Local Mispricing and Microstructural Noise: A Parametric Perspective*

Torben G. Andersen† Ilya Archakov‡
Gökhan Cebiroglu§ Nikolaus Hautsch¶

June 21, 2021

Abstract

We extend the classic "martingale-plus-noise" model for high-frequency returns to accommodate an error correction mechanism and endogenous pricing errors. It is motivated by (i) novel empirical evidence documenting that microstructure noise exhibits frequently changing patterns of serial dependence which are interwoven with innovations to the efficient price; (ii) building a bridge between high-frequency econometrics and market microstructure models. We identify temporal pricing error correction and noise endogeneity as complementary components driving high-frequency dynamics and inducing two separate regimes, characterized by the sign of the return serial correlation and an implied bias in realized variance estimates. We document frequent fluctuations between these regimes, which can be associated with price discovery in a setting with incomplete information and learning. The model links critical concepts from high-frequency statistics and market microstructure theory, suggesting new avenues for volatility estimation.

Keywords: volatility estimation; market microstructure noise; price reversal; momentum; contrarian trading; **JEL classification**: C58, C32, G14

*We thank participants of the "Jan Mossin Symposium," Bergen, 2016; FERM, Guangzhou, 2016; SoFiE, Aarhus, 2016; "New Developments in Measuring and Forecasting Financial Volatility," Durham, NC, 2016; MFA Meeting, Chicago, 2017; Stevanovich Conference, University of Chicago, 2018; IAAE Conference, Nicosia, 2019; and various seminar participants for comments and discussions. We are grateful to Peter R. Hansen for valuable comments. Andersen acknowledges support from CREATES (DNRF78), funded by the Danish National Research Foundation.

†Kellogg School of Management, Northwestern University, NBER, and CREATES. Email: t-andersen@kellogg.northwestern.edu.

‡Department of Statistics and Operations Research, University of Vienna. Email: ilya.archakov@univie.ac.at.

§Credit Suisse, London. Email: goekhan.cebi@gmail.com. The views, thoughts, and opinions expressed in this article belong solely to the author, and are not necessarily shared by the author's employer.

¶Department of Statistics and Operations Research, and the Research Platform "Data Science @ Uni Vienna," University of Vienna, Vienna Graduate School of Finance and Center for Financial Studies. Email: nikolaus.hautsch@univie.ac.at.

1 Introduction

The increasing availability of high-frequency asset price data has spurred two largely separate literatures. One focuses on model-free ex-post measurement of features concerning the realized return path, with the main attribute of interest being the realized volatility (RV), representing the empirical quadratic return variation. The premise is that asset prices reflect an underlying arbitrage-free "efficient" or "fundamental" return process polluted by market microstructure noise (MMN), due to rounding to a price grid, varying transaction and quote intensities, private information, bid-ask spreads, and other trading costs. Hence, the fundamental log-price process is a semimartingale, while MMN is viewed as a nuisance factor. In some specifications, the MMN process is serially correlated, but it is typically uncorrelated with the efficient price process. Implicitly, this literature presumes that all relevant information is impounded instantaneously so that, at all times, transaction prices and quotes embed the latent efficient equilibrium price, bracketed by the bid-ask spread. Consequently, at the micro level, the fundamental price is "hidden" through MMN distortions which, locally, are of non-trivial order. As a result, much work builds on such a "martingale-plus-noise" framework and is devoted to MMN-robust inference for realized components of the efficient price process.[1]

In contrast, a central topic in market microstructure is price discovery. This literature explores a variety of complications, including asymmetric information, heterogeneous trading motives, and market design. Learning, price impact, and strategic trading arise as natural phenomena. The interaction of such features generates a complex environment, so most theoretical microstructure models are fairly stylized, allowing for specific aspects of the price discovery process to be explored in isolation. Kyle (1985), Holden and Subrahmanyam (1992), Vives (1995) and Glosten and Milgrom (1985), among others, are seminal contributions analyzing strategic trading in environments with differently informed agents and deriving implications for (the speed of) information revelation, the formation of bid/ask quotes, and the role of noise.

Importantly, the specification of the high-frequency return dynamics naturally changes when asymmetric information is explicitly accommodated within the framework. As an illustration, consider the Glosten and Milgrom (1985) model, which incorporates the fact that agents may possess private information regarding the fundamental asset value. Hence, the active purchase or sale of an asset is informative and causes uninformed market participants to update their assessment of the fair asset price. Specifically, given current information, the efficient price reflects public information, including the full history of the order flow. Thinking ahead, the bid and ask quotes are then set by competitive market makers to reflect the update in asset value

[1] See, e.g., the survey by Andersen and Bollerslev (2018a), the classic articles in Andersen and Bollerslev (2018b), and the book treatment of Ait-Sahalia and Jacod (2014).

that will occur, if the next transaction is a sell or buy. Thus, conditional on a sale at the bid, this transaction price now reflects the expected fundamental asset value and, likewise, a purchase at the ask renders this transaction price the new efficient price, based on the available public information. That is, there are no "noisy" transactions, and traditional sources of noise are instead intrinsically linked to innovations regarding the fundamental asset value. While this is an extreme perspective, relaxing the underlying assumptions in this setting will not eliminate the forces that induce correlations between the noise and fundamental return innovations.

We conclude that, generally speaking, microstructure models imply that noise is endogenous, because innovations to transaction prices and quotes interact with the efficient price process through learning, inventory, and temporal feedback mechanisms. Meanwhile, microstructure theory does not focus on time variation in the return dynamics, which is of primary interest in econometric work on realized volatility. Building a bridge between these strands of the literature requires extending the classic "martingale-plus-noise" model by components accommodating temporal feedback and price endogeneity. Accordingly, we generalize the martingale-plus-noise model and analyze its statistical and empirical implications. We discuss the consequences for volatility estimation, the validity of the underlying semimartingale paradigm and identification of the underlying mechanisms. The analysis points to the potential for new insights as well as the pitfalls arising when econometric models for the high-frequency return dynamics are integrated with structural, or theoretically motivated, features from market microstructure theory.

1.1 Illustrative Empirical Evidence

As an initial step, for the Nasdaq 100 stocks, we document persistent, yet time-varying, return autocorrelations of alternating signs. These features induce locally monotone relations between realized volatility and the underlying sampling frequency. Accordingly, this relation also displays substantive fluctuation across time. Results along these lines have been reported before but, arguably, have not received the scrutiny they deserve, especially in terms of their implications for model specification and inference. Since our explorations do not involve the development of novel techniques, we defer the detailed review of methodology and associated findings to the Supplementary Appendix. Instead, this section reviews findings that motivate the theoretical developments and subsequent empirical approach adopted in the paper.

As a vivid illustration, Figure 1 presents the so-called volatility signature plots, introduced by Andersen et al. (2000), for two stocks constructed from intraday quotes over one trading day. The increasing RV measure at higher sampling frequencies for Ebay in the left panel (moving leftward along the horizontal axis) is consistent with i.i.d. noise, which generates an MA(1)

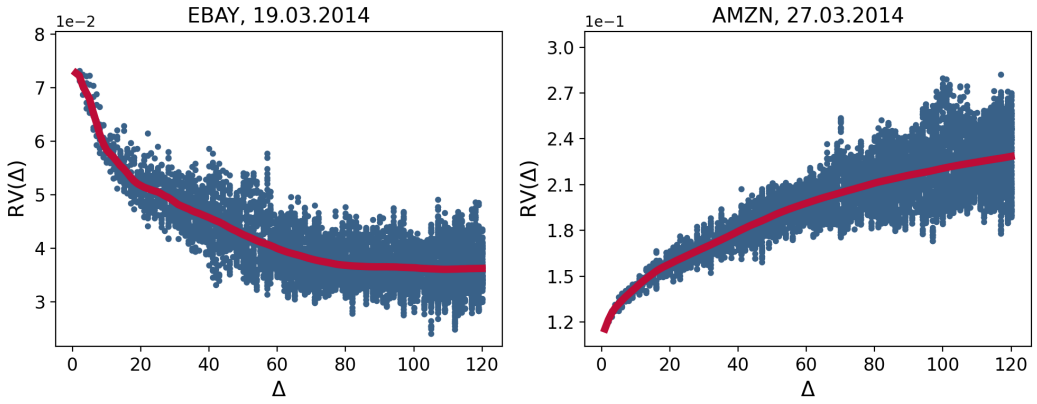

Figure 1: The figure depicts signature plots for the realized variances (RV) of Ebay and Amazon on selected trading days. The left and right panels illustrate cases with downward and upward sloping signature plots, respectively. For each sampling frequency Δ (1-120 seconds with one second increments), we compute the RV measure for multiple "grids" by initiating the grid at each second mark from 1 until Δ (thus, for Δ=1sec, we have one grid, and for Δ=120sec, we obtain 120 grids). Each blue dot corresponds to a daily RV estimate computed at a given sampling frequency and a specific initial grid point. The red solid lines represent the average value of RV for a given sampling frequency across all grid configurations (the so-called sub-sampled RV estimate).

structure and negative return autocorrelation, as noted by Roll (1984).[2] However, our Nasdaq sample contains many examples of the exact opposite behavior, exemplified by the right panel, where the signature plot for Amazon decreases monotonically and drops sharply at the highest frequencies.

The shape of signature plots is tied to the underlying autocorrelation structure. Figure 2 depicts the first-order return serial correlations on these particular days for all of the stocks.[3] They are computed over each of the 39 consecutive 10-minute intervals across the trading day at two different sampling frequencies, with the relevant Ebay and Amazon sample autocorrelations highlighted in red. Focusing initially on these observations, even if the return dependencies appear to vary across the trading day, they are predominantly negative and positive, respectively, for Ebay and Amazon. Moreover, it turns out that the sample statistics, as expected, are not perfectly correlated across the two sampling frequencies, because MMN has a differential impact in these scenarios - a feature verified by the corresponding signature plots in Figure 1.

We now turn to the intraday return autocorrelations for all remaining NASDAQ 100 constituents on the two selected days (grey dots) in Figure 2. The relative number of positive sample autocorrelations is close to, and often exceeds, the fraction of negative ones. This finding is stable across trading days and robust to the choice of sampling frequency. Moreover, there is a wide dispersion in the strength of the contemporaneous serial correlation across the

[2]The i.i.d. noise assumption was invoked in the early RV literature to identify and estimate the magnitude of the MMN induced return variation, see Zhang et al. (2005) and Bandi and Russell (2008).

[3]The statistics are computed imposing a mean return of zero, as detailed in the Supplementary Appendix.

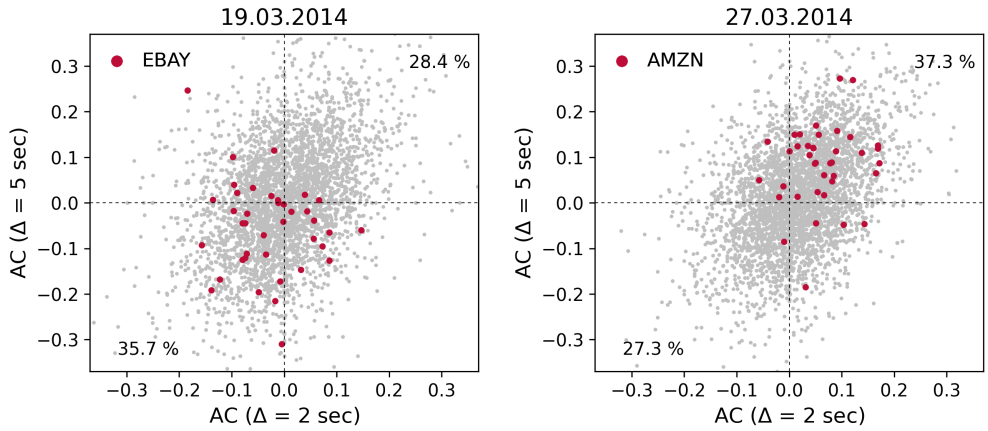

Figure 2: The figure shows return autocorrelations estimated within 10 min local intervals on four selected trading days for the NASDAQ 100 constituents. Each dot on a scatter plot corresponds to a single local interval and single stock (in total, 3900 dots for a trading day). A single dot reflects serial autocorrelation calculated with returns sampled at Δ=2 sec (horizontal axis) and Δ=5 sec (vertical axis). Red dots highlight the estimated local autocorrelations for a selected stock specified in the upper left corner (39 dots for a trading day). Percentage numbers in the upper right and lower left corners indicate the fractions of dots placed in the upper right and lower left quadrants, respectively.

stocks, even if the majority display negative dependencies in the left panel and positive in the right panel. The Supplementary Appendix provides comprehensive evidence for moderately persistent and significant return autocorrelation regimes of alternating sign across consecutive intraday intervals.[4] This implies that the return serial correlation for individual stocks usually display pronounced fluctuations over the course of the trading day. To summarize, we find about an equal amount of positive and negative high-frequency return autocorrelations which, in turn, are the ultimate drivers behind upward and downward sloping volatility signatures.

1.2 Implications for High-Frequency Return Dynamics

The empirical findings in Section 1.1 have important implications. First of all, the analysis of local first-order return autocorrelations across a variety of frequencies indicate that the MMN process is complex. Negative return autocorrelation is consistent with i.i.d. exogenous noise, orthogonal to the efficient price. In contrast, significant *positive* serial correlation and monotonic downward sloping signature plots imply that the noise is persistent and/or endogenous. The empirical evidence, detailed in the Supplementary Appendix, finds about an equal number of negative and positive intraday autocorrelation regimes, with frequent transitions between regimes for a given stock. That is, the MMN process is dynamic and strongly time-varying.

[4]Specifically, for every stock in the sample, we test if the local return serial correlation equals zero for a range of sampling frequencies and local interval lengths. To enhance statistical power, we apply one-sided heteroskedasticity-robust tests for the average absolute and signed (only positive/negative) autocorrelations over a sequence of consecutive intervals. In addition, we document a statistically significant persistence in the sign of the return autocorrelation across local intervals for most of the stocks throughout the sample period.

Secondly, monotonically declining volatility signature plots also imply that the MMN cannot be exogenous, even if it displays positive short-term autocorrelation. This finding corroborates the hypothesis of *endogenous* microstructure noise. Specifically, if MMN is stationary and uncorrelated with the efficient price process, then the variance of the observed high-frequency returns equals the sum of the variance of the underlying components, i.e., $Var(r_i) = Var(r_i^*) + Var(u_i)$, where r_i and r_i^* denote the observed and efficient returns over the i-th high-frequency interval, and u_i is the corresponding MMN increment. This decomposition trivially carries over to RV measures obtained as the cumulative sum of squared high-frequency returns over a local interval. Moreover, since the MMN increments are largely invariant with respect to the return frequency, i.e., they are of order $O_p(1)$, while the fundamental return variation is of order $O_p(\Delta)$ and grows linearly with the horizon Δ, the impact of MMN is large relative to the typical efficient price innovation over short high-frequency intervals, that is, $Var(r_i) \approx Var(u_i) > Var(r_i^*)$, but close to negligible over longer return horizons, where $Var(r_i) \approx Var(r_i^*)$. Consequently, if MMN is uncorrelated with the fundamental price process, the RV measure should be growing with the sampling frequency due to the larger relative impact of $Var(u_i)$.

