \bar{\eta})^{-1}\) (cf. (25)) and variance \(\bar{\eta}(1 - \bar{\eta})^{-3}\). The expected size of the total offspring over all generations, triggered by one immigrant is \(\bar{\eta}/(1 - \bar{\eta})\), cf. (28). As such, we see that \(\bar{\psi}(0)\) and \(\psi(0)\) are the expected value of the random variables representing respectively the cluster size including and without the triggering immigrant\(^1\).

Using this fact, the well-known result for the average intensity in the case of constant immigration \(\mu(t) \equiv \mu\) follows in a straight-forward manner as

\[
E[\lambda(t)] = \mu + \mu \ast \psi(t) = \mu + \mu \int_{\mathbb{R}^+} \psi(s)ds.
\]

(29)

By virtue of (28), this translates into

\[
\mu + \mu \bar{\psi}(0) = \mu + \mu \frac{\phi(0)}{1 - \phi(0)} = \frac{\mu}{1 - \bar{\eta}}.
\]

(30)

In the time domain, the function \(\psi(t)\) satisfies the equation

\[
\psi(t) = \phi(t) + \phi \ast \psi(t)
\]

(31)

and is often known as the “activity level” in systems with memory – in the language of branching processes the dispersion of all offspring of one immigrant over time. The function \(R(t) = \psi(t)/\bar{\eta} = h \ast \psi(0)\) is often called resolvent, renormalized kernel or response function and describes the response of the system to an impulsive (exogenous) perturbation. It integrates to \(1/(1 - \bar{\eta})\), representing the mean cluster size of the Galton-Watson branching process to which the Hawkes process can be mapped. Note again the link to time series analysis: the impulse-response function (to a unit shock in the innovations) in an AR(p) process is given by the corresponding MA(\(\infty\)) coefficients. Analogously – as already noted earlier – the function \(\psi(0)\) (or \(\psi(t)\) when not counting the immigrant) defines the activity of the Hawkes process following an immigrant, i.e. the response to an innovation.

In the case of an exponential offspring density, \(\psi(t)\) is known explicitly

\[
\phi(t) = \bar{\eta}/(\tau_0 + t)^{1 + \alpha} \Rightarrow \psi(t) = \bar{\eta}e^{-\beta(1 - \bar{\eta})t},
\]

(32)

which illustrates the name “renormalized kernel” as the characteristic time scale of the kernel is renormalized with \((1 - \bar{\eta})\), namely the characteristic time scale is amplified by the factor \(1/(1 - \bar{\eta})\), thus expressing the increased duration of the cluster due to the cascade of generations.

Of particular interest in many applications are kernels with power law memory

\[
\phi(t) = \bar{\eta} e^{-\alpha t} \frac{\eta}{(\tau_0 + t)^{1 + \alpha}} \Rightarrow \phi(z) = \bar{\eta} e^{\beta z} \eta \Gamma(-\alpha, \eta z)
\]

(33)

for \(\Gamma(a, x)\) the upper incomplete gamma function. In this case, the Laplace transform in (28) cannot be inverted analytically, but the asymptotic properties of the resulting resolvent have been studied in detail as well, see\(^25\).

To compute the expected number of points up to some time \(t\), we can use the solution (27) to obtain

\[
E[N(t)] = \int_0^t E[\lambda(s)]ds = \int_0^t \mu(s)ds + \int_0^t (\psi \ast \mu(s))ds.
\]

(34)

For the case of constant exogenous intensity, the fact that \(\psi \ast \mu = \mu \bar{\psi}(0)\) directly recovers \(E[N(t)] = \mu t/(1 - \bar{\eta})\).

For the case of time-dependent immigration, we notice that, by the Fubini theorem, the integral of the endogenous part of the compensator \(\Lambda(t) = \int_0^t \lambda(s)ds\) satisfies

\[
\int_0^t ds \int_0^s \phi(s-u)dN(u) = \int_0^t dN(u) \int_u^t \phi(s-u)ds = \bar{\eta} \int_0^t H(t-s)dN(s)
\]

(35)

