Journal of Econometrics 238 (2024) 105522

Contents lists available at ScienceDirect

Journal of Econometrics

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

Systematic staleness[☆]

Federico M. Bandi[a], Davide Pirino[b,*], Roberto Renò[c]

[a] Johns Hopkins University, Carey Business School, 555 Pennsylvania Avenue, Washington DC 20001, USA
[b] Università degli Studi di Roma "Tor Vergata", Dipartimento di Economia e Finanza, Via Columbia 2, 00173 Roma, Italy
[c] ESSEC Business School, 3 Avenue Bernard Hirsch, 95021 Cergy-Pontoise, France

ARTICLE INFO

Keywords:
Systematic staleness
Idiosyncratic staleness
Liquidity
Joint asymptotics

ABSTRACT

Asset prices are stale. We define a measure of systematic (market-wide) staleness as the percentage of small price adjustments across multiple assets. A notion of idiosyncratic (asset-specific) staleness is also established. For both systematic and idiosyncratic staleness, we provide a limit theory based on *joint* asymptotics relying on increasingly-frequent observations over a fixed time span and an increasing number of assets. Using systematic and idiosyncratic staleness as moment conditions, we introduce novel structural estimates of systematic and idiosyncratic measures of liquidity obtained from transaction prices only. The economic signal contained in the structural estimates is assessed by virtue of suitable metrics.

1. Introduction

Asset prices do not update as frequently as one may believe. In particular, they do not update as often as one would assume given traditional modeling in continuous time. The lack of updating may have a component which is systematic and, therefore, pervasive across stocks. We argue that lack of price updates, and their cross-sectional correlation structure, can be economically just as informative as long spells of erratic price dynamics as induced by volatility or jumps, i.e., virtually the exclusive focus of the current, successful high-frequency literature (see, e.g., the rich review in Aït-Sahalia and Jacod, 2014). Hence, the provision of a methodological framework designed to model price staleness – and shed light on its economic content – appears warranted.

Staleness is related to lack of volume or low volume (Bandi et al., 2020). It must therefore carry information about frictions in the trading process and their determinants, such as the extent of liquidity. In light of this premise, this article begins with a definition of notions of *systematic* (market-wide) staleness and *idiosyncratic* (asset-specific) staleness for which we formulate estimators, continues with the provision of a limit theory for the proposed estimators and concludes with the discussion of novel staleness-based structural

☆ We are grateful to the Editor, the Associate Editor and three anonymous reviewers for their very helpful comments. We are also grateful to René Garcia and Jean-Sébastien Fontaine for very kindly sharing their funding liquidity measure as well as other measures used in their work. We thank the conference participants at the Boston University Conference "Recent Advances in Financial Econometrics: Identifying New Risks from Financial Data", September 10 2016, the Duke University Conference "New Developments in Measuring and Forecasting Financial Volatility", Durham, September 16–17 2016, the Annual International Conference of the Thailand Econometric Society, Chiang Mai, January 11–13 2017, the Vienna-Copenaghen Conference, Vienna, March 9–11 2017, the Conference "New Methods for the Empirical Analysis of Financial Markets", Comillas, June 9-10 2017, and the "Second Conference on New Trends and Methods in Econometrics", Llhavo - Aveiro, June 22–23 2018. We are especially thankful to the discussant Jia Li. Davide Pirino acknowledges partial support through the grant RBSI14DDNN *"A new measure of liquidity"* financed within the program SIR of the Italian Ministry of Education and Research. Davide Pirino and Roberto Renò acknowledge support from the project HiDEA (Advanced Econometric Methods for High-Frequency Data) financed by the Italian Ministry of Education, University and Research under the program PRIN 2017, Prot. 2017RSMPZZ. Roberto Renò is also grateful to the Institut Europlace de Finance Louis Bachelier for the 2013 EIF Grant *"A new measure of liquidity in financial markets"*. All errors are our own.

* Corresponding author.
E-mail addresses: fbandi1@jhu.edu (F.M. Bandi), davide.erminio.pirino@uniroma2.it (D. Pirino), reno@essec.edu (R. Renò).

https://doi.org/10.1016/j.jeconom.2023.105522
Received 18 July 2020; Received in revised form 18 November 2022; Accepted 9 February 2023
Available online 16 October 2023

measurements of systematic and idiosyncratic liquidity using transaction prices *only*. These structural estimates are shown to contain revealing economic signal.

Intuitively, systematic staleness is defined as lack of (large) price updates across many assets. In this sense, it represents pervasive declines of normal market activity. Idiosyncratic staleness, on the other hand, is asset-specific and may therefore occur in the absence of market-wide effects. Our proposed estimators of systematic and idiosyncratic staleness make solely use of high-frequency transaction prices for a cross-section of stocks and are constructed based on price adjustments smaller than a suitable, vanishing (in the limit) threshold. The estimates are functionals of two frequencies: the frequency of *joint* stale prices across multiple assets ("multivariate staleness") and the frequency of stale prices for *individual* assets ("univariate staleness"). While multivariate staleness will be shown to be a consistent estimator of systematic staleness, an appropriately-defined nonlinear functional of multivariate staleness and univariate staleness will define a consistent estimator of idiosyncratic staleness for each asset.

We study the limiting properties of our proposed estimators for systematic staleness and idiosyncratic staleness in an asymptotic environment in which the number of observations increases without bound (over a fixed time span) jointly with the number of assets. This is done under a null hypothesis in which the underlying price processes evolve as a vector semi-martingale with some (asset-specific or idiosyncratic) likelihood of repeated prices. The alternative is one in which there is both a likelihood of systematic staleness (not allowed under the null) and a likelihood of idiosyncratic staleness (allowed under the null) around, again, a driving semi-martingale price process. We show convergence in probability of the estimates (under both the null and the alternative) to (the probability of) systematic and idiosyncratic staleness (whether zero or not) as well as, under the null, root-n weak convergence to asymptotically normal random variables with estimable limiting variances.

Staleness is in contradiction with traditional (frictionless) approaches to the modeling of asset prices. Our interest in notions of staleness is, therefore, justified by price formation processes which account for the microstructure of market dynamics (as in, e.g., Kyle, 1985, and Glosten and Milgrom, 1985). To this extent, we estimate structurally (using both notions of staleness as moment conditions) a microstructure model featuring market makers (learning from trades), uninformed traders as well as informed traders (modulating their trading decisions on the basis of the magnitude of execution costs). In the model, the informed traders compare the size of execution costs to deviations between the efficient price (assumed to be known, given their information set) and the mid-quote of the bid/ask prices. Should these deviations be smaller (in absolute value) than the size of execution costs, the informed traders will opt out of trades, thereby leading to low volume, lack of price updates and, therefore, staleness.

In the model, execution costs have a systematic and an idiosyncratic component. The systematic component has an interpretation in terms of the "shadow cost" of capital, or funding liquidity, as in Brunnermeier and Pedersen (2009). The idiosyncratic component has, instead, an interpretation in terms of asset-specific effective spread, which we formalize explicitly. In order to provide support for this logic in the data, we compare our structural estimates of funding costs to old and new proxies of the same quantity.[1] The same is done with our structural estimates of effective spreads, which are compared to bid–ask spreads as well as to traditional measures of *effective* spreads and *realized* spreads, both cross-sectionally and in the time series.[2]

We conclude this Introduction with an important observation. There is an empirical finance literature which employs low-frequency measures of staleness (such as the number of zeros in the data over a specific time period) as liquidity proxies (e.g., Lesmond, 2005; Bekaert et al., 2007). Even though the simplicity of these measures and their logic justifies doing so, accepted theories of price formation imply that – at the minimum – the interplay between liquidity (as represented by the cost(s) of execution), learning, information asymmetries and efficient price volatility is bound to affect them and add noise to signal. In light of this, we do not proxy for liquidity directly using the proposed staleness measures. Instead, the staleness measures assist in the extraction of liquidity proxies from the estimation of a suitable structural pricing model. Our proposed high-frequency structural approach is shown (both in the data and in simulation) to yield revealing signal(s) and provide effective identification of liquidity-related quantities, as well as of other features of the price formation process. In our framework, structural identification builds crucially on the informational content of preliminary high-frequency estimates of systematic and idiosyncratic staleness, estimates for which the present article provides both a definition and a limiting theory under joint asymptotics.

We proceed as follows. Section 2 introduces multivariate and univariate staleness, i.e., the inputs of systematic and idiosyncratic staleness. Section 3 contains asymptotic results for multivariate and univariate staleness leading to estimators of systematic and idiosyncratic staleness whose asymptotic properties are discussed in a *joint infill/large cross-section* asymptotic environment. While the relation between systematic and idiosyncratic volatility is well-understood (see, e.g., Herskovic et al., 2016), the corresponding staleness measures are novel and, hence, deserving of a purely descriptive analysis. Using 250 NYSE-listed stocks, we provide such an analysis in Section 4. Among other results, we show that systematic staleness is positively correlated with the mean of idiosyncratic staleness and the dynamics of idiosyncratic staleness are suggestive of the presence of a common factor. Both findings are reminiscent of analogous results in the volatility literature. We also show that both systematic staleness and the mean of idiosyncratic staleness are strongly negatively correlated with mean volumes, a fact which provides empirical support for the link between staleness and trading dynamics. Section 5 details a market microstructure model with asymmetries in information which we estimate structurally using systematic and idiosyncratic staleness as moment conditions. Day-by-day estimation of the model leads to daily proxies of effective spreads and funding liquidity, among other quantities. In Section 6, we compare these structural liquidity measures to

[1] Papers devoted to the analysis of systematic liquidity are, *inter alia*, Chordia et al. (2000, 2001), Hasbrouck and Seppi (2001), Pastor and Stambaugh (2003), Kamara et al. (2008), Næs et al. (2011), Fontaine and Garcia (2012) and Hu et al. (2013).

[2] Papers devoted to the analysis of idiosyncratic liquidity are, *inter alia*, Roll (1984), Lesmond et al. (1999), Amihud (2002), Hasbrouck (2004, 2009) and Holden (2009). Goyenko et al. (2009) provide a rich discussion, empirical comparisons and new estimators.

an array of existing proxies. Section 7 evaluates their economic content using a cross-sectional pricing metric. Section 8 concludes. Appendix A contains proofs. Appendix B provides Monte Carlo simulations.

2. Definitions

We begin with preliminaries. We define and justify univariate and multivariate staleness. These measures will be the key ingredients of the proposed estimators.

Definition 2.1. A smoother is an integrable function $\mathcal{S}(\cdot) : \mathbb{R}^+ \to (0,1]$ with a bounded first derivative and such that $\mathcal{S}(0) = 1$.

Definition 2.2. Let $t \in [0,1]$ and let $\mathbf{X}_t$ be a vector of N real-valued stochastic processes whose qth component is denoted by the process $X_t^{(q)}$. For all $q = 1,\dots,N$, let $\Theta_{t,n}^{(q,N)}$ be a threshold defined as

$$\Theta_{t,n}^{(q,N)} \doteq h_n^{(N)}\, \xi_{t,n}^{(q)}, \tag{1}$$

where $\xi_{t,n}^{(q)}$ is a bounded positive adapted stochastic process on $[0,1]$ and $h_n^{(N)}$ is a positive double sequence of real numbers infinitesimal both in n and N.[3] Assume that each of the N processes $X_t^{(q)}$ and the threshold $\Theta_{t,n}^{(q,N)}$ are observed on the evenly-spaced time grid $t_{j,n} = j\,\Delta_n$ with $\Delta_n = 1/n$ and $j = 0,\dots,n$. Denote by $\left(X_{j\,\Delta_n}^{(q)}\right)_{j=0,\dots,n}$ and $\left(\Theta_{j\,\Delta_n,n}^{(q,N)}\right)_{j=0,\dots,n}$ the qth process and the threshold sampled on the grid. Finally, let $\mathcal{S}(\cdot) : \mathbb{R}^+ \to (0,1]$ be the smoother in Definition 2.1.

The univariate staleness estimator of the generic qth process at frequency Δ_n is defined as

$$\mathbb{U}_n^{(q)} \doteq \Delta_n \sum_{j=1}^{n} \mathcal{S}\left(\frac{\left|X_{j\,\Delta_n}^{(q)} - X_{(j-1)\,\Delta_n}^{(q)}\right|}{\Theta_{j\,\Delta_n,n}^{(q,N)}}\right). \tag{2}$$

The N-multivariate staleness estimator at frequency Δ_n is defined as

$$\mathbb{M}_n^{(N)} \doteq \Delta_n \sum_{j=1}^{n} \prod_{q=1}^{N} \mathcal{S}\left(\frac{\left|X_{j\,\Delta_n}^{(q)} - X_{(j-1)\,\Delta_n}^{(q)}\right|}{\Theta_{j\,\Delta_n,n}^{(q,N)}}\right)^{1/N}. \tag{3}$$

Notwithstanding the use of a smooth function $\mathcal{S}(|x|)$, instead of an indicator function $\mathbb{I}_{\{|x|\le 1\}}$, Eq. (2) coincides with the notion of idle time[4] presented and studied in Bandi et al. (2017). (Here, and throughout the article, we denote by $\mathbb{I}_{\{A\}}$ the indicator function associated with the set A.) Replacing $\mathcal{S}(|x|)$ with $\mathbb{I}_{\{|x|\le 1\}}$, however, clarifies the logic of this estimator. If all of the price adjustments were below the threshold (a case of extreme staleness), then $\mathbb{U}_n^{(q)}$ would be exactly 1. On the other hand, if all of the price adjustments were outside of the threshold, the indicator would be zero, and the measure would be zero. Intermediate cases of sluggish price adjustments would give rise to values between 0 and 1. While the selection criterion implied by the indicator has a binary nature (since the price changes are either inside or outside the threshold), a smooth kernel would give rise to continuous transitions. The same logic would, however, apply. In light of these observations, the interpretation of $\mathbb{M}_n^{(N)}$ in Eq. (3) as a multivariate staleness estimator is natural. The estimator measures the joint probability of stale price updates.

We will show that $\mathbb{M}_n^{(N)}$ directly identifies our notion of systematic staleness (defined in Hypothesis $\mathcal{H}_A$ below) when $N, n \to \infty$ jointly (c.f., Theorem 1). Our proposed notion of idiosyncratic staleness (defined in Hypothesis $\mathcal{H}_0$ below) will, instead, require the use of a functional of $\mathbb{M}_n^{(N)}$ and $\mathbb{U}_n^{(q)}$ for consistent estimation, again as $N, n \to \infty$ jointly (c.f., the Corollary to Theorem 1).

3. Systematic and idiosyncratic staleness: limit theory

We assume existence of an efficient price vector process satisfying the following assumption.

Assumption 1 (*The Efficient Price Vector Process*). *There exist N real-valued (logarithmic) efficient price processes*

$$\left\{ \widetilde{X}_t^{(q)};\ t \ge 0;\ q = 1,\dots,N \right\},$$

defined on the usual probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{P}\right)$ (Protter, 2004) each of which is a Brownian semimartingale

$$d\widetilde{X}_t^{(q)} = \mu_t^{(q)}\,dt + \sigma_t^{(q)}\,dW_t^{(q)},$$

[3] Because of the presence of the term $\xi_{t,n}^{(q)}$, the definition of the threshold allows, e.g., for dependence on the spot volatility of the underlying efficient price process (defined in Assumption 1), something which is conceptually and empirically important.

[4] Bandi et al. (2017) define, without reference to volume levels, their measure of staleness as *idle time* or, in its demeaned version, *excess idle time*. In Bandi et al. (2020), the adopted nomenclature distinguishes between lack of price updates without conditioning on volume and lack of price updates conditional on the absence of volume. Specifically, the word "staleness" is associated with lack of price updates, whereas the word "idleness" (resp. "near idleness") is used to define lack of price updates without volume (resp. with very limited volume). While Bandi et al. (2020) show that the two notions are very closely related (in that staleness is associated with declines in trading activity), we abide strictly by the latter nomenclature and use the terminology staleness when information on volumes is not provided. As a consequence, for clarity, Eq. (2) is defined as (an estimator of) "univariate staleness" rather than as (an estimator of) "univariate idle time". The same reasoning applies to Eq. (3).

where $W_t^{(q)}$ is a standard $\mathcal{F}$-Brownian motion and $\mu_t^{(q)}$ and $\sigma_t^{(q)} > 0$ are càdlàg $\mathcal{F}$-adapted bounded processes satisfying the differentiability conditions in Bandi and Renò (2017). For each pair $\left(W_t^{(q)}, W_t^{(k)}\right)$ of Brownian motions, with $q \neq k$, there exists a process $\rho_t^{(q,k)} \in [-1, 1]$ such that

$$\mathbb{E}_t\left[dW_t^{(q)}\, dW_t^{(k)}\right] = \rho_t^{(q,k)}\, dt.$$

We now turn to the frictional process. Under the null, we allow for some likelihood of repeated prices. If a price is not repeated, the underlying efficient price in Assumption 1 is observed. Since, under the null, only idiosyncratic staleness is allowed, the likelihood of repeated prices is modeled as being independent across assets.

Hypothesis $\mathcal{H}_0$ (*The Observed Price Process Under the Null*)**.** The collection of N observed (logarithmic) price processes

$$\left\{X_{j\,\Delta_n}^{(q)};\ j = 0, \dots, n;\ q = 1, \dots, N\right\}$$

on the time grid $t_{j,n} = j\,\Delta_n$ is such that $X_0^{(q)} = \widetilde{X}_0^{(q)}$ and, for $j = 1, \dots, n$,

$$X_{j\,\Delta_n}^{(q)} = \widetilde{X}_{j\,\Delta_n}^{(q)}\left(1 - B_{j,n}^{(q)}\right) + B_{j,n}^{(q)}\, X_{(j-1)\,\Delta_n}^{(q)}, \tag{4}$$

where $B_{j,n}^{(q)}$, with $q = 1, \dots, N$, are N triangular arrays of $\mathcal{F}_{t_j}$-measurable Bernoulli variates. These variates are so that

$$p_n^{(q)} \doteq \mathbb{P}\left[B_{j,n}^{(q)} = 1\right] = \mathbb{E}\left[B_{j,n}^{(q)}\right] \underset{n\to\infty}{\to} p_\infty^{(q)} \in (0, 1)$$

and are pairwise independent, that is, for all $i_1 \neq i_2$ and for all $j, k = 1, \dots, n$,

$$\mathbb{P}\left[B_{j,n}^{(i_1)} = a,\ B_{k,n}^{(i_2)} = b\right] = \mathbb{P}\left[B_{j,n}^{(i_1)} = a\right]\, \mathbb{P}\left[B_{k,n}^{(i_2)} = b\right],$$

with $a \in \{0, 1\}$ and $b \in \{0, 1\}$. Also,[5]

$$\mathbb{V}\left[\frac{1}{n}\sum_{j=1}^{n} B_{j,n}^{(q)}\right] \underset{n\to\infty}{\to} 0 \quad q = 1, \dots, N, \qquad \mathbb{V}\left[\frac{1}{n}\sum_{j=1}^{n}\prod_{q=1}^{N} B_{j,n}^{(q)}\right] \underset{n\to\infty}{\to} 0 \quad \forall N. \tag{5}$$

Moreover, for each $j = 1, \dots, n$, and each $q = 1, \dots, N$, let $f_j^{(q)}$ be the number of consecutive flat trades before instant $t_{j-1,n} = (j-1)\,\Delta_n$ for the qth process. We assume that $F_n^{(q)} \doteq \max_{j=1,\dots,n} f_j^{(q)}$ is such that[6]

$$\frac{F_n^{(q)} \log n}{n^{\alpha}} \underset{n\to\infty}{\overset{p}{\longrightarrow}} 0, \quad q = 1, \dots, N, \tag{6}$$

with $\alpha < 1/2$.

For each asset, the null allows for staleness with probability equal to $p_n^{(q)}$. Since the Bernoulli variates are independent across assets, the type of staleness captured by the null is purely idiosyncratic. The probability of staleness (whether idiosyncratic or, as presented below, systematic) is assumed to be *frequency-specific*, a modeling device which replicates the dependence on frequency of the likelihood of price updates (or lack thereof) which has been observed in the data (Bandi et al., 2020). We note that the pricing model employed by Phillips and Yu (2023) to study realized variance can be expressed as the model in Eq. (4) with independent (over time) Bernoulli variates and a constant, rather than frequency-specific, probability of repeated prices.

The null in the present article corresponds to the alternative in Bandi et al. (2017). Bandi et al. (2017) provide a framework to test for absence of staleness. Should absence of staleness be rejected, something that Bandi et al. (2017) and Bandi et al. (2020) do forcefully for the market and a cross-section of stocks respectively, staleness should (at least in its idiosyncratic version) be a feature of the data generating process under the null. We take this view here.

The alternative hypothesis is, instead, constructed to allow for the presence of a systematic staleness component. This systematic component represents the substantive core of this article's contribution. In what follows, the superscript (S) stands for "systematic." Next, we formulate the alternative.

Hypothesis $\mathcal{H}_A$ (*The Observed Price Process Under the Alternative*)**.** There exists a triangular array of Bernoulli variates $C_{j,n}^{(\mathrm{S})}$ such that the observed price process sampled on the evenly-spaced time grid $t_{j,n} = j\,\Delta_n$ is given by

$$X_{j\,\Delta_n}^{(q)} = \left(1 - C_{j,n}^{(\mathrm{S})}\right)\left(\widetilde{X}_{j\,\Delta_n}^{(q)}\left(1 - B_{j,n}^{(q)}\right) + X_{(j-1)\,\Delta_n}^{(q)}\, B_{j,n}^{(q)}\right) + C_{j,n}^{(\mathrm{S})}\, X_{(j-1)\,\Delta_n}^{(q)}, \tag{7}$$

with $X_0^{(q)} = \widetilde{X}_0^{(q)}$ for $q = 1, \dots, N$. The variates $C_{j,n}^{(\mathrm{S})}$ are independent of $B_{j,n}^{(q)}$ for all $j = 1, \dots, n$ (and for all $q = 1, \dots, N$) and

$$p_n^{(\mathrm{S})} \doteq \mathbb{P}\left[C_{j,n}^{(\mathrm{S})} = 1\right] = \mathbb{E}\left[C_{j,n}^{(\mathrm{S})}\right] \underset{n\to\infty}{\to} p_\infty^{(\mathrm{S})} \in (0, 1)\,.$$

5 The conditions in Eq. (5), and the analogous conditions in Eq. (8) and (9), are trivially satisfied if the Bernoulli variates are temporally independent. With these conditions, we are solely excluding extreme forms of dependence.

6 The condition in Eq. (6) is easily satisfied by iid Bernoulli sequences since, in this case, $F_n^{(q)} = O_p(\log n)$ (Schilling, 1990).

Also,

$$\mathbb{V}\left[\frac{1}{n}\sum_{j=1}^{n} C_{j,n}^{(\mathrm{S})}\right] \underset{n\to\infty}{\to} 0, \tag{8}$$

and

$$\mathbb{V}\left[\frac{1}{n}\sum_{j=1}^{n} C_{j,n}^{(\mathrm{S})} B_{j,n}^{(q)}\right] \underset{n\to\infty}{\to} 0 \quad q=1,\ldots,N, \quad \mathbb{V}\left[\frac{1}{n}\sum_{j=1}^{n} C_{j,n}^{(\mathrm{S})} \prod_{q=1}^{N} B_{j,n}^{(q)}\right] \underset{n\to\infty}{\to} 0 \quad \forall N. \tag{9}$$

Finally, we assume $C_{j,n}^{(\mathrm{S})}$ is such that the property on the maximum of consecutive flat trades in Eq. (6) is satisfied.

Under the alternative, when $C_{j,n}^{(\mathrm{S})} = 1$, $X_{j\Delta_n}^{(q)} = X_{(j-1)\Delta_n}^{(q)}$ for *all* $q = 1,\ldots,N$. When $C_{j,n}^{(\mathrm{S})} = 0$, the observed price process $X_{j\Delta_n}^{(q)}$ may or may not equal $X_{(j-1)\Delta_n}^{(q)}$ depending on the idiosyncratic Bernoulli variates $B_{j,n}^{(q)}$. Thus, the alternative has a hierarchical structure. Prices may repeat themselves due to systematic effects. If they do not, there continues to be an asset-specific likelihood of repeated prices due to idiosyncratic effects. We denote by $p_n^{(\mathrm{S})}$ the *frequency-specific* probability of systematic staleness.

Because (systematic and/or idiosyncratic) Bernoulli shocks drive the limiting results under both the null and the alternative, one could dispense with the use of a threshold $\Theta_{t,n}^{(q,N)}$ (see Eq. (1)) in the definition of both $\mathbb{U}_n^{(q)}$ and $\mathbb{M}_n^{(N)}$. While the use of a threshold adds technical subtleties to the proofs, we believe it is warranted for two reasons. First, it preserves the applicability of the estimators in situations in which the null model is continuous (as in Bandi et al., 2017) and the null hypothesis of absence of staleness is tested. Second, empirically, it allows the estimates to capture *near* staleness, rather than solely *exact* staleness, irrespective of the null model. We emphasize that, should one dispense with the use of a threshold, Assumption 1 could be relaxed to an assumption of "existence" of the continuous (vector) efficient price process without emphasis on its dynamic evolution, and all of our results would go through unchanged.[7]

We note that, *without* a threshold, we would be measuring joint and individual "zero" returns. There is a successful literature in finance which has used zero low-frequency returns as proxies for liquidity (e.g., Lesmond, 2005; Bekaert et al., 2007). In the absence of a threshold, our work may be viewed as making four contributions to the literature on "zeros". First, we provide high-frequency multidimensional and unidimensional counterparts for these (low-frequency) estimates in continuous time. Second we illustrate how our high-frequency estimates can be re-combined to obtain genuinely systematic and idiosyncratic measures. Third, we offer an inferential theory for these measures. Finally, after recognizing that accepted market microstructure models provide a logical link between "zeros" and liquidity, we discuss how they do so in imperfect ways due to the presence of additional effects (learning and asymmetric information, *inter alia*). We, however, document how our measures can be used to extract structural information about liquidity, jointly with an array of other quantities. The latter contribution is logically related with the LOT measure of Lesmond et al. (1999) who, differently from our approach in Section 5, use low-frequency "zeros" and Probit-type estimation on a threshold model for returns to identify deeper parameters representing percentage transaction costs.

We now turn to notions of consistency for both multivariate and univariate staleness. The threshold will be required to vanish at a faster rate than the modulus of continuity of the underlying Brownian shocks. The requirement is intuitive. Should the condition not be satisfied, then the Brownian updates would be deemed to be "small," something which would be contrary to the type of friction-induced staleness that we are aiming to capture.