Contrary to the predictions outlined above, exploring a large cross-section of equity returns, we find no systematic evidence that realized return volatility increases with the sampling frequency. Instead, we encounter a striking variety of growing and declining slopes for the volatility signature plots, both across time for a given stock and across stocks at a given point in time. Since declining volatility signature plots and negatively biased RV measures are only feasible, if the noise and fundamental return innovations are *negatively* correlated, the noise leaves a distinct footprint in the data, signifying the presence of endogeneity. Consistent with these findings, studies – going back at least as far as Hansen and Lunde (2006) – conclude that such high-frequency return dynamics are inconsistent with an exogenous noise representation.[5]

Hence, our empirical evidence – along with insights founded in market microstructure theory – poses a challenge for all representations assuming orthogonality between MMN and the efficient price process. This is true even if the i.i.d. assumption for MMN is relaxed. Recent examples of sophisticated econometric work developing nonparametric inference while allowing for a flexible MMN dependency structure include Jacod et al. (2017), Li and Linton (2020), Da and Xiu (2020), Li (2020) and Li et al. (2020). These approaches, however, all maintain the assumption of orthogonality between the MMN increments and the fundamental martingale price component. In fact, this exogeneity assumption is essential, because nonparametric identification of the separate latent martingale and noise components becomes infeasible, when MMN is endogenous, as we discuss in more detail later.

[5]Hasbrouck and Ho (1987) is among the first studies noting the inadequacy of the basic MA(1) representation.

1.3 A Parsimonious Framework Accommodating Endogeneity

Our discussion of microstructure theory suggests that MMN is endogenous, while the evidence in Sections 1.1-1.2, and further detailed in the Supplementary Appendix, corroborates this conjecture. Consequently, we seek a tractable framework for high-frequency asset prices that accommodates MMN endogeneity and diverse serial correlation patterns in the return dynamics.

To capture these empirical features and connect high-frequency volatility estimation to market microstructure theory, we adopt a parametric model facilitating inference over short intraday windows. It is cast in discrete time, allowing for temporal feedback effects which persist, even if returns are sampled at (very) high frequencies. Within this setup, we extend the classic "martingale-plus-noise" model by components accommodating an error correction mechanism and price endogeneity. These ingredients arise naturally, once one recognizes that the investors' information set is imperfect and subject to acquisition and processing delays. Under such circumstances, investors cannot determine, with certainty, whether a given price increase is due to the arrival of positive (private) news or random buying orders from liquidity traders (noise). In the former case, the price is likely to continue its upward trajectory while, in the second case, the price tends to revert. Even if astute traders avoid *systematic* mispricing, the valuation is unbiased only *on average*. During episodes with sustained imbalances in net liquidity demand or unusually intense information-based trading, uninformed agents tend to err in the same direction. Nonetheless, rationality does imply that simple trading strategies cannot earn significant profits. The fact that we find positive and negative return autocorrelation patterns to be about equally common is consistent with this prediction.

The ingredients of our model are not new to the literature. The classical decomposition of observed prices into random-walk and stationary components is discussed already by Hasbrouck (1993). However, no existing tractable framework simultaneously accommodates noise endogeneity and error correction effects, treating them as interacting features operating at varying intensity, and shaping the diverse and frequently changing return autocorrelation patterns.

We retain tractability by assuming our model applies only locally, and thus allow the model parameters to vary across intraday intervals. This flexible approach is aligned with our evidence, showing that the sign of return autocorrelations alternates frequently across the trading day. It effectively allows the shifting market environment to guide the evolution of the system. Moreover, the local approximating model with constant coefficients enables us to sidestep the difficult task of jointly modeling the intraday patterns and the persistent activity dynamics. By cumulating local volatility estimates, we obtain – as a byproduct – an alternative return variation measure that generally is consistent with the RV measures, but deviates systematically under specific market conditions, reflecting the relative strength of different MMN distortions.

Our model includes four parameters; (i) the volatility of the random walk process (the "fundamental" volatility); (ii) the "magnitude of the noise," or noise-to-signal ratio; (iii) the speed of reversal from mispricing, which is closely related to the persistent return autocorrelation or momentum effects; and (iv) the instantaneous response to efficient price innovations, capturing the extent of mispricing due to informational frictions. This parsimonious representation captures various return dependencies implied by microstructure models. In particular, the classic "martingale-plus-noise" model and variants, allowing for endogenous noise, emerge as special cases. Moreover, it embeds the partial adjustment dynamics of Amihud and Mendelson (1987), which, combined with our accommodation of heterogeneity across intraday intervals, renders the setting reminiscent of the endogenous noise specification of Kalnina and Linton (2008).[6] It further embeds the illustrative model of Chan (1993), who emphasizes return correlation patterns resulting from signal extraction in a context of noise and fundamental information. Our approach is also related to studies linking the properties of noise more explicitly to the trading process and the underlying microstructure, e.g., Hasbrouck (1993), Madhavan et al. (1997), Diebold and Strasser (2013), Li et al. (2016), and Chaker (2017). However, none of these studies consider the type of high-frequency variation in mispricing and temporal feedback effects that are crucial in explaining the return dynamics underlying Figures 1 and 2.

While our framework covers fundamental microstructure price mechanisms in a minimalistic way, it remains sufficiently rich to capture distinct market states. In particular, the speed of price reversion interacts with the noise-to-signal ratio to determine the sign of the return autocorrelation. We classify the market environment accordingly into two regimes. In one, the impact of "mispricing," induced by idiosyncratic noise shocks, is dominated by the strength of the price reversal, inducing negative return serial correlation, or "contrarian" traits. In the other regime, the feedback from mispricing generates "momentum," or positive return autocorrelation.

In summary, we identify the difficulty of aligning temporal feedback effects with infill asymptotics, where pricing errors (in the limit) are corrected instantaneously, as a prime reason for the lack of commonality between high-frequency based volatility estimation and market microstructure models. Moreover, as discussed by Hasbrouck (1993), in a discrete-time random-walk-plus-noise model, separate identification of the noise variance and the degree of noise endogeneity is generally infeasible. In our extended model with temporal feedback effects, these identification issues become even more involved. In response, we provide an analysis of (partial) identification restrictions and explore conditions that ensure identification of the full parameter vector. As such, our approach represents a step towards the development of high-frequency volatility estimators that retain a link to the market microstructure literature and allow for a more structural treatment of the impact originating from noise shocks.

[6] See also Sheppard (2013) for a related notion intended to measure market "speed."

Conveniently, our model may be cast in a linear state-space form, enabling estimation via the Kalman filter and ensuring consistent estimates for the (locally constant) volatility, albeit with non-negligible error for short horizons. We estimate the model over intraday intervals using high-frequency mid-quote returns. The results imply a high degree of informational efficiency, although the price process also contains a non-trivial element of sluggishness. We identify strong intraday periodicities in the speed of price discovery and find that *fundamental* volatility is highly elevated at the market open and flat through the remainder of the trading day. The increasing activity towards the market close is, in contrast, driven by uninformative trading and idiosyncratic noise. This illustrates the potential of our approach to build a foundation for new quantitative measures of market efficiency and fundamental volatility. Formal tests favor our flexible parameter approach relative to popular stylized microstructure models.

1.4 Overview

The rest of the paper is structured as follows. Section 2 introduces the parametric model and explores its dynamic properties along with its relation to existing models. Section 3 focuses on identification issues in our general specification and reviews our estimation strategy. Section 4 presents empirical findings, and Section 5 concludes. Auxiliary materials are deferred to the Supplementary Appendix, including comprehensive evidence for statistically significant local return autocorrelations of varying sign as well as a large-scale Monte Carlo study that corroborates the robustness of our empirical findings.

2 A Model for High-Frequency Asset Price Dynamics

This section introduces a parsimonious parametric model accommodating the salient features of the price dynamics illustrated in Section 1 and more systematically documented in the Supplementary Appendix. The objective is to assist in the identification and interpretation of the drivers behind the time-variation in the shape of the volatility signature plots and return autocorrelation patterns. This should provide a starting point for building more realistic models of real-time price discovery and market fluctuations, in which we can gauge the impact of learning, information heterogeneity and various market microstructure frictions.

2.1 The Discrete-Time Local Regime Model

We observe the logarithm of the quote midpoint for a given asset, p_i, across an equidistant time grid, $i \in \{0, 1, 2, \ldots, T\}$, yielding a total of T log-returns, $r_i = p_i - p_{i-1}$, $i = 1, \ldots, T$. Each return covers a short interval, on the order of multiple seconds, but not fractions of a second.

An important consideration for our model design is the identification of channels for price discovery, as opposed to MMN, allowing for significant return autocorrelation over non-trivial time intervals. Generally, we cannot disentangle such features by nonparametric techniques, so we impose identifying structure through parametric assumptions, subject to empirical scrutiny.

The "efficient" (full information and frictionless) log price at time i, p_i^*, follows a random walk, consistent with a no-arbitrage representation for the return over short intraday intervals,

$$p_i^* \;=\; p_{i-1}^* + \varepsilon_i^* \qquad \text{and} \qquad r_i^* \;=\; \varepsilon_i^* \;=\; p_i^* - p_{i-1}^*\,, \tag{1}$$

where $r_i^* = \varepsilon_i^*$ is the "efficient" return with $\varepsilon_i^* \sim \text{i.i.d.}\big(0, \sigma_*^2\big)$.

The evidence in Section 1.1 suggests that the mid-quote price dynamics deviates from that of the efficient price in significant ways. We propose a simple representation that accommodates both price endogeneity and correlation in the pricing errors via readily identifiable components. Specifically, the return dynamics, locally, evolves according to the scheme,

$$r_i \;=\; p_i \;-\; p_{i-1} \;=\; -\,\alpha\,(p_{i-1} \;-\; p_{i-1}^*) \;+\; (\gamma\,\varepsilon_i^* \;+\; \varepsilon_i)\,, \quad 0 < \alpha < 2, \tag{2}$$

where ε_i is an i.i.d. component, uncorrelated with ε_i^*, $\mathbb{E}[\varepsilon_i\,] = \mathbb{E}[\varepsilon_i\,\varepsilon_i^*] = 0$ and $\mathbb{V}[\varepsilon_i] = \sigma_\varepsilon^2$.

In the martingale-plus-noise model, the degree of noise impacts the optimal sampling frequency, see Ait-Sahalia et al. (2005) and Bandi and Russell (2006). It is also critical for the short-run dynamics in our framework. Hence, we define the noise-to-signal ratio as,

$$\lambda \;=\; \sigma_\varepsilon^2/\sigma_*^2.$$

For $\alpha = \gamma = 1$, we obtain the "classic" representation, $p_i = p_i^* + \varepsilon_i$. In contrast, if $\gamma \neq 1$ or $\alpha \neq 1$, we introduce noise endogeneity and error correction, potentially generating downward-sloped volatility signatures and persistent return autocorrelation. Below we separately discuss the effects of price endogeneity and temporal feedback by, in turn, letting $\alpha = 1$ or $\gamma = 1$.

Uncorrelated Endogenous Pricing Errors

For $\alpha = 1$, we eliminate persistent return serial dependence. Specifically,

$$p_i \;=\; p_i^* \;+\; (\gamma \;-\; 1)\,\varepsilon_i^* \;+\; \varepsilon_i\,, \tag{3}$$

so the system features i.i.d. noise but, in general, there is endogeneity, as the error term is correlated with the fundamental price innovation ε_i^*, whenever $\gamma \neq 1$.

Letting γ differ across local intervals is one way to mimic intertemporal heterogeneity in the information environment. Traders with incomplete information seek to infer the fundamental

value from the concurrent market dynamics, including signals derived from price innovations, evolving order book imbalances, consummated trades, and incoming news items. At a point in time, rational agents with incomplete information tend to draw similar conclusions from available public signals. Hence, cumulative random shocks in either direction generate correlated, albeit short-lived, pricing errors. Even if investors price assets correctly on average, they will, for short intervals, induce $\gamma > 1$ in some scenarios and $\gamma < 1$ in others. Such temporary periods of over- or under-reaction to latent news impact the price dynamics. To formalize the discussion, we explicitly characterize the return variation and first-order return dependence in the system.

From equation (3), the return variance and first-order auto-covariance take the forms,

$$\mathbb{V}[r_i] \,=\, \sigma_*^2 \,+\, 2\,[\,\gamma(\gamma-1) \,+\, \lambda\,]\,\sigma_*^2 \quad \text{and} \quad \mathbb{C}ov[r_i, r_{i-1}] \,=\, [\,\gamma(1-\gamma) \,-\, \lambda\,]\,\sigma_*^2\,, \qquad (4)$$

and all higher order return autocorrelation coefficients ($h > 1$) are zero.

Both the return variation and covariation depend on the quadratic term $\gamma\,(\gamma-1)$. For $\gamma = 1$, we obtain the classic i.i.d. noise result of variation inflation and negative serial correlation. For $\gamma > 1$, the price overreacts to fundamental news, reinforcing the excess volatility and exacerbating the negative correlation induced by idiosyncratic noise. In contrast, if $\gamma < 1$, the concurrent return innovation fails to fully incorporate the efficient price innovation, smoothing the price path and counteracting the irregularity stemming from i.i.d. noise. In this case, the return variation may even drop below the value associated with the fundamental return, $\mathbb{V}[r_i] < \sigma_*^2$, and the return autocorrelation turns positive. Because $\alpha = 1$, the adjustment to the efficient price innovation is completed over the subsequent interval, and the smoothing effect is maximized, when the price change is evenly distributed over the two intervals, i.e., $\gamma = 1/2$.

These results reflect the symmetry of the second-order return moments around $\gamma = 1/2$. The degree of price smoothing is controlled by $|\gamma - 1/2|$, so there are two distinct values of γ that imply the identical return variance and auto-covariance. For example, the martingale plus i.i.d. noise model arises in two separate cases, $\gamma = 1$ and $\gamma = 0$. More generally, this symmetry feature induces an identification problem, that we analyze in further detail below.

Error Correction Dynamics

Focusing instead on the error correction mechanism, the scenario $\gamma = 1$ and $\alpha \neq 1$ yields,

$$p_i \,=\, p_{i-1} \,-\, \alpha\,(p_{i-1} \,-\, p_{i-1}^*) \,+\, \varepsilon_i^* \,+\, \varepsilon_i \,\;=\; p_i^* \,+\, (1 \,-\, \alpha)\,(p_{i-1} \,-\, p_{i-1}^*) \,+\, \varepsilon_i\,. \qquad (5)$$

Now the i.i.d. error is independent of the efficient return innovation but, for $0 < \alpha < 1$, the response to pricing errors is sluggish, inducing longer-run return autocorrelation.

This parsimonious reduced-form error correction mechanism is motivated by the intuition that risk-averse agents with incomplete information form (unconditionally) unbiased, yet not error-free, expectations about the efficient price. Investors are aware that temporary mispricing creates opportunities for (risky) speculative trading which generates a pull back towards efficiency. Prior studies, e.g., Kyle (1985) and Vives (1995), identify various settings and economic factors, that render prolonged price reaction patterns consistent with sequential learning.