\(^1\)For example, we can compute the average duration of a cluster as \(\int_0^1 t \frac{\phi(0)}{\mu(0)\psi(0)} dt\), which e.g. in the case of an exponential kernel like in (32) yields \(\frac{1}{\bar{\eta} - \eta}\).
and integration by parts$^2$ with $H(t)$ shows that this last expression equals $\phi \ast N(t)$ and thus we get $\Lambda(t) = I(t) + \phi \ast N(t)$ for $I(t) = \int_0^t \mu(s)ds$. Taking expectations, a renewal equation for the expected event count is found, which is similar to (20), as
\[ \mathbb{E}[N(t)] = I(t) + \phi \ast \mathbb{E}[N(t)], \]
where we have used the fact that, at fixed $t$, $N(t)\Lambda(t) \sim \text{Poi}(\Lambda(t))$ and thus $\mathbb{E}[N(t)] = \mathbb{E}[\Lambda(t)]$. The solution of this equation is analogously
\[ \mathbb{E}[N(t)] = I(t) + \psi \ast I(t). \]

As a final remark underlining the conceptual similarities between autoregressive point processes and time series, it is worth noting that the scaling limits of the Hawkes process derived in$^{18,29}$ also find their natural analogue in the time series case$^{30}$.

### 4 Effect of observation error on variance estimates of $X$ and bias correction

A general result from stochastic calculus says that for a random partition$^3$ of $[0,t]$, $\pi = \{0 = t_0 \leq \cdots \leq t_n = t\}$, and two general semimartingales $W$ and $Z$, we have that$^31$
\[ \text{plim} \sum_{i \in \pi \leq t} (W(t_i) - W(t_{i-1}))(Z(t_i) - Z(t_{i-1})) = [W, Z]_t \]
and the convergence in probability to the quadratic covariation $[\cdot, \cdot]_t$ is for the mesh size $\sup_{i \geq 1} (t_i - t_{i-1}) \to 0$ as $n \to \infty$. Furthermore, in the case of the specific dynamics we postulate for our fundamental price $X$, the fundamental identity for stochastic integrals$^4$ yields
\[ [X, X]_t = \int_0^t \sigma_X^2(s) d[B, B]_s = \int_0^t \sigma_X^2(s) ds = v(t). \]
As a consequence, a common way to estimate the integrated variance of a process is to use the sum of squared returns, or realized variance. Assuming we are able to observe the latent process $X$ without error at times $t_i$, the ideal estimator
\[ \sum_{i \in \pi \leq t} (X(t_i) - X(t_{i-1}))^2 \]
converges to $v(t)$ in probability as the sampling frequency increases$^{32}$. However, in the presence of observation error, we are forced to consider instead
\[ [Y, Y]_t = [X, X]_t + 2[X, U]_t + [U, U]_t, \]
so the sum of squared observed returns is in general inconsistent for $v(t)$.

This argument immediately makes clear that $\sigma^2$ from equation (4) in the main text, based on observed returns, is subject to the same biases as the realized variance estimator (40), as soon as there is observation error. We make the following assumptions:

1. The noise process $U$ is stochastically independent of $X$. This is a standard assumption in the literature, even though there is some evidence for more intricate noise dynamics at tick frequencies$^{33}$.

2. $U(t)$ is a covariance stationary process with $\mathbb{E}[U(t)] = 0$.

3. We sample the processes at the times $t_i$ where the observed price changes to compute our variance estimates.

Under assumptions 1. - 3. and defining $e_i = U(t_i) - U(t_{i-1})$ with $\mathbb{E}[e_i^2] = \sigma^2$, the bias in equation (14) from the main text is found by taking expectations as
\[ \mathbb{E}[	ilde{m} \Lambda(t) - v(t)] = \mathbb{E}[(U, U)] = \sum_{i=1}^{N(t)} \mathbb{E}[e_i^2] = N(t) \sigma^2. \]

Furthermore, under the assumptions made, (42) makes clear that the relative contributions of the exogenous and endogenous variances from $\Lambda(t)$ are unchanged by the observation error – only their absolute magnitudes are inflated when using observed returns, due to the presence of $U$.

---

$^2$We can write $H(t)N(t) - H(T)N(0) = \int_0^t H(t-s) dN(s) - \int_0^t h(t-s) N(s) ds$

$^3$It should be noted that, so as to not complicate notation, we have omitted that in fact we are dealing with an increasing sequence of partitions, so that $n \to \infty$.

$^4$For a semimartingale process $X$ and an $X$-integrable process $\xi$, we have the identity $\int \xi dX, \int \xi dX = \int \xi^2 d[X]$. 