Theorem 1 (*The Systematic Staleness Estimator*). *Under Assumption 1 and if, for a fixed N, $h_n^{(N)}$ in Eq. (1) of Definition 2.2 is such that $h_n^{(N)}\sqrt{n} \to 0$, we have, as $n \to \infty$, that*

$$\mathbb{U}_n^{(q)} \xrightarrow{p} \begin{cases} p_\infty^{(q)} & \text{under } \mathcal{H}_0 \\ p_\infty^{(\mathrm{S})} + (1-p_\infty^{(\mathrm{S})})p_\infty^{(q)} & \text{under } \mathcal{H}_A \end{cases}, \quad q=1,\ldots,N,$$

and

$$\mathbb{M}_n^{(N)} \xrightarrow{p} \begin{cases} \prod_{q=1}^{N} p_\infty^{(q)} & \text{under } \mathcal{H}_0 \\ p_\infty^{(\mathrm{S})} + (1-p_\infty^{(\mathrm{S})})\prod_{q=1}^{N} p_\infty^{(q)} & \text{under } \mathcal{H}_A \end{cases}.$$

In addition, if $h_n^{(N)}$ is such that

$$h_n^{(N)} \sim \frac{\beta^{N^{1+\eta}}}{n^{\gamma}} \text{ with } n \sim a^N \tag{10}$$

and

$$0<\beta<1,\ \eta>1,\ \gamma>\frac{1}{2},\ a>1, \tag{11}$$

we have, as $N \to \infty$, that

$$\mathbb{M}_n^{(N)} \xrightarrow{p} \begin{cases} 0 & \text{under } \mathcal{H}_0 \\ p_\infty^{(\mathrm{S})} & \text{under } \mathcal{H}_A \end{cases}. \tag{12}$$

[7] In particular, one could easily accommodate rich forms of market microstructure noise without affecting the asymptotic findings. We analyze the finite-sample impact of frictions on our smoother-based estimates in Appendix B.

Proof. *See Appendix A.*

Under the null, each estimator $\mathbb{U}_n^{(q)}$ (with $q = 1,\dots,N$) is consistent for the corresponding idiosyncratic probability of staleness. Similarly, for any fixed N, the product estimator $\mathbb{M}_n^{(N)}$ is consistent for the product of these probabilities (given the independence of the driving Bernoulli variates).

Under the alternative, the unconditional (limiting, for $n \to \infty$) probability of staleness for each individual stock is given by $\mathbb{P}^{(q)}(\text{staleness}) = p_\infty^{(\mathrm{S})} + \left(1 - p_\infty^{(\mathrm{S})}\right) p_\infty^{(q)}$. Specifically, staleness could be induced by systematic effects (whose limiting probability is $p_\infty^{(\mathrm{S})}$). If it is not induced by systematic effects, there is a chance $p_\infty^{(q)}$ that it may be induced by idiosyncratic effects. Each $\mathbb{U}_n^{(q)}$ is consistent for $\mathbb{P}^{(q)}(\text{staleness})$. For any fixed N, the product estimator is analogously consistent for $\mathbb{P}^{(prod)}(\text{staleness}) = p_\infty^{(\mathrm{S})} + \left(1 - p_\infty^{(\mathrm{S})}\right) \prod_{q=1}^N p_\infty^{(q)}$.

Thus, for a fixed N, univariate and multivariate staleness do not permit separation of systematic and idiosyncratic staleness. However, taking also limits with respect to an increasing number of assets (i.e., as $N \to \infty$, jointly with $n \to \infty$) is effective – under conditions – in identifying $p_\infty^{(\mathrm{S})}$. Theorem 1, in fact, states that, as $n, N \to \infty$ jointly and at the right rate, $\mathbb{M}_n^{(N)}$ is a consistent estimator of the probability of systematic staleness under both the null (in which case such probability is zero) and the alternative. A consequence of Theorem 1, spelled out in the following corollary, is that the idiosyncratic probabilities of staleness $p_\infty^{(q)}$ can, also, be estimated consistently both under the null and the alternative.

We note that, because the number of intra-period observations n is modeled as a function of the number of assets N in Eq. (10), from now on we will dispense with the statement "$N, n \to \infty$ jointly" and just write "$N \to \infty$," in order to avoid clutter.

Corollary to Theorem 1 (*The Idiosyncratic Staleness Estimator(s)*). *Under Assumption 1, Eqs. (10) and (11), as $N \to \infty$, Slutsky's theorem implies that*

$$\mathbb{Id}_n^{(q,N)} \doteq \frac{\mathbb{U}_n^{(q)} - \mathbb{M}_n^{(N)}}{1 - \mathbb{M}_n^{(N)}} \xrightarrow{p} p_\infty^{(q)}, \tag{13}$$

under both $\mathcal{H}_0$ and $\mathcal{H}_A$, for all $q = 1,\dots,N$.

We now turn to the limiting distributions of both $\mathbb{M}_n^{(N)}$ and $\mathbb{Id}_n^{(q,N)}$ for $q = 1,\dots,N$.

Theorem 2 (*Weak Convergence of the Systematic and Idiosyncratic Staleness Estimators*). *Let Assumption 1 hold. Assume independence of $C_{j,n}^{(\mathrm{S})}$ and $B_{j,n}^{(q)}$ across j values. Let, also, $h_n^{(N)}$ be defined as in Eqs. (10) and (11) of Theorem 1 with*

$$\sqrt{a} < 1/\overline{p}_\infty^\infty \tag{14}$$

and

$$a > 1/\underline{p}_\infty^\infty, \tag{15}$$

where $\overline{p}_\infty^\infty = \limsup_{N\to\infty} \max_{q=1,\dots,N} p_\infty^{(q)} \in (0,1)$ and $\underline{p}_\infty^\infty = \liminf_{N\to\infty} \min_{q=1,\dots,N} p_\infty^{(q)} \in (0,1)$ with $\left(\overline{p}_\infty^\infty\right)^2 < \underline{p}_\infty^\infty$. Then, under $\mathcal{H}_0$, as $N \to \infty$, we have that

$$\mathbb{Z}_n^{(N)} \doteq \frac{\mathbb{M}_n^{(N)} - P_n^{(N)}}{\sigma_n^{(N)}} \xrightarrow{d} N(0,1), \tag{16}$$

where

$$P_n^{(N)} \doteq \prod_{q=1}^N p_n^{(q)}, \quad \sigma_n^{(N)} \doteq \left(\frac{P_n^{(N)} - \left(P_n^{(N)}\right)^2}{n}\right)^{1/2}.$$

Under $\mathcal{H}_A$, as $N \to \infty$, we instead have that

$$\mathbb{Z}_n^{(N)} \xrightarrow{p} \infty. \tag{17}$$

Also, under $\mathcal{H}_0$, as $N \to \infty$,

$$\left(\frac{n}{p_n^{(q)}\left(1 - p_n^{(q)}\right)}\right)^{1/2} \left(\mathbb{Id}_n^{(q,N)} - p_n^{(q)}\right) \xrightarrow{d} N(0,1), \tag{18}$$

and, under $\mathcal{H}_A$, as $N \to \infty$,

$$\left(\frac{n\left(1 - p_n^{(\mathrm{S})}\right)}{p_n^{(q)}\left(1 - p_n^{(q)}\right)}\right)^{1/2} \left(\mathbb{Id}_n^{(q,N)} - p_n^{(q)}\right) \xrightarrow{d} N(0,1). \tag{19}$$

Proof. *See Appendix A.*

Limiting variance estimation is discussed in a remark.

Remark 1. Using Theorem 1, it is immediate to derive the feasible versions of the central limit theorems stated in Theorem 2. Specifically, Eqs. (13) and (16) imply that, under $\mathcal{H}_0$,

$$\sqrt{n}\,\frac{\mathbb{M}_n^{(N)}-\prod_{q=1}^{N}\mathbb{Id}_n^{(q,N)}}{\sqrt{\prod_{q=1}^{N}\mathbb{Id}_n^{(q,N)}-\left(\prod_{q=1}^{N}\mathbb{Id}_n^{(q,N)}\right)^2}}\xrightarrow{d}N(0,1)\,.$$

Under both $\mathcal{H}_0$ and $\mathcal{H}_A$, given Eqs. (12), (13), (18) and (19), we have that

$$\left(\frac{n\left(1-\mathbb{M}_n^{(N)}\right)}{\mathbb{Id}_n^{(q,N)}\left(1-\mathbb{Id}_n^{(q,N)}\right)}\right)^{1/2}\left(\mathbb{Id}_n^{(q,N)}-p_n^{(q)}\right)\xrightarrow{d}N(0,1)\,.$$

The relation between systematic and idiosyncratic volatility has been studied (see, e.g., Herskovic et al., 2016, and the references therein). The corresponding staleness measures are novel and, hence, deserving of a preliminary descriptive analysis, to which we now turn. The accuracy of the reported limiting results is evaluated in Appendix B by simulation.

4. A descriptive look at staleness

We employ a dataset whose constituents are the 250 most liquid (in terms of average transaction volume during the period considered) NYSE-listed stocks. We are focusing on large and liquid stocks, i.e., the universe of stocks which should be affected by staleness *the least*. Because our goal is to illustrate the informational content of staleness, our emphasis on the most liquid NYSE-listed stocks leads to conservative implications and is, therefore, economically revealing about the relevance of staleness, in general. For each stock, we have trades from January 2006 to December 2014.

Idiosyncratic and systematic staleness are computed, for each day, using 10-second returns constructed from transaction prices using previous tick interpolation. The number of 10-second returns n per day is, therefore, 2,340 and the number of days N_D is 2,265. The thresholds are chosen as $\Theta_{t,n}^{(q,N)}=\Theta_n^{(q)}=\alpha\,\sigma^{(q)}/n^{1/2}$, for all $q=1,2,\dots,N$, where $\alpha=\frac{1}{10}$ and $\sigma^{(q)}$ is the estimated volatility over the day for the stock indexed by q (i.e., the square root of 5-minute daily realized variance). The smoother $\mathcal{S}(\cdot)=\exp(-|\cdot|)$ is employed to obtain all estimates.

For each day $d=1,\dots,N_D$ in the sample, we proceed in the following way. Let $t_{j,n}=j\Delta_n=j/n$, $j=0,\dots,n$, be the 10-second partition for the day $[d-1,d]$. At each instant $t_{j,n}$ we compute the product

$$\zeta_{d,j}\doteq\prod_{q=1}^{N_{d,j}}\mathcal{S}\left(\frac{\left|X_{j/n}-X_{(j-1)/n}\right|}{\Theta_n^{(q)}}\right)^{1/N_{d,j}}=\prod_{q=1}^{N_{d,j}}\exp\left(-\frac{\left|X_{j/n}-X_{(j-1)/n}\right|}{N_{d,j}\,\Theta_n^{(q)}}\right),$$

where $N_{d,j}$ is the number of stocks available[8] in the time interval $[t_{j-1,n},t_{j,n}]$ of day d. We estimate systematic staleness for day d by computing

$$\hat{p}_d^{(\mathrm{S})}\doteq\frac{1}{n_d}\sum_{j=1}^{n_d}\zeta_{d,j}\,\mathbb{I}_{\{N_{d,j}\geq 60\}},\quad n_d\doteq\sum_{j=1}^{n}\mathbb{I}_{\{N_{d,j}\geq 60\}},\tag{20}$$

where, in order to reduce finite-sample distortions, we have imposed to consider only the instants of the 10-second partition during which at least 60 stocks are available.[9]

Having estimated $p_{d,\infty}^{(\mathrm{S})}$ for day d by virtue of $\hat{p}_d^{(\mathrm{S})}$ we derive, for each available asset, the corresponding estimate of $p_{d,\infty}^{(q)}$ using Eq. (13). More precisely, we compute the estimator

$$\hat{p}_d^{(q)}\doteq\frac{\overline{\mathbb{U}}_{d,n}^{(q)}-\hat{p}_d^{(\mathrm{S})}}{1-\hat{p}_d^{(\mathrm{S})}},\quad q=1,\dots,N,\tag{21}$$

where $\overline{\mathbb{U}}_{d,n}^{(q)}$ is the day d *rounding-adjusted*[10] percentage of stale returns at the 10-second frequency.

8 The cross section may vary from day to day due to rare missing stocks.

9 This precautionary rule is far from binding. With the exception of a handful of days in the sample, we discard, on average, only 0.17% of the returns in the 10-second partition.

10 On a 10-second partition, rounding may inflate the percentage of zero returns, thereby inducing a bias in the estimation of the idiosyncratic probability of staleness. In light of this observation, we adopt the procedure described in Bandi et al. (2020) to purge the observed percentage of stale returns from finite-sample contaminations due to rounding. We discuss rounding explicitly in Appendix B.

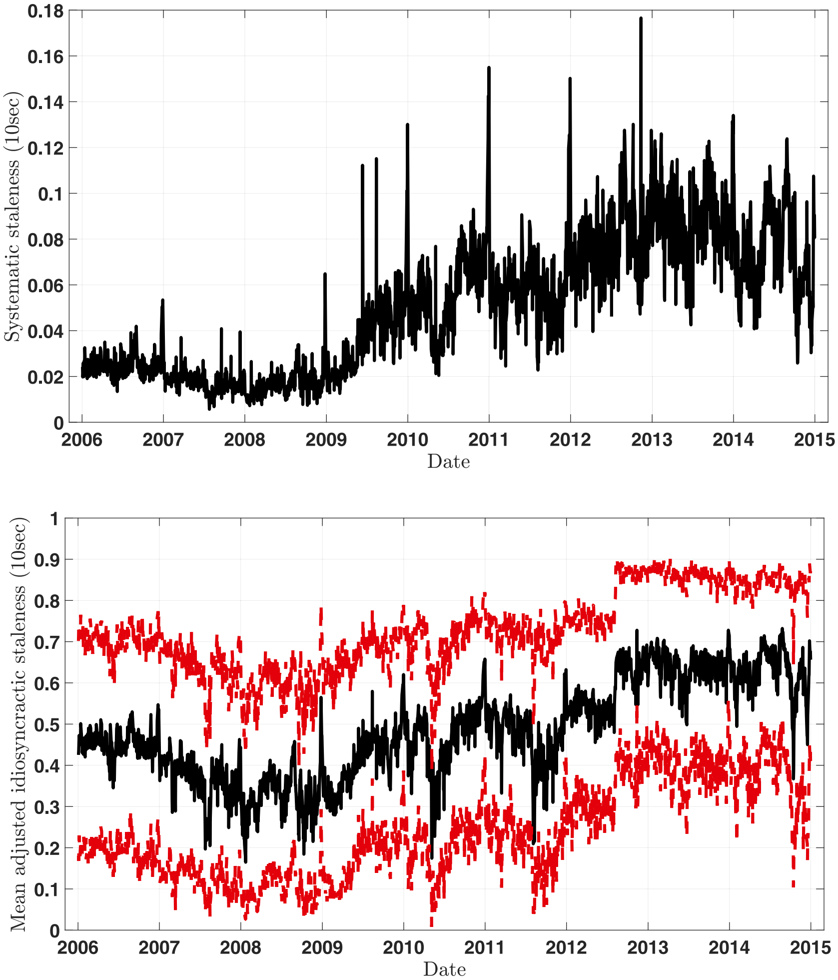

Fig. 1. (Top panel.) Daily time series of $\hat{p}_d^{(S)}$. (Bottom panel.) Daily time series of $\text{mean}_q \hat{p}_d^{(q)}$, i.e. the mean, across all stocks, of daily idiosyncratic staleness. The idiosyncratic staleness estimates are adjusted for rounding using the procedure in Bandi et al. (2020). The red lines are the 5th and the 95th percentiles of the cross-sectional estimates. (For interpretation of the references to color in this caption, the reader is referred to the web version of this article.)

Fig. 1 plots values of $\hat{p}_d^{(S)}$ in the top panel and cross-sectional averages of $\hat{p}_d^{(q)}$ in the bottom panel.[11] We do not report standard errors as they are extremely tight. In the idiosyncratic case, we however report, along with the cross-sectional averages, the 5th and the 95th percentiles of the cross-sectional values. Systematic staleness features a low-frequency upward trend directly related to the low-frequency downward trend in average volume over the same time period. Because the low-frequency trend in average volume is known to be due to factors (like reductions in bank's proprietary trading) other than increased illiquidity,[12] we remove the trend when conducting structural estimation in the next section. Systematic staleness also appears to be more volatile in recent periods. We do not find compelling evidence suggesting that this increased volatility is due to volatility spill-overs from volume. We, however, recall that the systematic staleness estimator is, in essence, a sample frequency intended to measure a daily probability. Its variability, as an estimator, is therefore bound to increase as the underlying true probability – and its estimate – rise towards 0.5, something which is consistent with the documented increases in more recent times.

The series are persistent (with first-order autocorrelations near 90%) and highly correlated (with correlation equal to 87%). Table 1 displays the regression coefficients associated with past daily staleness, past staleness over a week (i.e., the average over the past 5 days) and past staleness over a month (i.e., the average over the past 22 days), along with the corresponding t-statistics,

[11] In principle, one could extract systematic staleness from an index, like the S&P 500. After doing so, we found an R^2 between the staleness of the index and our notion of systematic staleness of 1.7%. Theoretically, while times of market-wide staleness (as implied by our measure) must necessarily be associated with staleness of the index, the index could be stale even when individual assets are not because of suitable cancellations in its portfolio structure. Consistent with this observation, we document larger values of staleness for the S&P 500 index impacting levels as well as dynamics (as evidenced by the low R^2) relative to our measure.

[12] We thank the Editor for pointing this out to us.

Table 1

HAR estimates. We estimate a HAR model for systematic staleness and for the cross-sectional average of idiosyncratic staleness. The parameters correspond to the intercept as well as to past daily, weekly and monthly values of the series. T-statistics are in the last column.

Panel A: $\hat{p}_d^{(S)}$

Parameters	Values	t-stats
β_0	0.001***	3.289
β_d	0.353***	7.298
β_w	0.403***	5.385
β_m	0.223***	4.536

Panel B: $\text{mean}_q \hat{p}_d^{(q)}$

Parameters	Values	t-stats
β_0	0.007**	2.165
β_d	0.591***	17.454
β_w	0.280***	6.356
β_m	0.116***	3.629

***, ** and * denote significance at the 1% level, the 5% level and the 10% level, respectively.

Table 2

HAR estimates with volume and variance. We estimate a HAR model for systematic staleness and for the cross-sectional average of idiosyncratic staleness. The first four parameters correspond to the intercept as well as to past daily, weekly and monthly values of the series. The last two parameters correspond to the one-day lagged (logarithm of the) daily cross-sectional average of dollar-weighted volume and the one-day lagged (logarithm of the) daily cross-sectional average of 5-minute realized variance, respectively. T-statistics are in the last column.

Panel A: $\hat{p}_d^{(S)}$

Parameters	Values	t-stats
β_0	0.008***	5.607
β_d	0.326***	6.844
β_w	0.386***	5.267
β_m	0.208***	4.322
β_{dwv}	−1.589***	−5.725
β_{v}	−0.831**	−2.379

Panel B: $\text{mean}_q \hat{p}_d^{(q)}$

Parameters	Values	t-stats
β_0	0.001	0.217
β_d	0.602***	16.636
β_w	0.280***	6.371
β_m	0.112***	3.439
β_{dwv}	0.410	0.546
β_{v}	1.510	1.430

***, ** and * denote significance at the 1% level, the 5% level and the 10% level, respectively.

for both measures. In the regressions, persistence is, therefore, captured by virtue of the (heterogeneous auto-regressive) HAR model proposed by Corsi (2009). The estimates have a familiar (from the volatility literature) look. The R^2s are close to 85% for systematic staleness and 93% for mean idiosyncratic staleness. Some observations are in order:

1. The substantial correlation between systematic staleness and *mean* idiosyncratic staleness (87%) is reminiscent of the equally large correlation between systematic volatility and *mean* idiosyncratic volatility (see, e.g., Herskovic et al., 2016, Fig. 5, Panel A).
2. The persistent dynamics of *mean* idiosyncratic staleness are suggestive of the presence of a common factor in idiosyncratic staleness. This is, again, reminiscent of analogous findings in the volatility literature (Herskovic et al., 2016).

In Fig. 2 we plot systematic staleness (top panel) and *mean* idiosyncratic staleness (bottom panel) against mean daily dollar-weighted volume. Volume is strongly negatively correlated with both measures. The correlation coefficient is, in both cases, around −70%. This finding is in line with the logic behind both measures: the lower volume, the smaller the price updates, the larger staleness. The microstructure model presented in Section 5 provides a structural justification for changing levels of trading activity, whether systematic or idiosyncratic. In the model, decreased trading (as induced, e.g., by higher execution costs) will translate into more price staleness. This effect is strongly in the data, as also suggested by the color-coded break-down by year in Fig. 2.

Fig. 3 plots systematic staleness (top panel) and *mean* idiosyncratic staleness (bottom panel) against the logarithm of mean 5-minute realized variance. The correlation between both notions of staleness and variance is negative and larger than −80%, particularly in the case of idiosyncratic staleness. Once more, the microstructure model in Section 5 provides a framework to conceptualize this effect. Because realized variance is designed to measure the variance of the underlying efficient price and the efficient price is defined as a conditional expectation of future payoffs given all available information, high efficient price variance is symptomatic of changes in the information set of the informed agents. In the model, all else equal, these changes in information may lead to trades. In particular, the larger the changes, the further away efficient prices will be from mid-quotes (for every level of execution costs), the more likely information-based trades. Trading, however, reduces staleness.

It is well-known that volume and variance are correlated. Our sample is no exception. The correlation between (mean, across stocks) log volume and (mean, across stocks) log variance is about 60%. In order to separate variance effects from volume effects, we add both variables (with a lag) to the HAR specification estimated previously. The results are in Table 2.

In the next section, we discuss a microstructure model which provides a framework to (1) rationalize our notions of systematic and idiosyncratic staleness and (2) extract *structural* information from their measurements.

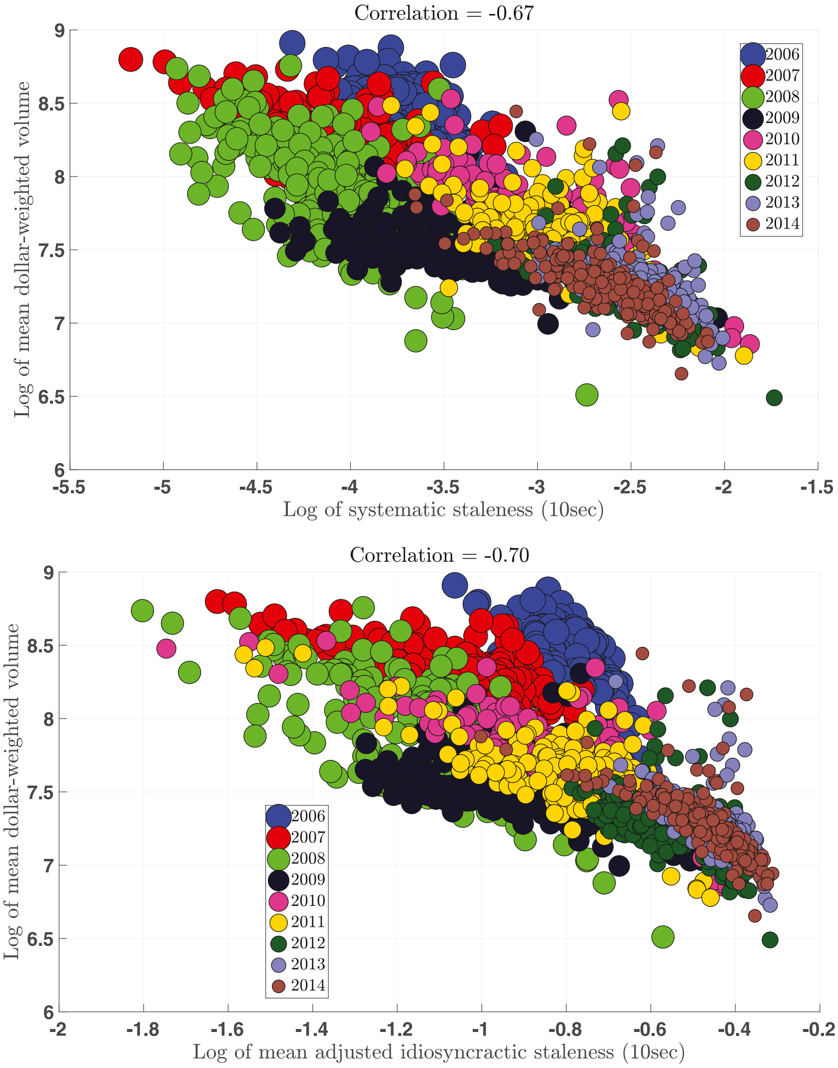

Fig. 2. (Top panel.) Scatter plot of the logarithm of daily systematic staleness vs. the logarithm of daily mean dollar-weighted volume. (Bottom panel.) Scatter plot of the logarithm of daily mean idiosyncratic staleness vs. the logarithm of daily mean dollar-weighted volume.

5. Micro-founding price staleness

We consider an N-variate price formation process in which private information (and its interaction with illiquidity, through execution costs) plays a key role in driving transaction prices, in the spirit of Kyle (1985) and Glosten and Milgrom (1985). In a nutshell, if the value of the information signal is larger than execution costs, informed traders will act on it and trade. Otherwise, they will choose not to trade, thereby leading to price staleness.

Empirically, we will show how measured (systematic and idiosyncratic) staleness from the data can be put to work to identify key features of the assumed price formation process, to which we now turn.

5.1. The price formation process

The model has three sets of N-variate prices: unobserved efficient prices, mid-points of bid/ask prices (i.e., mid-quotes) and transaction prices. Let t be a discrete-time index running across 10-second intervals[13] and let $q = 1, \ldots, N$ be, as earlier, the

[13] We note that, in Section 2, t defined a continuous-time index running across a trading day of length 1, i.e., $t \in [0, 1]$.