Formally, we readily obtain the unconditional return variance and h^{th}-order auto-covariance,

$$\mathbb{V}[r_i] \;=\; \sigma_*^2 \;+\; \frac{2\,\lambda}{2\;-\;\alpha}\,\sigma_*^2 \quad \text{and} \quad \mathbb{C}ov[r_i, r_{i-h}] \;=\; -\,(1\;-\;\alpha)^{h-1}\,\frac{\alpha\,\lambda}{2\;-\;\alpha}\,\sigma_*^2\,. \tag{6}$$

Absent endogeneity, the error term inflates the return variation and renders the autocorrelation negative. A small α leads to a more protracted error correction, more return smoothing and less price variation, while $1 < \alpha < 2$ generates overshooting and increased return variation.

The "Random-Walk-plus-Noise" Representation

To facilitate comparisons to the extant literature, we couch our model (1)-(2) in a random-walk-plus-noise format by labeling the mispricing component, its innovation, and its variance,

$$\mu_i \;=\; p_i \;-\; p_i^*, \qquad \varepsilon_i^{\mu} \;=(\gamma-1)\,\varepsilon_i^* \;+\; \varepsilon_i \quad \text{and} \quad \mathbb{V}[\varepsilon_i^{\mu}] \;=\; \sigma_{\mu}^2 \;=\; (\gamma-1)^2\sigma_*^2 \;+\; \sigma_{\epsilon}^2\,. \tag{7}$$

The system (1)–(2) now implies,

$$p_i \;=\; p_i^* \;+\; \mu_i, \tag{8}$$

$$\mu_i \;=\; (1-\alpha)\,\mu_{i-1} \;+\; \varepsilon_i^{\mu}\,. \tag{9}$$

Equation (8) identifies μ_i as a MMN component, generating a wedge between p_i and p_i^*. Correlation between the innovations in model (1)-(2) along with temporal feedback imply, that the pricing error is affected by both efficient price and noise shocks. Econometrically, $\alpha\,\mu_{i-1}$ is an error correction term, pushing the mid-quote back towards "equilibrium," $p_i = p_i^*$.

The role of ε_i^* in the error ε_i^{μ} resembles features discussed in the case of uncorrelated endogenous errors. If $\gamma = 1$, the efficient price innovation is fully incorporated in p_i and has no impact on ε_i^{μ}. In contrast, if $\gamma = 0$, new information has no immediate impact, implying a lagged price response, and ε_i^* is fully embedded in ε_i^{μ}. We review this scenario in Section 2.4.5.

The mispricing component in Equations (8)-(9) takes the form of a mean-zero AR(1) process. It is stationary for $0 < \alpha < 2$ with unconditional variance,

$$\mathbb{V}[\mu_i] \;=\; \frac{\sigma_{\mu}^2}{\alpha\,(2-\alpha)} \;=\; \frac{(\gamma-1)^2 \;+\; \lambda}{\alpha\,(2-\alpha)}\,\sigma_*^2. \tag{10}$$

Expression (10) may be interpreted as a measure of (average) mispricing, or "inefficiency." The mid-quote equals the efficient price in expectation, i.e., $\mathbb{E}[\mu_i] = 0$ and $\mathbb{E}[p_i] = \mathbb{E}[p_i^*]$, so the average squared pricing error equals the variance, $\mathbb{V}[\mu_i]$. It is evident that $\mathbb{V}[\mu_i]$ is minimized for $\alpha = 1$ implying, rather intuitively, that rapid error correction enhances price efficiency.

The second factor controlling the degree of inefficiency is the pricing error innovation in the numerator of Equation (10). This term is small if, all else equal, there is no (local) over- or under-reaction to latent news ($\gamma = 1$), and the idiosyncratic error is minimal (low λ).

In summary, the price dynamics reflects the covariance structure of the innovations to the mispricing component and the fundamental price, and how these shocks are propagated through the temporal adjustment process. The covariance structure takes the form,

$$\Sigma \;=\; \begin{bmatrix} \mathbb{E}[(\varepsilon_i^{\mu})^2] & \mathbb{E}[\varepsilon_i^{\mu}\varepsilon_i^*] \\ \mathbb{E}[\varepsilon_i^{\mu}\varepsilon_i^*] & \mathbb{E}[(\varepsilon_i^*)^2] \end{bmatrix} \;=\; \begin{bmatrix} (\gamma-1)^2+\lambda & \gamma-1 \\ \gamma-1 & 1 \end{bmatrix} \sigma_*^2 \,. \tag{11}$$

Equation (11) shows the off-diagonal (noise endogeneity) entries are non-zero, whenever $\gamma \neq 1$.

We reiterate that our model, deliberately, is of reduced form and intended to apply only for short intervals over which the dynamics remains stable. Allowing model parameters to shift across adjacent intervals accommodates time-varying local return autocorrelations.[7]

Persistent Endogenous Noise and Local Trends

The system (8)-(9) recasts model (1)-(2) as a discrete-time MMN (or pricing error) process, with mispricing being persistent and/or endogenous for $\alpha \neq 1$ and/or $\gamma \neq 1$. These properties are complementary and both critical in generating realistic serial correlations for high-frequency returns in our setting. Persistent pricing errors allow for non-zero return autocorrelation beyond order one, while endogeneity (correlation between pricing errors and fundamental price shocks) allows the observed autocorrelations to be *positive* across *all* return horizons. In contrast, if noise is modeled as i.i.d. and orthogonal to the efficient price ($\alpha = \gamma = 1$), only the first-order autocorrelation is non-zero, and necessarily negative. Likewise, in "martingale-plus-noise" models, where noise components can be dependent, but uncorrelated with the efficient price (i.e., no endogeneity), return autocorrelations can be positive, but not jointly across all horizons.

Through the ability to induce patterns of protracted positive return autocorrelations, persistent endogenous noise is consistent with stylized features in observed price dynamics, such as occasional local trending. The mechanism behind this trending behavior in our model is simple. Suppose, for example, that the noise-to-signal ratio, λ, is moderate and the observed

[7] The spirit of this approach mimics the assumptions often invoked for developing inference techniques with high-frequency data, see, e.g., Mykland and Zhang (2009) and Bibinger et al. (2014), who explicitly rely on local windows in which the quantity of interest, in their case return volatility, may be assumed to remain fixed.

price (locally) underreacts to efficient price changes (i.e., endogeneity with $\gamma < 1$). This induces severe mispricing, if there is a large shift in the efficient price. Moreover, if the error correction is sluggish (low α, i.e., noise persistence), the initial mispricing is followed by a prolonged adjustment to the new equilibrium level (i.e., slow price discovery). This tends to generate a sequence of unidirectional price movements, so we observe a locally trending dynamic, even though the efficient price has zero drift.

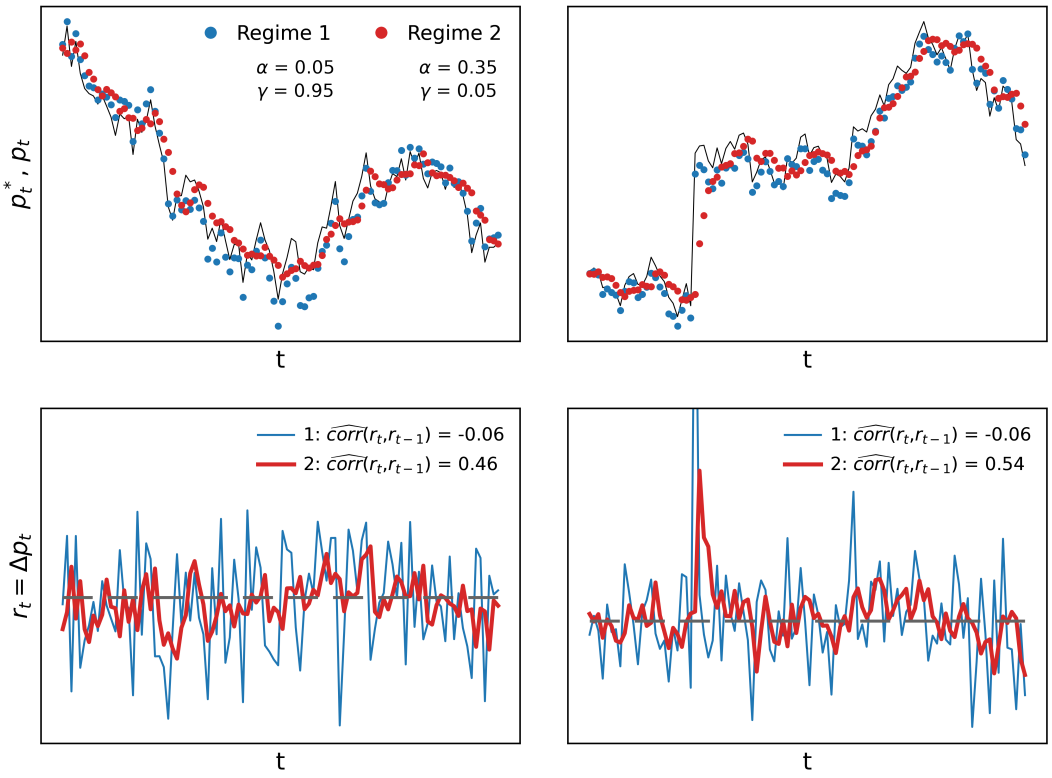

Figure 3: Simulated price (upper plots) and return processes (bottom plots) for alternative parameter values. The first regime (blue) features a nearly unbiased instantaneous reaction to fundamental shocks (γ=0.95), implying low noise endogeneity. The second regime (red) displays severe instantaneous underreaction to fundamental news (γ=0.05), implying substantial noise endogeneity. The error correction coefficients are calibrated to generate persistent pricing errors (Regime 1: α=0.05, Regime 2: α=0.35). Noise-to-signal ratios are identical (λ=0.2).

The plots for Regime 2 in Figure 3 illustrate how trending episodes emerge naturally, when the price is simulated for model parameters mimicking persistent endogenous noise. In the left panel, the fundamental price path (full-drawn line), modeled as a random walk without drift, displays random oscillations across the interval. Regime 2 generates the corresponding observed price path – a heavily "smoothed" version of the efficient path due to systematic mispricing (low γ) followed by slow error correction (low α). In the right panel, the efficient price exhibits a positive jump so, in Regime 2, the pricing error is particularly severe, generating a temporary adjustment process akin to "gradual" jumps, as described by Barndorff-Nielsen et al. (2008a). Importantly, even if the pricing error is very persistent, such local trends are largely absent

within Regime 1, for which the degree of endogeneity is low - the corresponding estimated return serial correlations are, in fact, slightly negative. This reflects the near instantaneous incorporation of the fundamental shock ($\gamma \approx 1$), ensuring a strong temporal coherence between the efficient and observed returns (for reasonable λ values). Nonetheless, the price may still, locally, differ non-trivially from the efficient level, as is, indeed, the case for Regime 1.

2.2 Second Order Return Moments

This section provides model-implied conditions for excess volatility and return serial correlation. These features are closely related and reflect the impact of MMN versus smoothing in the return variation process. The latter also affects our ability to identify the drivers behind the price dynamics from the observed return moments in the absence of auxiliary assumptions.

2.2.1 Excess Return Variation

We recall that the log-return in Equation (2) may be stated as $r_i = -\alpha\mu_{i-1} + (\varepsilon_i^* + \varepsilon_i^\mu)$. The following lemma provides an explicit expression for the return variation in this general case.

Lemma 1 (Return Variance)**.** *Assume* $\sigma_\varepsilon^2 > 0,\ \ 0 < \alpha \leq 1,\ \ 0 \leq \gamma < 2.$ *Then,*

$$\mathbb{V}[r_i] \ = \ \alpha^2\,\mathbb{V}[\mu_i] \,+\, (\gamma^2 + \lambda)\,\sigma_*^2 \ = \ \sigma_*^2 \,+\, \frac{2}{2\,-\,\alpha}\,F_\alpha(\gamma,\lambda)\ \sigma_*^2\ , \tag{12}$$

where, for later convenience, for given α, we define $F_\alpha(\gamma,\lambda)$ as a function of γ and λ,

$$F_\alpha(\gamma,\lambda) \ = \ \gamma^2 \,+\, (\,1\,-\,\gamma\,)\,\alpha \,+\, \lambda \,-\, 1.$$

Proof. Follows by a sequence of straightforward calculations from equation (10). □

Equation (12) implies the return volatility is increasing in σ_*^2 and λ. Likewise, due to the "smoothing" of the price path for low α, the volatility grows with α (for $\gamma < 1$). Finally, as before, the return variation is minimized for $\gamma = \alpha/2$, increasing for $\gamma > \alpha/2$, and decreasing for $\gamma < \alpha/2$, generalizing the result for the uncorrelated endogenous noise setting ($\alpha = 1$).

The following corollary summarizes the conditions for excess volatility in our general setting, featuring both error correction and endogeneity.

Corollary 1. *Assume* $\sigma_\varepsilon^2 > 0,$ *and* $0 < \alpha \leq 1.$ *Then,*

$$\mathbb{V}[r_i] \ \leq \ \mathbb{V}[r_i^*] \qquad \textit{if} \qquad F_\alpha(\gamma,\lambda) \ \leq \ 0\,, \tag{13}$$

$$\mathbb{V}[r_i] \ > \ \mathbb{V}[r_i^*] \qquad \textit{otherwise}. \tag{14}$$

Consequently, the return volatility is lower than the fundamental volatility if price updating is slow, the extent of noise is low (i.e., α and λ are small), and γ is close to $\alpha/2$, implying a variance-minimizing degree of smoothing. The condition $F_\alpha(\gamma, \lambda) < 0$, alternatively, can be stated as $\gamma^2 + \lambda < 1 - \alpha(1 - \gamma)$. Hence, the two regimes are governed by the relative size of the *effective return smoothing*, $1 - \alpha(1 - \gamma)$, and the size of the return innovations, $\gamma^2 + \lambda$.

Thus, we have two distinct regimes. If $\gamma = 1$, the presence of noise, $\lambda > 0$, trivially leads to excess volatility. In contrast, with partial price adjustment, the volatility may drop below the fundamental return variation and generate inverted volatility signature plots.

2.3 Return Auto-Covariances

The determinants of whether returns display excess volatility are also critical for the sign of the return auto-covariances. This is a consequence of the following lemma and corollary.

Lemma 2 (Return Auto-Covariances)**.** *Assume* $\sigma_\varepsilon^2 > 0$, $0 < \alpha \leq 1$, *and* $h \geq 1$. *Then,*

$$\mathbb{C}ov[r_i, r_{i-h}] \;=\; \psi(h-1)\,\left[1 - \alpha(1-\gamma) - (\gamma^2 + \lambda)\right]\,\sigma_*^2 \;=\; -\,F_\alpha(\gamma, \lambda)\,\psi(h-1)\,\sigma_*^2 \quad (15)$$

with $\psi(h-1) \;=\; \frac{\alpha}{2-\alpha}\,(1-\alpha)^{h-1}$ *and* $\psi(0) = 1$, *if* $\alpha = 1$.