9/12
5 Data sets and pre-processing

The main data sets used here are mid-prices of two highly liquid markets. First, we study the EUR/USD currency pair traded on Electronic Broking Services (EBS), one of the major interbank markets for foreign exchange and key market place for the EUR/USD currency pair. The time stamps of the EBS data set have a resolution of 100ms, and there is an entry whenever either the bid, offer or both changed within the last 100ms interval. We use data for the entire calendar year of 2016. During this period, there were around 30’000 mid-price changes per day on average. Since foreign exchange markets operate continuously, we partition each week of the data into five 24 hour trading days, starting at the market opening on Sunday evening and disregard the 48h weekend period as there is no activity during that time.

The second main data set we use is from the S&P 500 E-mini futures contract traded on the Globex electronic platform at the Chicago Mercantile Exchange (CME), considered a key venue for equity price discovery in the United States. Trading on the Globex platform is possible almost continuously – only halted temporarily between 3:15pm and 3:30pm CT (Chicago Time) and for a maintenance window between 4pm and 5pm CT. We work with three months of data from January to March 2019 and define each trading day to start at 5pm CT and end at 3:15pm CT, discarding the 30 minutes of trading between the end of the temporary trading halt and the start of the maintenance window. This results in around 85’000 mid price changes per trading day on average in the sample. The time stamps of the E-mini mid-price changes have millisecond resolution.

For some intraday case studies, we will additionally revisit a data set for the GBP/USD currency pair from7, traded on Thomson Reuters Matching (now known as Refinitiv FX), which was obtained from Refinitiv Tick History and also has millisecond resolution.

As customary for point process calibration14,34, we randomize the time stamps of all data sets within an interval of uncertainty (here 100 milliseconds) and exclude any holidays from the analysis.

6 Estimation of the observation error

To complete our variance estimator requires specification of a particular model for the observation error. Under the common assumption of a serially independent noise process $U$ with variance $\mathbb{E}[U^2] = \omega_1^2$, the increments $e_t$ follow an MA(1) structure. In this case, an estimate for the observation error variance can be obtained from the relation $\mathbb{E}[y_t y_{t-1}] = -\omega_1^2$ as

$$\omega_1^2 = -\frac{1}{N(t) - 1} \sum_{i=2}^{N(t)} y_i y_{i-1},$$

where $y_t = Y(t) - Y(t_{i-1})$. Under the independent observation error assumption, the bias correction term thus becomes $\sigma_e^2 = 2\omega_1^2$. To account for more complex serial dependence in the observation error, we will consider the generalization of $e_t$ to an MA($q$) structure as

$$e_t = \sum_{j=1}^{q} \rho_j (v_{t+1-j} - v_{t-j}),$$

where $v_t \sim i.i.d.(0, \omega_q^2)$ and $\rho_1 = 1$ for identification purposes. In this case, the bias correction term is

$$\sigma_e^2 = 2\omega_q^2 \left( \sum_{j=1}^{q} \rho_j^2 - \sum_{j=2}^{q} \rho_j \rho_{j-1} \right).$$

Estimation of the noise model (45) can be done using the quasi maximum likelihood (QML) approach of37. This procedure allows one to select the order $q$ of the MA model consistently using information criteria. Here, we rely on the Akaike Information Criterion (AIC), which was shown to yield consistent estimates in37. Important for our purposes, the QML scheme also provides estimates for the autocovariance function of the noise process, which can then be factorized to estimate the $\rho_j$.

Estimates for $\hat{\mu}$ and the observation error variance according to the different specifications in the previous subsection are presented in Table S2. The fact that $\omega_q^2$ is frequently found to be negative, indicates significant departure from the standard i.i.d. assumption for the observation error. This is especially pronounced for the E-mini price process, but also happens for EUR/USD prices. This pattern is also confirmed by the AIC-selected orders of the MA($q$) observation error model, where the median selected order is 23 for the E-mini and 8 for the EUR/USD. Finally, when accounting for dependence in the observation error, its variance $\omega_q^2$ is approximately one order of magnitude larger for the E-mini than for the EUR/USD, suggesting a more significant role of observation error in E-mini variance estimation. For empirical applications presented in the main text, the bias correction applied is (45).
Table S2. Shows the median daily estimates of the respective quantities defined in previous subsections and (5%, 95%)-quantiles in brackets below. The units of the return parameters are in $\text{bp}^2$. $q^{\text{AIC}}$ indicates the AIC-optimal order of the MA(q) noise (45) according to the QML procedure in$^{37}$.