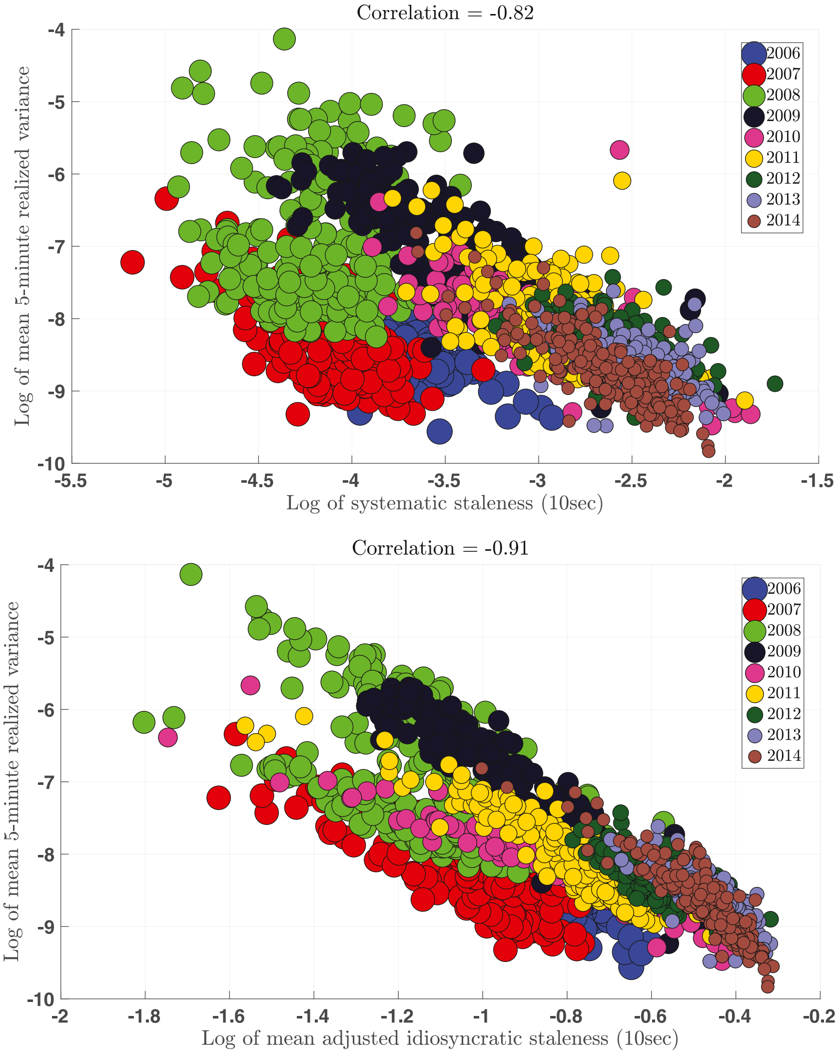

Fig. 3. (Top panel.) Scatter plot of the logarithm of daily systematic staleness vs. the logarithm of 5-minute daily realized variance. (Bottom panel.) Scatter plot of the logarithm of daily mean idiosyncratic staleness vs. the logarithm of 5-min daily realized variance.

index denoting stocks. The latent efficient (logarithmic) price process is assumed to be unpredictable given past information and is expressed as follows:

$$\mathfrak{e}_{q,t} = \mathfrak{e}_{q,t-1} + \beta_q \, \sigma_{\mathsf{M}} \sqrt{\Delta} \, \varepsilon_{\mathsf{M},t} + \sigma_q^{(\mathfrak{e})} \sqrt{\Delta} \, \varepsilon_{q,t}^{(\mathfrak{e})}, \tag{22}$$

where the $\varepsilon_{q,t}^{(\mathfrak{e})}$s are zero mean, unit variance, white noise shocks independent across stocks and the $\varepsilon_{\mathsf{M},t}$s are zero mean, unit variance, white noise shocks common across stocks. The quantities $\sigma_q^{(\mathfrak{e})}\sqrt{\Delta}$ and $\sigma_{\mathsf{M}}\sqrt{\Delta}$ define the standard deviations of the same shocks (over any 10-second interval Δ). The efficient prices are private information.

The mid-points of bid/ask prices are, instead, common knowledge. Consistent with the work of Hasbrouck and Ho (1987) and Amihud and Mendelson (1987), the mid-quotes evolve as

$$\mathfrak{m}_{q,t} = \mathfrak{m}_{q,t-1} + \delta_q \left(\mathfrak{e}_{q,t} - \mathfrak{m}_{q,t-1}\right) + (1-\delta_q)\sigma_q^{(\mathfrak{m})} \sqrt{\Delta} \, \varepsilon_{q,t}^{(\mathfrak{m})}, \tag{23}$$

with the initial conditions $\mathfrak{m}_{q,0} = \mathfrak{e}_{q,0}$, where the $\varepsilon_{q,t}^{(\mathfrak{m})}$s are, again, zero mean, unit variance, white noise shocks independent across stocks and independent of the efficient price shocks. The quantity $\sigma_q^{(\mathfrak{m})}\sqrt{\Delta}$ defines the 10-second standard deviation of the shocks to mid-quotes.

The market maker reacts to the order flow. The partial adjustment model in Eq. (23) captures his/her learning (and price setting behavior) without modeling directly the direction of the order flow. The efficient price is unobserved to him/her but, if $\mathfrak{e}_{q,t} - \mathfrak{m}_{q,t-1} > 0$

(resp. $\mathfrak{e}_{q,t} - \mathfrak{m}_{q,t-1} < 0$), the market maker is more likely to receive buy (resp. sell) orders (the behavior of the informed traders will later justify this statement). Hence, the mid-quote will be adjusted upward (resp. downward). The speed of this adjustment (i.e., the speed of learning of the market maker) is given by δ_q. The closer δ_q to one, the faster the market-maker learns and the closer the mid-quote is to the unobserved efficient price. Consistent with this logic, when $\delta_q = 0$, the mid-quotes are random walks independent of the efficient prices. When $\delta_q = 1$, instead, the mid-quotes track the efficient prices exactly. Relative to the mid-quote specification in Bandi et al. (2017), the current specification permits a more transparent interpretation of the role of δ_q. We note that the expected mid-quote coincides with the expected efficient price, i.e., $\mathbb{E}\left[\mathfrak{m}_{q,t}\right] = \mathbb{E}\left[\mathfrak{e}_{q,t}\right]$.

We assume that informed traders arrive to the market with a fixed probability which, following Bandi et al. (2017), is dubbed PAIT (Probability of Arrival of Informed Traders). At any point in time t, trading is conducted by an informed trader, rather than by a noise trader, if $U_t \leq$ PAIT, where the U_ts are iid shocks uniformly distributed in $(0, 1)$. By definition, the informed traders know the prevailing values of the efficient price processes $\mathfrak{e}_{q,t}$, for all $q = 1, \ldots, N$, and make their decision (to buy, sell or remain idle) by comparing the wedge between mid-quotes and efficient prices to the cost of execution c_q.[14] We think of c_q as the half spread s_q plus per-trade funding costs f, i.e., $c_q = s_q + f$. As in Brunnermeier and Pedersen (2009), f may be viewed as the "shadow cost" of capital. Importantly, it is assumed to be the same across assets and, therefore, systematic in nature. Because this parameter is designed to capture market-wide ability to raise funds, the assumption of constancy across assets is warranted. The spread s_q is, instead, idiosyncratic. While a literal reading of the model would relate it to the half bid/ask spread (an interpretation that may now be helpful to understand the logic of the model), a more formal interpretation in terms of (scaled) *effective* spread is provided in Section 6. Because trades can occur within the quoted bid/ask spread, our formalization will be more reflective of the actual costs incurred by traders as compensation for immediacy. We refer the reader to Roll (1984) for an early discussion and to Hasbrouck (2009) for a subsequent approach and references.[15]

We now turn to the determination of transaction prices. If $U_t \leq$ PAIT, trading – or lack thereof – is due to information. Informed traders will trade only when they receive compensation for both the cost of immediacy s_q and the cost of funding f. Specifically, if $\left|\mathfrak{e}_{q,t} - \mathfrak{m}_{q,t}\right| \leq c_q = s_q + f$, informed traders do not trade and prices are stale. If, instead, $|\mathfrak{e}_{q,t} - \mathfrak{m}_{q,t}| > c_q$, the observed traded prices are

$$\mathfrak{p}_{q,t} = \mathfrak{m}_{q,t} + s_q \mathbb{I}_{\{\mathfrak{e}_{q,t} - \mathfrak{m}_{q,t} > c_q\}} - s_q \mathbb{I}_{\{\mathfrak{e}_{q,t} - \mathfrak{m}_{q,t} < -c_q\}}. \tag{24}$$

Informed traders only trade in the "right" direction, i.e., they buy when the mid-quote is "sufficiently" lower than the efficient price and they sell when the mid-quote is "sufficiently" higher. This behavior justifies the assumed reaction of the market maker, as represented by the partial adjustment model in Eq. (23). If trades are uninformed (that is, if $U_t \geq$ PAIT), we write

$$\mathfrak{p}_{q,t} = \mathfrak{m}_{q,t} + \eta_{q,t} s_q,$$

where $\eta_{q,t}$ can be either $+1$ or -1 with equal likelihood and s_q represents, again, the cost of immediacy for the qth asset.

We emphasize that the model can give rise to systematic staleness through the market-wide funding cost f or through increases in PAIT, for all levels of execution costs. It could also give rise to idiosyncratic staleness through, e.g., changes in the asset-specific spreads s_q. While we do not model explicitly a common (market-wide) factor in the spreads s_q, we will let the proposed staleness measures identify it, for each day in our sample. For further discussions of a single-asset version of the model *without* any distinction between market-wide and idiosyncratic effects, a distinction which represents the substantive core of this article, we refer the reader to Bandi et al. (2017).

While rich, the model specification can be extended. One may allow for the presence of intermittent *noise* trading, something which is unconventional in microstructure theory (with notable exceptions, such as the work of Easley and O'Hara, 1992) but empirically believable. Because efficient prices would be unobservable to these (noise) traders, they would not compare deviations between efficient prices and mid-quotes to execution costs. They would, instead, respond to the *absolute* size of the execution costs, as opposed to their *relative* size, and opt out of trades when the absolute size is large. Their decisions would, however, also lead to lower volume, and increased staleness, as a function of higher costs of execution. Lack of trading on the part of all traders has the potential to yield enhanced signal about the cost of trading (be it idiosyncratic or systematic) and less loading on PAIT. Because the evaluation of the trigger for noise trade inaction is, however, not obvious (what is a "large" execution cost in absolute terms?), and no guidance is offered by the existing literature, in this article we follow the literature convention of only allowing for information-based inaction. A second extension has to do with PAIT. This probability, itself, could be asset specific, rather common across stocks. A stock-specific PAIT would, of course, add parameters to the model specification (149 additional parameters in our case, c.f. Section 5.2) and render identification more cumbersome. Also, because we are interested in systematic effects, it would be meaningful for us to model a systematic component of PAIT but this choice would entail ambiguities in model specification on which microstructure theory and empirical evidence provide, again, no guidance. Relatedly, one may view the current (common across stocks) PAIT as an average (across stocks) probability of arrival. We emphasize that having a single PAIT does not imply that the probability of informed trading is the same across stocks. Given, in fact, their arrival (which is controlled by PAIT), the informed traders may or may not decide to trade. Thus, the probability of informed trading is stock-specific and time-varying. A third, potential, extension relates the magnitude of the spreads s_q to the amount of information-based trading. While we could specify them as being a function of the distance between mid-quotes and efficient prices (a larger distance being, as discussed, conducive

[14] The informed traders "know" the efficient prices because these prices are conditional expectations of future cash flows given their information set.

[15] The notion of effective spread is important to regulators as well. SEC's Reg NMS Rule 605 requires market venues to provide estimates of effective spreads.

to more information-based trading), the same effects are currently captured non-parametrically. As we detail in Section 5.2, we, in fact, produce daily estimates (time-varying across days) of all quantities (including s_q). The daily s_q estimates are expected to be affected by everything that drives the specialist's decisions, from the probability of informed trading (which depends on PAIT, efficient price volatility and more) to the specialist's learning.

5.2. Structural estimation

The N-variate model contains $5N + 3$ structural parameters: β_q, $\sigma_q^{(e)}$, $\sigma_q^{(m)}$, δ_q and s_q, for $q = 1, \dots, N$, plus σ_M, PAIT and f. For each stock, we estimate the β_qs using a rolling window of 22 days, thereby reducing the number of structural parameters to $4N + 3$.[16]

After quality cuts on the initial set of stocks,[17] the model is estimated on 150 stocks. Using the selected stocks, we identify the parameters by simulated method of moments (Gourieroux et al., 1993, and Duffie and Singleton, 1993). Specifically, for each stock and for each day in the sample, we compute nine stock-specific realized measures:

1. 10-second and 300-second realized variances,
2. 10-second auto-correlations at lags 1, 4, 8 and 16,
3. 10-second, 30-second and 1-minute idiosyncratic staleness measures, as defined in Eq. (21).

For each day in the sample, we also compute the following economy-wide realized measures:

4. 10-second, 30-second and 1-minute systematic staleness, computed as in Eq. (20),
5. The mean of the means (time-series and cross-sectional) of the 300-second realized variances.

The $9N + 4$ daily realized measures are used as moment conditions and matched to the corresponding daily moments simulated from data. To generate simulated data, all model's shocks in Section 5.1 are drawn from a Gaussian distribution.

Importantly, because both the empirical and the simulated moments are defined *for each day* in the sample, something that only estimates based on high-frequency data can afford, they lead to daily estimates of the parameters of interest and, therefore, daily time series of the same parameters. These daily time series are helpful to understand dynamics, interactions and feedback effects. In Fig. 4, we report 10-day moving averages of the daily estimates of all model parameters. For all of the non-systematic quantities, we also report the 10th and 90th largest value associated with the 150 assets used for estimation. The estimates of the "variance"-related parameters ($\sigma_q^{(e)}$, $\sigma_q^{(m)}$ and s_q) have a familiar pattern with apparent increases around the 2008/2009 financial crisis. Consistent with their economic logic, we document a similar increase for PAIT and f.

The proposed identification method is evaluated in Appendix B by simulation. Here, we provide further details regarding the role played by our suggested staleness measures. Each of the realized moments is, of course, a nonlinear function of the set of parameters and is, therefore, responsive to them. We show that the realized staleness measures are key to the identification of the deep parameters associated with granular trading dynamics, from PAIT to the speed of learning of the market maker, δ_q, to the "shadow" cost, f. For illustration, in Fig. 5 we plot the objective function of the model as a function of all model parameters, considered one-by-one. These parameters are varied across the horizontal axis keeping all other parameters fixed at their (time-series and cross-sectional) average. The moments to be matched are those of data artificially generated (for 10 assets) with values which are reported, in all panels, as vertical blue dotted lines. Specifically, we report the average (across 100 simulations) of two different objective functions: one (represented by the black continuous line) obtained including all of the staleness moments (i.e., the systematic staleness measure and the N idiosyncratic staleness measures) and a second (the red continuous line with circles) obtained by excluding them. The dotted lines define the $5\% - 95\%$ confidence band associated with each average. The plots document that, absent the information in staleness, the objective function would be excessively flat to identify credibly PAIT, δ_q and, in particular, f. While the staleness moments improve identification across the board, solely the "variance"-related parameters (again, $\sigma_q^{(e)}$, $\sigma_q^{(m)}$ and s_q) appear to be estimable reliably using "variance" moments *only*.

6. Structural idiosyncratic and systematic liquidity

In this section, we dive more deeply into measures that (i) are naturally related to liquidity, (ii) have accepted benchmarks in the literature and (iii) allow us to evaluate the ability of our proposed methods to extract both idiosyncratic and systematic information. Specifically, we focus on the idiosyncratic (half) spreads, s_q, and on systematic funding costs, f. We begin with the former.

[16] In unreported experiments, we have also considered non-zero correlations between the shocks to the mid-quotes $\epsilon_{q,t}^{(m)}$ and the (asset-specific) shocks to the efficient prices $\epsilon_{q,t}^{(e)}$. We have found that these correlations do not affect our findings in any meaningful way. In order not to add a large number of additional parameters (one per asset) that are hard to identify, and are less critical in terms of economic logic, we opted for keeping the model simpler.

[17] For each of the 250 stocks and on each day, we compute (1) the total volume traded, (2) the total number of transactions and (3) the longest time interval with no trading, obtaining three 250 × 2265 matrices, one for each characteristic. We flag all of the daily entries of the three matrices with total log-volume smaller than 12.5, or with a number of transactions smaller than 500, or with a maximum time length of no trading larger than ten minutes. We keep only those stocks that, after removing all flagged days, have daily returns over at least 97% of the sample.

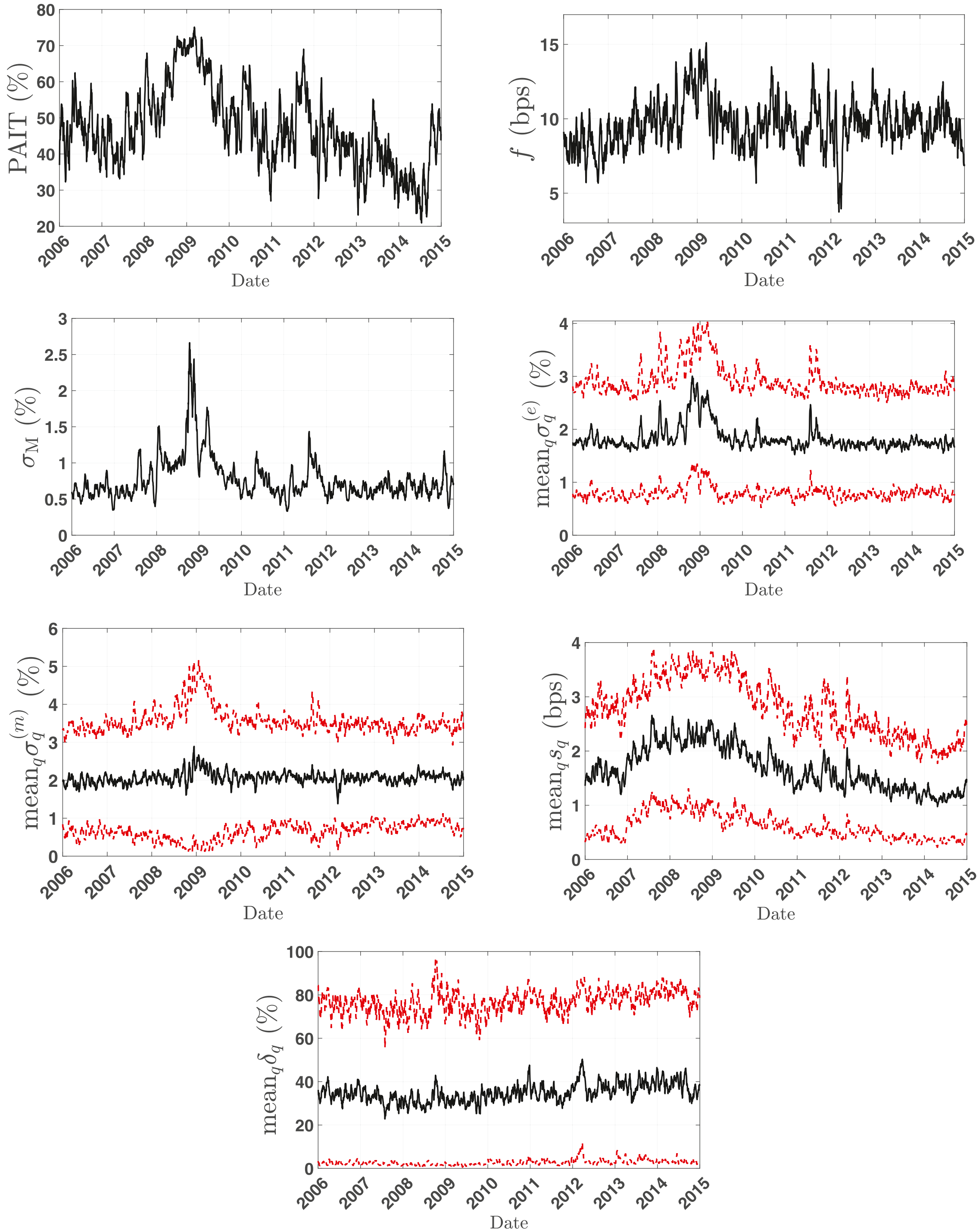

Fig. 4. We report the 10-day moving averages of the daily estimates of the model parameters: the probability of arrival of informed traders PAIT, the funding cost f (expressed in basis points), the market volatility σ_{M} (expressed in percentage), the mean (across the 150 stocks) of the efficient price volatility $\sigma_q^{(\mathrm{e})}$ (expressed in percentage), the mean (across the 150 stocks) of the mid-quote volatility $\sigma_q^{(\mathrm{m})}$ (expressed in percentage), the mean (across the 150 stocks) of the half bid–ask spread s_q (expressed in basis points), and the mean (across the 150 stocks) of the speed of learning δ_q. The red dotted lines represent the 10th and 90th largest value associated with the 150 stocks used for estimation. (For interpretation of the references to color in this caption, the reader is referred to the web version of this article.)

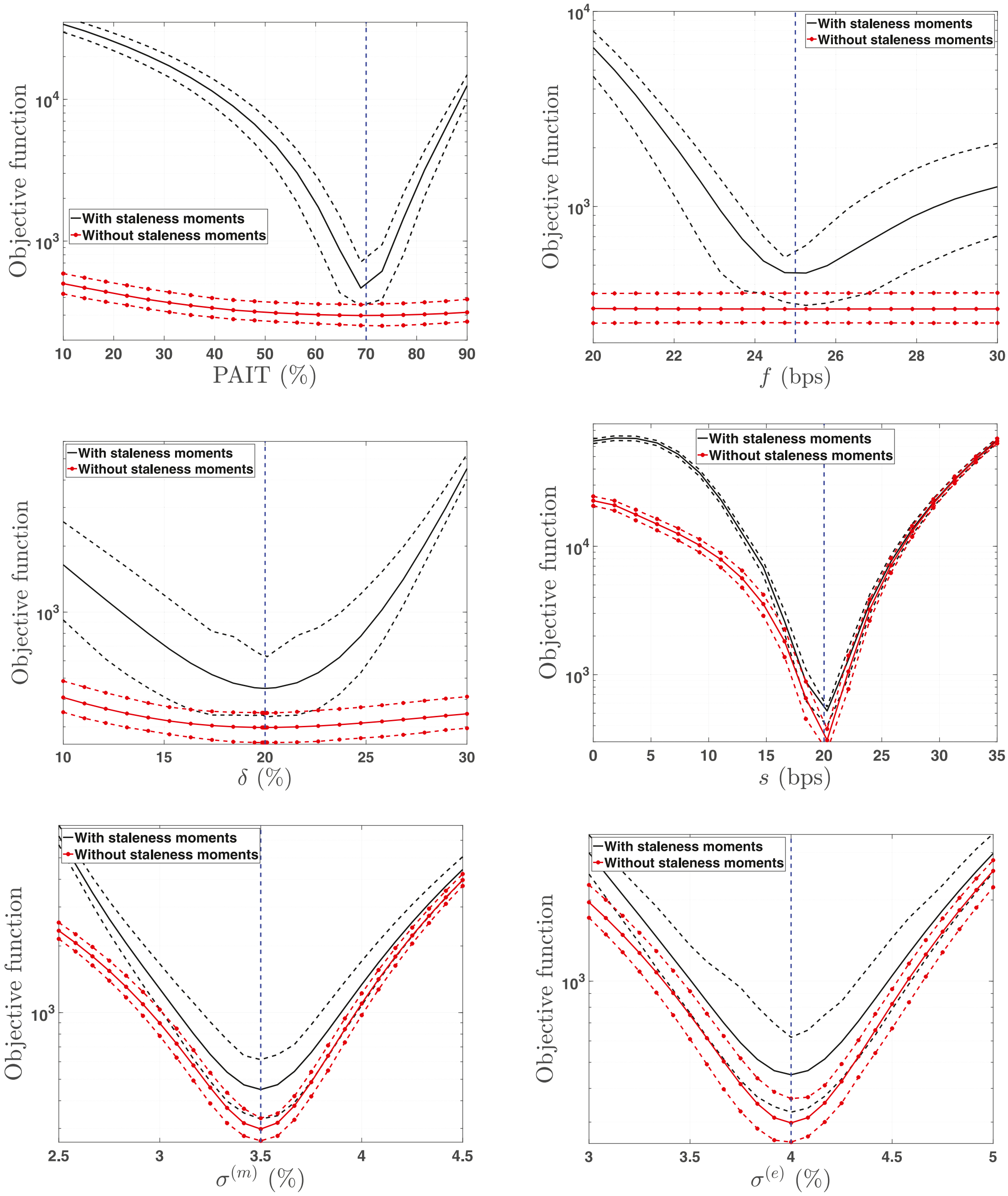

Fig. 5. In each plot we report the average, across 100 simulations, of the objective function computed with (black continuous line) and without (red continuous line with circles) the staleness moments. The dashed lines define $5\% - 95\%$ confidence bands. The moments to be matched are those of artificially-generated data for a sample of $N = 10$ assets. The horizontal axis of each panel reports the parameter that is varied (all other parameters are set equal to their cross-sectional and time-series average). The vertical dashed blue lines coincide with the true values of the parameters, i.e., the values used to generate the artificial samples. To facilitate comparison, the vertical axis is in logarithmic scale. (For interpretation of the references to color in this caption, the reader is referred to the web version of this article.)

6.1. Execution costs

When taken literally, the model specification in Eq. (24) leads to a direct association of the parameter s_q with the half bid/ask spread. A superior interpretation, one which takes into account the fact that the model is estimated with transaction prices *only* (i.e., without any information on bid/ask spreads) and transacted prices may lay within the spread, suggests that s_q should, instead, by viewed as a proxy for the *effective* spread. We now formalize this logic within the model.