Proof. See the Supplementary Appendix. □

Hence, the relation between $1 - \alpha\,(1 - \gamma)$ and $\gamma^2 + \lambda$ also governs the sign of the auto-covariance for the observed returns. We summarize this result in the form of a corollary.

Corollary 2. *Assume* $\sigma_\varepsilon^2 > 0$, $0 < \alpha \leq 1$, *and* $h \geq 1$. *Then,*

$$\mathbb{C}ov[r_i, r_{i-h}] \;\geq\; 0 \qquad \textit{if} \qquad F_\alpha(\gamma, \lambda) \;\leq\; 0\,, \quad (16)$$

$$\mathbb{C}ov[r_i, r_{i-h}] \;\leq\; 0 \qquad \textit{otherwise}. \quad (17)$$

We note that, for $0 < \alpha < 1$, the sign of the auto-correlation is identical for all $h \geq 1$, so both "contrarian" and "momentum" regimes display persistent return dependence. In Section 4, we seek to identify this feature empirically over short trading intervals.

Without endogeneity, i.e., $\gamma = 1$, return momentum is infeasible here – a point further discussed in Section 2.4.1. However, contrary to the pure price endogeneity setting of Section 2.1, momentum effects may arise for $\gamma = 0$, if $\lambda < 1 - \alpha$.

It is useful to relate these model implications to empirical evidence. The Supplementary Appendix documents that, on average, the stocks display only limited return serial dependence. For low γ values, it implies that $1 - \alpha \approx \lambda$. Our subsequent review of the estimation results

suggests that this arises because the MMN component is small, while prices impound information swiftly, i.e., $1-\alpha$ is small as well. More generally, zero return correlation is consistent with MMN, as long as the impact is mitigated by a certain sluggishness in market prices.

A natural application of these results is realized volatility estimation. Traditional realized variance measures, obtained from cumulative high-frequency squared returns, will be biased and inconsistent, unless the market environment generates a return dynamic that satisfies the condition for no excess volatility, namely, $\gamma^2 + \lambda \approx 1 - \alpha\,(1-\gamma)$ or $F_\alpha(\gamma,\lambda) \approx 0$.

Finally, note that for given α and λ, scenarios for which $|\gamma - \alpha/2|$ are identical, e.g., the cases $\gamma = 0$ and $\gamma = \alpha$, yield the same return variances and autocovariances. We discuss this issue in more detail in the context of statistical identification in Section 3.1.

2.4 Nested Models

Our model captures basic structural mechanisms from alternative market microstructure approaches within a uniform setting, providing a starting point for identifying strengths and limitations of existing paradigms. Because we nest several important special cases, our empirical work should help shed light on which models best approximate salient features of the high-frequency return process. A second objective is to highlight the identification issues that arise in microstructure models with a latent endogenous noise component. This often-overlooked feature becomes evident, as we summarize the empirical implications of various models below.

2.4.1 The Classical Model with Idiosyncratic Noise Autocorrelation

A critical feature of our model (1)–(2) is the accommodation of noise endogeneity via $\gamma \neq 1$. Specifically, $\gamma = 1$ implies that the innovation of the mispricing component, ε_i^μ, is uncorrelated with the (efficient price) increment ε_i^*.[8] In this case, we obtain,

$$\Sigma \;=\; \begin{bmatrix} \mathbb{E}[(\varepsilon_i^\mu)^2] & \mathbb{E}[\varepsilon_i^\mu \varepsilon_i^*] \\ \mathbb{E}[\varepsilon_i^\mu \varepsilon_i^*] & \mathbb{E}[(\varepsilon_i^*)^2] \end{bmatrix} \;=\; \begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix} \sigma_*^2 \,.$$

The absence of correlation among the innovations has important ramifications. Letting $\Delta\mu_i = \mu_i - \mu_{i-1}$, for *any* stationary noise dynamics and integers i, j, we have $\mathbb{E}[\,\varepsilon_i^*\,\Delta\mu_j\,] = 0$ and,

$$r_i \;=\; p_i \;-\; p_{i-1} \;=\; \varepsilon_i^* \;+\; \Delta\mu_i \,. \tag{18}$$

It follows that the actual return variation always exceeds the efficient return variation,

[8] As discussed by Hasbrouck (1993), this condition corresponds to a standard identification restriction in the macroeconomic literature, see, e.g., the discussion in Watson (1986).

$$\mathbb{E}[\, r_i^2] \;=\; \sigma_*^2 \;+\; \mathbb{E}[\, (\Delta \mu_i)^{\,2}] \;\geq\; \sigma_*^2,$$

and the return variance and autocovariances take the form in equation (6). It implies that, uniformly, the volatility signature plots should lie *above* the efficient variance. This is at odds with the empirical evidence, as the signature plots often drop sharply at higher frequencies.

We reiterate that this feature arises, even with dependence in the return series induced by serial correlation in the noise process. Our model specification (8) - (9) now implies,

$$p_i \;=\; p_i^* \;+\; \mu_i \,, \tag{19}$$

$$\mu_i \;=\; (1-\alpha)\, \mu_{i-1} \;+\; \varepsilon_i \,, \tag{20}$$

with $\mathbb{E}[\varepsilon_i \, \varepsilon_i^*] = 0$ and $\mathbb{E}[\, \varepsilon_i \; \mu_j \,] = 0$ for all integers i, j. Hence, in our parametric setting, the mispricing component, μ_i, follows an AR(1) process, inducing return dependence for $\alpha \neq 1$. This is relevant for volatility measures obtained from data sampled at the very highest frequencies, as features such as the bid-ask spread, price grid and order splitting induce serial dependence in both transaction and quote returns. Recognizing this issue, the original realized volatility estimation procedures employ sparse sampling, e.g., Andersen and Bollerslev (1998). Subsequently, a variety of robust approaches were proposed, including realized kernels, pre-averaging, multi-scale, or spectral procedures, e.g., Barndorff-Nielsen et al. (2008b), Jacod et al. (2009), Zhang et al. (2005), and Bibinger et al. (2014). A recent development of simple procedures attaining the same objective is given by Da and Xiu (2020).

2.4.2 The Classical Model

The i.i.d. noise model arises from the above by imposing $\alpha = 1$, so that $\mathbb{E}[\varepsilon_i \, \varepsilon_i^*] = 0$ and,

$$p_i \;=\; p_i^* \;+\; \varepsilon_i \,. \tag{21}$$

Thus, fundamental news are embodied instantaneously and past pricing errors are corrected without delay, so shocks fully dissipate by the next observation, yielding $\mathbb{V}[r_i] = \sigma_*^2(1 + 2\lambda)$ and $\mathbb{C}ov[r_i \,, r_{i-1}] = -\lambda \sigma_*^2$. That is, instant error correction implies negative first-order serial correlation, but also no return correlation beyond lag one.

This "classic" random-walk-plus-iid-noise model, see, e.g., Zhou (1996), Ait-Sahalia et al. (2005) and Bandi and Russell (2008), still provides a basic reference for empirical microstructure effects, yet rules out both feedback effects and noise endogeneity, which arise for $0 < \alpha < 1$, and $\gamma \neq 1$. In fact, these conditions seem to apply routinely across a large set of subintervals, as documented in Section 4.4.

2.4.3 The Uncorrelated Endogenous Noise Model

As noted in Section 2.4.2, imposing $\alpha = 1$ yields a scenario with endogenous, but uncorrelated noise, $p_i = p_i^* + (\gamma - 1)\,\varepsilon_i^* + \varepsilon_i$, and first-order autocovariance, $\mathbb{C}ov(r_i\,, r_{i-1}) = [(1 - \gamma)\,\gamma - \lambda]\,\sigma_*^2$. The specification is consistent with positive return dependence, if $\gamma < 1$, i.e., the concurrent return innovation does not fully incorporate the efficient price innovation. In this setting, the fundamental return variation may be estimated consistently through a first-order lag adjustment to the realized volatility estimator, as originally suggested by Zhou (1996).

The point is that noise endogeneity and information feedback are distinct features. Endogeneity can alter the sign of the return autocorrelation, while persistent noise generates longer lasting correlation effects. One of our objectives is to determine conditions under which we can identify the underlying source of noise from the observed high-frequency return dependencies.

2.4.4 The Amihud-Mendelson Model

Our model (1) - (2) modifies the standard representation to allow for economic mechanisms that generate longer-run return dependence. Meanwhile, the finance literature contains several models designed to accommodate low-frequency serial correlation. Prominent examples include Amihud and Mendelson (1987) and Hasbrouck and Ho (1987). Specifically, the Amihud-Mendelson model emerges as a special case by setting $\gamma = \alpha$, implying,

$$p_i \;=\; (1 \,-\, \alpha)\, p_{i-1} \,+\, \alpha\, p_i^* \,+\, \varepsilon_i\,. \tag{22}$$

In this scenario, the mid-quote price is a weighted average of the past price and the *current* efficient price, with an innovation term, that is uncorrelated with the efficient price innovation.

The price dynamics in equation (22) mimics the one in equation (2), with the notable difference that the price adjustment in the latter is based on the discrepancy between the lagged observed and *past* efficient price. Importantly, the model allows for both noise endogeneity and feedback effects through an error correction mechanism, while providing a parsimonious and readily identifiable model. The main drawback is the tight link between the strength of endogeneity and speed of error correction via the restriction $\gamma = \alpha$.

Note that Amihud and Mendelson (1987) seek to capture *daily* return dynamics. In that context, updating based on the *current* efficient price (established over a full trading day) is sensible relative to one-day-old information. In a high-frequency setting, featuring incessant order book revisions, multiple trades per second, and nearly continuous news feeds, it is less plausible that traders have the capability and all relevant information to gauge the efficient price every instant. Instead, updates may occur with a slight delay. As such, equation (2) may provide the more suitable representation at very high frequencies.

2.4.5 The Information Delay Model

An alternative specification, yielding the identical first- and second-order return moments, as in the Amihud and Mendelson (1987) model, may be obtained via a different economic mechanism. As noted in Section 2.2, $\gamma = 0$ implies that even well-informed agents respond to newly arriving information with a minor delay. This leads to a lagged price response and feedback – features that appear empirically plausible. Moreover, the delay and feedback effects can be stronger or weaker depending on the information and market environment at the time, motivating our approach of keeping model parameters fixed only over short intraday intervals.

In this model, the only contemporaneous shock to the price is idiosyncratic noise, but corrections to past pricing errors are also ongoing, so the efficient price innovations govern the intermediate dynamics, as the market disentangles the noise and fundamental shocks to prices.

Letting $\gamma = 0$, equations (1)–(2) generate the following price dynamics,

$$p_i \;=\; (1 \,-\, \alpha)\, p_{i-1} \,+\, \alpha\, p^*_{i-1} \,+\, \varepsilon_i \,. \tag{23}$$

A few comments are in order. First, since ε^*_i is latent and uncorrelated with ε_i, we obtain an equivalent representation by relabeling p^*_{i-1} as p^*_i. But this renders Equations (22) and (23) identical. In other words, we cannot separately identify the two models – they generate the identical process for the observed returns. This complicates identification and inference in general – an issue we discuss in depth in Section 3.1.

Second, the covariance structure for the innovations now takes the simple form,

$$\Sigma \;=\; \begin{bmatrix} \mathbb{E}[(\varepsilon^\mu_i)^2] & \mathbb{E}[\varepsilon^\mu_i \varepsilon^*_i] \\ \mathbb{E}[\varepsilon^\mu_i \varepsilon^*_i] & \mathbb{E}[(\varepsilon^*_i)^2] \end{bmatrix} \;=\; \sigma^2_* \begin{bmatrix} 1+\lambda & -1 \\ -1 & 1 \end{bmatrix}.$$

This result reflects the incorporation of efficient innovations into prices with a delay, so the mispricing component must absorb the full impact of any concurrent ε^*_i shock, rendering the noise endogenous. In this scenario, the return dynamics, reflecting the error correction mechanism, is the only source of coherence between market and efficient prices. Hence, the market is in a perpetual state of transition, driven by an ongoing process of price discovery. Conceptually, this separates our specification from standard microstructure models in which prices are in equilibrium, except for short-lived distortions induced by exogenous noise shocks.

Third, the "information delay" and Amihud and Mendelson (1987) models differ, as the latter has a covariance structure identical to the general representation (11), but with $\gamma = \alpha$. For either model, information processing proceeds at a rate governed by α. The latency of the efficient price implies, that we cannot determine if this occurs with a temporal delay or not. This notwithstanding, the incomplete and delayed price adjustment renders the noise endogenous.

Typically, such endogeneity is imposed via explicit statistical assumptions, see, e.g., Kalnina and Linton (2008), but it is endowed with a more structural interpretation in our specification.

In summary, our information delay model – or high-frequency reinterpretation of Amihud and Mendelson (1987) – featuring just three free parameters, (α, λ, σ_*), endows the asset pricing process with dynamic properties reflecting distinct underlying economic factors. The empirical evidence in Section 4 illustrates how this fully identifiable variant of our model captures salient features of the short-run dynamics, providing a framework suitable for assessing the separate economic forces at play relative to traditional microstructure representations.

2.5 Realized Volatility Measures

This section reviews the implications of the preceding results for realized volatility measures computed from return data generated by model (1)–(2). The realized volatility for interval $[0, T]$, using equidistant log-price observations, $p_{i\Delta}\,, i = 0, \ldots, n = n(\Delta) = \lfloor T/\Delta \rfloor$, is obtained as,

$$RV_T(\Delta) \;=\; \sum_{i=1}^{n(\Delta)} \left(p_{i\Delta} - p_{(i-1)\Delta} \right)^2 \;=\; \sum_{i=1}^{n(\Delta)} \left(r_{i\Delta} \right)^2 . \tag{24}$$

In the frictionless variant of our model with $\alpha = \gamma = 1$ and $\lambda = 0$, we have,

$$\mathbb{E}\left[\, RV_T(\Delta) \,\right] \;=\; T\,\sigma_*^2 . \tag{25}$$

In the case of general exogenous noise, $r_i \;=\; \varepsilon_i^* + (\mu_i - \mu_{i-1}) \;=\; \varepsilon_i^* + \Delta\mu_i$, as discussed in Section 2.4.1, we find, in analogy to the result established there, that,

$$\mathbb{E}\left[\, RV_T(\Delta) \,\right] \;=\; T\,\left[\, \sigma_*^2 + \mathbb{E}[\,(\Delta\mu_i)^2\,]\,\right] \;=\; T\left[\, 1 + \frac{2\lambda}{2-\alpha} \,\right] \sigma_*^2 \;\geq\; T\,\sigma_*^2 . \tag{26}$$

This result reiterates a point stressed by Hansen and Lunde (2006) – the RV measure, *absent endogenous noise*, cannot be uniformly downward biased. Thus, a monotone declining signature plot indicates the presence of endogenous noise. Moreover, for exogenous noise, the extent and variation in any upward bias is governed by the properties of the noise process. The usual assumption is that each price observation contains a noise component unrelated to the sampling interval, so $\Delta\mu_i = (\mu_i - \mu_{i-1})$ constitutes an $O_p(1)$ term which, for $\Delta \to 0$, dominates the $O_p(\sqrt{\Delta})$ term of the efficient price innovation. This generates an increasing upward bias at higher sampling frequency, and a positively sloped volatility signature plot, as $\Delta \to 0$.