|     | $\hat{m}$  | $\hat{\omega}_1^2$ | $q^{\text{AIC}}$ | $\hat{\omega}_q^2$ |
|-----|-----------|---------------------|------------------|---------------------|
| E-mini | 0.20617  | -0.00033            | 23               | 0.01923            |
|      | (0.19473, 0.23518) | (-0.01202, 0.03213) | (6, 50)          | (0.00038, 0.04569) |
| EUR/USD | 0.09306  | 0.00427             | 8                | 0.00156            |
|      | (0.07671, 0.14265) | (-0.00672, 0.01578) | (3, 50)          | (0.00001, 0.01136) |

References

1. McLachlan, G. & Krishnan, T. The EM algorithm and extensions. Wiley series in probability and statistics (John Wiley & Sons, 2008), 2nd edn.
2. Marsan, D. & Lengline, O. Extending Earthquakes’ Reach Through Cascading. Science 319, 1076–1079, DOI: 10.1126/science.1148783 (2008).
3. Veen, A. & Schoenberg, F. Estimation of space-time branching process models in seismology using an EM-type algorithm. J. Amer. Stat. Assoc., 103 614–624, DOI: 10.1198/01621450800000148 (2008).
4. Lewis, E. & Mohler, G. A nonparametric EM algorithm for multiscale Hawkes processes. J. Nonparametric Stat. (2011).
5. Wheatley, S., Wehrli, A. & Sornette, D. The endo-exo problem in high frequency financial price fluctuations and rejecting criticality. Quant. Finance 19, 1165–1178, DOI: 10.1080/14697688.2018.1550266 (2020).
6. Wehrli, A., Wheatley, S. & Sornette, D. Scale-, time- and asset-dependence of Hawkes process estimates on high frequency price changes. Quant. Finance 21, 729–752, DOI: 10.1080/14697688.2020.1838602 (2021).
7. Wehrli, A. & Sornette, D. Classification of flash crashes using the Hawkes(p,q) framework. Quant. Finance 22, DOI: 10.1080/14697688.2021.1941212 (2021).
8. Wu, C. On the convergence properties of the EM algorithm. The Annals Stat. 11.1 95–103, DOI: 10.1214/aos/1176346060 (1983).
9. Ogata, Y. On Lewis’ simulation method for point processes. IEEE Transactions on Inf. Theory 27, 23–31, DOI: 10.1109/TIT.1981.1056305 (1981).
10. Kooperberg, C. & Stone, C. J. A study of logspline density estimation. Comput. Stat. & Data Analysis 12, 327–347, DOI: 10.1016/0167-9473(91)90115-I (1991).
11. Schatz, M., Wheatley, S. & Sornette, D. The ARMA Point Process and its Estimation. Econom. Stat. in press DOI: 10.1016/j.ecosta.2021.11.002 (2021).
12. Kim, D. K. & Taylor, J. M. The restricted EM algorithm for maximum likelihood estimation under linear restrictions on the parameters. J. Am. Stat. Assoc. 90, 708–716, DOI: 10.1080/01621459.1995.10476564 (1995).
13. Møller, J. & Waagepetersen, R. Statistical Inference and Simulation for Spatial Point Processes. Monographs on Statistics and Applied Probability 100 (Chapman & Hall/CRC: Boca Raton, FL, 2003).
14. Hardiman, S. J., Bercot, N. & Bouchaud, J. P. Critical reflexivity in financial markets: A Hawkes process analysis. Eur. Phys. J. B 86, 442, DOI: 10.1140/epjb/e2013-40107-3 (2013).
15. Bacry, E., Dayri, K. & Muzy, J.F. Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. Eur. Phys. J. B 85, 157, DOI: 10.1140/EPJB/E2012-21005-8 (2012).
16. Bacry, E., Delattre, S., Hoffmann, M. & Muzy, J. Some limit theorems for Hawkes processes and application to financial statistics. Stoch. Process. Appl. 123, 2475–2499, DOI: 10.1016/j.spa.2013.04.007 (2013).
17. Bacry, E. & Muzy, J.-F. First- and Second-Order Statistics Characterization of Hawkes Processes and Non-Parametric Estimation. IEEE Transactions on Inf. Theory 62, 2184–2202, DOI: 10.1109/TIT.2016.2533397 (2016).
18. Bacry, E., Mastromatteo, I. & Muzy, J.-F. Hawkes processes in finance. Mark. Microstruct. Liq. 1, 155005, DOI: 10.1142/S2382626615500057 (2015).
19. Hawkes, A. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**, 83–90, DOI: 10.1093/biomet/58.1.83 (1971).