Consider the gap process $\mathfrak{g} = \mathfrak{e} - \mathfrak{m}$. Given Eqs. (22) and (23), and writing $\varphi_q = 1 - \delta_q$ for conciseness, we have

$$\mathfrak{g}_{q,t} = \mathfrak{e}_{q,t} - \mathfrak{m}_{q,t} = \varphi_q \left(\mathfrak{e}_{q,t-1} - \mathfrak{m}_{q,t-1}\right) + \varphi_q \, \sigma_q^{(\mathfrak{e})} \sqrt{\Delta}\, \varepsilon_{q,t}^{(\mathfrak{e})} - \varphi_q \, \sigma_q^{(\mathfrak{m})} \sqrt{\Delta}\, \varepsilon_{q,t}^{(\mathfrak{m})} + \varphi_q \, \beta_q \, \sigma_{\mathsf{M}} \sqrt{\Delta}\, \varepsilon_{\mathsf{M},t}$$

or

$$\mathfrak{g}_{q,t} = \varphi_q \, \mathfrak{g}_{t-1} + \varepsilon_{q,t}^{(\mathfrak{g})},$$

where $\varepsilon_{q,t}^{(\mathfrak{g})} = \varphi_q \, \sigma_q^{(\mathfrak{e})} \sqrt{\Delta}\, \varepsilon_{q,t}^{(\mathfrak{e})} - \varphi_q \, \sigma_q^{(\mathfrak{m})} \sqrt{\Delta}\, \varepsilon_{q,t}^{(\mathfrak{m})} + \varphi_q \, \beta_q \, \sigma_{\mathsf{M}} \sqrt{\Delta}\, \varepsilon_{\mathsf{M},t}$. Assume, now, normality of the shocks, an assumption which is solely made to arrive at operational closed-form expressions. Then,

$$\mathfrak{g}_{q,t} = \sum_{j=0}^{+\infty} \varphi_q^j \, \varepsilon_{q,t-j}^{(\mathfrak{g})} \sim N\left(0, \frac{\sigma^2_{\varepsilon^{(\mathfrak{g})}}}{1-\varphi_q^2}\right)$$

with

$$\sigma^2_{\varepsilon^{(\mathfrak{g})}} = \varphi_q^2 \, \Delta \left(\left(\sigma_q^{(\mathfrak{e})}\right)^2 + \left(\sigma_q^{(\mathfrak{m})}\right)^2 + \beta_q^2 \, \sigma_{\mathsf{M}}^2 \right).$$

Thus, the probability of having a zero return for the qth asset is given by

$$p_q^{\emptyset} = \text{PAIT} \cdot \mathbb{P}\left[\left|\mathfrak{g}_{q,t}\right| \le f + s_q\right] = \text{PAIT} \cdot \frac{\sqrt{1-\varphi_q^2}}{\sqrt{2\pi}\,\sigma_{\varepsilon^{(\mathfrak{g})}}} \int_{-c_q}^{c_q} \exp\left(-\frac{\left(1-\varphi_q^2\right) x^2}{2\,\sigma^2_{\varepsilon^{(\mathfrak{g})}}}\right) dx, \tag{25}$$

where $c_q = f + s_q$ is the total cost of trading. Denoting by D_q the direction of a trade ($D_q = +1$ for a buy, $D_q = -1$ for a sell), the model-implied effective spread is:

$$\begin{aligned}
\mathbb{E}\left[D_{q,t}\left(\mathfrak{p}_{q,t} - \mathfrak{m}_{q,t}\right)\right] &= \underbrace{\text{PAIT}\left(\mathbb{E}\left[\mathfrak{m}_{q,t} + s_q - \mathfrak{m}_{q,t}\right]\mathbb{P}\left[\mathfrak{g}_t > c_q\right] - \mathbb{E}\left[\mathfrak{m}_{q,t} - s_q - \mathfrak{m}_{q,t}\right]\mathbb{P}\left[\mathfrak{g}_t < -c_q\right]\right)}_{\text{informed trader}} + \\
&\quad \underbrace{+ (1-\text{PAIT})\left(+\frac{1}{2}s_q - \left(-\frac{1}{2}s_q\right)\right)}_{\text{noise trader}} \\
&= s_q \, \text{PAIT}\left(\mathbb{P}\left[\mathfrak{g}_t > c_q\right] + \mathbb{P}\left[\mathfrak{g}_t < -c_q\right]\right) + (1-\text{PAIT})\, s_q \\
&= s_q \, \text{PAIT}\left(1 - \mathbb{P}\left[\left|\mathfrak{g}_{q,t}\right| \le c_q\right]\right) + (1-\text{PAIT})\, s_q \\
&= s_q \left(1 - p_q^{\emptyset}\right) \\
&= s_{q,\star}.
\end{aligned}$$

In light of this discussion, we note that while viewing s_q as a proxy for the prevailing effective spread is reasonable, an even superior proxy would multiply s_q by a factor equal to 1 minus the probability of zero returns. Such a probability can be computed in closed form (given an assumption on the distribution of the shocks and the model parameters), as evidenced by Eq. (25).

In Fig. 6, we compare model-implied measures to empirical spread measures from the TAQ data set. Specifically, we compare estimates of the spread measure s_q (left column plots) and the effective spread measure $s_{q,\star}$ (right column plots) to traditional benchmarks (c.f. Goyenko et al., 2009): daily averages of logarithmic bid–ask spreads, daily averages of *effective* spreads and daily averages of *realized* spreads (first, second and third row, respectively). The realized spreads are defined as in Bessembinder (2003).

The time-series correlations (for each stock and each pair of measures, over time) is represented using colors, from blue (correlation 0) to dark red (correlation 1). The cross-sectional correlation is, instead, reported in each of the panel's title. As emphasized, the structural estimates only hinge on transaction prices (i.e., no information on quotes is needed). In addition, we do not require the use of any algorithm designed to sign the order flow as in, e.g., the classical work of Lee and Ready (1991). Yet, the model-implied structural estimates and the three high-frequency execution cost proxies are, remarkably, very highly correlated both cross-sectionally and in the time series. The structural estimates lay within the recorded bid/ask spreads, an empirical finding which is in line with the observation that substantial price improvements, as compared to recorded bid/ask spreads, are ubiquitous (see, e.g., Bessembinder, 2003).

Table 3 reports average time-series correlations between the model-implied estimates (s_q and $s_{q,\star}$) and the same empirical estimates as in Fig. 6, namely recorded bid/ask spreads, effective spreads and realized spreads. Importantly, we also report average time-series correlations between idiosyncratic staleness (as defined in Eq. (21)) and the three empirical measures. We note that the latter are negative. While idiosyncratic staleness contains revealing signal about execution costs – a signal that structural estimation can bring to light – the impact of features of the trading process other than execution costs make it a rather noisy proxy. We will return to this point.

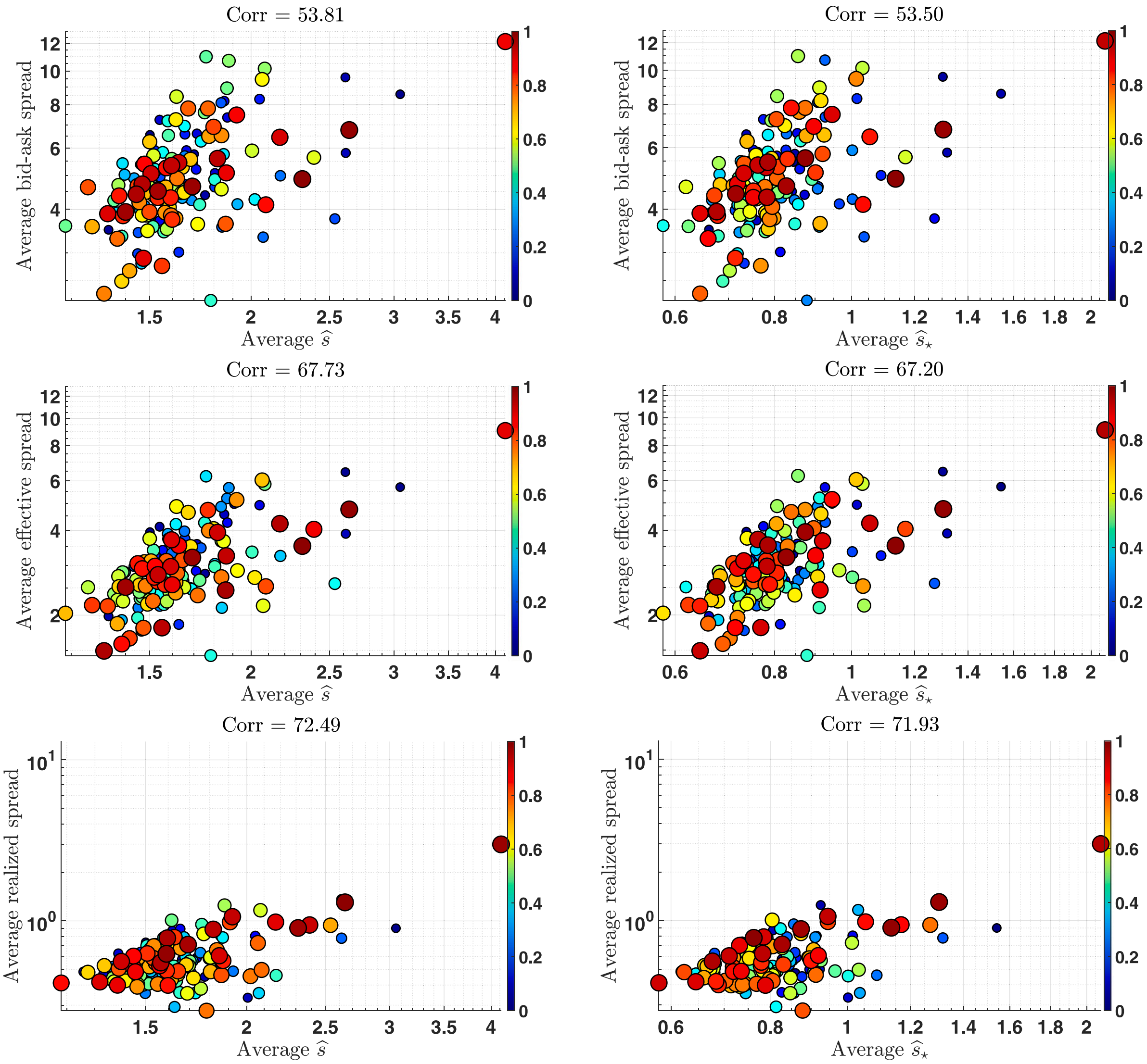

Fig. 6. In each plot a circle corresponds to a different stock in the data. For each stock, the horizontal axes report (in logarithmic scale of basis points) the daily time-series averages of the model-implied spread parameters $\widehat{s}_q$ (left column) and the daily time-series averages of the model-implied effective spread parameters $\widehat{s}_{q,\star} = \widehat{s}_q \left(1 - \widehat{p}_q^{\emptyset}\right)$ (right column). For each stock, the vertical axes report (in logarithmic scale of basis points) the daily time-series averages, computed using TAQ data, of the logarithmic bid–ask spreads (first row), the daily time-series averages of the effective spreads (second row) and the daily time-series averages of the realized spreads (third row). The color-coding, whose legend appears in the color map next to each plot, provides information on the sample correlation between each pair of daily time series. The closer to red the color (and, also, the larger the circle), the higher the correlation. Finally, the title in each plot reports the percentage cross-sectional correlation between the averages. (For interpretation of the references to color in this caption, the reader is referred to the web version of this article.)

6.2. *Funding liquidity*

As emphasized earlier, given the workings of the model, it is natural to interpret the parameter f as a market-wide (per unit of trade) "shadow" cost of capital. Consistent with the logic in Brunnermeier and Pedersen (2009), our daily structural estimate of f may, therefore, be viewed as a proxy for daily systematic funding costs.

In order to support this idea empirically, we run regressions of the daily estimates of f on three alternative proxies used in the recent work of Fontaine et al. (2020), namely the bond-based funding liquidity measure in Fontaine and Garcia (2012), the classical TED spread and the measure in Hu et al. (2013). Table 4 contains our findings. In agreement with our logic, our structural proxy is highly correlated with all three alternatives, the measure provided by Hu et al. (2013) having the largest correlation with $\widehat{f}$ (59.1%, corresponding to an R^2 of 35%). We emphasize that the reported R^2s are in line with what one would obtain when regressing the three benchmark proxies on each other. Regressing, for instance, the measure in Fontaine and Garcia (2012) on the TED spread and

Table 3
Average time-series correlations. This table reports the average (across stocks) time series (percentage) correlations between $\hat{s}_q$ (model-implied daily spread), $\hat{s}_{q,\star} = \hat{s}_q \left(1 - \hat{p}_q^{\emptyset}\right)$ (model-implied daily effective spread), $\hat{p}_d^{(q)}$ (daily idiosyncratic staleness) and BA, ES and RS, i.e., the daily logarithmic bid–ask, effective and realized spreads computed using TAQ data, respectively.

	BA	ES	RS
$\hat{s}$	59.22	64.89	69.30
$\hat{s}_\star$	16.83	22.95	51.33
$\hat{p}_d$	−39.15	−49.40	−52.90

Table 4
The relation between $\hat{f}$ and alternative proxies of funding costs. We compare our structural estimate of funding costs $\hat{f}$ to three alternative proxies: the measure in Fontaine and Garcia (2012) (FUND_1), the TED spread (FUND_2) and the measure in Hu et al. (2013) (FUND_3). T-statistics are in parenthesis.

	FUND_1	FUND_2	FUND_3	R^2
(1)	0.572***			0.134
	(4.057)			
(2)		0.439***		0.138
		(4.127)		
(3)			0.573***	0.350
			(7.558)	
(4)	0.423***	0.330***		0.204
	(2.929)	(3.018)		
(5)	0.183		0.521***	0.361
	(1.328)		(6.101)	
(6)		0.100	0.530***	0.355
		(0.920)	(5.945)	
(7)	0.169	0.083	0.489***	0.364
	(1.217)	(0.758)	(5.132)	

***, ** and * denote significance at 1% level, 5% level and 10% level, respectively.

the measure of Hu et al. (2013) would yield R^2s of 11.6% and 21.6%, respectively. Regressing the last two proxies on each other would, instead, deliver an R^2 of 27.6%.

In Table 5 we run the same regressions but replace $\hat{f}$ with systematic staleness. All betas are now negative and sometimes significantly so. Once more, while systematic staleness contains signal about systematic liquidity, the nature of the trading process makes it a rather noisy proxy, thereby justifying our structural approach. Again, we will return to this observation.

7. An economic lens: Cross-sectional pricing

We examine whether alternative systematic measures of liquidity contain cross-sectional pricing signal. We use the 150 stocks employed for structural estimation and price *monthly* returns. Because the test assets are individual stocks rather than portfolios, allowing for time-varying factor loadings is particularly relevant. We implement a classical Fama–MacBeth procedure (Fama and MacBeth, 1973) and report results for two rolling windows used in the estimation of the factor loadings, 30 months and 50 months. The latter is a typical length in the literature, the former allows for more time variation (at the cost of less efficiency) in the estimated betas. We report risk prices for returns in excess of the 5-factor model in Fama and French (2015). In other words, we control for state-of-the-art systematic risk factors when evaluating the pricing impact of alternative systematic liquidity proxies.

The liquidity proxies are systematic staleness ($\hat{p}_m^{(S)}$),[18] the structural estimate of funding liquidity ($\hat{f}$), the average (across stocks of) idiosyncratic staleness ($\text{mean}_q \hat{p}_m^{(q)}$), the average (across stocks of) the structural spread measure ($\text{mean}_q \hat{s}_q$), the reversal measure (PS) of Pastor and Stambaugh (Pastor and Stambaugh, 2003) and the three proxies of funding costs in the previous section, namely the measure in Fontaine and Garcia (2012) (FUND_1), the TED spread (FUND_2) and the measure in Hu et al. (2013) (FUND_3).

We include $\hat{f}$ and $\text{mean}_q \hat{s}_q$ because they represent, given the structural model in Section 5 and the comparisons in Section 6, meaningful proxies of systematic liquidity, the former in particular. While rather noisy, $\hat{p}_m^{(S)}$ and $\text{mean}_q \hat{p}_m^{(q)}$ are added due to their computational simplicity. The reversal measure PS has been shown to be priced in other cross sections. Finally, we include FUND_1,

18 Here and below, the subscript m is used to define "monthly" measures obtained as the average of daily measures over the corresponding month.

Table 5
The relation between $\widehat{p}_d^{(S)}$ and alternative proxies of funding costs.
We compare the systematic staleness estimates $\widehat{p}_d^{(S)}$ to three alternative proxies of funding costs: the measure in Fontaine and Garcia (2012) ($FUND_1$), the TED spread ($FUND_2$) and the measure in Hu et al. (2013) ($FUND_3$). T-statistics are in parenthesis.

	$FUND_1$	$FUND_2$	$FUND_3$	R^2
(1)	−0.002			0.005
	(−0.764)			
(2)		−0.015***		0.477
		(−9.829)		
(3)			−0.009***	0.231
			(−5.649)	
(4)	0.005**	−0.016***		0.506
	(2.503)	(−10.321)		
(5)	0.005**		−0.010***	0.260
	(2.010)		(−6.008)	
(6)		−0.013***	−0.003**	0.496
		(−7.429)	(−2.010)	
(7)	0.008***	−0.014***	−0.005***	0.552
	(3.611)	(−8.243)	(−3.271)	

***, ** and * denote significance at 1% level, 5% level and 10% level, respectively.

$FUND_2$ and $FUND_3$ because of their interpretation in terms of funding liquidity measures and, therefore, for comparison with our central proxy for systematic liquidity, i.e., $\widehat{f}$.

Tables 6 and 7 contain estimated prices of risk along with Newey–West t-statistics for 16 models. Model 1 through 8 focus on each of the systematic liquidity proxies one at a time (in addition to the five Fama–French factors). Model 9 through 16 expand on the pricing of our reduced-form systematic staleness measure $\widehat{p}_m^{(S)}$, a noisy proxy for systematic liquidity, and on the pricing of our structural systematic liquidity measure, $\widehat{f}$. They do so by considering each measure and by controlling for PS, $FUND_1$, $FUND_2$ and $FUND_3$ one at a time (in addition, once more, to the five Fama–French factors).

We emphasize that all measures are, technically, *illiquidity* (rather than *liquidity*) proxies. As such, they should increase when liquidity dries up. Because higher illiquidity is associated with lower returns, the illiquidity factor loadings are, in general, negative. We, therefore, expect the prices of risk to also be negative in that adverse states of the world (those associated with increases in illiquidity and lower returns) should yield higher expected compensations. Having made this point, the measures leading to consistently negative and largely statistically-significant prices of risk (across the reported choices of rolling window length) are $\widehat{f}$, PS and $FUND_1$. Yet, while the statistical significance of PS and $FUND_1$ is affected somewhat by the choice of window length, it is stable in the case of $\widehat{f}$.[19] Other proxies yield economically-unreasonable *positive* prices of risk, insignificant estimates, or both.

Our intent in this section cannot be to study the pricing of systematic illiquidity broadly defined. Our more limited objective is, in fact, solely to evaluate, in our data, the relative (to popular measures) pricing signal contained in our estimates. Consistent with this observation, we find that $\widehat{f}$ performs satisfactorily and is generally more robust than successful illiquidity proxies, like PS and $FUND_1$. We also find that, while the average (across stocks) of the structural spread measure ($\text{mean}_q \widehat{s}_q$) may also have an interpretation in terms of market-wide illiquidity proxy, its signal is statistically and economically considerably more feeble than that associated with $\widehat{f}$, by construction a truly systematic measure. Finally, and importantly for our purposes, the use of noisier proxies, like $\widehat{p}_m^{(S)}$ and $\text{mean}_q \widehat{p}_m^{(q)}$, translates into less robust performance as compared to $\text{mean}_q \widehat{s}_q$ and, in particular, to $\widehat{f}$.

8. Conclusions

Asset prices do not update as frequently as implied by traditional modeling in continuous time. Lack of price updates may, however, be economically as informative as long spells of volatility or jumps. The latter have been the focus of the successful high-frequency literature. The former is the subject of this article.

We provide a new conceptual framework to define notions of (market-wide) systematic staleness and (asset-specific) idiosyncratic staleness and a methodological framework to measure staleness, in its various forms. The methods result in a complete theory of inference which builds on joint asymptotics relying on increasingly-frequent observations over the period along with an increasing number of assets.

Text-book liquidity measures, like quoted bid/ask spreads, are known to be highly correlated with volatility. Alternative liquidity measures, like effective spreads, are known to be affected by asymmetries in information. There is tendency in the empirical finance literature to associate the number of zero returns (a form of staleness) to illiquidity. Yet, similar to other proxies, our

[19] The pricing performance of $\widehat{f}$ is, in fact, robust to window lengths ranging from 20 months to 100 months.

Table 6
Pricing with shorter windows. We run Fama–MacBeth regressions (with 30-month rolling windows) of the returns on the 150 stocks used for structural estimation on alternative illiquidity proxies as well as on the five Fama–French factors. The illiquidity proxies are: systematic staleness ($\hat{p}_m^{(S)}$), the structural estimate of funding liquidity ($\hat{f}$), the average (across stocks) idiosyncratic staleness (mean$_q \hat{p}_m^{(q)}$), the average (across stocks) structural spread measure (mean$_q \hat{s}_q$), the reversal measure (PS) of Pastor and Stambaugh (Pastor and Stambaugh, 2003) and the three proxies of funding costs in the previous section, namely the measure in Fontaine and Garcia (2012) ($FUND_1$), the TED spread ($FUND_2$) and the measure in Hu et al. (2013) ($FUND_3$). T-statistics are in parenthesis.

	$\hat{p}_m^{(S)}$	$\hat{f}$	mean$_q \hat{p}_m^{(q)}$	mean$_q \hat{s}_q$	PS	$FUND_1$	$FUND_2$	$FUND_3$	$\overline{R^2}$
(1)	−0.001								0.126
	(−0.646)								
(2)		−0.420***							0.128
		(−3.079)							
(3)			−0.001						0.126
			(−0.160)						
(4)				−0.003					0.130
				(−0.133)					
(5)					−0.007*				0.123
					(−1.720)				
(6)						−0.543**			0.129
						(−2.062)			
(7)							0.064		0.120
							(1.131)		
(8)								−0.057	0.125
								(−0.774)	
(9)	−0.001				−0.007				0.133
	(−0.605)				(−1.554)				
(10)	−0.001					−0.520**			0.136
	(−0.457)					(−2.015)			
(11)	−0.001						0.063		0.129
	(−0.565)						(1.067)		
(12)	−0.001							−0.055	0.134
	(−0.680)							(−0.741)	
(13)		−0.427***			−0.007				0.134
		(−3.129)			(−1.632)				
(14)		−0.440***				−0.485*			0.139
		(−3.197)				(−1.789)			
(15)		−0.401***					0.046		0.131
		(−2.877)					(0.775)		
(16)		−0.410***						−0.062	0.137
		(−3.019)						(−0.841)	

***, ** and * denote significance at 1% level, 5% level and 10% level, respectively.

proposed staleness measures should be viewed as being driven by more than the extent of execution costs and, therefore, that of illiquidity. The inevitable interaction between liquidity, information, volatility and other effects (such as learning) makes, in our view, the disentangling afforded by structural approaches to price formation particularly appealing. In one such approach, we have documented that our proposed (reduced-form) high-frequency estimates of staleness have the potential to provide revealing structural identification. Much remains to be done, our results being an initial step into structural explorations based on suitable realized measures.

We emphasize that work on structural estimation using high-frequency realized measures is, at best, sparse. In this early study, it was, therefore, natural for us to follow the progression of the high-frequency literature: our reduced-form estimates of systematic and idiosyncratic staleness are viewed as being "integrated" measures over the period (the trading day) and are, therefore, matched to simulated moments over the same period. A computationally expensive, but intriguing, next step is to localize estimation and focus on intra-daily effects, rather than on daily (albeit time-varying across days) dynamics. Intra-daily periodicities in volatility are well-documented. Intra-daily periodicities in staleness have been reported in Bandi et al. (2020). In principle, localization would require a devoted theory of *spot* staleness estimation and opportune selection of a truly time-varying threshold $\Theta_{t,n}^{(q,N)}$, among other issues. While this is a topic better left for future work, we trust the identification potential of high-frequency data will stimulate the transition from purely reduced-form investigations – the norm in the current high-frequency literature – to studies that are more genuinely structural in nature.

Table 7
Pricing with longer windows. We run Fama–MacBeth regressions (with 50-month rolling windows) of the returns on the 150 stocks used for structural estimation on alternative illiquidity proxies as well as on the five Fama–French factors. The illiquidity proxies are: systematic staleness ($\hat{p}_m^{(S)}$), the structural estimate of funding liquidity ($\hat{f}$), the average (across stocks) idiosyncratic staleness ($\text{mean}_q \hat{p}_m^{(q)}$), the average (across stocks) structural spread measure ($\text{mean}_q \hat{s}_q$), the reversal measure (PS) of Pastor and Stambaugh (Pastor and Stambaugh, 2003) and the three proxies of funding costs in the previous section, namely the measure in Fontaine and Garcia (2012) (FUND_1), the TED spread (FUND_2) and the measure in Hu et al. (2013) (FUND_3). T-statistics are in parenthesis.

	$\hat{p}_m^{(S)}$	$\hat{f}$	$\text{mean}_q \hat{p}_m^{(q)}$	$\text{mean}_q \hat{s}_q$	PS	FUND_1	FUND_2	FUND_3	$\overline{R^2}$
(1)	0.001 (0.734)								0.112
(2)		−0.303*** (−2.642)							0.114
(3)			0.002 (0.515)						0.110
(4)				−0.001 (−0.039)					0.113
(5)					−0.008* (−1.857)				0.111
(6)						−0.283 (−0.723)			0.123
(7)							0.038 (0.467)		0.111
(8)								0.156* (1.824)	0.115
(9)	0.002 (1.210)				−0.006* (−1.741)				0.117
(10)	0.001 (0.403)					−0.296 (−0.751)			0.128
(11)	0.001 (0.798)						0.020 (0.254)		0.116
(12)	0.001 (0.836)							0.170* (1.918)	0.121
(13)		−0.303*** (−2.599)			−0.008* (−1.946)				0.118
(14)		−0.281** (−2.416)				−0.337 (−0.859)			0.129
(15)		−0.289** (−2.402)					0.025 (0.296)		0.118
(16)		−0.311*** (−2.782)						0.161* (1.852)	0.124

***, ** and * denote significance at 1% level, 5% level and 10% level, respectively.

Appendix A. Proofs

We begin by proving a lemma which will be used throughout.

Lemma 1. *Suppose that, for a fixed N, $h_n^{(N)}$ in Eq. (1) of Definition 2.2 is such that $h_n^{(N)}\sqrt{n} \to 0$, as $n \to \infty$. Then,*

$$\Delta_n \sum_{j=1}^{n} \mathcal{S}\left(\frac{\left|\widetilde{X}_{j\,\Delta_n}^{(q)} - X_{(j-1)\,\Delta_n}^{(q)}\right|}{\Theta_{j\,\Delta_n,n}^{(q,N)}}\right)^{1/N} = O_p\left(h_n^{(N)}\sqrt{n}\right) \xrightarrow{p} 0, \tag{26}$$

under both $\mathcal{H}_0$ and $\mathcal{H}_A$, as $n \to \infty$.

Suppose, now, that $h_n^{(N)}$ is such that $h_n^{(N)} \sim \beta^{N^{1+\eta}}/n^{\gamma}$ with $0 < \beta < 1$, $\eta > 1$ and $\gamma > \frac{1}{2}$. Assume, also, that $n \sim a^N$ with $a > 1$. Thus, for $N \to \infty$, we have

$$\Delta_n \sum_{j=1}^{n} \mathcal{S}\left(\frac{\left|\widetilde{X}_{j\,\Delta_n}^{(q)} - X_{(j-1)\,\Delta_n}^{(q)}\right|}{\Theta_{j\,\Delta_n,n}^{(q,N)}}\right)^{1/N} = O_p\left(\left(\frac{h_n^{(N)}\sqrt{n}}{\sqrt{\log(n)}}\right)^{1/N}\right) = O_p\left(\beta^{N^{\eta}}\right) \xrightarrow{p} 0, \tag{27}$$

under both $\mathcal{H}_0$ and $\mathcal{H}_A$.