As observed previously, Figure 1 illustrates the general point that, as the sampling frequency increases, the volatility signature plot may be sharply upward sloping in some periods and sharply downward sloping in other cases. In Section 2.4.1, we documented that the latter is

possible only if the noise is endogenous. For our full parametric model, with $\gamma \neq 1$,

$$\mathbb{E}\left[\, RV_T(\Delta) \,\right] \;=\; T\left[1 + \frac{2}{2-\alpha}\, F_\alpha(\gamma,\lambda)\right]\sigma_*^2\,. \tag{27}$$

As before, the sign of $F_\alpha(\gamma,\lambda)$ determines whether the return volatility measure is upward or downward biased. Both cases are empirically important, as each occur with regularity, showing that a (time-varying) degree of exogenous and endogenous noise is required to capture the observed features of the volatility process for individual stock returns.

3 Statistical Inference

3.1 The Identification Problem

Section 2 shows that the general model (1)–(2) is not uniquely identified from the second-order return moments. This identification issue has been recognized previously. In the literature on high-frequency based volatility estimation, researchers typically impose two distinct sets of assumptions on the noise to achieve identification. One approach casts the model in continuous time and conducts inference via an in-fill asymptotic scheme, where the number of observations (ticks) diverges within a fixed interval. In this case, the noise dependence is identified, because the tick-by-tick innovations in the efficient price shrink with the length of a sampling interval, while the magnitude of the noise shocks is invariant with respect to sampling frequency, so the noise is asymptotically "big," ensuring nonparametric identification.[9]

In contrast, if the noise dependence is fixed in discrete time, asymptotic inference requires an increasing time horizon. In this case, identification hinges on the stipulation that the fundamental price process is a (semi-)martingale, while the noise is stationary. The latter implies an element of mean reversion but, without auxiliary assumptions, this can only be ascertained with precision over longer time intervals. Moreover, dependence between the noise and fundamental price process complicates identification even further, as noted in Ait-Sahalia et al. (2006). Typically, one can only accomplish this task via careful parametric modeling.

The general model in Section 2, involving frictions and temporal price corrections, naturally falls into the second category. In fact, absent either endogeneity ($\gamma = 1$) or noise in the mispricing component ($\sigma_\varepsilon^2 = 0$),[10] the system is identified through the parametric specification,

[9] A thorough development is provided by Jacod et al. (2017); see also the discussion in Li and Linton (2020). Likewise, Da and Xiu (2020) obtain consistent inference from a tick-time MA(q) noise process in a continuous-time setting with independence between noise and efficient price, but otherwise very general assumptions.

[10] This restriction, used in the Beveridge and Nelson (1981) decomposition for macroeconometric time series models, is invoked by Hasbrouck (1993) to identify and parameterize the pricing error in terms of present and past returns as well as auxiliary explanatory variables.

but once such dependence is present, we have only partial identification. We can estimate the key parameter determining the fundamental volatility, σ_*^2, but the finer details of the noise dynamics are elusive, as the second-order return moments fail to disentangle the delay with which the price incorporates the fundamental innovation versus the relative size (variance) of the noise component. Specifically, separate identification of λ and γ fails. Both impact the return variation (12) and autocovariances (15) only through the function $F_\alpha(\gamma,\lambda) = \gamma^2 - \alpha\gamma + \alpha + \lambda - 1$. Thus, any combinations of γ and λ yielding the same value for $F_\alpha(\gamma,\lambda)$ imply identical second-order moments. In contrast, it is readily shown that α and σ_*^2 are identifiable. Henceforth, in this section, we often abbreviate $F_\alpha(\gamma,\lambda)$ by F_α.

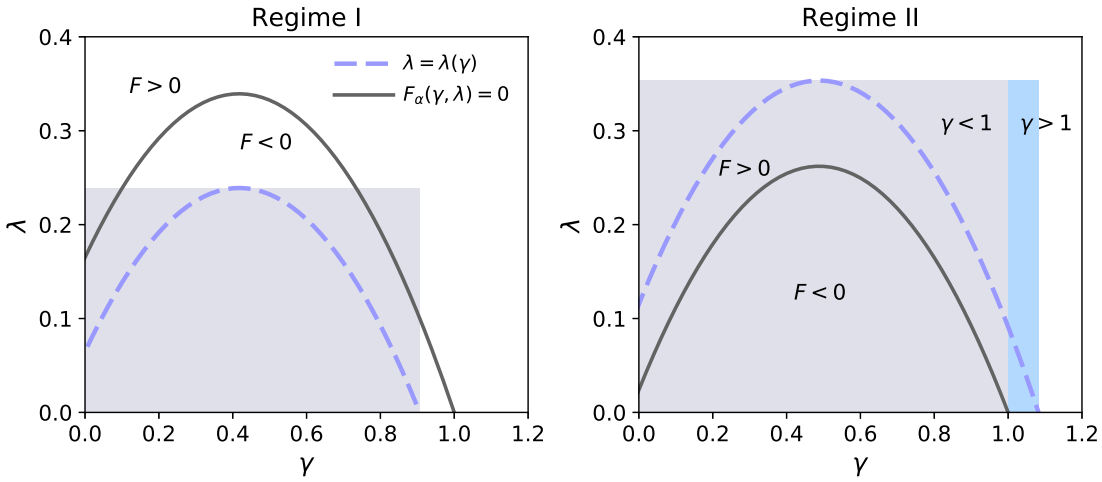

Figure 4: The dashed blue lines describe admissible values of (γ,λ), corresponding to Regime I (left plot) and Regime II (right plot). Regime I has $\alpha = 0.835$, $\sigma_*^2 = 0.986 \cdot 10^{-8}$ and $F_\alpha = -0.100$ and generates positive return autocorrelation. Regime II has $\alpha = 0.976$, $\sigma_*^2 = 0.524 \cdot 10^{-8}$ and $F_\alpha = 0.091$ and generates negative autocorrelation. The solid black lines represent values of (γ,λ) which solve equation $F_\alpha(\gamma,\lambda) = 0$ for a given α.

We may impose additional constraints on γ and λ, such as non-negativity. Yet, whenever the parameters are not on the boundary, any small shift in γ may be offset by a compensatory shift in λ, leaving F_α unaltered. Since F_α is linear in λ and quadratic in γ, fixing γ identifies λ, but fixing λ does *not* generally identify γ. In particular, scenarios with identical values for $|\gamma - \alpha/2|$ are observationally equivalent. When estimating the general model, we initially focus on the lower-dimensional vector $(\alpha, \sigma_*^2, F_\alpha)$, with values for F_α consistent with a range of (γ,λ) combinations for given return dynamics. Auxiliary economic restrictions can help identify all parameters, as illustrated in Section 4.3, enabling full model estimation.

To illustrate the geometry of non-identification regions for (γ,λ), we consider two scenarios associated with positive and negative return autocorrelations, respectively. In Regime I, we choose the parameter values $\alpha = 0.835$, $\sigma_*^2 = 0.986 \cdot 10^{-8}$ and $F_\alpha = -0.100$ producing a *positive* first-order autocorrelation of 0.087. In Regime II, we set $\alpha = 0.976$, $\sigma_*^2 = 0.524 \cdot 10^{-8}$ and $F_\alpha = 0.091$, resulting in a *negative* first-order autocorrelation of -0.074. These values for

α, σ_*^2 and F_α correspond to the median estimates $\hat{\alpha}$, $\hat{\sigma}_*^2$ and $\hat{F}_\alpha$ obtained across local 10-minute intervals with positive and negative return autocorrelations, respectively, in Section 4.

Figure 4 depicts the admissible loci of (γ, λ) for Regimes I (left) and II (right), imposing non-negativity. The shaded rectangular area indicates the range of values γ and λ can take, given the remaining parameters. The admissible (γ, λ) combinations are indicated by the dashed blue line, with λ treated as a function of γ and the value of (α, F_α) for the corresponding regime.

The solid black line depicts the loci of (γ, λ) solving $F_\alpha(\gamma, \lambda) = 0$ for a given α. All points below this line imply $F_\alpha(\gamma, \lambda) < 0$, whereas all points above indicate $F_\alpha(\gamma, \lambda) > 0$. Accordingly, for Regime I (positive autocorrelation) and negative F_α, all admissible pairs (γ, λ) are located below the parabolic curve $F_\alpha(\gamma, \lambda) = 0$. In contrast, for Regime II (negative autocorrelation), all admissible (γ, λ) combinations lie above $F_\alpha(\gamma, \lambda) = 0$.

This analysis delivers maximum admissible values for γ and λ in a given regime. For λ, it is attained at the level of γ that minimizes the return variance, i.e., $\gamma = \alpha/2$,

$$\lambda_{max} = \frac{\alpha^2}{4} - \alpha + F_\alpha + 1. \tag{28}$$

Correspondingly, it is readily seen that γ_{max} is attained for $\lambda = 0$, so that,

$$\gamma_{max} = \frac{\alpha}{2} + \sqrt{\lambda_{max}}. \tag{29}$$

This provides algebraic upper bounds given estimates for the identified parameters $(\alpha, \sigma_*^2, F_\alpha)$.

We further note that the maximum γ satisfying the boundary condition $F_\alpha(\gamma, \lambda) = 0$ is attained for $\lambda = 0$ and equals unity. Therefore, $F_\alpha < 0$ implies $\gamma_{max} < 1$, and $F_\alpha > 0$ implies $\gamma_{max} > 1$. This feature is nicely illustrated in Figure 4. If $F_\alpha < 0$ (Regime I), the dashed curve $\lambda(\gamma)$ is below the black curve ($F_\alpha(\gamma, \lambda) = 0$), crossing the horizontal axis below 1, and $\gamma_{max} < 1$. If $F_\alpha > 0$ (Regime II), the situation is reversed. Here, overreactions to return innovations ($\gamma > 1$) only occur for $F_\alpha > 0$, implying negative first-order autocorrelation.

These identification relations have direct implications for the nested models discussed in Section 2.4. Firstly, the equivalence between the Amihud-Mendelson and Information Delay models is due to $\gamma = \alpha$ and $\gamma = 0$ both implying $F_\alpha(\gamma, \lambda) = \alpha + \lambda - 1$. Secondly, the i.i.d. noise model implies $\gamma = 1$ and $F_\alpha(1, \lambda) = \lambda > 0$, excluding positive serial correlation. As demonstrated in the empirical analysis, this constraint is often counterfactual, and the associated estimates of the efficient return variance, σ_*^2, will generally be biased.

The endogenous uncorrelated noise model of Section 2.4.3 with $\alpha = 1$ implies $F_1(\gamma, \lambda) = \gamma^2 - \gamma + \lambda$. In contrast to the classical model with idiosyncratic noise, this allows for a positive first-order return autocorrelation. Due to the constraints discussed above, we have $F_1(\gamma, \lambda) \geq -\frac{1}{4}$, thus the first-order autocorrelation must fall within $(-\frac{1}{2}, \frac{1}{2})$.

Finally, by fixing γ, the information delay (equivalently, Amihud-Mendelson) model allows for identification of all remaining parameters without imposing additional constraints on the return dynamics. This specification allows for only partial instant incorporation of fundamental information along with the feedback mechanism associated with dynamic price discovery.

3.2 Estimation Under Identification

For estimation purposes, it is convenient to restate the model as,

$$\begin{aligned} r_i &= -\alpha\,\mu_{i-1} + \widetilde{\varepsilon}_i, \\ \mu_i &= (1-\alpha)\,\mu_{i-1} + \varepsilon_i^{\mu}, \end{aligned}$$

where $\widetilde{\varepsilon}_i = \gamma\varepsilon_i^* + \varepsilon_i$ and $\varepsilon_i^{\mu} = \widetilde{\varepsilon}_i - \varepsilon_i^* = (\gamma-1)\varepsilon_i^* + \varepsilon_i$. Then, returns r_i are driven by a latent variable μ_i following an AR(1) process with covariance structure given by,

$$\begin{aligned} \mathbb{E}[\,\widetilde{\varepsilon}_i\mu_{i+h}\,] &= (1-\alpha)^h\,\left(\gamma^2+\lambda-\gamma\right)\,\sigma_*^2\,, \qquad \forall h \geq 0, \\ \mathbb{E}[\,\widetilde{\varepsilon}_i\mu_{i-h}\,] &= 0, \qquad \forall h > 0, \\ \mathbb{E}[\,\widetilde{\varepsilon}_i\varepsilon_i^{\mu}\,] &= [\,\gamma(\gamma-1)+\lambda\,]\,\sigma_*^2, \\ \mathbb{E}[\,\widetilde{\varepsilon}_i\varepsilon_{i-h}^{\mu}\,] &= 0, \qquad \forall h \neq 0. \end{aligned}$$

In the Supplementary Appendix, we show the model can be estimated by maximum likelihood (ML) via the prediction error decomposition, see e.g., Harvey (1989), yielding the log-likelihood,

$$\ell(Y,\theta) \;=\; -\frac{1}{2}\sum_{i=1}^{T}\ln(s_i) \;-\; \frac{1}{2}\sum_{i=1}^{T}\frac{\nu_i^2}{s_i^2}, \tag{30}$$

where ν_i denotes the (optimal) linear prediction of r_i given the past returns $(r_{i-1},\ldots,r_1)$, and $s_i^2 = \mathbb{E}[\nu_i^2|r_{i-1},\ldots,r_1]$ is the conditional variance. Assuming ε_i and ε_i^* are Gaussian, $\ell(Y,\theta)$ is readily computed by the Kalman filter by recasting the model in terms of a linear state-space system. If the errors ε_i^* and ε_i are *not* normally distributed, equation (30) instead has the interpretation of a *quasi* likelihood function yielding consistent parameter estimates, as long as the second order return moments are correctly specified. In this case, the linear predictions of the state variables are not optimal among all prediction functions, but remain the best *linear* ones, see the Supplementary Appendix, Section 3, for details.

We reiterate that our estimation approach differs fundamentally from the infill asymptotic procedures applied routinely for realized volatility estimation. This reflects our finding of prolonged periods of persistent return serial correlation and the corresponding model feature

involving a temporal feedback mechanism. To capture such temporal effects, the persistence is most naturally viewed as given for a fixed interval length $\Delta > 0$. In this context, ML inference provides a standard $\sqrt{T}$ convergence rate for $T \to \infty$.

As noted in Section 3.1, γ and λ are not separately identifiable in our general model. Instead, we can determine $F_\alpha = F_\alpha(\gamma, \lambda)$ – a specific function of the underlying coefficients. Thus, the parameter vector θ effectively contains only three identifiable coefficients, $\theta = (\alpha, \sigma_*^2, F_\alpha)'$. Of course, in many market microstructure models, the absence of noise endogeneity or temporally persistent pricing errors imply specific parameter restrictions, that ensure identification. Below, we propose mild identification assumptions, which allow us to recover all four primary model parameters, α, σ_*^2, σ_ε^2 (or equivalently λ) and γ, in the general setting.