20. Hawkes, A. Point Spectra of Some Mutually Exciting Point Processes. *J. Royal Stat. Soc. Ser. B (Methodological)* **33**, 438–443, DOI: 10.1111/j.2517-6161.1971.tb01530.x (1971).

21. Sornette, D. & Sornette, A. Renormalization of earthquake aftershocks. *Geophys. Res. Lett.* **26**, 1981–1984, DOI: 10.1029/1999GL900394 (1999).

22. Helmstetter, A. & Sornette, D. Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *J. Geophys. Res. Solid Earth* **107**, ESE 10–1 – ESE 10–21, DOI: 10.1029/2001JB001580 (2002).

23. Helmstetter, A. & Sornette, D. Diffusion of epicenters of earthquake aftershocks, Omori’s law, and generalized continuous-time random walk models. *Phys. Rev. E* **66**, 061104, DOI: 10.1103/PhysRevE.66.061104 (2002).

24. Sornette, D. & Helmstetter, A. Endogeneous Versus Exogeneous Shocks in Systems with Memory. *Phys. A* **318**, 577–591, DOI: 10.1016/S0378-4371(02)01371-7 (2003).

25. Saichev, A. & Sornette, D. Generation-by-generation dissection of the response function in long memory epidemic processes. *The Eur. Phys. J. B-Condensed Matter Complex Syst.* **75**, 343–355, DOI: 10.1140/epjb/e2010-00121-7 (2010).

26. Saichev, A., Maillart, T. & Sornette, D. Hierarchy of Temporal Responses of Multivariate Self-Excited Epidemic Processes. *Eur. Phys. J. B* **86**, 1–19, DOI: 10.1140/epjb/e2013-30493-9 (2013).

27. Hawkes, A. & Oakes, D. A Cluster Process Representation of a Self-Exciting Process. *J. Appl. Probab.* **11**, 493–503, DOI: 10.2307/3212693 (1974).

28. Jaisson, T. & Rosenbaum, M. Limit theorems for nearly unstable Hawkes processes. *The Annals Appl. Probab.* **25**, 600–631, DOI: 10.1214/14-AAP1005 (2015).

29. Jaisson, T. & Rosenbaum, M. Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *The Annals Appl. Probab.* **26**, 2860–2882, DOI: 10.1214/15-AAP1164 (2016).

30. Jaisson, T. & Rosenbaum, M. The Different Asymptotic Regimes of Nearly Unstable Autoregressive Processes. In Podolskij, M., Stelzer, R., Thorbjørnsen, S. & Veraart, A. E. D. (eds.) *The Fascination of Probability, Statistics and their Applications: In Honour of Ole E. Barndorff-Nielsen*, 283–301, DOI: 10.1007/978-3-319-25826-3_13 (Springer International Publishing, Cham, 2016).

31. Protter, P. E. *Stochastic Integration and Differential Equations* (Springer-Verlag Berlin Heidelberg, 2005), 2nd edn.

32. Barndorff-Nielsen, O. & Shephard, N. Estimating quadratic variation using realized variance. *J. Appl. Econom.* **17**, 457–477, DOI: 10.1002/jae.691 (2002).

33. Hansen, P. R. & Lunde, A. Realized Variance and Market Microstructure Noise. *J. Bus. & Econ. Stat.* **24**, 127–161, DOI: 10.1198/073500106000000071 (2006).

34. Filimonov, V. & Sornette, D. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E* **85**, 056108, DOI: 10.1103/PhysRevE.85.056108 (2012).

35. Oomen, R. C. Comment on 2005 JBES invited address “Realized variance and market structure noise” by Peter R. Hansen and Asger Lunde. *J. Bus. Econ. Stat.* **2**, 195–202 (2006).

36. Oomen, R. C. A. Properties of Realized Variance under Alternative Sampling Schemes. *J. Bus. & Econ. Stat.* **24**, 219–237, DOI: 10.1198/073500106000000044 (2006).

37. Da, R. & Xiu, D. When Moving-Average Models Meet High-Frequency Data: Uniform Inference on Volatility. *Econometrica* **89**, 2787–2825, DOI: 10.3982/ECTA15593 (2021).