Finally, define $P_n^{(N)} \doteq p_n^{(1)} p_n^{(2)} \cdots p_n^{(N)}$, $\underline{p}_n \doteq \min\left(p_n^{(1)}, \ldots, p_n^{(N)}\right)$ and

$$\sigma_n^{(N)} \doteq \sqrt{\frac{P_n^{(N)} - \left(P_n^{(N)}\right)^2}{n}}.$$

Given the same conditions on $h_n^{(N)}$, *we have*

$$\frac{\Delta_n}{\sigma_n^{(N)}} \sum_{j=1}^{n} \mathcal{S}\left(\frac{\left|\widetilde{X}^{(q)}_{j\,\Delta_n} - X^{(q)}_{(j-1)\,\Delta_n}\right|}{\Theta^{(q,N)}_{j\,\Delta_n,n}}\right)^{1/N} = O_p\left(\beta^{N^\eta}\left(\frac{a}{\underline{p}_n}\right)^{N/2}\right) \xrightarrow{p} 0, \tag{28}$$

under both $\mathcal{H}_0$ *and* $\mathcal{H}_A$, *as* $N \to \infty$.

Proof of Lemma 1. In order to simplify notation, we will denote by C all constants – even when they change from place to place – whenever it does not cause confusion. Other constants will be denoted by C with superscripts or subscripts. Define

$$\widetilde{\psi}^{(q)}_{j,n} \doteq \frac{\left|\widetilde{X}^{(q)}_{j\,\Delta_n} - X^{(q)}_{(j-1)\,\Delta_n}\right|}{\Theta^{(q,N)}_{j\,\Delta_n,n}}, \quad \mathrm{IT}^{(q)}_n \doteq \Delta_n \sum_{j=1}^{n} \mathbb{I}_{\left\{\widetilde{\psi}^{(q)}_{j,n} \le 1\right\}}. \tag{29}$$

We note that $\widetilde{X}^{(q)}_{j\,\Delta_n}$ is the efficient price process sampled on a discrete time grid (c.f. Assumption 1) whereas $X^{(q)}_{(j-1)\,\Delta_n}$ is the (lagged) observed price process (either under $\mathcal{H}_0$ or under $\mathcal{H}_A$). In fact, $\widetilde{X}^{(q)}_{j\,\Delta_n} - X^{(q)}_{(j-1)\,\Delta_n} = \widetilde{X}^{(q)}_{j\,\Delta_n} - \widetilde{X}^{(q)}_{\left(j-1-f_j^{(q)}\right)\Delta_n}$, where $f_j^{(q)}$ is the random waiting time for a price update. This waiting time was characterized formally under $\mathcal{H}_0$. Under $\mathcal{H}_A$, $f_j^{(q)}$ can be defined analogously, since the dynamics of the observed price process have the same structure as under $\mathcal{H}_0$. This is easy to see. Under $\mathcal{H}_A$, the observed price process behaves as in Eq. (7), which can be re-written as

$$X^{(q)}_{j\,\Delta_n} = \left(1 - S^{(q)}_{j,n}\right) \widetilde{X}^{(q)}_{j\,\Delta_n} + S^{(q)}_{j,n}\, X^{(q)}_{(j-1)\,\Delta_n}, \tag{30}$$

where $S^{(q)}_{j,n}$ is a functional of a triangular array of Bernoulli processes (and a triangular array of Bernoulli processes itself) defined as

$$S^{(q)}_{j,n} \doteq C^{(\mathrm{S})}_{j,n} + B^{(q)}_{j,n} - C^{(\mathrm{S})}_{j,n}\, B^{(q)}_{j,n}, \tag{31}$$

with $\mathbb{P}\left[S^{(q)}_{j,n} = 1\right] = \mathbb{E}\left[S^{(q)}_{j,n}\right] \underset{n\to\infty}{\to} p^{(\mathrm{S})}_\infty + p^{(q)}_\infty - p^{(\mathrm{S})}_\infty\, p^{(q)}_\infty$. We have $0 \le 1 - S^{(q)}_{j,n} \le 1$. Given $\sup(f+g) \le \sup f + \sup g$, $S^{(q)}_{j,n}$ also satisfies the property on the maximum of consecutive flat trades stated in Eq. (6). We conclude that the orders derived below will apply both under $\mathcal{H}_0$ and under $\mathcal{H}_A$.

We begin with a fixed N. Define the event $E^{(q)}_{j,n} \doteq \left\{\left|\widetilde{X}^{(q)}_{j\,\Delta_n} - \widetilde{X}^{(q)}_{\left(j-1-f_j^{(q)}\right)\Delta_n}\right| \le \Theta^{(q,N)}_{j\,\Delta_n,n}\right\}$. We have

$$\mathbb{E}\left[\mathrm{IT}^{(q)}_n\right] = \frac{1}{n}\mathbb{E}\left[\sum_{j=1}^{n} \mathbb{I}_{\left\{E^{(q)}_{j,n}\right\}}\right] = \frac{1}{n}\mathbb{E}\left[\mathbb{E}\left[\sum_{j=1}^{n} \mathbb{I}_{\left\{E^{(q)}_{j,n}\right\}} \middle| f_j^{(q)}, \Theta^{(q,N)}_{j\,\Delta_n,n}\right]\right].$$

Applying the expansion of the indicator function on page 1842 of Bandi et al. (2017), we write

$$\mathbb{E}\left[\mathrm{IT}^{(q)}_n\right] \sim \mathbb{E}\left[\sqrt{\frac{2}{\pi}}\,\frac{h_n^{(N)}\,\xi^{(q)}_{j,n}}{\sqrt{\Delta_n\left(f_j^{(q)}+1\right)}}\,\frac{1}{\sigma^{(q)}_{\left(j-1-f_j^{(q)}\right)\Delta_n}}\right] \le C\, h_n^{(N)}\sqrt{n},$$

for some constant C, which gives $\mathrm{IT}^{(q)}_n = O_p\left(h_n^{(N)}\sqrt{n}\right)$. Now, consider the sets of indexes Ψ^- and Ψ^+ defined as

$$\Psi^- = \left\{1 \le j \le n : \widetilde{\psi}^{(q)}_{j,n} \le 1\right\}, \quad \Psi^+ = \left\{1 \le j \le n : \widetilde{\psi}^{(q)}_{j,n} > 1\right\}, \tag{32}$$

and write the absolute value of the difference between $\Delta_n \sum_{j=1}^{n} \mathcal{S}\left(\frac{\left|\widetilde{X}^{(q)}_{j\,\Delta_n} - X^{(q)}_{(j-1)\,\Delta_n}\right|}{\Theta^{(q,N)}_{j\,\Delta_n,n}}\right)^{1/N}$ and $\mathrm{IT}^{(q)}_n$ as

$$\begin{aligned}
\left|\Delta_n \sum_{j=1}^{n} \mathcal{S}\left(\frac{\left|\widetilde{X}^{(q)}_{j\,\Delta_n} - X^{(q)}_{(j-1)\,\Delta_n}\right|}{\Theta^{(q,N)}_{j\,\Delta_n,n}}\right)^{1/N} - \mathrm{IT}^{(q)}_n\right| &= \Delta_n \left|\sum_{j=1}^{n}\left(\mathcal{S}\left(\widetilde{\psi}^{(q)}_{j,n}\right)^{1/N} - \mathbb{I}_{\left\{\widetilde{\psi}^{(q)}_{j,n}\le 1\right\}}\right)\right| \\
&= \Delta_n \left|\sum_{j\in\Psi^-}\left(\mathcal{S}\left(\widetilde{\psi}^{(q)}_{j,n}\right)^{1/N} - 1\right) + \sum_{j\in\Psi^+}\mathcal{S}\left(\widetilde{\psi}^{(q)}_{j,n}\right)^{1/N}\right| \\
&= \Delta_n \left|\sum_{j\in\Psi^-}^{n} \frac{\mathcal{S}'\left(c^{(q)}_{j,n}\right)}{N\,\mathcal{S}\left(c^{(q)}_{j,n}\right)^{\frac{N-1}{N}}}\,\widetilde{\psi}^{(q)}_{j,n} + \sum_{j\in\Psi^+}^{n}\mathcal{S}\left(\widetilde{\psi}^{(q)}_{j,n}\right)^{1/N}\right| \\
&\le \Delta_n \left|\sum_{j\in\Psi^-}^{n} \frac{C}{N\,\mathcal{S}\left(c^{(q)}_{j,n}\right)}\,\widetilde{\psi}^{(q)}_{j,n} + \sum_{j\in\Psi^+}^{n}\mathcal{S}\left(\widetilde{\psi}^{(q)}_{j,n}\right)^{1/N}\right|
\end{aligned}$$

$$\leq \Delta_n \left| \sum_{j\in\Psi^-}^{n} \frac{C^*}{N} \widetilde{\psi}_{j,n}^{(q)} + \sum_{j\in\Psi^+}^{n} \mathcal{S}\left(\widetilde{\psi}_{j,n}^{(q)}\right)^{1/N} \right|$$
$$\leq C^* \frac{\mathrm{IT}_n^{(q)}}{N} + \underbrace{\Delta_n \sum_{j=1}^{n} \mathcal{S}\left(\widetilde{\psi}_{j,n}^{(q)}\right)^{1/N} \mathbb{I}_{\left\{\widetilde{\psi}_{j,n}^{(q)}>1\right\}}}_{\mathsf{A}_n^{(q,N)}} .$$

The first equality is definitional. The second equality derives from breaking the summation down into two pieces and recognizing that the indicator is either 1 or zero over the two sets in Eq. (32). The third equality uses $\mathcal{S}(0) = 1$, c.f. Definition 2.1, and the mean value theorem with $c_{j,n}^{(q)} \in \left(0, \widetilde{\psi}_{j,n}^{(q)}\right)$. The first inequality follows from (*i*) the boundedness of $\mathcal{S}'(x)$ and (*ii*) the inequality $\frac{1}{\mathcal{S}(x)^{\frac{N-1}{N}}} \leq \frac{1}{\mathcal{S}(x)}$ which, in turn, derives from $0 < \mathcal{S}(x) \leq 1$ (c.f., again, Definition 2.1). Now, because $\mathcal{S}(x) \geq \varepsilon > 0$, for ε small enough, we have $\frac{C}{\mathcal{S}\left(c_{j,n}^{(q)}\right)} \leq \frac{C}{\varepsilon} = C^*$, which yields the second inequality. The last inequality is based on the triangle inequality and the fact that $\Delta_n \sum_{j\in\Psi^-}^{n} \widetilde{\psi}_{j,n}^{(q)} \leq \mathrm{IT}_n^{(q)}$.

Since $\mathrm{IT}_n^{(q)} \geq 0$, the last inequality also implies that[20]

$$\Delta_n \sum_{j=1}^{n} \mathcal{S}\left(\frac{\left|\widetilde{X}_{j\Delta_n}^{(q)} - X_{(j-1)\Delta_n}^{(q)}\right|}{\Theta_{j\Delta_n,n}^{(q,N)}}\right)^{1/N} = \left|\Delta_n \sum_{j=1}^{n} \mathcal{S}\left(\frac{\left|\widetilde{X}_{j\Delta_n}^{(q)} - X_{(j-1)\Delta_n}^{(q)}\right|}{\Theta_{j\Delta_n,n}^{(q,N)}}\right)^{1/N}\right| \leq \left(\frac{C^*}{N} + 1\right) \mathrm{IT}_n^{(q)} + \mathsf{A}_n^{(q,N)}.$$

Turning, now, to $\mathsf{A}_n^{(q,N)}$, define

$$\widetilde{Y}_{j,n}^{(q)} \doteq \left|\widetilde{X}_{j\Delta_n}^{(q)} - \widetilde{X}_{\left(j-1-f_j^{(q)}\right)\Delta_n}^{(q)}\right|,$$

and $t_{j-1,n}^{(q)} \doteq t_{j-1-f_j^{(q)},n}$. We have

$$\mathsf{A}_n^{(q,N)} = \Delta_n \sum_{j=1}^{n} \mathcal{S}\left(\frac{\widetilde{Y}_{j,n}^{(q)}}{\Theta_{j\Delta_n,n}^{(q,N)}}\right)^{1/N} \mathbb{I}_{\left\{\widetilde{Y}_{j,n}^{(q)}>\Theta_{j\Delta_n,n}^{(q,N)}\right\}} \leq \Delta_n \sum_{j=1}^{n} \mathcal{S}\left(\frac{\widetilde{Y}_{j,n}^{(q)}}{\Theta_{j\Delta_n,n}^{(q,N)}}\right)^{1/N}$$
$$= \frac{1}{n} \sum_{j=1}^{n} \mathcal{S}\left(\frac{1}{\Theta_{j\Delta_n,n}^{(q,N)}} \left|\int_{t_{j-1,n}^{(q)}}^{t_{j,n}} \mu_s^{(q)}\, ds + \int_{t_{j-1,n}^{(q)}}^{t_{j,n}} \sigma_s^{(q)}\, dW_s^{(q)}\right|\right)^{1/N}. \tag{33}$$

Now, consider that, for any arbitrary h, it holds that, a.s.,

$$\left|\int_t^{t+h} \mu_s\, ds + \int_t^{t+h} \sigma_s\, dW_s\right| = \left|\int_t^{t+h} \mu_s\, ds + W_{\int_0^{t+h}\sigma_s^2 ds} - W_{\int_0^{t}\sigma_s^2 ds}\right|$$
$$\geq \left|W_{\int_0^{t+h}\sigma_s^2 ds} - W_{\int_0^{t}\sigma_s^2 ds}\right| - \left|\int_t^{t+h} \mu_s\, ds\right|$$
$$\geq \left|W_{\int_0^{t+h}\sigma_s^2 ds} - W_{\int_0^{t}\sigma_s^2 ds}\right| - Ch$$
$$\geq C\left(\sqrt{\left(\int_0^{t+h} \sigma_s^2\, ds - \int_0^{t} \sigma_s^2\, ds\right) \log\left(\frac{1}{\left(\int_0^{t+h} \sigma_s^2\, ds - \int_0^{t} \sigma_s^2\, ds\right)}\right)} - h\right)$$
$$= C\left(\sqrt{\left(\int_t^{t+h} \sigma_s^2\, ds\right) \log\left(\frac{1}{\left(\int_t^{t+h} \sigma_s^2\, ds\right)}\right)} - h\right),$$

where we have temporarily omitted, to avoid clutter, the superscript q in μ, σ and W. In the equality, the Dambis-Dubins–Schwarz Theorem (see, e.g., Theorem 1.6 of Revuz and Yor, 2001) is used to express the stochastic integral $\int_t^{t+h} \sigma_s\, dW_s$ as a time-changed Brownian motion. The first inequality is a triangle inequality. In the second inequality, we employ the boundedness of the drift, c.f. Assumption 1. In the third inequality, for the first term, we use Theorem 1.13 in Mörters and Peres (2010).

[20] The implication is easy to see. For non-negative a, b, C and C' real numbers, it holds

$$|a-b| \leq C\,b + C' \Leftrightarrow -C\,b - C' \leq a - b \leq C\,b + C' \Leftrightarrow -C\,b + b - C' \leq a \leq C\,b + b + C'.$$

But, since $b \geq 0$, then $-C\,b + b - C' \geq -C\,b - b - C'$. Hence, $|a| \leq C\,b + b + C' = (C+1)\,b + C'$.

Because spot variance is bounded from above and from below, i.e., $C_\ell \le \sigma_s^2 \le C_u$ from Assumption 1, we obtain

$$\sqrt{\left(\int_t^{t+h} \sigma_s^2\, ds\right) \log\left(\frac{1}{\left(\int_t^{t+h} \sigma_s^2\, ds\right)}\right)} \ge \left(\sqrt{C_\ell h \log\left(\frac{1}{C_u h}\right)}\right).$$

After absorbing all constants in one as earlier, there exist suitable constants C such that, uniformly in j, for a sufficiently large n, a.s.,

$$\begin{aligned}
\frac{1}{\Theta^{(q,N)}_{j\,\Delta_n,n}} \left| \int_{t^{(q)}_{j-1,n}}^{t_{j,n}} \mu_s^{(q)}\, ds + \int_{t^{(q)}_{j-1,n}}^{t_{j,n}} \sigma_s^{(q)}\, dW_s^{(q)} \right| &\ge \frac{C}{h_n^{(N)}} \left(\sqrt{\frac{1+f_j}{n} \log\left(\frac{n}{C_u\,(1+f_j)}\right)} - \frac{1+f_j}{n} \right) \\
&\ge \frac{C}{h_n^{(N)}} \left(\sqrt{\frac{1}{n} \log\left(\frac{n}{C_u(1+\frac{n^\alpha}{\log n})}\right)} - \frac{1+\frac{n^\alpha}{\log n}}{n} \right) \\
&\ge \frac{C}{h_n^{(N)}} \left(\frac{1}{\sqrt{n}} \sqrt{\log\left(\frac{n \log n}{\log n + n^\alpha}\right)} - \frac{1}{n} - \frac{n^{\alpha-1}}{\log n} \right) \longrightarrow \infty, \qquad (34)
\end{aligned}$$

since $h_n^{(N)}\sqrt{n} \to 0$ and $\alpha < 1/2$. Given the integrability of the smoother, $\lim_{x\to\infty} x\,\mathcal{S}(x) = 0$, which implies that, for x sufficiently large, $\mathcal{S}(x) \le \frac{C}{x}$, where C defines, as always, a suitable constant. This gives $\mathcal{S}(x)^{1/N} \le \left(\frac{C}{x}\right)^{1/N}$, for x large enough. Hence, from Eqs. (33)–(34) and the continuous mapping theorem, we obtain

$$0 \le \mathsf{A}_n^{(q,N)} \le \left(C\, h_n^{(N)} \sqrt{n} \left(\log\left(\frac{n \log n}{\log n + n^\alpha}\right)\right)^{-1/2} \right)^{1/N} \sim \left(\frac{h_n^{(N)}\sqrt{n}}{\sqrt{\log n}} \right)^{1/N},$$

for a fixed N and a large n, where the order of the bound is the result of the following observation:

$$\log\left(\frac{n \log n}{\log n + n^\alpha}\right) = \log(n \log n) - \log(\log n + n^\alpha) = \log n + \log(\log n) - \log(\log n + n^\alpha) \sim \log n.$$

In sum,

$$0 \le \Delta_n \sum_{j=1}^n \mathcal{S}\left(\frac{\left|\widetilde{X}^{(q)}_{j\,\Delta_n} - X^{(q)}_{(j-1)\,\Delta_n}\right|}{\Theta^{(q,N)}_{j\,\Delta_n,n}} \right)^{1/N} \le \underbrace{O_p\left(h_n^{(N)}\sqrt{n}\right)}_{\text{from } \mathsf{IT}_n^{(q)}} + \underbrace{O_p\left(\left(\frac{h_n^{(N)}\sqrt{n}}{\sqrt{\log n}}\right)^{1/N}\right)}_{\text{from } \mathsf{A}_n^{(q,N)}}. \qquad (35)$$

Thus, if N is fixed, it is enough to have $h_n^{(N)}\sqrt{n} \to 0$ in order to obtain

$$\Delta_n \sum_{j=1}^n \mathcal{S}\left(\frac{\left|\widetilde{X}^{(q)}_{j\,\Delta_n} - X^{(q)}_{(j-1)\,\Delta_n}\right|}{\Theta^{(q,N)}_{j\,\Delta_n,n}} \right)^{1/N} = O_p\left(h_n^{(N)}\sqrt{n}\right) \xrightarrow{p} 0$$

as $n \to \infty$, which is, e.g., guaranteed by $h_n^{(N)} \sim \frac{\beta^{N^{1+\eta}}}{n^\gamma}$ with $\gamma > \frac{1}{2}$. This proves Eq. (26).

Let us now turn to the case $N \to \infty$. Writing, again, $h_n^{(N)} \sim \frac{\beta^{N^{1+\eta}}}{n^\gamma}$, we obtain

$$\mathsf{A}_n^{(q,N)} = O_p\left(\left(\frac{h_n^{(N)}\sqrt{n}}{\sqrt{\log n}}\right)^{1/N}\right) = O_p\left(\beta^{N^\eta}\left(\frac{n^{1/2-\gamma}}{\sqrt{\log n}}\right)^{1/N}\right).$$

If $n \sim a^N$, with $a > 1$, we also have, for large N,

$$\beta^{N^\eta}\left(\frac{a^{N\,(1/2-\gamma)}}{\sqrt{N\,\log a}}\right)^{1/N} \sim \beta^{N^\eta}. \qquad (36)$$

In order to establish which among the two terms in Eq. (35) dominate as $N \to \infty$, consider that

$$h_n^{(N)}\sqrt{n}\,\frac{(\log n)^{1/(2N)}}{\left(h_n^{(N)}\sqrt{n}\right)^{1/N}} = \left(h_n^{(N)}\sqrt{n}\right)^{1-1/N} (\log n)^{1/(2N)} \sim \left(\frac{\beta^{N^{1+\eta}}}{a^{N\left(\gamma-\frac{1}{2}\right)}}\right)^{1-1/N} (N\,\log a)^{1/(2N)} \to 0.$$

Hence, $\mathsf{A}_n^{(q,N)}$ is of lower asymptotic order than $\mathsf{IT}_n^{(q)}$. This finding, along with Eqs. (35) and (36), yields Eq. (27).

To prove Eq. (28), define

$$\underline{p}_n \doteq \min\left(p_n^{(1)},\ldots,p_n^{(N)}\right), \qquad \overline{p}_n \doteq \max\left(p_n^{(1)},\ldots,p_n^{(N)}\right) \qquad (37)$$

so that the product $P_n^{(N)} \doteq p_n^{(1)} p_n^{(2)} \cdots p_n^{(N)}$ of all of the idiosyncratic probabilities verifies

$$\underline{p}_n^N \le P_n^{(N)} \le \overline{p}_n^N.$$

Thus, for large N,

$$\sqrt{\frac{\underline{p}_n^N}{n}} \le \sigma_n^{(N)} \le \sqrt{\frac{\overline{p}_n^N}{n}}.$$

Finally,

$$\begin{aligned} \frac{1}{\sigma_n^{(N)}} \Delta_n \sum_{j=1}^n \mathcal{S}\left(\frac{\left|\widetilde{X}_{j\Delta_n}^{(q)} - X_{(j-1)\Delta_n}^{(q)}\right|}{\Theta_{j\Delta_n,n}^{(q,N)}}\right)^{1/N} &\le O_p\left(\frac{\sqrt{n}}{\underline{p}_n^{N/2}} \beta^{N^\eta}\right) \\ (\text{recalling } \; n \sim a^N) &= O_p\left(\beta^{N^\eta}\left(\frac{a}{\underline{p}_n}\right)^{N/2}\right) \xrightarrow{p} 0, \end{aligned}$$

provided $\eta > 1$. The condition on η depends on the following observation. Taking logs, one can write $\phi(N) = N^\eta \log \beta + \frac{N}{2} \log \frac{a}{\underline{p}_n} = N\left(N^{\eta-1} \log \beta + \frac{1}{2} \log \frac{a}{\underline{p}_n}\right)$. Because $\beta < 1$ and $\frac{a}{\underline{p}_n} > 1$, for $\eta > 1$, $\phi(N) \to -\infty$ as $N \to \infty$ and $\exp(\phi(N)) \to 0$. This proves Eq. (28). □

Proof of Theorem 1. First, note that, under $\mathcal{H}_0$,

$$X_{j\Delta_n}^{(q)} - X_{(j-1)\Delta}^{(q)} = \left(\widetilde{X}_{j\Delta_n}^{(q)} - X_{(j-1)\Delta_n}^{(q)}\right)\left(1 - B_{j,n}^{(q)}\right).$$

The key of the proof is to recognize that

$$\begin{aligned} \mathbb{U}_n^{(q)} &= \Delta_n \sum_{j=1}^n \mathcal{S}\left(\frac{\left|\left(\widetilde{X}_{j\Delta_n}^{(q)} - X_{(j-1)\Delta_n}^{(q)}\right)\left(1 - B_{j,n}^{(q)}\right)\right|}{\Theta_{j\Delta_n,n}^{(q,N)}}\right) \\ &= \Delta_n \sum_{j=1}^n B_{j,n}^{(q)} \mathcal{S}(0) + \Delta_n \sum_{j=1}^n \left(1 - B_{j,n}^{(q)}\right) \mathcal{S}\left(\frac{\left|\widetilde{X}_{j\Delta_n}^{(q)} - X_{(j-1)\Delta_n}^{(q)}\right|}{\Theta_{j\Delta_n,n}^{(q,N)}}\right) \\ &= \underbrace{\Delta_n \sum_{j=1}^n B_{j,n}^{(q)}}_{\widetilde{\mathbb{U}}_n^{(q)}} + \underbrace{\Delta_n \sum_{j=1}^n \left(1 - B_{j,n}^{(q)}\right) \mathcal{S}\left(\frac{\left|\widetilde{X}_{j\Delta_n}^{(q)} - X_{(j-1)\Delta_n}^{(q)}\right|}{\Theta_{j\Delta_n,n}^{(q,N)}}\right)}_{\mathsf{R}_n^{(q)}}, \end{aligned} \tag{38}$$

since $\mathcal{S}(0) = 1$. Now, as $n \to \infty$, we have $\widetilde{\mathbb{U}}_n^{(q)} \xrightarrow{p} p_\infty^{(q)}$, given Eq. (5). Because $0 \le 1 - B_{j,n}^{(q)} \le 1$, Lemma 1 gives

$$\mathsf{R}_n^{(q)} = O_p\left(h_n^{(N)} \sqrt{n}\right) \xrightarrow{p} 0.$$

Thus, under $\mathcal{H}_0$, we have, as $n \to \infty$, $\mathbb{U}_n^{(q)} \xrightarrow{p} p_\infty^{(q)}$.