Our identification approach relies on the hypothesis that traders form their perception of the state of the market from the information flow and the past trade and quote activity. This learning process is updated over time, as new data is obtained. In a noisy environment, the investors' inferred state only changes significantly after a sizable set of signals, indicating a shift in the dynamics, is obtained. As such, any shift in the intensity of the fundamental return shocks must play out over some time, before the majority of the investors learn and adapt. Likewise, a change in the speed with which pricing errors is corrected only manifests itself in the (conditional) mean of the return. Therefore, we stipulate that the (exogenous) fundamental return variance, σ_*^2, and the hard-to-measure speed of error correction, α, vary at a higher frequency than the parameters governed more directly by the majority of the trader population, namely the intensity of the noise variance, σ_ε^2, and the assessment of fundamental news, as captured by γ.

We operationalize the above hypothesis by splitting our sequence of intraday returns for every asset into non-overlapping groups of $K \geq 2$ consecutive intraday intervals, each of length T. We assume that the two (unidentified) parameters, σ_e^2 and γ, are fixed across the $k = 1, \ldots, K$ consecutive intervals, while the two remaining (identified) parameters, α and σ_*^2 are free to vary across the K samples. We denote the sub-vector of fixed parameters across the K intervals, by $\theta_0' = (\sigma_e^2, \gamma)$, and the remaining interval-k-specific parameters by $\theta_k' = (\alpha_k, \sigma_{*\,k}^2)$.

The second-order return moments, computed from the K consecutive intervals, now determine all $2K + 2$ parameters. The coefficients in θ_k are identified through the sample moments in interval k, while the parameters in θ_0 are obtained from moments generated across all K intervals. The principle is akin to the identification through heteroskedasticity, detailed by Rigobon (2003), with temporal variation in the system serving to separate the distinct features. Specifically, the fluctuations in θ_k induce shifts in F_{α_k} for $k = 1, ..., K$, which generate K equations with only two unknowns, σ_e^2 and γ, given that the remaining parameters are identified from other moment conditions. If the system is correctly specified, the parameter vector is uniquely identified, whenever at least two samples generate distinct second-order moments.

The total log-likelihood function based on the K intervals can be written as a sum of log-likelihood functions for the individual samples in equation (30),

$$\sum_{k=1}^{K} \ell\big(Y_k, (\theta_0', \theta_k')'\big),$$

where Y_k denotes the return series for interval k. Assuming the second-order return moments are correctly specified, and the return series across the intervals generate non-identical moments, the parameter vectors $\theta_0, \theta_1, ..., \theta_K$ are consistently estimated through this quasi-ML approach. We rely on this design for empirical illustrations at the end of Section 4.

4 Evidence from Model Estimation

4.1 Data

We estimate the model as outlined in Section 2.1 using high-frequency mid-quote returns of the Nasdaq 100 equity index constituents, obtained from LOBSTER.[11] We split the sample of 100 equities into quintiles based on the average daily number of mid-quote revisions, with Group 1 being the 20% of stocks with most frequent revisions and Group 5 being the 20% with the fewest.[12] Consequently, for a typical stock in Group 5, we usually have fewer quote revisions than intra-day returns, whereas this ratio is much more favorable for stocks in Groups 1 to 4. While we concentrate our discussion on the most liquid groups, we convey a sense of robustness by reporting results for the full spectrum of stocks.

We calculate the mid-quote price as the average of bid and ask prices observed at the top of the order book. To keep the analysis manageable, we restrict our sample to the first three months of 2014 yielding a total of 61 trading days. For each day, we consider the full 6.5 hour trading period, which is split into 39 non-overlapping local intervals of T = 10 min. Consequently, we employ up to a total of 237,900 local intervals. The underlying returns are sampled at the fixed frequency of Δ= 2 seconds. The model is estimated by maximum likelihood via the Kalman filter, as described in Section 3.2.

4.2 Parameter Estimates With Partial Identification

We first focus on parameter estimates without imposing additional identification restrictions, i.e., estimates of α, σ_*^2 and $F_\alpha(\gamma, \lambda)$. From estimates of $F_\alpha(\gamma, \lambda)$ we back out corresponding

[11] The LOBSTER database (https://lobsterdata.com) builds on Nasdaq's historical TotalView-ITCH data. It provides information on all trade and quote activity on Nasdaq at nano-second time stamp precision.

[12] The quintiles include an average of 17,485, 10,365, 8,166, 5,531 and 2,497 daily mid-quote revisions, respectively.

upper bounds, γ_{max} and λ_{max} as given by equations (28) and (29).

Figure 5 displays the empirical distributions of the parameter estimates[13] obtained across all stocks in the first, fourth and fifth quintile (with 20 stocks in each), respectively, sorted according to the intensity of quote-midpoint revisions. All distributions appear unimodal and approximately symmetric, apart from the distinct right skew in the σ_*^2 distribution.

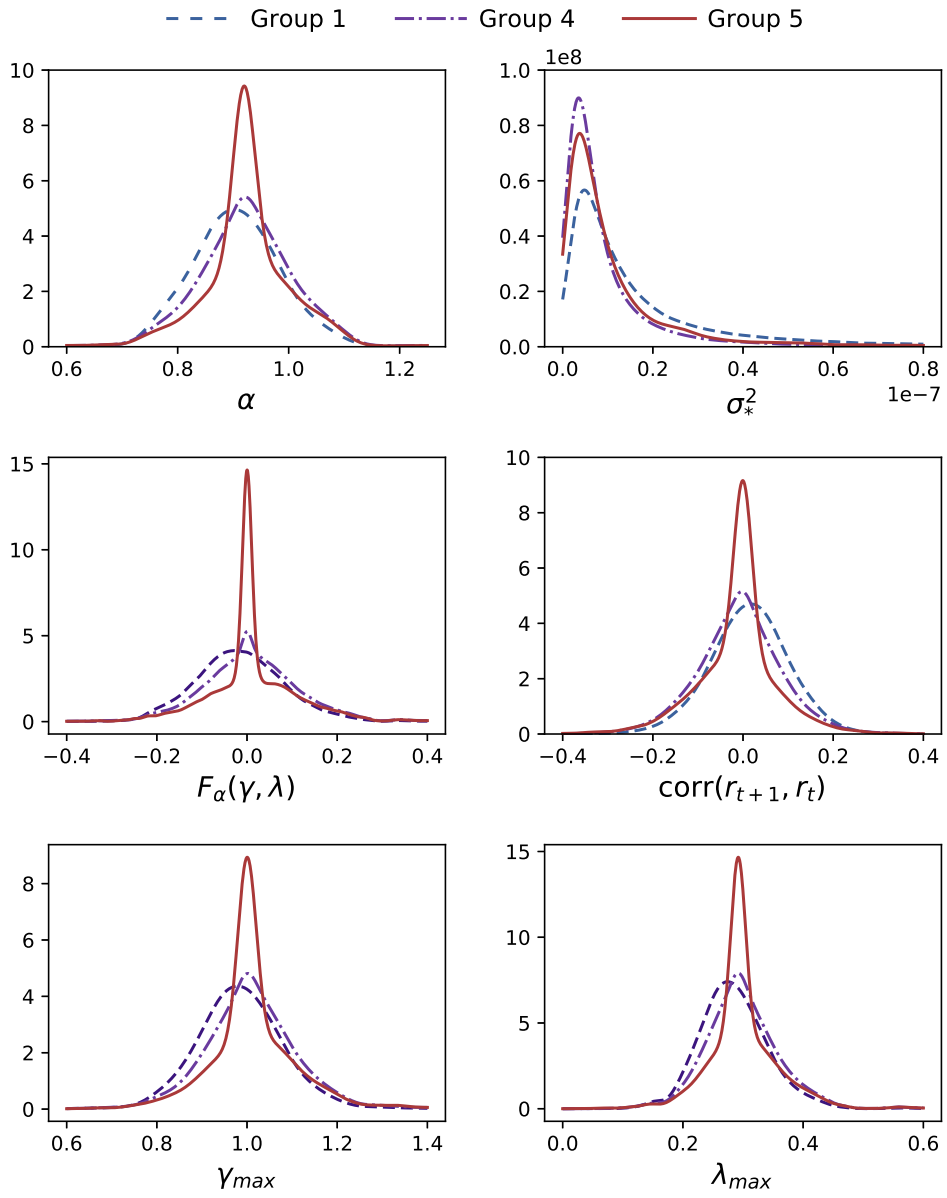

Figure 5: The figure depicts the distribution of estimated parameters from T = 10 min intervals sampled at the frequency of Δ= 2 sec. Results are reported for the stocks in the 1st, 4th and 5th group sorted by the average number of daily mid-quote revisions for the first 61 trading days in 2014. The densities are constructed using Gaussian kernels from parameter estimates obtained across 47,580 local intervals.

[13]These are empirical histograms of parameter estimates smoothed by a standard Gaussian kernel estimator.

In Group 1, the α estimates have a mode around 0.90, corroborating the hypothesis of a feedback effect and mildly sluggish price adjustments. The distribution of γ_{max} is centered on 1, consistent with a first-order return autocorrelation around zero, but with a significant number of positive and negative coefficients (see Table 1 in the Supplementary Appendix). The two regimes imply, respectively, that $\gamma_{max} < 1$ and $\gamma_{max} > 1$, with these values attained only for $\lambda = 0$, as illustrated by the left and right panels of Figure 4. Because the returns, invariably, embody some degree of high-frequency noise, we presume $\lambda > 0$. Hence, the relevant case involves $\gamma < \gamma_{max}$, reflecting a leftward movement along the dashed curves in Figure 4, starting from $(\gamma, \lambda) = (\gamma_{max}, 0)$. We conclude that, for a majority of the local intervals, $\gamma < 1$.

This discussion reflects the fundamental identification issue for microstructure models with latent and endogenous noise. Our model has two separate sources of non-martingale return behavior. Idiosyncratic noise is "corrected" over time, inducing negative return dependence, while incomplete incorporation of fundamental innovations – endogeneity – induces positive dependence, as prices subsequently adjust to the shift in p^* (for $\alpha/2 < \gamma < 1$). Hence, an increase in both the idiosyncratic noise and the initial underreaction to fundamental news have offsetting effects, potentially leaving the return autocorrelation pattern unchanged.

Our results imply that, for most intervals, the return dynamics is affected by endogenous noise. This induces a smoother price path, with return autocorrelation being positive or negative depending on the relative size of the idiosyncratic noise component, λ. Note also that some local intervals featuring negative return autocorrelation may imply $\gamma > 1$, indicating an information environment, where agents are (overly) concerned about asymmetric information and overreact to information signals, generating "overshooting." For such episodes, there is an additional source behind the price reversals, reinforcing the impact of idiosyncratic noise.

Our upper bound estimates for λ_{max} mostly fall below 0.4. This confirms recent findings of a relatively small noise component for large cap stocks. Consistent with the empirical evidence presented in Introduction and in the Supplementary Appendix, point estimates for the Group 1 autocorrelation have a small positive mode and a right skew, indicating regimes with positive return autocorrelation are common and, perhaps, even in the majority.

For the less liquid stocks in Group 4, we find a slightly lower degree of return dependence, a lower level of fundamental volatility, and more rapid error correction. To interpret these findings, it is useful to have actual estimates for the relative degree of idiosyncratic noise, but this is complicated due to the lack of separate identification for λ. We return to this issue below.

In Group 5, the estimates of α and λ_{max} are higher, on average, have smaller dispersion around the mode and are distinctly right-skewed. Hence, less liquid stocks with fewer quote revisions and larger relative tick size generate noisier returns and react more strongly to past pricing errors. Consequently, the model-implied return autocorrelations are predominantly

negative, as also indicated by the slight left skew in the associated empirical distribution. Table 4 of the Supplementary Appendix provides summary statistics for the parameter estimates in each liquidity quintile. The identical qualitative patterns are present in all quintiles and the relevant statistics shift monotonically across the quintiles, corroborating our earlier findings. However, Group 5 seems to constitute an outlier and may be unduly impacted by illiquidity.[14]

The fully identified Information Delay model from Section 2.4.5 or, equivalently, the Amihud-Mendelson model in Section 2.4.4, enables us – at the cost of one additional restriction – to shed light on the trade-off between γ and λ. This model retains both noise endogeneity and error correction, but α now governs the strength of each. The model is fully consistent with our general setup, as the permissible combinations of (γ, λ) always allow for $\gamma = 0$ or, equivalently, $\gamma = \alpha$. In this scenario, the fundamental price innovation is incorporated only partially, and possibly with a delay, into the current market price. The distribution of the estimates for the relative size of the idiosyncratic noise component, λ, in this model is displayed in Figure 6.

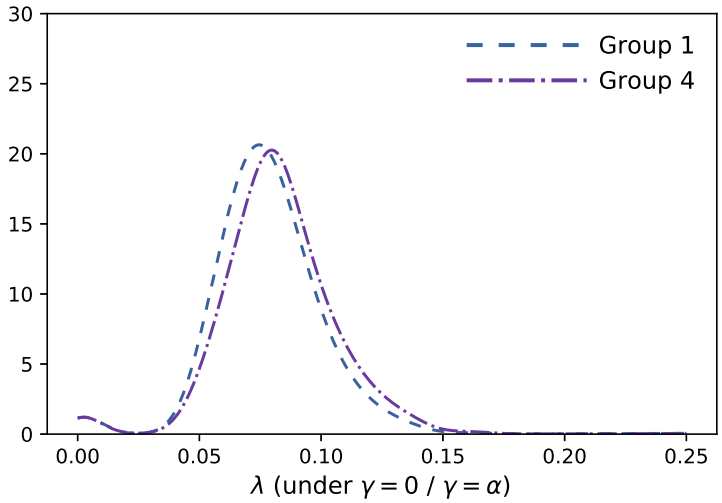

Figure 6: The figure depicts the distribution of the λ estimates (under the restriction $\gamma = 0$, or $\gamma = \alpha$) obtained from T = 10 min intervals sampled at the frequency of Δ= 2 sec. Results are reported for the stocks in the 1st and 4th group sorted by the average number of daily mid-quote revisions for the first 61 trading days in 2014. In total, each density is constructed based on Gaussian kernels from parameter estimates across 47,580 local intervals.

We observe an upward shift in λ for Group 4 versus Group 1. The group 2 and 3 stocks have λ-distributions between these two, reflecting a monotonic pattern and suggesting more active stocks display less idiosyncratic noise. For higher λ, any given price change is more likely to reflect noise, enhancing the willingness to trade against recent price moves. This raises

[14]The results for the initial 10 minutes of trading, reported in Figure 1 of the Supplementary Appendix, corroborates this conjecture. The trading around the open is very active, and the Group 5 stocks are also liquid over this interval. Figure 1 of the Supplementary Appendix confirms that the Group 5 distribution of the parameter estimates are closely aligned with those in the other groups. This confirms that the general estimates for Group 5 likely are heavily impacted by microstructure frictions, and we focus our attention on the other groups in the sequel.

α, pulling the return autocorrelation lower. Finally, the large, active stocks are likely subject to a more intense flow of fundamental information, consistent with the patterns in Figure 5.

These results are all obtained through our model-based translation of specific price-path realizations into an interpretable quantitative representation. Section 3 of the Supplementary Appendix provides a detailed graphical illustration of this process, mapping realized high-frequency price observations into parameter estimates across a number of local intraday intervals.