The N-multivariate staleness estimator can be studied analogously. We begin, as always, with N fixed. To clarify the logic, let us focus on the bivariate case first, i.e., the object $\mathbb{M}_n^{(2)}$. Write

$$\begin{aligned} \mathbb{M}_n^{(2)} = \; & \underbrace{\Delta_n \sum_{j=1}^n B_{j,n}^{(1)} B_{j,n}^{(2)}}_{\widetilde{\mathbb{M}}_n^{(2)}} + \underbrace{\Delta_n \sum_{j=1}^n B_{j,n}^{(1)} \left(1 - B_{j,n}^{(2)}\right) \mathcal{S}\left(\widetilde{\psi}_{j,n}^{(2)}\right)^{1/2}}_{\mathsf{R}_n} + \underbrace{\Delta_n \sum_{j=1}^n \left(1 - B_{j,n}^{(1)}\right) B_{j,n}^{(2)} \mathcal{S}\left(\widetilde{\psi}_{j,n}^{(1)}\right)^{1/2}}_{\mathsf{R}'_n} + \\ & \underbrace{+ \Delta_n \sum_{j=1}^n \left(1 - B_{j,n}^{(1)}\right)\left(1 - B_{j,n}^{(2)}\right) \mathcal{S}\left(\widetilde{\psi}_{j,n}^{(1)}\right)^{1/2} \mathcal{S}\left(\widetilde{\psi}_{j,n}^{(2)}\right)^{1/2}}_{\mathsf{R}''_n}, \end{aligned}$$

where the quantity $\widetilde{\psi}_{j,n}^{(q)}$ is defined in Eq. (29). The Bernoulli variates $B_{j,n}^{(1)}$ and $B_{j,n}^{(2)}$ in $\widetilde{\mathbb{M}}_n^{(2)}$ are independent. Thus, we have $\mathbb{E}\left[B_{j,n}^{(1)} B_{j,n}^{(2)}\right] = p_n^{(1)} p_n^{(2)}$ and $\widetilde{\mathbb{M}}_n^{(2)} \xrightarrow[n\to\infty]{p} p_\infty^{(1)} p_\infty^{(2)}$, given Eq. (5). Regarding the terms R_n, R'_n and R''_n, using Lemma 1 and the fact that each Bernoulli variates can be either zero or one, it is readily shown that, just like the remainder term $\mathsf{R}_n^{(q)}$ above, they all converge to zero in probability, when $n \to \infty$, provided $h_n^{(N)} \sqrt{n} \to 0$. Hence, under $\mathcal{H}_0$, we have $\mathbb{M}_n^{(2)} \xrightarrow{p} p_\infty^{(1)} p_\infty^{(2)}$, as $n \to \infty$.

The bivariate result is straightforwardly extended to the situation in which N is generic. In this case,

$$\mathbb{M}_n^{(N)} = \underbrace{\Delta_n \sum_{j=1}^{n} \prod_{q=1}^{N} B_{j,n}^{(q)}}_{\widetilde{\mathbb{M}}_n^{(N)}} + \underbrace{\Delta_n \sum_{\substack{\mathsf{F}_j^{(1)},\dots,\mathsf{F}_j^{(N)} \\ \mathsf{F}_j^{(q)}=B_{j,n}^{(q)} \text{ or } \mathsf{F}_j^{(q)}=1-B_{j,n}^{(q)} \\ \left(\mathsf{F}_j^{(1)},\dots,\mathsf{F}_j^{(N)}\right)\neq\left(B_{j,n}^{(1)},\dots,B_{j,n}^{(N)}\right)}} \sum_{j=1}^{n} \mathsf{F}_j^{(1)} \cdots \mathsf{F}_j^{(N)} \chi\left(\mathsf{F}_j^{(1)}\right) \cdots \chi\left(\mathsf{F}_j^{(N)}\right)}_{\mathsf{R}_n^{(N)}}, \tag{39}$$

where the sum defining $\mathsf{R}_n^{(N)}$ is over all possible N-tuples $\left(\mathsf{F}_j^{(1)},\dots,\mathsf{F}_j^{(N)}\right)$ such that the qth element of the N-tuple, i.e. $\mathsf{F}_j^{(q)}$, can be either $1-B_{j,n}^{(q)}$ or $B_{j,n}^{(q)}$, but the N-tuple in which $\mathsf{F}_j^{(q)} = B_{j,n}^{(q)}$ for all q is excluded, and where the functional $\chi(\cdot)$ is defined as

$$\chi\left(\mathsf{F}_j^{(q)}\right) \doteq \begin{cases} \mathcal{S}\left(\widetilde{\psi}_{j,n}^{(q)}\right)^{1/N} & \text{if } \mathsf{F}_j^{(q)} = 1 - B_{j,n}^{(q)} \\ 1 & \text{if } \mathsf{F}_j^{(q)} = B_{j,n}^{(q)} \end{cases}.$$

For a given N, because of Eq. (5), the quantity $\widetilde{\mathbb{M}}_n^{(N)}$ is consistent in probability for $\prod_{q=1}^{N} p_\infty^{(q)}$. The remaining $2^N - 1$ terms in $\mathsf{R}_n^{(N)}$ vanish in probability by virtue of Lemma 1. In fact, a generic term in the summation over the N-tuples $\left(\mathsf{F}_j^{(1)},\dots,\mathsf{F}_j^{(N)}\right)$ is bounded from above by

$$0 \leq \Delta_n \sum_{j=1}^{n} \mathsf{F}_j^{(1)} \cdots \mathsf{F}_j^{(N)} \chi\left(\mathsf{F}_j^{(1)}\right) \cdots \chi\left(\mathsf{F}_j^{(N)}\right) \leq \Delta_n \sum_{j=1}^{n} \mathcal{S}\left(\widetilde{\psi}_{j,n}^{(\overline{q})}\right)^{1/N} = O_p\left(h_n^{(N)}\sqrt{n}\right) \to 0$$

having assumed, without loss of generality, that $\overline{q} \in \{1,\dots,N\}$ is such that $\mathsf{F}_j^{(\overline{q})} = 1 - B_{j,n}^{(\overline{q})}$ and having used the properties $0 \leq \mathsf{F}_j^{(q)} \leq 1$ and $0 < \mathcal{S}(x) \leq 1$. We conclude that, for N fixed, $\mathbb{M}_n^{(N)} \xrightarrow{p} \prod_{q=1}^{N} p_\infty^{(q)}$.

When $N \to \infty$, being $\eta > 1$, we have

$$\mathsf{R}_n^{(N)} = O_p\left(2^N \left(\frac{h_n^{(N)}\sqrt{n}}{\sqrt{\log n}}\right)^{1/N}\right) = O_p\left(2^N \beta^{N^\eta}\right) \xrightarrow{p} 0,$$

given the choice of bandwidth in Eqs. (10) and (11). Also, in this case, trivially, $\prod_{q=1}^{N} p_\infty^{(q)} \to 0$, when $N \to \infty$. Thus, $\mathbb{M}_n^{(N)} \xrightarrow{p} 0$, as $N \to \infty$.

In order to compute the limit in probability of $\mathbb{U}_n^{(q)}$ under $\mathcal{H}_A$, we re-arrange the representation in Eq. (30). Write

$$X_{j\Delta_n}^{(q)} - X_{(j-1)\Delta_n}^{(q)} = \left(\widetilde{X}_{j\Delta_n}^{(q)} - X_{(j-1)\Delta_n}^{(q)}\right)\left(1 - S_{j,n}^{(q)}\right),$$

with $S_{j,n}^{(q)}$ defined in Eq. (31). As earlier, this expression leads to

$$\mathbb{U}_n^{(q)} = \underbrace{\Delta_n \sum_{j=1}^{n} S_{j,n}^{(q)}}_{\widetilde{\mathbb{U}}_n^{\prime(q)}} + \underbrace{\Delta_n \sum_{j=1}^{n}\left(1 - S_{j,n}^{(q)}\right) \mathcal{S}\left(\frac{\left|\widetilde{X}_{j\Delta_n}^{(q)} - X_{(j-1)\Delta_n}^{(q)}\right|}{\Theta_{j\Delta_n,n}^{(q,N)}}\right)}_{\mathsf{R}_n^{\prime(q)}}. \tag{40}$$

Now, as $n \to \infty$, $\widetilde{\mathbb{U}}_n^{\prime(q)} \xrightarrow{p} p_\infty^{(\mathsf{S})} + p_\infty^{(q)} - p_\infty^{(\mathsf{S})} p_\infty^{(q)}$, given Eq. (5), Eq. (8), and Eq. (9). Also, by virtue of Lemma 1,

$$\mathsf{R}_n^{\prime(q)} = O_p\left(h_n^{(N)}\sqrt{n}\right) \xrightarrow{p} 0,$$

as $n \to \infty$. In sum,

$$\mathbb{U}_n^{(q)} \xrightarrow{p} p_\infty^{(\mathsf{S})} + p_\infty^{(q)} - p_\infty^{(\mathsf{S})} p_\infty^{(q)} \quad \text{as } n \to \infty.$$

Because the processes $S_{j,n}^{(q)}$ are not pairwise (i.e., for different pairs of q) independent, in order to evaluate $\mathbb{M}_n^{(N)}$, it is convenient to employ the original representation in Eq. (7). To clarify the logic, we use again the quantities $\widetilde{\psi}_{j,n}^{(q)}$ defined in Eq. (29) and consider, first, the case $N = 2$:

$$\begin{aligned}
\mathbb{M}_n^{(2)} &= \Delta_n \sum_{j=1}^{n} \mathcal{S}\left(\frac{\left|X_{j\Delta_n}^{(1)} - X_{(j-1)\Delta_n}^{(1)}\right|}{H_{j,n}^{(1)}}\right)^{1/2} \mathcal{S}\left(\frac{\left|X_{j\Delta_n}^{(2)} - X_{(j-1)\Delta_n}^{(2)}\right|}{H_{j,n}^{(2)}}\right)^{1/2} \\
&= \Delta_n \sum_{j=1}^{n} \mathcal{S}\left(\frac{\left|\widetilde{X}_{j\Delta_n}^{(1)} - X_{(j-1)\Delta_n}^{(1)}\right|}{H_{j,n}^{(1)}}\left(1 - C_{j,n}^{(\mathsf{S})}\right)\left(1 - B_{j,n}^{(1)}\right)\right)^{1/2} \mathcal{S}\left(\frac{\left|\widetilde{X}_{j\Delta_n}^{(2)} - X_{(j-1)\Delta_n}^{(2)}\right|}{H_{j,n}^{(2)}}\left(1 - C_{j,n}^{(\mathsf{S})}\right)\left(1 - B_{j,n}^{(2)}\right)\right)^{1/2} \\
&= \Delta_n \sum_{j=1}^{n} C_{j,n}^{(\mathsf{S})} + \Delta_n \sum_{j=1}^{n}\left(1 - C_{j,n}^{(\mathsf{S})}\right)\left[B_{j,n}^{(1)} B_{j,n}^{(2)} + \left(1 - B_{j,n}^{(1)}\right) B_{j,n}^{(2)} \mathcal{S}\left(\widetilde{\psi}_{j,n}^{(1)}\right)^{1/2} + B_{j,n}^{(1)}\left(1 - B_{j,n}^{(2)}\right) \mathcal{S}\left(\widetilde{\psi}_{j,n}^{(2)}\right)^{1/2} + \right.
\end{aligned}$$

$$+\left(1-B_{j,n}^{(1)}\right)\left(1-B_{j,n}^{(2)}\right)\,\mathcal{S}\left(\widetilde{\psi}_{j,n}^{(1)}\right)^{1/2}\mathcal{S}\left(\widetilde{\psi}_{j,n}^{(2)}\right)^{1/2}\Bigg]$$

$$= \Delta_n\sum_{j=1}^n C_{j,n}^{(S)} + \Delta_n\sum_{j=1}^n B_{j,n}^{(1)}B_{j,n}^{(2)} - \Delta_n\sum_{j=1}^n C_{j,n}^{(S)}B_{j,n}^{(1)}B_{j,n}^{(2)} + \Delta_n\sum_{j=1}^n\left(1-C_{j,n}^{(S)}\right)\left[\left(1-B_{j,n}^{(1)}\right)B_{j,n}^{(2)}\mathcal{S}\left(\widetilde{\psi}_{j,n}^{(1)}\right)^{1/2}+\right.$$

$$\left.+B_{j,n}^{(1)}\left(1-B_{j,n}^{(2)}\right)\mathcal{S}\left(\widetilde{\psi}_{j,n}^{(2)}\right)^{1/2}+\left(1-B_{j,n}^{(1)}\right)\left(1-B_{j,n}^{(2)}\right)\,\mathcal{S}\left(\widetilde{\psi}_{j,n}^{(1)}\right)^{1/2}\mathcal{S}\left(\widetilde{\psi}_{j,n}^{(2)}\right)^{1/2}\right].$$

As $n\to\infty$, the first three terms satisfy $\Delta_n\sum_{j=1}^n C_{j,n}^{(S)}\xrightarrow{p} p_\infty^{(S)}$, $\Delta_n\sum_{j=1}^n B_{j,n}^{(1)}B_{j,n}^{(2)}\xrightarrow[n\to\infty]{p} p_\infty^{(1)}p_\infty^{(2)}$, and $\Delta_n\sum_{j=1}^n C_{j,n}^{(S)}B_{j,n}^{(1)}B_{j,n}^{(2)}\xrightarrow{p} p_\infty^{(S)}p_\infty^{(1)}p_\infty^{(2)}$, respectively, given Eq. (5), Eq. (8), and Eq. (9). The other terms vanish in probability, given Lemma 1, since $h_n^{(N)}\sqrt{n}\to 0$. Hence,

$$\mathbb{M}_n^{(2)}\xrightarrow[n\to\infty]{p} p_\infty^{(S)}+p_\infty^{(1)}p_\infty^{(2)}-p_\infty^{(S)}p_\infty^{(1)}p_\infty^{(2)}=p_\infty^{(S)}+\left(1-p_\infty^{(S)}\right)p_\infty^{(1)}p_\infty^{(2)}.$$

For the generic N-multivariate case, we may easily extend the representation used for $N=2$, thereby obtaining

$$\mathbb{M}_n^{(N)}=\Delta_n\sum_{j=1}^n C_{j,n}^{(S)}+\Delta_n\sum_{j=1}^n\left(1-C_{j,n}^{(S)}\right)\prod_{q=1}^N B_{j,n}^{(q)}+\mathsf{R}_n^{\prime(N)}\xrightarrow[n\to\infty]{p}p_\infty^{(S)}+\left(1-p_\infty^{(S)}\right)\prod_{q=1}^N p_\infty^{(q)},\tag{41}$$

for fixed N. For $N\to\infty$ and $h_n^{(N)}$ defined as in Eqs. (10) and (11), we obtain

$$\mathbb{M}_n^{(N)}\xrightarrow{p}p_\infty^{(S)}$$

since, by Lemma 1 and the same reasoning used for $\mathsf{R}_n^{(N)}$, we have

$$\mathsf{R}_n^{\prime(N)}=O_p\left(2^N\beta^{N^\eta}\right)\xrightarrow{p}0$$

and since

$$\Delta_n\sum_{j=1}^n\left(1-C_{j,n}^{(S)}\right)\prod_{q=1}^N B_{j,n}^{(q)}=O_p\left(\overline{p}_n^N\right)\xrightarrow{p}0.$$

The last result depends on the fact that

$$\mathbb{P}\left[\frac{\prod_{q=1}^N B_{j,n}^{(q)}}{\overline{p}_n^N}>C\right]\le\frac{\mathbb{E}\left[\prod_{q=1}^N B_{j,n}^{(q)}\right]}{\overline{p}_n^N\,C}\le\frac{\overline{p}_n^N}{\overline{p}_n^N\,C}=\frac{1}{C}=\varepsilon,\tag{42}$$

where $\overline{p}_n$ is defined in Eq. (37) as the maximum probability among $\left\{p_n^{(1)},\dots,p_n^{(N)}\right\}$. □

Proof of Theorem 2. Under $\mathcal{H}_0$, using the expansion in Eq. (39), we may write, for a generic N,

$$\mathbb{Z}_n^{(N)}=\frac{\mathbb{M}_n^{(N)}-P_n^{(N)}}{\sigma_n^{(N)}}=\underbrace{\frac{\Delta_n\sum_{j=1}^n B_{j,n}^{(1)}\cdots B_{j,n}^{(N)}-P_n^{(N)}}{\sigma_n^{(N)}}}_{\widetilde{\mathbb{Z}}_n^{(N)}}+\frac{1}{\sigma_n^{(N)}}\mathsf{R}_n^{(N)}.$$

Define, now, $Y_{j,n}\doteq B_{j,n}^{(1)}\cdots B_{j,n}^{(N)}-P_n^{(N)}$. Note that $\mathbb{E}\left[Y_{j,n}\right]=0$ and $\mathbb{E}\left[\left(Y_{j,n}\right)^2\right]=P_n^{(N)}\left(1-P_n^{(N)}\right)$. After, also, defining

$$s_n^2\doteq\sum_{j=1}^n\mathbb{V}\left[Y_{j,n}\right]=nP_n^{(N)}\left(1-P_n^{(N)}\right),$$

we have that,

$$\frac{1}{s_n^4}\sum_{j=1}^n\mathbb{E}\left[\left(Y_{j,n}\right)^4\right]\sim\frac{nP_n^{(N)}}{n^2(P_n^{(N)})^2}\sim\frac{1}{nP_n^{(N)}}=O\left(\frac{1}{a^N\underline{p}_n^N}\right)\to 0$$

by Eq. (15). The Lindeberg–Feller Theorem (c.f. Theorem 27.2 in Billingsley, 1995), now, yields

$$\frac{1}{s_n}\sum_{j=1}^n Y_{j,n}\xrightarrow{d}N(0,1),$$

or, equivalently,

$$\widetilde{\mathbb{Z}}_n^{(N)}=\frac{\Delta_n\sum_{j=1}^n B_{j,n}^{(1)}\cdots B_{j,n}^{(N)}-P_n^{(N)}}{\sigma_n^{(N)}}\xrightarrow{d}N(0,1).$$

By virtue of Eq. (28) in Lemma 1, we also have

$$\frac{1}{\sigma_n^{(N)}}\,\mathsf{R}_n^{(N)} = O_p\left(2^N\left(\frac{a}{\underline{p}_n}\right)^{N/2}\beta^{N^\eta}\right) = O_p\left(\left(\frac{4\,a}{\underline{p}_n}\right)^{N/2}\beta^{N^\eta}\right)\xrightarrow{p} 0,$$

provided $\eta > 1$. We conclude that

$$\mathbb{Z}_n^{(N)} \xrightarrow{d} N\left(0,1\right),$$

which proves Eq. (16). Under $\mathcal{H}_A$, because of Theorem 1, we obtain

$$\mathbb{Z}_n^{(N)} = \frac{\overbrace{\mathbb{M}_n^{(N)}}^{\xrightarrow{p} p_\infty^{(\mathbb{S})}} - \overbrace{P_n^{(N)}}^{\xrightarrow{p} 0}}{\underbrace{\sigma_n^{(N)}}_{\to 0}} \xrightarrow{p} \infty,$$

which proves Eq. (17).

In order to verify Eqs. (18) and (19), write

$$\mathbb{Id}_n^{(q,N)} - p_n^{(q)} = \frac{\mathbb{U}_n^{(q)} - \mathbb{M}_n^{(N)}}{1-\mathbb{M}_n^{(N)}} - p_n^{(q)} = \frac{\mathbb{U}_n^{(q)} - \mathbb{M}_n^{(N)} - p_n^{(q)} + p_n^{(q)}\,\mathbb{M}_n^{(N)}}{1-\mathbb{M}_n^{(N)}} \doteq \frac{\mathsf{N}_n^{(q,N)}}{\mathsf{D}_n^{(N)}}. \tag{43}$$

Under $\mathcal{H}_0$, the numerator $\mathsf{N}_n^{(q,N)}$ in Eq. (43) is given by

$$\begin{aligned}\mathsf{N}_n^{(q,N)} &= \Delta_n\sum_{j=1}^n B_{j,n}^{(q)} - p_n^{(q)} + \mathsf{R}_n^{(q)} - \left(1-p_n^{(q)}\right)\Delta_n\sum_{j=1}^n\prod_{q=1}^N B_{j,n}^{(q)} - \left(1-p_n^{(q)}\right)\mathsf{R}_n^{(N)}\\ &= \Delta_n\sum_{j=1}^n\left(B_{j,n}^{(q)} - p_n^{(q)} - \left(1-p_n^{(q)}\right)\prod_{q=1}^N B_{j,n}^{(q)}\right) + O_p\left(2^N\,\beta^{N^\eta}\right),\end{aligned} \tag{44}$$

where the remainders $\mathsf{R}_n^{(q)}$ and $\mathsf{R}_n^{(N)}$ are defined, respectively, in Eqs. (38) and (39) and their asymptotic order in Eq. (44) follows from Lemma 1. Thus, using $\prod_{q=1}^N B_{j,n}^{(q)} = O_p\left(\overline{p}_n^N\right)$, which is justified in Eq. (42), we have that

$$\mathsf{N}_n^{(q,N)} = \Delta_n\sum_{j=1}^n\left(B_{j,n}^{(q)} - p_n^{(q)}\right) + O_p\left(\overline{p}_n^N\right) + O_p\left(2^N\,\beta^{N^\eta}\right).$$

Define, now, $Y_{j,n}^{\prime(q)} \doteq B_{j,n}^{(q)} - p_n^{(q)}$. Note that $\mathbb{E}\left[Y_{j,n}^{\prime(q)}\right] = 0$ and $\mathbb{E}\left[\left(Y_{j,n}^{\prime(q)}\right)^2\right] = p_n^{(q)}\left(1-p_n^{(q)}\right)$. After, also, defining

$$(s_n^{\prime})^2 \doteq \sum_{j=1}^n\mathbb{V}\left[Y_{j,n}^{\prime(q)}\right] = n\,p_n^{(q)}\,\left(1-p_n^{(q)}\right),$$

we have that

$$\frac{1}{(s_n^{\prime})^4}\sum_{j=1}^n\mathbb{E}\left[\left(Y_{j,n}^{\prime(q)}\right)^4\right] \sim \frac{np_n^{(q)}}{n^2(p_n^{(q)})^2} \sim \frac{1}{np_n^{(q)}} \to 0 \tag{45}$$

as $N\to\infty$, since $p_n^{(q)} \to p_\infty^{(q)} \in (0,1)$. The Lindeberg–Feller Theorem, now, gives

$$\frac{1}{s_n^{\prime}}\sum_{j=1}^n Y_{j,n}^{\prime(q)} \xrightarrow{d} N\left(0,1\right),$$

which implies

$$\sqrt{n}\,\frac{\mathsf{N}_n^{(q,N)}}{\sqrt{p_n^{(q)}\left(1-p_n^{(q)}\right)}} = \frac{1}{\sqrt{n\,p_n^{(q)}\left(1-p_n^{(q)}\right)}}\sum_{j=1}^n Y_{j,n}^{\prime(q)} + O_p\left(\sqrt{a^N}\,\overline{p}_n^N\right) + O_p\left(\sqrt{a^N}\,2^N\,\beta^{N^\eta}\right)\xrightarrow{d} N\left(0,1\right),$$

by Eq. (14). Finally, since, under $\mathcal{H}_0$, $\mathbb{M}_n^{(N)} \xrightarrow{p} 0$, we obtain $\mathsf{D}_n^{(N)} \xrightarrow{p} 1$ and Slutsky's theorem yields

$$\sqrt{n}\,\frac{\mathbb{Id}_n^{(q,N)} - p_n^{(q)}}{\sqrt{p_n^{(q)}\left(1-p_n^{(q)}\right)}} \xrightarrow{d} N\left(0,1\right),$$

which proves Eq. (18). Under $\mathcal{H}_A$, we have that the numerator $\mathsf{N}_n^{(q,N)}$, defined in Eq. (43), can be expressed as

$$\begin{aligned}\mathsf{N}_n^{(q,N)} &= \Delta_n \sum_{j=1}^{n} \left(C_{j,n}^{(\mathrm{S})} + B_{j,n}^{(q)} - C_{j,n}^{(\mathrm{S})} B_{j,n}^{(q)} - p_n^{(q)} \right) + \\ &+ \mathsf{R}_n'^{(q)} - \left(1 - p_n^{(q)}\right) \left(\Delta_n \sum_{j=1}^{n} C_{j,n}^{(\mathrm{S})} + \Delta_n \sum_{j=1}^{n} \left(1 - C_{j,n}^{(\mathrm{S})}\right) \prod_{q=1}^{N} B_{j,n}^{(q)} + \mathsf{R}_n'^{(N)} \right),\end{aligned}$$

where $\mathsf{R}_n'^{(q)}$ is defined in Eq. (40) and $\mathsf{R}_n'^{(N)}$ is defined in Eq. (41). Given the asymptotic orders, i.e.,

$$\mathsf{R}_n'^{(q)} = O_p\left(h_n^{(N)} \sqrt{n}\right), \quad \mathsf{R}_n'^{(N)} = O_p\left(2^N \beta^{N^\eta}\right), \quad \Delta_n \sum_{j=1}^{n} \left(1 - C_{j,n}^{(\mathrm{S})}\right) \prod_{q=1}^{N} B_{j,n}^{(q)} = O_p\left(\overline{p}_n^N\right),$$

since $\frac{\mathsf{R}_n'^{(q)}}{\mathsf{R}_n'^{(N)}} \to 0$ immediately as $N \to \infty$, we only retain the last two orders:

$$\begin{aligned}\mathsf{N}_n^{(q,N)} &= \Delta_n \sum_{j=1}^{n} \left(C_{j,n}^{(\mathrm{S})} + B_{j,n}^{(q)} - C_{j,n}^{(\mathrm{S})} B_{j,n}^{(q)} - p_n^{(q)} - \left(1 - p_n^{(q)}\right) C_{j,n}^{(\mathrm{S})} \right) + O_p\left(2^N \beta^{N^\eta}\right) + O_p\left(\overline{p}_n^N\right) \\ &= \Delta_n \sum_{j=1}^{n} \left(C_{j,n}^{(\mathrm{S})} \left(1 - B_{j,n}^{(q)} - \left(1 - p_n^{(q)}\right)\right) + B_{j,n}^{(q)} - p_n^{(q)} \right) + O_p\left(2^N \beta^{N^\eta}\right) + O_p\left(\overline{p}_n^N\right).\end{aligned}$$

Because the expected value of $Y_{j,n}''^{(q)} \doteq \left(C_{j,n}^{(\mathrm{S})} + B_{j,n}^{(q)} - C_{j,n}^{(\mathrm{S})} B_{j,n}^{(q)} - p_n^{(q)} - \left(1 - p_n^{(q)}\right) C_{j,n}^{(\mathrm{S})} \right)$ is zero, we have

$$\begin{aligned}\mathbb{V}\left[Y_{j,n}''^{(q)}\right] &= \mathbb{E}\left[\left(C_{j,n}^{(\mathrm{S})} \left(1 - B_{j,n}^{(q)} - \left(1 - p_n^{(q)}\right)\right) + B_{j,n}^{(q)} - p_n^{(q)}\right)^2\right] \\ &= \mathbb{E}\left[\left(C_{j,n}^{(\mathrm{S})}\right)^2\right] \mathbb{E}\left[\left(1 - B_{j,n}^{(q)} - \left(1 - p_n^{(q)}\right)\right)^2\right] + \mathbb{E}\left[\left(B_{j,n}^{(q)} - p_n^{(q)}\right)^2\right] + \\ &\quad +2\,\mathbb{E}\left[C_{j,n}^{(\mathrm{S})} \left(1 - B_{j,n}^{(q)} - \left(1 - p_n^{(q)}\right)\right) \left(B_{j,n}^{(q)} - p_n^{(q)}\right)\right] \\ &= p_n^{(\mathrm{S})} \left(1 - p_n^{(q)}\right) p_n^{(q)} + \left(1 - p_n^{(q)}\right) p_n^{(q)} + \\ &\quad +2\, p_n^{(\mathrm{S})}\, \mathbb{E}\left[-B_{j,n}^{(q)} \left(1 - p_n^{(q)}\right) - p_n^{(q)} \left(1 - B_{j,n}^{(q)}\right) + p_n^{(q)} \left(1 - p_n^{(q)}\right)\right] \\ &= p_n^{(\mathrm{S})} \left(1 - p_n^{(q)}\right) p_n^{(q)} + \left(1 - p_n^{(q)}\right) p_n^{(q)} - 2\, p_n^{(\mathrm{S})} p_n^{(q)} \left(1 - p_n^{(q)}\right) \\ &= p_n^{(q)} \left(1 - p_n^{(q)}\right) \left(1 - p_n^{(\mathrm{S})}\right).\end{aligned}$$

Thus, after defining

$$\left(s_n''\right)^2 \doteq \sum_{j=1}^{n} \mathbb{V}\left[Y_{j,n}''^{(q)}\right] = n\, p_n^{(q)} \left(1 - p_n^{(q)}\right) \left(1 - p_n^{(\mathrm{S})}\right),$$

we obtain

$$\frac{1}{\left(s_n''\right)^4} \sum_{j=1}^{n} \mathbb{E}\left[\left(Y_{j,n}''^{(q)}\right)^4\right] \to 0,$$

by the same argument used for Eq. (45). The Lindeberg–Feller Theorem, now, yields

$$\frac{1}{s_n''} \sum_{j=1}^{n} Y_{j,n}''^{(q,N)} \xrightarrow{d} N(0,1),$$

and, in turn,

$$\begin{aligned}&\sqrt{n}\, \frac{\mathsf{N}_n^{(q,N)}}{\sqrt{p_n^{(q)} \left(1 - p_n^{(q)}\right) \left(1 - p_n^{(\mathrm{S})}\right)}} \\ &= \frac{1}{\sqrt{n\, p_n^{(q)} \left(1 - p_n^{(q)}\right) \left(1 - p_n^{(\mathrm{S})}\right)}} \sum_{j=1}^{n} Y_{j,n}''^{(q)} + O_p\left(\sqrt{a^N} \overline{p}_n^N\right) + O_p\left(\sqrt{a^N} 2^N \beta^{N^\eta}\right) \xrightarrow{d} N(0,1),\end{aligned}$$

by Eq. (14). Finally, since, under $\mathcal{H}_A$, $\mathbb{M}_n^{(N)} \xrightarrow{p} p_\infty^{(\mathrm{S})}$, we have $\mathsf{D}_n^{(N)} \xrightarrow{p} 1 - p_\infty^{(\mathrm{S})}$ and Slutsky's theorem implies

$$\sqrt{\frac{n\left(1 - p_n^{(\mathrm{S})}\right)}{p_n^{(q)} \left(1 - p_n^{(q)}\right)}} \left(\mathbb{Id}_n^{(q,N)} - p_n^{(q)}\right) \xrightarrow{d} N(0,1),$$

thereby verifying Eq. (19). □

Appendix B. Monte Carlo simulations

This Appendix reports two Monte Carlo studies. The first is devoted to the exploration of the finite-sample accuracy of the limiting results reported in the theorems. The second is an assessment of the reliability of the numerical optimization used to perform the simulated method-of-moment estimation, as described in Section 5.2.

B.1. The finite-sample properties of the staleness measures

We simulate 10,000 paths of N stochastic processes both under the null (c.f. Eq. (4)) and under the alternative (c.f. Eq. (7)). In both cases, the paths of the efficient logarithmic prices $\widetilde{X}_t^{(q)}$, with $q = 1, \ldots, N$, are generated as the solution to the stochastic differential equation

$$\begin{cases} d\widetilde{X}_t^{(q)} = & c_4\, dt + U_t \sqrt{\exp\left(V_t^{(q)}\right)} \left(\sqrt{1-c_5^2}\, dW_t^{(q)} + c_5\, dZ_t^{(q)}\right) + \beta_q\, \sigma_{\mathsf{M}}\, d\mathsf{M}_t, \\ dV_t^{(q)} = & \left(c_1 - c_2\, V_t^{(q)}\right) dt + c_3\, dZ_t^{(q)}, \end{cases} \qquad q = 1, \ldots, N,$$

where $W_t^{(q)}$ and $Z_t^{(q)}$ are independent asset-specific Brownian motions and M_t is a Brownian motion, independent of all others, which is common to all assets. The process M_t is, thus, a systematic component, say the market. The assets' sensitivities to market variation are set equal to $\beta_q = 1.06$, for all $q = 1, \ldots, N$. The chosen value (1.06) corresponds to the (time-series and cross-sectional) average for our data. The volatility of the common factor is set to $\sigma_{\mathsf{M}} = 0.0103$, which is the market volatility in our data. The percentage values of the other parameters (in daily units) are $c_1 = -0.0120$, $c_2 = 0.0145$, $c_3 = 0.1153$, $c_4 = 0.0304$ and $c_5 = -0.3$. These are estimates from Andersen et al. (2002) based on S&P500 historical data. The deterministic function of time $U_t = (0.1271\, t^2 - 0.1260\, t + 0.1239)/0.1033$, with $t \in [0, 1]$, is estimated on the intra-daily path of the S&P500 and reproduces the typical U-shaped intra-daily pattern of spot volatility.

We allow for rounding and microstructure noise. To do so, we simulate the paths of the processes $\widetilde{X}_t^{(q)}$ on an equi-spaced partition $t_{j,n_s} = j/n_s$ of the time interval $[0, 1]$ with $n_s = 7 \times 60 \times 60$, which corresponds to a frequency of one second for a trading day of seven hours. On the grid, noise is added to the efficient price process (before rounding) using the specification of Barndorff-Nielsen et al. (2011). The efficient logarithmic price is, thus, expressed as

$$\overline{X}_{t_{j,n_s}}^{(q)} = \widetilde{X}_{t_{j,n_s}}^{(q)} + \varepsilon_{j,n_s},$$

where ε_{j,n_s} is a $N\left(0, \omega^2\right)$-iid shock, $\omega^2 = \xi^2 \sqrt{\frac{1}{N} \sum_{q=1}^{N} \frac{1}{n_s} \sum_{j=1}^{n_s} \exp\left(V_{t_{j,n_s}}^{(q)}\right)}$ and ξ^2 is the noise-to-signal ratio at the one-second frequency. We will consider, below, the cases $\xi^2 = 0$ (no noise), $\xi^2 = 10$ bp (medium noise) and $\xi^2 = 50$ bp (large noise). The noisy prices $\overline{X}_{t_{j,n_s}}^{(q)}$ are then sampled on a coarser sampling grid $t_{j,n} = j/n$, with $n = 1260, 1680, 2520$ and 5040, corresponding, respectively, to a sampling frequency of 20, 15, 10 and 5 s.

The *observed* logarithmic prices are generated (on the four coarser partitions $\left\{t_{j,n} = j/n; j = 0, \ldots, n\right\}$) according to the recursive equation

$$\begin{cases} X_0^{(q)} = \overline{X}_0^{(q)}, \\ X_{j/n}^{(q)} = \left(1 - C_{j,n}^{(\mathsf{S})}\right) \left(\overline{X}_{j/n}^{(q)} \left(1 - B_{j,n}^{(q)}\right) + X_{(j-1)/n}^{(q)}\, B_{j,n}^{(q)}\right) + C_{j,n}^{(\mathsf{S})}\, X_{(j-1)/n}^{(q)}, \; j = 1, \ldots, n, \end{cases} \tag{46}$$

where $C_{j,n}^{(\mathsf{S})}$ and $B_{j,n}^{(q)}$, $q = 1, \ldots, N$, are $N + 1$ triangular arrays of iid Bernoulli random variables. We assume the following scaling laws for the N idiosyncratic probabilities of stale trading

$$\mathbb{E}\left[B_{j,n}^{(q)}\right] = p_n^{(q)} = p_\infty^{(q)} \left(1 - e^{-\phi\, n}\right), \quad q = 1, \ldots, N, \tag{47}$$

with $\phi = 0.001$ and $p_\infty^{(q)} = 0.99\ \forall q$. Under the alternative, in addition to the N scaling laws in Eq. (47), we assume a similar scaling law for the probability of systematic staleness. In particular, we set

$$\mathbb{E}\left[C_{j,n}^{(\mathsf{S})}\right] = p_n^{(\mathsf{S})} = p_\infty^{(\mathsf{S})} \left(1 - e^{-\phi\, n}\right), \tag{48}$$

with $\phi = 0.001$ and $p_\infty^{(\mathsf{S})} = 0.99$. In order to justify Eqs. (47) and (48), we recall that both the probability of idiosyncratic staleness and the probability of systematic staleness are modeled, in the theorems, as changing across scales and, therefore, as being *frequency-specific*. This modeling device is a reflection of the higher likelihood of staleness associated with increased sampling in the data (Bandi et al., 2020).

Finally, the observed logarithmic prices in Eq. (46) are transformed back into prices (through the exponential function), rounded to one cent, and put back into a logarithmic form.

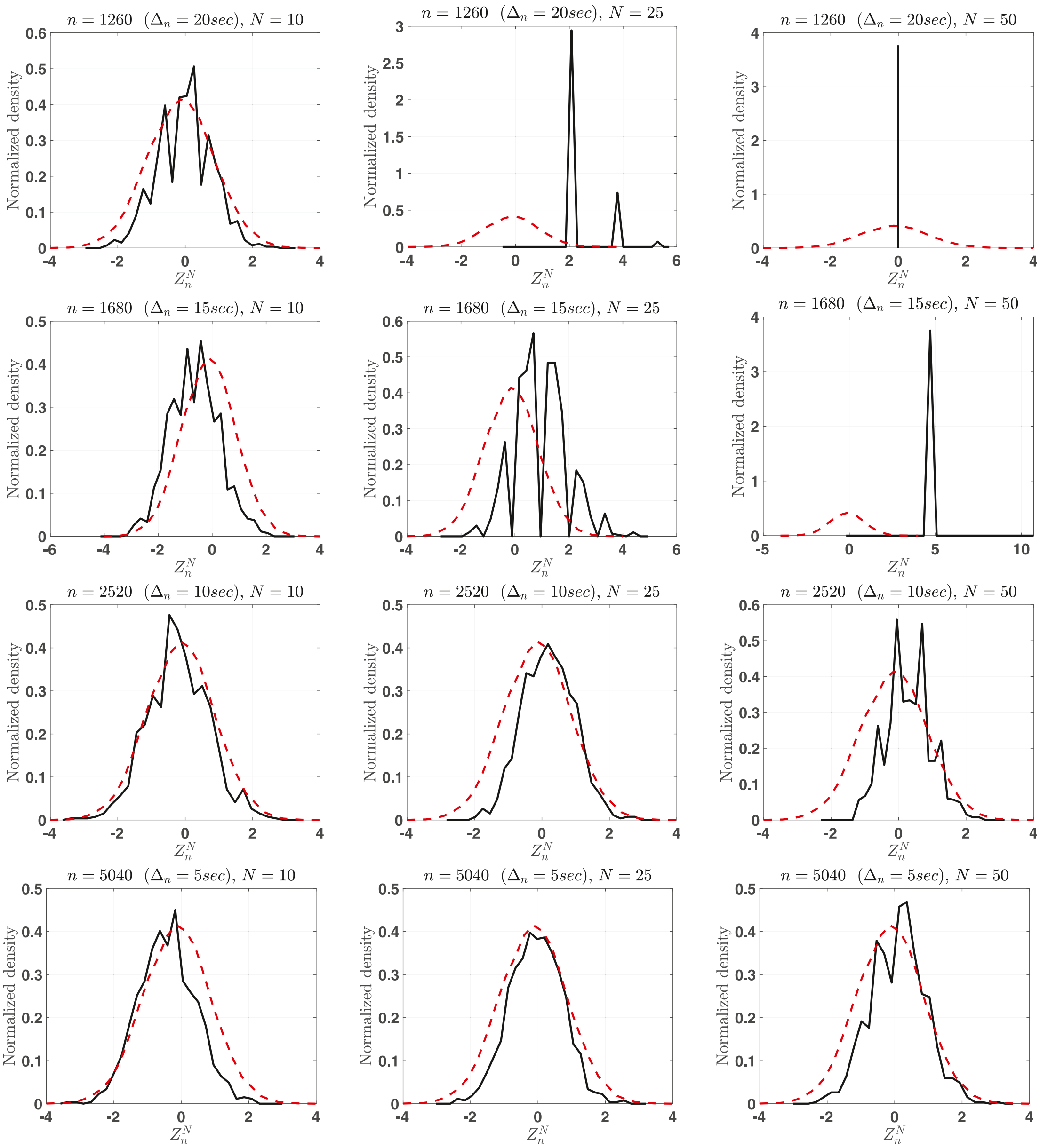

Fig. 7. The figure reports the simulated normalized density of the random variable $\mathbb{Z}_n^{(N)}$ under $\mathcal{H}_0$. The threshold $\Theta_n^{(q,N)}$, with $q = 1,\dots,N$, is chosen as $\Theta_n^{(q,N)} = h_n^{(N)}\hat{\sigma}/n^{1/2}$, where $h_n^{(N)}$ is selected as $h_n^{(N)} = \frac{0.8^{N^{2.0001}}}{n^{0.9}}$ and $\hat{\sigma}$ is the square root of the five-minute realized variance. The red dashed lines correspond to the standard Gaussian density. The title of each plot reports the number of intra-daily observations n (and the corresponding time-distance between observations, i.e., Δ_n) and the number of assets N. Noise and rounding are switched off since this experiment is meant to be illustrative of the asymptotic theory. (For interpretation of the references to color in this caption, the reader is referred to the web version of this article.)

For a given sampling frequency n, a given number of stocks N and a specific asset q, the threshold $\Theta_{t,n}^{(q,N)}$ in Eq. (1) is set equal to $\Theta_n^{(q,N)} = h_n^{(N)}\hat{\sigma}/n^{1/2}$, where $h_n^{(N)}$ is chosen as

$$h_n^{(N)} = \frac{\beta^{N^{1+\eta}}}{n^{\gamma}},$$

with $\beta = 0.8$, $\eta = 1.0001$ and $\gamma = 0.9$ and $\hat{\sigma}$ is the square-root of the five-minute realized variance of the observed price.

The panels in Fig. 7 report, for several values of $\Delta_n = \frac{1}{n}$, the finite-sample distributions, under the null, of the random variable

$$\mathbb{Z}_n^{(N)} \doteq \frac{\mathbb{M}_n^{(N)} - P_n^{(N)}}{\sigma_n^{(N)}}$$

Table 8
Mean and standard deviation of the relative error of idiosyncratic staleness. We report the mean and, in parentheses, the standard deviation of the relative error of idiosyncratic staleness as an estimator of true idiosyncratic staleness. Results are shown for different values of the number of observations n, the number of assets N, the noise-to-signal ratio ξ^2 and the level of rounding.

n	$\xi^2 = 0$			$\xi^2 = 10$ bps			$\xi^2 = 50$ bps		
	$N = 10$	$N = 25$	$N = 50$	$N = 10$	$N = 25$	$N = 50$	$N = 10$	$N = 25$	$N = 50$
No rounding									
168	0.031	0.085	0.184	0.034	0.076	0.179	0.038	0.082	0.177
	(0.191)	(0.198)	(0.197)	(0.198)	(0.192)	(0.197)	(0.191)	(0.201)	(0.196)
420	−0.001	0.022	0.051	0.006	0.019	0.050	0.003	0.019	0.049
	(0.083)	(0.083)	(0.083)	(0.081)	(0.082)	(0.079)	(0.082)	(0.082)	(0.082)
840	−0.007	0.002	0.011	−0.006	0.003	0.012	−0.006	0.004	0.011
	(0.045)	(0.046)	(0.044)	(0.046)	(0.046)	(0.044)	(0.045)	(0.045)	(0.045)
1260	−0.019	−0.003	0.001	−0.017	−0.002	0.001	−0.017	−0.003	−0.000
	(0.034)	(0.033)	(0.033)	(0.034)	(0.033)	(0.033)	(0.034)	(0.033)	(0.033)
1680	−0.037	−0.007	−0.003	−0.037	−0.006	−0.003	−0.037	−0.006	−0.004
	(0.030)	(0.027)	(0.026)	(0.030)	(0.027)	(0.027)	(0.030)	(0.027)	(0.027)
2520	−0.077	−0.020	−0.009	−0.076	−0.020	−0.009	−0.078	−0.020	−0.009
	(0.035)	(0.023)	(0.020)	(0.033)	(0.022)	(0.021)	(0.033)	(0.023)	(0.021)
Rounding: small tick									
168	0.085	0.109	0.195	0.075	0.110	0.194	0.073	0.110	0.186
	(0.195)	(0.203)	(0.202)	(0.211)	(0.205)	(0.200)	(0.194)	(0.207)	(0.199)
420	0.036	0.042	0.062	0.035	0.041	0.059	0.034	0.037	0.056
	(0.085)	(0.085)	(0.082)	(0.081)	(0.083)	(0.083)	(0.083)	(0.082)	(0.083)
840	0.010	0.013	0.019	0.008	0.013	0.018	0.008	0.012	0.016
	(0.046)	(0.047)	(0.044)	(0.047)	(0.045)	(0.044)	(0.047)	(0.045)	(0.045)
1260	−0.011	0.003	0.004	−0.012	0.002	0.004	−0.013	0.003	0.004
	(0.034)	(0.033)	(0.033)	(0.034)	(0.033)	(0.032)	(0.034)	(0.033)	(0.033)
1680	−0.037	−0.003	−0.002	−0.034	−0.002	−0.001	−0.034	−0.004	−0.002
	(0.029)	(0.027)	(0.027)	(0.030)	(0.027)	(0.027)	(0.030)	(0.027)	(0.027)
2520	−0.076	−0.020	−0.009	−0.076	−0.020	−0.009	−0.077	−0.019	−0.009
	(0.032)	(0.023)	(0.021)	(0.032)	(0.023)	(0.021)	(0.032)	(0.023)	(0.021)
Rounding: large tick									
168	1.682	1.677	1.678	1.657	1.669	1.662	1.623	1.625	1.630
	(0.260)	(0.269)	(0.265)	(0.261)	(0.259)	(0.261)	(0.263)	(0.262)	(0.266)
420	0.745	0.755	0.754	0.738	0.748	0.744	0.708	0.712	0.713
	(0.083)	(0.083)	(0.083)	(0.082)	(0.083)	(0.082)	(0.083)	(0.082)	(0.084)
840	0.277	0.304	0.305	0.276	0.300	0.302	0.263	0.286	0.286
	(0.038)	(0.037)	(0.037)	(0.038)	(0.037)	(0.038)	(0.041)	(0.039)	(0.038)
1260	0.099	0.138	0.141	0.100	0.137	0.139	0.093	0.130	0.132
	(0.029)	(0.026)	(0.026)	(0.029)	(0.026)	(0.026)	(0.029)	(0.027)	(0.027)
1680	0.013	0.062	0.066	0.012	0.061	0.066	0.011	0.058	0.062
	(0.027)	(0.021)	(0.021)	(0.027)	(0.022)	(0.021)	(0.026)	(0.022)	(0.021)
2520	−0.068	−0.006	0.008	−0.067	−0.006	0.008	−0.069	−0.006	0.008
	(0.032)	(0.021)	(0.018)	(0.032)	(0.020)	(0.018)	(0.031)	(0.021)	(0.018)

defined in Eq. (16). Under the null, in order to illustrate the validity of the asymptotic theory, both rounding and noise are turned off. Fig. 7 shows that $\mathbb{Z}_n^{(N)}$ is approximately Gaussian. More specifically, as implied by theory, $\mathbb{Z}_n^{(N)}$ is closer to a Gaussian, the larger the number of intra-daily observations n and the larger the number of assets N. Consistent with the bandwidth condition in the statement of Theorem 1, the tuning parameter $h_n^{(N)}$ plays a role.

The left and right panels of Fig. 8, Fig. 9 and Fig. 10 report, respectively, the average value, under the alternative, of $\mathbb{I}\mathbb{d}_n^{(q,N)}$ (defined in Eq. (13)) and $\mathbb{M}_n^{(N)}$ (defined in Eq. (3)) for different values of N and as a function of the sampling frequency n. Fig. 8 illustrates the ideal case of no rounding and no noise and, separately, the effect of noise without rounding and rounding without noise. In Figs. 9 and 10, which correspond, respectively, to the case $\xi^2 = 10$ bps and $\xi^2 = 50$ bps, we report three rounding scenarios: small tick size (graphs in the top panel, average price level ≈ 100\$, daily efficient price volatility ≈ 1% and prices rounded to one cent), medium tick size (graphs in the middle panel, average price level ≈ 50\$, daily efficient price volatility ≈ 1% and prices rounded to one cent) and large tick size (graphs in the bottom panel, average price level ≈ 5\$, daily efficient price volatility ≈ 1% and prices rounded to one cent). In each scenario, we allow for different intra-daily sampling frequencies (corresponding to a total number of observations reported along the horizontal axis) and three possible values for the total number of assets (more specifically, $N = 10$, $N = 50$ and $N = 100$). The black continuous line represents the true value of the estimated quantities, either idiosyncratic staleness (graphs in the left column) or systematic staleness (graphs in the right column). The average value of the estimated quantities is reported, for each sampling frequency, as a red cross ($N = 10$), a blue star ($N = 50$) or a black empty circle ($N = 100$). Tables 8 and 9 report the mean and the standard deviation of the relative error of, respectively, $\mathbb{I}\mathbb{d}_n^{(q,N)}$ and $\mathbb{M}_n^{(N)}$ when estimating the corresponding true quantities. Again, results are shown for different values of the number of observations n, the number of assets N, the noise-to-signal ratio ξ^2 and the level of rounding.

Table 9
Mean and standard deviation of the relative error of systematic staleness. We report the mean and, in parentheses, the standard deviation of the relative error of systematic staleness as an estimator of true systematic staleness. Results are shown for different values of the number of observations n, the number of assets N, the noise-to-signal ratio ξ^2 and the level of rounding.

n	$\xi^2 = 0$			$\xi^2 = 10$ bps			$\xi^2 = 50$ bps		
	$N = 10$	$N = 25$	$N = 50$	$N = 10$	$N = 25$	$N = 50$	$N = 10$	$N = 25$	$N = 50$
No rounding									
168	−0.009	−0.006	−0.008	−0.022	−0.002	−0.001	−0.018	−0.003	−0.021
	(0.183)	(0.162)	(0.171)	(0.188)	(0.187)	(0.179)	(0.176)	(0.182)	(0.154)
420	−0.002	−0.002	−0.007	−0.008	0.004	−0.009	0.001	−0.002	0.001
	(0.065)	(0.069)	(0.072)	(0.078)	(0.072)	(0.072)	(0.080)	(0.072)	(0.064)
840	0.003	−0.005	−0.005	0.000	−0.006	−0.003	−0.005	−0.006	0.000
	(0.033)	(0.031)	(0.033)	(0.029)	(0.028)	(0.032)	(0.030)	(0.029)	(0.030)
1260	0.011	−0.004	−0.002	0.011	−0.001	−0.000	0.012	−0.002	−0.001
	(0.017)	(0.017)	(0.018)	(0.017)	(0.022)	(0.019)	(0.018)	(0.018)	(0.019)
1680	0.026	0.004	−0.002	0.027	0.001	−0.002	0.028	0.001	−0.000
	(0.011)	(0.010)	(0.012)	(0.010)	(0.012)	(0.012)	(0.013)	(0.011)	(0.013)
2520	0.037	0.009	0.000	0.038	0.009	−0.000	0.037	0.008	0.001
	(0.005)	(0.006)	(0.007)	(0.005)	(0.005)	(0.006)	(0.004)	(0.006)	(0.006)
Rounding: small tick									
168	−0.028	−0.013	0.029	−0.005	−0.000	−0.047	−0.005	−0.025	−0.009
	(0.185)	(0.138)	(0.192)	(0.198)	(0.179)	(0.178)	(0.192)	(0.162)	(0.190)
420	0.001	0.004	0.000	−0.005	−0.007	−0.005	−0.024	0.010	0.000
	(0.064)	(0.069)	(0.067)	(0.070)	(0.070)	(0.068)	(0.068)	(0.065)	(0.066)
840	−0.002	−0.000	−0.006	0.001	−0.003	−0.003	0.007	−0.006	−0.004
	(0.030)	(0.028)	(0.033)	(0.030)	(0.031)	(0.031)	(0.029)	(0.031)	(0.030)
1260	0.012	−0.001	0.002	0.015	−0.002	0.001	0.011	−0.003	0.001
	(0.016)	(0.018)	(0.018)	(0.018)	(0.019)	(0.020)	(0.017)	(0.017)	(0.014)
1680	0.028	−0.000	−0.001	0.028	−0.001	−0.002	0.027	0.001	−0.000
	(0.012)	(0.012)	(0.012)	(0.011)	(0.012)	(0.011)	(0.011)	(0.013)	(0.012)
2520	0.037	0.009	0.000	0.038	0.009	0.001	0.039	0.009	0.001
	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.007)	(0.005)	(0.006)	(0.007)
Rounding: large tick									
168	0.007	0.009	−0.014	−0.001	−0.007	−0.006	−0.000	−0.043	−0.030
	(0.184)	(0.160)	(0.186)	(0.179)	(0.186)	(0.178)	(0.181)	(0.173)	(0.193)
420	0.008	0.002	−0.002	0.014	−0.018	0.002	0.010	−0.008	−0.011
	(0.066)	(0.071)	(0.076)	(0.070)	(0.078)	(0.083)	(0.068)	(0.070)	(0.073)
840	0.037	−0.002	−0.002	0.034	−0.002	−0.001	0.034	−0.001	0.002
	(0.030)	(0.032)	(0.030)	(0.030)	(0.027)	(0.030)	(0.032)	(0.031)	(0.028)
1260	0.054	0.003	0.001	0.050	0.002	−0.004	0.047	0.000	−0.000
	(0.014)	(0.019)	(0.018)	(0.015)	(0.017)	(0.020)	(0.020)	(0.019)	(0.016)
1680	0.056	0.005	−0.001	0.055	0.006	−0.002	0.053	0.005	−0.001
	(0.009)	(0.013)	(0.012)	(0.010)	(0.013)	(0.011)	(0.009)	(0.013)	(0.013)
2520	0.046	0.015	0.002	0.045	0.013	0.001	0.045	0.015	0.001
	(0.004)	(0.005)	(0.006)	(0.004)	(0.005)	(0.006)	(0.004)	(0.006)	(0.006)

We show that rounding and noise are not an issue for the systematic staleness estimator. The impact of noise is immaterial and the limit $N \to \infty$ is effective in removing the impact of rounding. This property is easily explained: the probability of a simultaneous round-off error across multiple assets goes to zero exponentially fast as the number of assets diverges. For idiosyncratic staleness, the effect of rounding is marginal in the case of small and medium tick stocks and can be impactful only in the case of large tick stocks which are, however, only a small fraction of the stocks considered in our analysis. In addition, regimes with higher volatility than the one considered here will generate, *ceteris paribus*, a smaller rounding-induced bias in the estimation of idiosyncratic staleness.