4.3 Full System Estimation via Identification through Heterogeneity

To corroborate our findings from Section 4.2, we provide estimates using the identification strategy in Section 3.1. We estimate the model parameters jointly from samples covering K = 3 consecutive T = 10 min intervals. The parameters γ and σ_e^2 are assumed constant for the full 30 minutes, while the remaining parameters may differ across each 10 minute interval.

The key identification restriction is that γ and σ_e^2 change more slowly than other parameters, because they reflect ongoing learning on the part of investors. However, the return volatility and trading activity is known to shift rapidly after the market open and prior to the close. This fact invalidates the rationale behind the identification restriction, as investors will be aware of these periodic shifts in the environment. Hence, we exclude 30 minutes after the market opening and 30 minutes before the closure, yielding 11 sequences of 3 consecutive intervals per trading day. We defer a detailed review of the results to Section 5 of the Supplementary Appendix, but summarize a few relevant findings. First, more than 90% of the $\widehat{\gamma}$'s fall between 0 and 1. Recall from Section 2.1 that, for $\gamma \in (0, 1)$, the price innovation only partially accommodates concurrent fundamental signals, ranging from no immediate reaction ($\gamma = 0$) to perfect instantaneous incorporation ($\gamma = 1$). Thus, the estimates indicate that market "underreaction" prevails in most local intervals, implying an incomplete information environment, where signals initially are ascribed to noise. Second, there is a distinct separation of the estimates of γ into two clusters. More than 67% fall in the interval $[-0.1, 0.2]$, while more than 20% of estimates take values in $[0.7, 1]$. In particular, there is no evidence of *statistically significant* negative estimates for $\widehat{\gamma}$.

Third, $\widehat{\lambda}$ is below 0.25 in more than 90% of the local intervals, with a median of 0.069. These estimates are generally consistent with the loci for the partially identified (γ, λ) in Section 4.2, and confirm that the noise-to-signal ratio is moderate, but not zero, as implied, e.g., by the Beveridge-Nelson type identification restriction invoked in Hasbrouck (1993). Figure 7 compares the distribution of $\widehat{\lambda}$ for stocks in Group 1 and 4. It is broadly consistent with the estimates under the restriction $\gamma = 0$ in Figure 6, although it exhibits a degree of bi-modality, with a larger fraction located near zero. Thus, while $\gamma = 0$ constitutes a reasonable approximation in most cases, it is not universally valid. Finally, the distribution associated with Group 4 is again shifted upward, indicating that more actively traded stocks exhibit less noise.

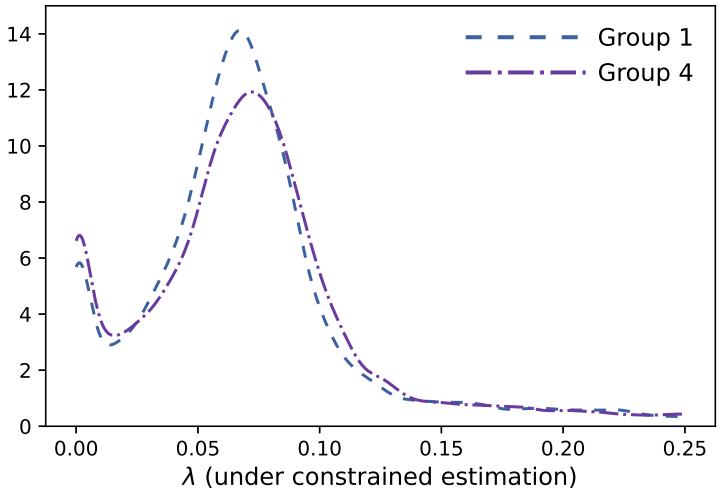

Figure 7: The figure depicts the distribution of estimates for λ under constrained estimation from the sequences of 3 consecutive local intervals (T = 10×3 min) sampled at the frequency of Δ= 2 sec. Results are reported for stocks in the 1st and 4th group sorted by the average number of daily mid-quote revisions for the first 21 trading days in 2014 (excluding 30 minutes after the market open and before the market close). In total, each density is constructed based on Gaussian kernels from parameter estimates across 4,620 local intervals.

4.4 Testing Popular Nested Model Specifications

Testing the restrictions $\alpha = 1$ and $\gamma = 1$ allows us to gauge whether our extension of models for high-frequency asset prices, routinely used in the literature, is warranted statistically.

We first test the null hypothesis $H_0 : \alpha = 1$ against the two-sided alternative, $H_1' : \alpha < 1$ and H_1'': $\alpha > 1$. We employ a Wald test based on QML ("sandwich") estimates of the asymptotic covariance. Table 1 reports the rejection frequencies based on a 5% significance level.

Rejection side	Group 1	Group 2	Group 3	Group 4	Group 5	All stocks
$\alpha < 1$	17.44	16.62	14.64	15.76	34.52	19.80
$\alpha > 1$	0.22	0.30	0.32	0.40	0.25	0.30

Table 1: Percentage of local intervals for which the hypothesis H_0: $\alpha = 1$ is rejected at 5% level by the Wald test (in favor of $\alpha \neq 1$). The rejection rates are computed for model estimates for all stocks and local intervals (T = 10 min, Δ = 2 sec) over the first 61 trading days of 2014. Results are provided both separately for stocks from different liquidity groups and for all stocks together (right column). The number of considered local intervals equals 47,580 for each liquidity group and 237,900, in total, for all considered stocks.

Despite low power due to short time intervals, we reject the null hypothesis for about 20% of all local intervals. Interestingly, while the rejections are equally common across the liquid stocks, the rejection rate for the least liquid ones (Group 5) is considerably higher, reaching almost 35%. Moreover, the rejections are almost exclusively due to point estimates of $\widehat{\alpha} < 1$, suggesting only partial corrections of concurrent pricing errors. Given the comparably high number and

one-sided nature of the rejections in a setting with limited power leads us to conclude that temporal feedback effects are prevalent in the high-frequency return dynamics.

Formal tests of restrictions on γ are possible under the assumptions ensuring full identification. Table 2 provides rejection frequencies of the Wald test for $\gamma = 0$ (upper panel) and $\gamma = 1$ (lower panel). The former restriction is compatible with the Information Delay model from Section 2.4.5, whereas the latter is implied by the "classical" model with i.i.d. noise, reviewed in Section 2.4.2. We find $\gamma = 1$ to be rejected about 63% of the time, indicating a delayed price reaction to fundamental news and the presence of endogenous pricing errors. Hence, at high frequencies, this model appears inadequate. In contrast, the hypothesis $\gamma = 0$ is rejected for less than 30% of the tests, which, to some degree, may be seen as supportive of the Information Delay model. Interestingly, for liquid stocks, the hypothesis $\gamma = 1$ is rejected slightly more often than for the illiquid ones, while the situation reverses for the hypothesis $\gamma = 0$.

	Rejection side	Group 1	Group 2	Group 3	Group 4	Group 5	All stocks
γ vs 0	$\gamma < 0$	1.67	1.80	1.62	1.77	1.90	1.75
	$\gamma > 0$	24.33	24.83	24.96	26.13	28.66	25.78
γ vs 1	$\gamma < 1$	63.16	63.48	62.23	62.03	61.23	62.43
	$\gamma > 1$	0.06	0.24	0.11	0.28	0.28	0.19

Table 2: The table reports the percentage of local intervals for which the hypotheses H_0: $\gamma = 0$ (upper panel) and H_0: $\gamma = 1$ (lower panel) are rejected at 5% level of significance by the Wald test (in favor of $\gamma \neq 0$ and $\gamma \neq 1$, respectively). The rejection rates are computed for estimates of the restricted model over the sequences of 3 consecutive local intervals (T = 10×3 min, Δ = 2 sec) for all stocks over the first 21 trading days of 2014 (excluding 30 minutes after the market opening and before the market closure). Results are provided both separately for stocks from different liquidity groups and for all stocks together (right column). The number of considered sequences of local intervals is equal to 4,620 for each liquidity group and 23,100, in total, for all considered stocks.

The findings of incomplete instantaneous incorporation of fundamental information are also consistent with the indirect evidence against the restriction $\gamma = 1$ obtained in Section 3.1. Moreover, $\gamma < 1$ is a necessary, albeit not sufficient, condition for positive return autocorrelation in this setting. Specifically, the empirical analysis provided in the Supplementary Appendix rejects the hypothesis of zero return correlation for almost all stocks, irrespective of liquidity and sampling frequency. In addition, between 27% (Group 5) and 57% (Group 1) of the local intervals display positive serial correlation across a variety of lower sampling frequencies, reinforcing the conclusion that the evidence favors the hypothesis of $\gamma < 1$ across a large fraction of the local intervals for all of the stocks in our sample.

4.5 Parameter Variation and Time-of-Day Effects

Figure 8 depicts the average time-of-day estimates for parameters and implied model features across all stocks in Groups 1 and 4 within the partially identified setting, also explored in Section 4.2. Qualitatively identical patterns are present for the other liquidity quintile groups.

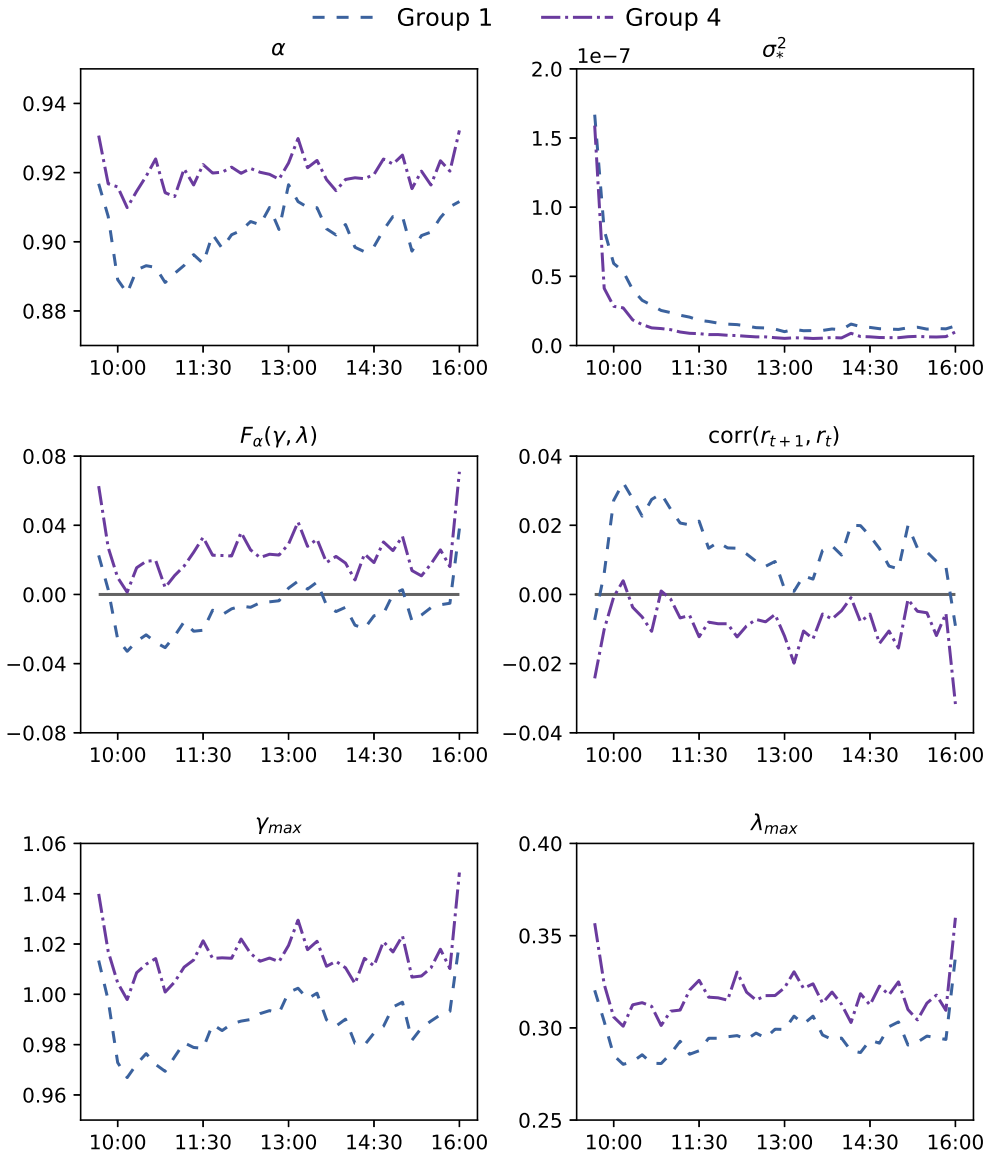

Figure 8: Intraday plots of average model parameters estimated over the period January - March, 2014 (61 trading days in total). Results are reported for the Nasdaq stocks from the 1st and 4th groups sorted by the average number of daily mid-quote revisions (with 20 stocks in each group) for Δ= 2 sec on intraday intervals T = 10 min.

Several of the panels reveal a pronounced intraday pattern. In the top right panel, the $\widehat{\sigma}_*^2$ values form a sharp L-shape, with high volatility in the morning followed by a quick decay and a largely flat pattern for the rest of the trading day. The picture is similar for both groups,

although the decay is more gradual in Group 1. The differences are even more evident for the return serial correlation. Both display sharply lower autocorrelation at the open and close, while the level across the middle of the day resembles a mild U-shape. At the same time, the Group 1 values are consistently above those for Group 4, with the former being positive throughout except at the open and close, while the latter are almost uniformly negative.

These features also impact the remaining parameters. For α, we observe an elevation at the open and close, combined with a mild reverse U-shape across the main phase of the trading session. The bottom two panels display the maximum γ and λ values consistent with the fully identified parameters. They merely provide upper bounds and cannot be attained simultaneously. Specifically, γ_{max} applies for $\lambda = 0$, while λ_{max} is achieved for $\gamma = \alpha/2$. Since we, unquestionable, must have $\lambda > 0$ to capture idiosyncratic noise stemming from market frictions, the average γ must be below unity almost uniformly for Group 1 and, likewise, most of the time in Group 4. Hence, endogenous noise is ubiquitous. These considerations aside, the spike in the upper bound for both parameters at the boundary of the trading day are noteworthy, pointing again to an unusual market environment at those times.

To shed additional light on the asymmetry between fundamental and idiosyncratic volatility, we depict the intraday pattern for average realized volatility and trading volume for Group 1 and 4 in Figure 9. After the few opening intervals, Group 1 uniformly displays higher volatility. Moreover, the qualitative pattern for realized volatility aligns well with the one for fundamental volatility. On the contrary, trading volume spikes dramatically at the end of trading, mimicking the late day upswing of the $F_\alpha(\gamma, \lambda)$ coefficient in Figure 8. To further illustrate the likely path of idiosyncratic volatility, we portray the intraday evolution of λ in the Amihud-Mendelson model, which retains all main structural features of our general model. Figure 10 documents a sharp increase in the relative size of the idiosyncratic volatility at the market close.