B.2. The finite-sample properties of the structural estimates

We now discuss the efficacy of the simulated method of moments, described in Section 5.2, in estimating the parameters of the microstructure-founded model of price formation discussed in Section 5.

For three different values of the number of assets N, i.e., $N = 2$, $N = 5$ and $N = 10$, we set, for all assets, the parameters as the (cross-sectional and time-series) average of the estimates obtained from data (i.e. the estimates reported in Fig. 4). We simulate 100 trading days of 7 h (each of which is divided into $\Delta_n^{-1} = 840$ intervals of 30 seconds) of the N-variate price process.

The plots in Fig. 11 display, for the parameters $\sigma_q^{(e)}$, $\sigma_q^{(m)}$, δ_q and s_q – pooled across all assets – and for f and PAIT, the densities of the relative error with respect to the true value, i.e., the densities of $\left(\hat{\vartheta} - \vartheta_0\right)/\vartheta_0$, where $\hat{\vartheta}$ and ϑ_0 are, respectively, estimated and true value. All parameters are well identified but two observations are in order. First, we conservatively present results for low values of N. A larger number of assets (as in our empirical work) would generally lead to further improvements. Second, relatedly, it may appear that some of the parameter estimates (such as PAIT) do not benefit in an obvious way from increasing the number of

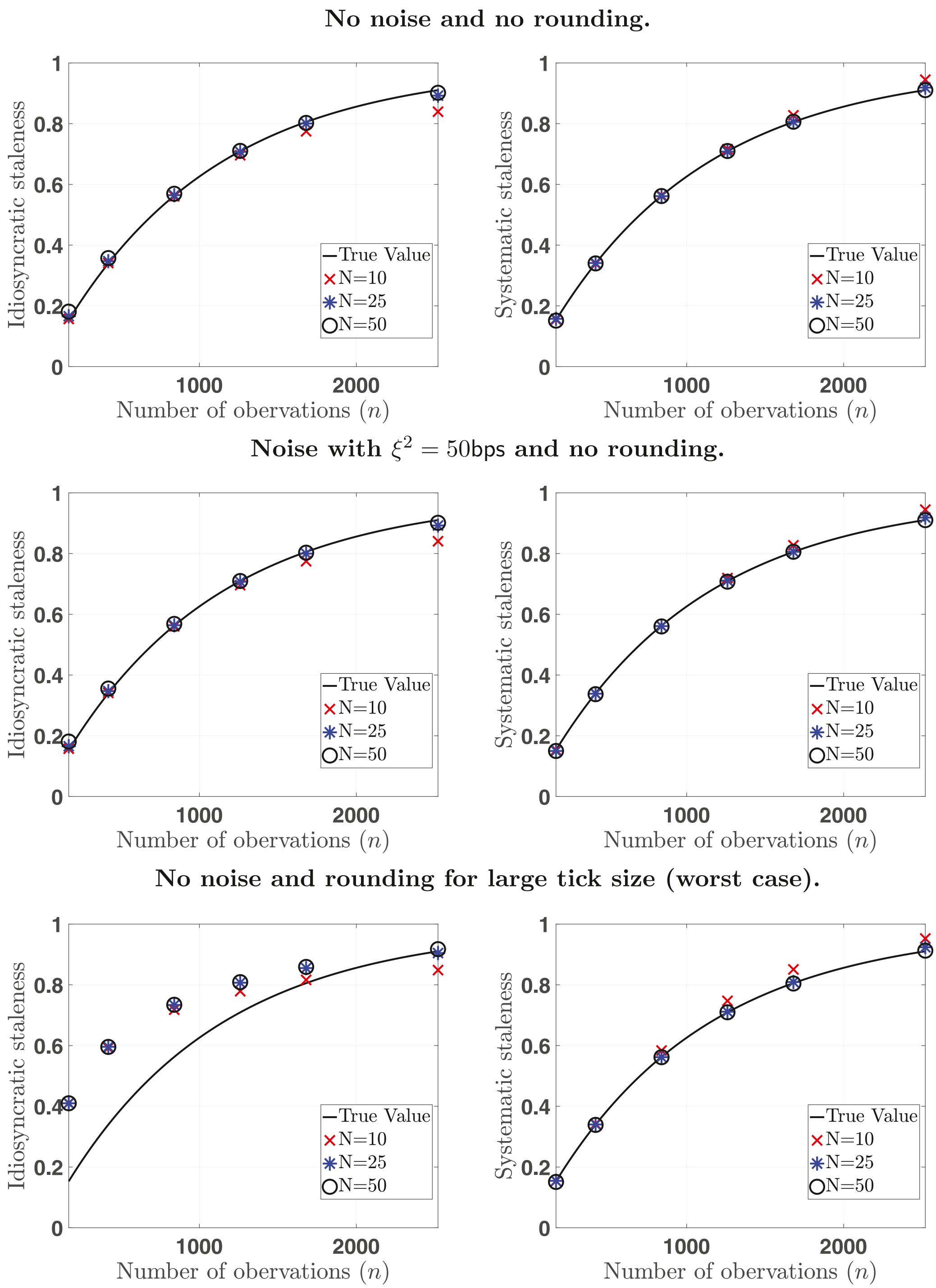

Fig. 8. Results of a simulated experiment designed to assess the impact of rounding on idiosyncratic and systematic staleness. In the top, middle and bottom panels we consider, respectively, the following three scenarios: (1) no noise and no rounding, (2) noise (added with a noise-to-signal ratio of $\xi^2 = 50$ bps) without rounding and (3) rounding for a large tick size stock (average price level ≈ 5$, daily efficient price volatility ≈ 1% and prices rounded to one cent) without noise. In each panel, the black continuous line represents the true quantity to be estimated (idiosyncratic staleness, assumed equal for all assets, in the left column, and systematic staleness, in the right column). Red crosses, blue stars and empty circles represent the estimated quantities when $N = 10$, $N = 50$ and $N = 100$ assets are simulated, respectively. The horizontal axis reports the inverse of the sampling frequency, i.e. the total number of observations. (For interpretation of the references to color in this caption, the reader is referred to the web version of this article.)

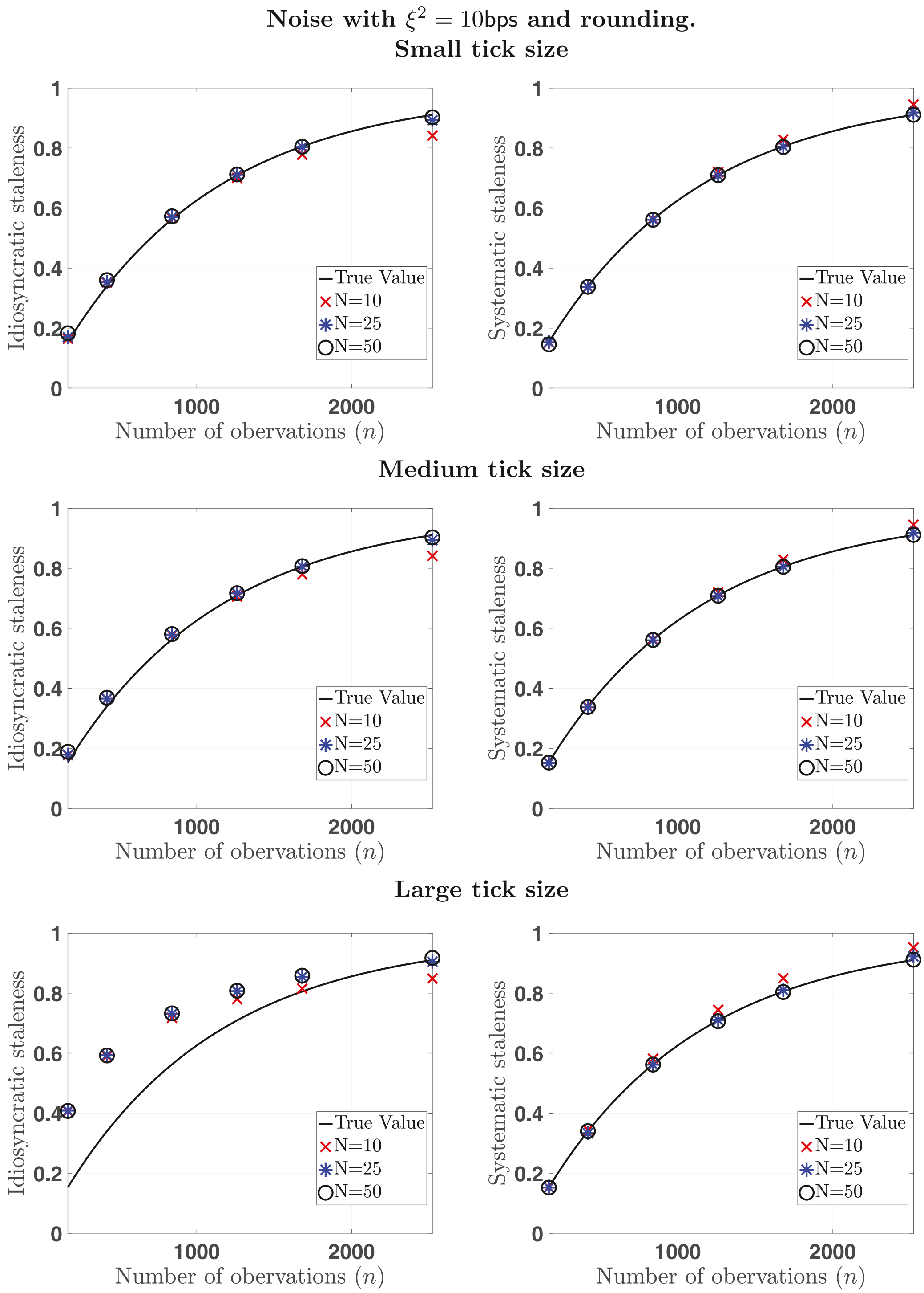

Fig. 9. Results of a simulated experiment designed to assess the impact of rounding on idiosyncratic and systematic staleness. In the top, middle and bottom panels we consider, respectively, the following three scenarios: (1) small tick size (average price level ≈ 100$, daily efficient price volatility ≈ 1% and prices rounded to one cent), (2) medium tick size (average price level ≈ 50$, daily efficient price volatility ≈ 1% and prices rounded to one cent) and (3) large tick size (average price level ≈ 5$, daily efficient price volatility ≈ 1% and prices rounded to one cent). In each panel, the black continuous line represents the true quantity to be estimated (idiosyncratic staleness, assumed equal for all assets, in the left column, and systematic staleness, in the right column). Red crosses, blue stars and empty circles represent the estimated quantities when $N = 10$, $N = 50$ and $N = 100$ assets are simulated, respectively. The horizontal axis reports the inverse of the sampling frequency, i.e. the total number of observations. Microstructure noise is added with a noise-to-signal ratio of $\xi^2 = 10$ bps. (For interpretation of the references to color in this caption, the reader is referred to the web version of this article.)

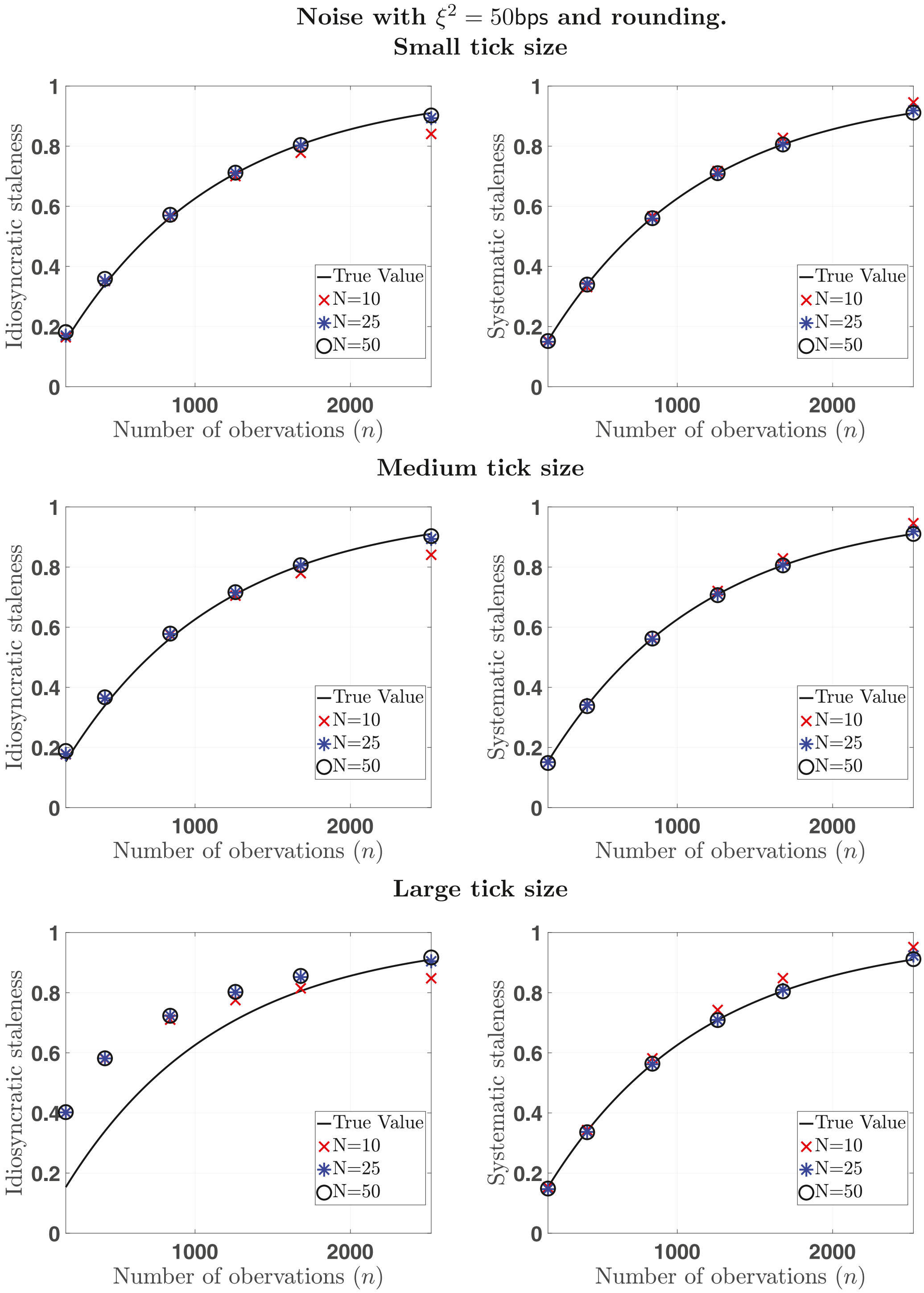

Fig. 10. Results of a simulated experiment designed to assess the impact of rounding on idiosyncratic and systematic staleness. In the top, middle and bottom panels we consider, respectively, the following three scenarios: (1) small tick size (average price level ≈ 100$, daily efficient price volatility ≈ 1% and prices rounded to one cent), (2) medium tick size (average price level ≈ 50$, daily efficient price volatility ≈ 1% and prices rounded to one cent) and (3) large tick size (average price level ≈ 5$, daily efficient price volatility ≈ 1% and prices rounded to one cent). In each panel, the black continuous line represents the true quantity to be estimated (idiosyncratic staleness, assumed equal for all assets, in the left column, and systematic staleness, in the right column). Red crosses, blue stars and empty circles represent the estimated quantities when $N = 10$, $N = 50$ and $N = 100$ assets are simulated, respectively. The horizontal axis reports the inverse of the sampling frequency, i.e. the total number of observations. Microstructure noise is added with a noise-to-signal ratio of $\xi^2 = 50$ bps. (For interpretation of the references to color in this caption, the reader is referred to the web version of this article.)

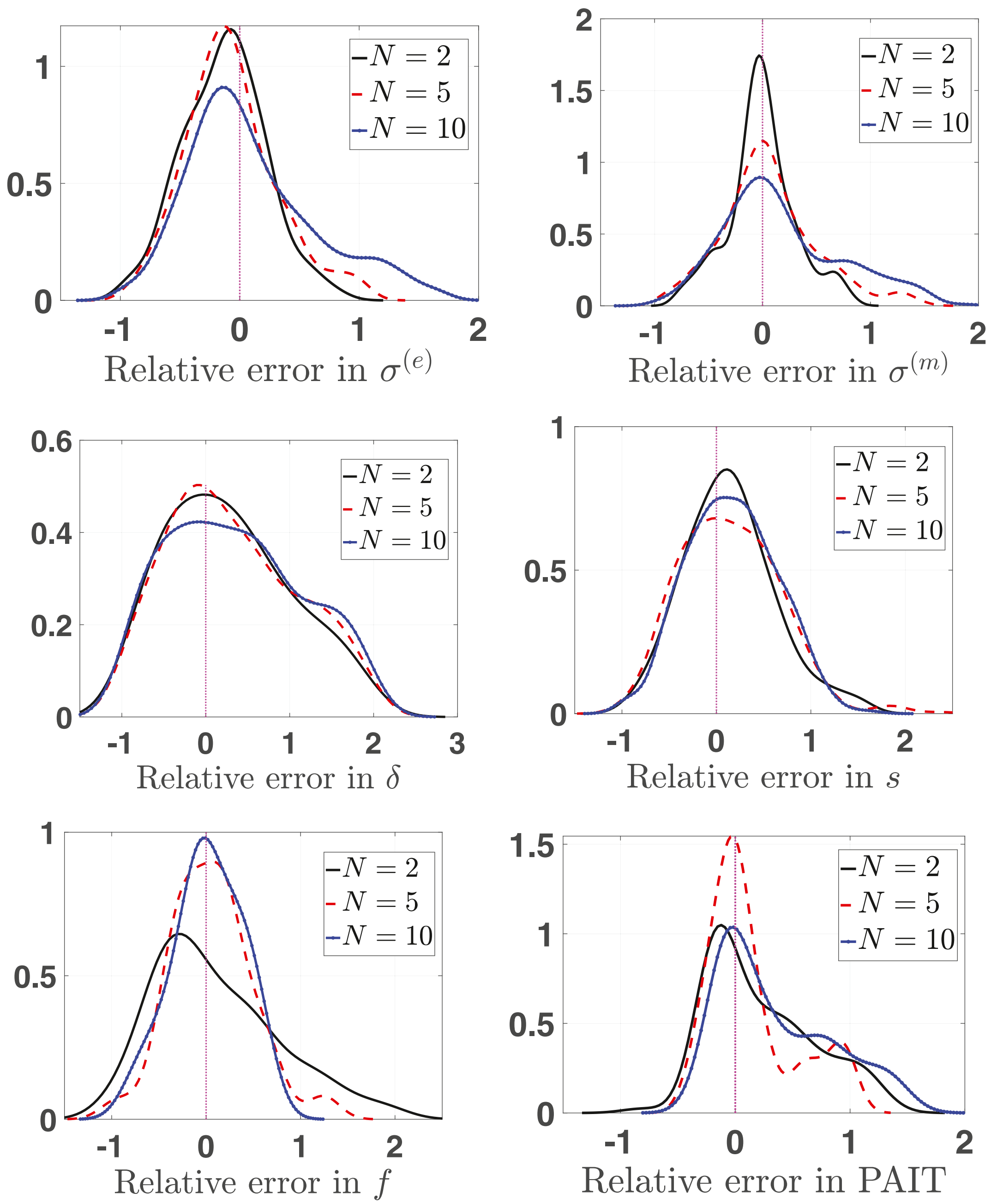

Fig. 11. The figure reports the kernel density of the relative error (computed over 100 replications) for estimates of the parameters of the model described in Section 5. For each parameter, we report three densities obtained by using three different numbers of assets, i.e. $N = 2$ (black continuous line), $N = 5$ (red dashed line) and $N = 10$ (blue line with stars). The densities for the asset-specific parameters $\sigma_q^{(e)}$, $\sigma_q^{(m)}$, δ_q and s_q are computed by pooling across all the assets. (For interpretation of the references to color in this caption, the reader is referred to the web version of this article.)

assets. This is due to the fact that the threshold $\Theta_{t,n}^{(q,N)}$ in Eq. (1), a threshold which, e.g., defines the multivariate staleness estimator in Eq. (3), may be further optimized (through $h_n^{(N)}$) with respect to the number of observations n and the number of assets N. Our selection rule is consistent with the one described in the main text. It is effective, as shown in these simulations, but it can likely be improved. The study of an optimal threshold selection rule is beyond the objectives of this article and we leave it for future work. Having made these points, the estimates of the funding cost parameter f, which are central to this study, are remarkably accurate and further improve when larger values of the number of assets are considered.

References

Aït-Sahalia, Y., Jacod, J., 2014. High-Frequency Financial Econometrics. Princeton University Press.
Amihud, Y., 2002. Illiquidity and stock returns: Cross-section and time-series effects. J. Financial Markets 5 (1), 31–56.
Amihud, Y., Mendelson, H., 1987. Trading mechanism and stock retuns: An empirical investigation. J. Finance 42 (3), 533–553.
Andersen, T., Benzoni, L., Lund, J., 2002. An empirical investigation of continuous-time equity return models. J. Finance 57, 1239–1284.
Bandi, F.M., Kolokolov, A., Pirino, D., Renò, R., 2020. Zeros. Manage. Sci. 66 (8), 3466–3479.
Bandi, F.M., Pirino, D., Renò, R., 2017. Excess idle time. Econometrica 85 (6), 1793–1846.
Bandi, F.M., Renò, R., 2017. Local Edgeworth Expansions. Working paper.
Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2011. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. J. Econometrics 162 (2), 149–169.
Bekaert, G., Harvey, C.R., Lundblad, C., 2007. Liquidity and expected returns: Lessons from emerging markets. Rev. Financial Stud. 20 (6), 1783–1831.
Bessembinder, H., 2003. Issues in assessing trade execution costs. J. Financial Markets 6 (3), 233–257.
Billingsley, P., 1995. Probability and Measure, third ed. Wiley Series in Probability and Mathematical Statistics.
Brunnermeier, M.K., Pedersen, L.H., 2009. Market liquidity and funding liquidity. Rev. Financial Stud. 22 (6), 2201–2238.
Chordia, T., Roll, R., Subrahmanyam, A., 2000. Commonality in liquidity. J. Financial Econ. 56 (1), 3–28.
Chordia, T., Roll, R., Subrahmanyam, A., 2001. Market liquidity and trading activity. J. Finance 56 (2), 501–530.
Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Econom. 7, 174–196.
Duffie, D., Singleton, K.J., 1993. Simulated moments estimation of Markov models of asset prices. Econometrica 61 (4), 929–952.
Easley, D., O'Hara, M., 1992. Time and the process of security price adjustment. J. Finance 47 (2), 577–605.
Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. J. Financial Econ. 116 (1), 1–22.
Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. J. Political Econ. 81 (3), 607–636.
Fontaine, J.-S., Garcia, R., 2012. Bond liquidity premia. Rev. Financ. Stud. 25 (4), 1207–1254.
Fontaine, J.-S., Garcia, R., Gungor, S., 2020. Intermediary leverage shocks and funding conditions. Available at SSRN 3649065.
Glosten, L., Milgrom, P., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. J. Financ. Econ. 14 (1), 71–100.
Gourieroux, C., Monfort, A., Renault, E., 1993. Indirect inference. J. Appl. Econometrics 8, 85,118.
Goyenko, R.Y., Holden, C.W., Trzcinka, C.A., 2009. Do liquidity measures measure liquidity? J. Financial Econ. 92 (2), 153–181.
Hasbrouck, J., 2004. Liquidity in the futures pits: Inferring market dynamics from incomplete data. J. Financ. Quant. Anal. 39 (2), 305–326.
Hasbrouck, J., 2009. Trading costs and returns for US equities: Estimating effective costs from daily data. J. Finance 64 (3), 1445–1477.
Hasbrouck, J., Ho, T., 1987. Order arrival, quote behavior, and the return-generating process. J. Finance 42 (4), 1035–1048.
Hasbrouck, J., Seppi, D.J., 2001. Common factors in prices, order flows, and liquidity. J. Financial Econ. 59 (3), 383–411.
Herskovic, B., Kelly, B., Lustig, H., Van Nieuwerburgh, S., 2016. The common factor in idiosyncratic volatility: Quantitative asset pricing implications. J. Financ. Econ. 119 (2), 249–283.
Holden, C.W., 2009. New low-frequency spread measures. J. Financial Mark. 12 (4), 778–813.
Hu, G.X., Pan, J., Wang, J., 2013. Noise as information for illiquidity. J. Finance 68 (6), 2341–2382.
Kamara, A., Lou, X., Sadka, R., 2008. The divergence of liquidity commonality in the cross-section of stocks. J. Financ. Econ. 89 (3), 444–466.
Kyle, P., 1985. Continuous auctions and insider trading. Econometrica 43, 1315–1335.
Lee, C.M., Ready, M.J., 1991. Inferring trade direction from intraday data. J. Finance 46 (2), 733–746.
Lesmond, D.A., 2005. Liquidity of emerging markets. J. Financial Econ. 77 (2), 411–452.
Lesmond, D.A., Ogden, J.P., Trzcinka, C.A., 1999. A new estimate of transaction costs. Rev. Financial Stud. 12 (5), 1113–1141.
Mörters, P., Peres, Y., 2010. Brownian Motion. Cambdrige University Press.
Næs, R., Skjeltorp, J.A., Ødegaard, B.A., 2011. Stock market liquidity and the business cycle. J. Finance 66 (1), 139–176.
Pastor, L., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. J. Political Econ. 111 (3), 642–685.
Phillips, P.C.B., Yu, J., 2023. Information loss in volatility measurement with flat price trading. Empir. Econ. 64 (6), 2957–2999.
Protter, P., 2004. Stochastic Integration and Differential Equations. Springer.
Revuz, D., Yor, M., 2001. Continuous Martingales and Brownian Motion. Springer, Berlin.
Roll, R., 1984. A simple measure of the implicit bid-ask spread in an efficient market. J. Finance 39, 1127–1139.
Schilling, M.F., 1990. The longest run of heads. College Math. J. 21 (3), 196–207.