How do we interpret this evidence? The most striking feature is the huge asymmetry between the open and close for fundamental volatility, coupled with the near symmetric behavior of all other series. The trading volume is highly elevated both at the open and close, but our model associates only the opening with the assimilation of an unusual amount of fundamental information. The results suggest this raises the speed by which pricing errors are corrected (α), and perhaps also the instantaneous response of market prices to new information (γ) as well as the relative amount of the idiosyncratic noise (λ). The high level of fundamental volatility at the start of trading is consistent with active price discovery associated with information processing of overnight news and orders. Given the (anticipated) rapid arrival rate of fundamental price signals following the open, market makers commit less capital to intermediation, so the bid-ask spread widens, the order book thins out, and the sensitivity to every piece of market news is magnified. These features enhance the role of idiosyncratic noise in the price discovery process.

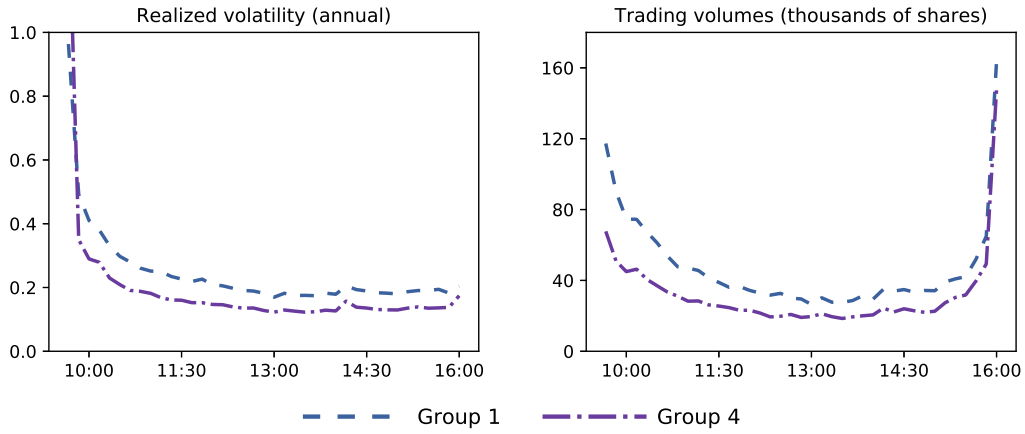

Figure 9: Intraday plots of average realized variance (left panel) and trading volumes (right panel) computed over the period January - March, 2014 (61 trading days in total). Results are reported for the Nasdaq stocks from the 1st and 4th groups sorted by the average number of daily mid-quote revisions (with 20 stocks in each group) for Δ= 2 sec on intraday intervals T = 10 min.

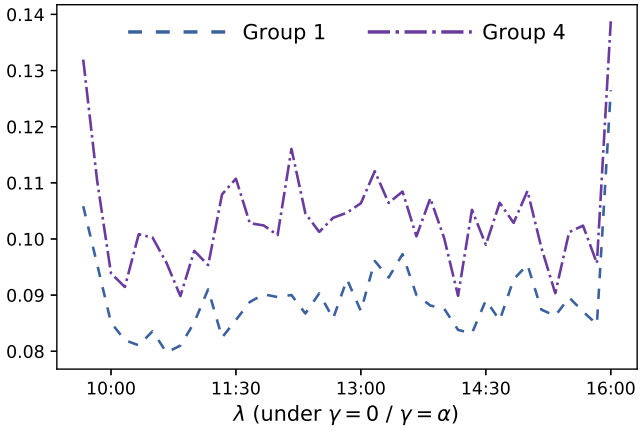

Figure 10: Intraday plots of average estimates of parameter λ (under the restriction $\gamma = 0$, or $\gamma = \alpha$) over the period January - March, 2014 (61 trading days in total). Results are reported for the Nasdaq stocks from the 1st and 4th groups sorted by the average number of daily mid-quote revisions (with 20 stocks in each group) for Δ= 2 sec on intraday intervals T = 10 min.

Finally, understanding that the efficient price is shifting quickly and noise is relatively prevalent, traders expend informational resources to adjust rapidly to incoming signals, so α rises. In all cases, we expect the return autocorrelation to drop, as we observe more immediate incorporation of fundamental news, a larger presence of idiosyncratic noise and less price smoothing.

The period prior to the close differs in that customer orders must be executed before trading terminates to avoid exposure to price uncertainty overnight. Likewise, popular incentive schemes for trade execution, such as the "volume-weighted average price" (VWAP) criterion, motivate risk-averse intermediaries to transact towards the close, as they trade closer to the benchmark they are measured against. In contrast, executing trades early carries exposure to subsequent price changes impacting the VWAP. Hence, there are also non-informational rationales for an

afternoon elevation in the trading intensity, enhancing market depth. Thus, a rising, but largely uninformative, order flow elevates the responsiveness of informed market participants to price signals, generating quicker price corrections, yet also additional idiosyncratic noise.

In summary, we find evidence that liquid market conditions early and late in the trading day foster efficient processing of fundamental news. At such times, the main component behind temporary misalignments in prices is microstructural noise, consistent with negative return serial correlation. In contrast, when the information and order flow are slower, the resources devoted to price discovery shrink, i.e., α drops. This effect enhances price sluggishness, generating conditions favoring momentum in the price formation during the middle of the trading day.

The Supplementary Appendix provides summary statistics for (unconstrained) parameter estimates over the first and last 10-minute trading interval, 9:30 to 9:40, and 15:50 to 16:00. During these periods, the return dynamics deviates substantially from the rest of the trading day. Furthermore, summary statistics across all intraday intervals, excluding the first and last 20 minutes of trading, confirm that evidence obtained across the full trading day is representative of the general market conditions and not unduly distorted by the dynamics at the open and close.

4.6 Model-Implied versus Realized Return Volatility

For robustness, we compare the relationship between our model-implied estimates of fundamental variance and realized volatility (RV) measures obtained at distinct frequencies. This is helpful in gauging whether our 2-second sampling and stylized model generate biased measures. The top panel of Figure 11 displays the mean estimate of σ_*^2 along with corresponding averages of daily RV obtained from sampling at the 2-second, 2- and 10-minute frequencies as well as squared open-to-close returns. Aggregating over the full sample and all stocks, we obtain presumably accurate (average) volatility estimates from 10-minute sampling and squared open-to-close returns, while a bias may be present in the standard RV measure based on 2-second and 2-minute sampling. The logic is similar to that of volatility signature plots, where the mean level of volatility may be approximated by averaging (nearly) unbiased, but noisy, estimates.

To interpret the results, recall that negative high-frequency return autocorrelation – akin to the bid-ask bounce effect – tends to generate an upward bias in the RV estimate, while positive return dependence – by smoothing out rapid price moves – tends to induce a negative bias in the empirical RV measure. Hence, the fact that the average model-implied RV in the top panel of Figure 11 is slightly above the 2-sec RV for Group 1 and below it for Group 5 reflects the model-based bias correction induced by the predominance of positive return autocorrelations in Group 1 and the reverse in Group 5. It is also reassuring that the model-implied daily RV is very close to the mean estimates from the 10-min as well as the open-to-close based RV measures for all liquidity categories. Overall, the main impression is that these measures, effectively, are

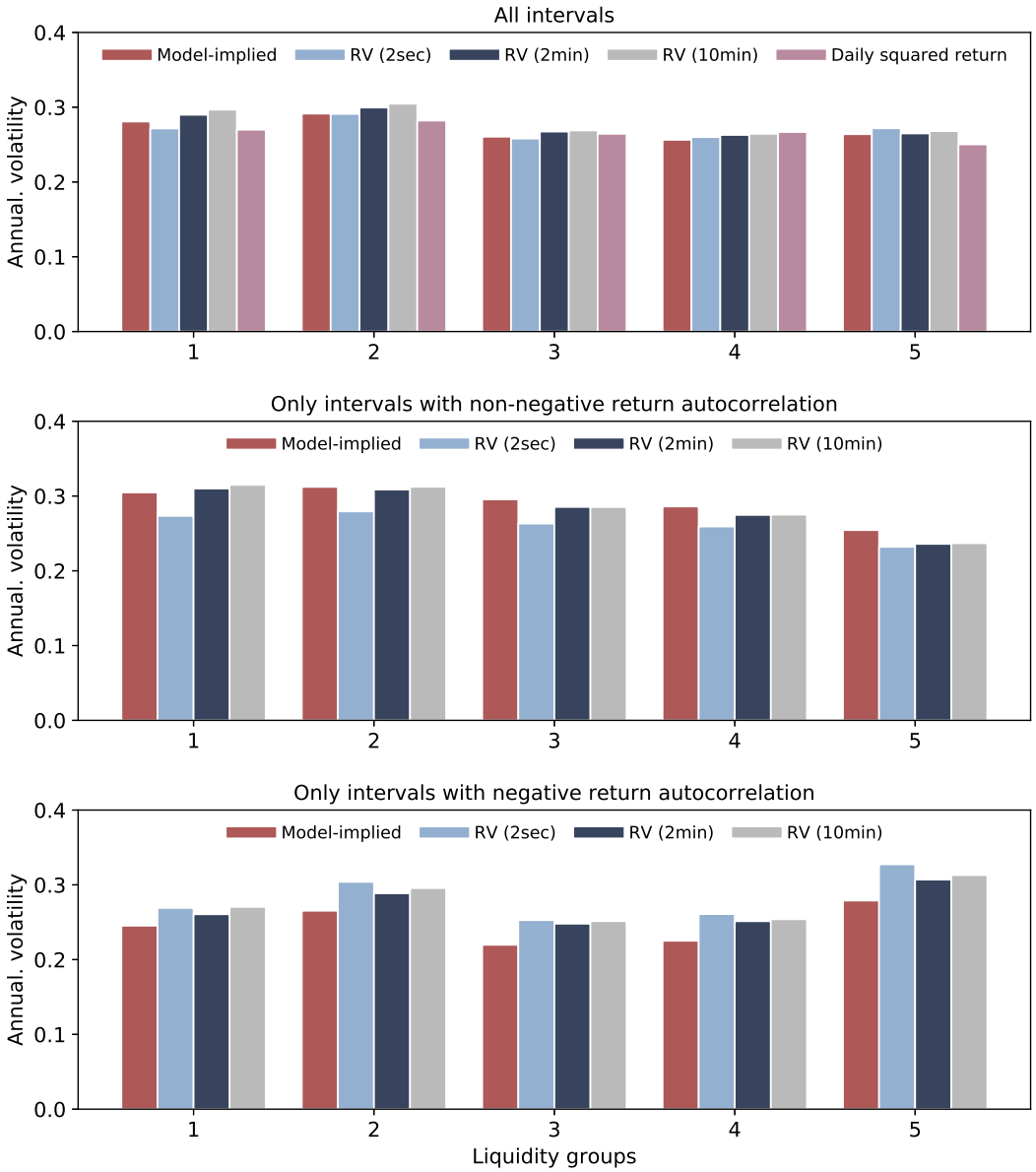

Figure 11: The barplots show the average estimates of the return variance (represented in annualized volatility units) computed by the model-implied and several realized variance measures. The model-implied measures are estimates of the fundamental variance parameter σ_*^2 obtained from T = 10 min intervals sampled at frequency Δ= 2 sec. The realized variance are computed at frequencies 2 sec, 2 min and 10 min. In addition, we consider the squared open-to-close return as a proxy for the daily return variance. All variance measures are averaged over the entire period (the first 61 trading days in 2014) and over the stocks of a given liquidity group, where groups are sorted by the average number of daily mid-quote revisions (20 stocks per group).

equivalent and do not indicate biases, on average, even for the 2-second RV estimates.

The two other panels of Figure 11 illustrate the model-implied bias correction more explicitly by averaging the RV measures only for the local intervals with positive, respectively only negative, sample return autocorrelations. For the positive return dependencies in the middle panel, the model-implied RV largely aligns with that of the 2-min and 10-min RV measure for

the liquid Groups 1 and 2. For the less liquid categories, the model-based estimates "adjust" the RV measures, so that it is more in line with the volatility in Groups 1 and 2. Since there is no fundamental reason why volatility should be particularly low among less liquid stocks in local intervals with positive return autocorrelation, this adjustment appears appropriate. Likewise, for the local intervals displaying negative return correlation in the bottom panel, the model-based estimates mitigate the discrepancies in the average RV measures across the liquidity groups. In particular, there is a large downward correction for Group 5. Since wide spreads and bid-ask bouncing are particularly prevalent in illiquid markets, it is sensible to expect the RV measures of volatility to be artificially inflated for the stocks with relatively few quote revisions.

To summarize, our microstructure model estimated over short intraday intervals produces return variation estimates consistent with those obtained from standard RV measures. However, they also deviate systematically, whenever there is a notable autocorrelation in the return series at the highest frequencies. This corroborates the hypothesis, that MMN effects generate biases in RV estimates of varying sign. For longer periods, like a full trading day, such effects are mitigated by cancellations induced by the randomly shifting sign of the bias, yet it remains non-trivial over shorter periods, and it manifests itself significantly for days, where the predominant MMN features induce persistent spells of return serial dependence in a given direction. As such, we bridge the MMN and RV literatures: the main microstructure theories imply a form for noise endogeneity, that violates the standard infill asymptotic RV approach, while the RV measures are robust to time-varying return heteroskedasticity, that is absent in microstructure models.

5 Concluding Remarks

MMN leaves a distinct imprint on high-frequency returns, manifesting itself through non-trivial return autocorrelations for individual stocks which (i) are significant for local intervals, and (ii) display persistence, yet also undergo rapidly alternating signs. The fact that a large proportion of local intervals displays nontrivial positive return autocorrelation and induces a negative bias in the associated RV estimates is a new finding, inconsistent with the common assumption of MMN increments being uncorrelated with the efficient price process. Notably, endogeneity is implied by many market microstructure models, e.g., Glosten and Milgrom (1985), suggesting that price discovery and noise innovations are intrinsically related, and Kyle (1985) or Vives (1995), who point towards longer-lasting pricing errors induced by strategic trading and learning.

Consequently, the portrayal of the price formation process in the market microstructure literature is at odds with typical high-frequency volatility estimation procedures. Specifically, RV measures usually rely on "semimartingale-plus-noise" models, that preclude persistent structural interaction between efficient price innovations and noise. Instead, the key assumption is that the

noise component becomes dominant, and thus locally identifiable by nonparametric means, as the sampling interval shrinks to zero. This defies the idea of significant longer-run dependencies in the error process itself and its interaction with the fundamental price innovations.

We seek to bridge this gap by extending the martingale-plus-noise model so it aligns better with the market microstructure literature and the empirical footprint left by MMN through the high-frequency return autocorrelogram. Our modification introduces noise endogeneity and an error correction mechanism, capturing partial adjustment dynamics along the lines of Amihud and Mendelson (1987) although, contrary to their model, we separate these effects so they serve as complementary sources of local mispricing. Our parsimonious model captures salient features of high-frequency return dynamics, as the interplay of the model's four parameters - the volatility of the fundamental martingale process, the noise variance, the strength of temporal error correction, and the endogeneity between the efficient price innovations and noise - triggers local regimes of "momentum" and "contrarian" trading. Moreover, it helps identify the complementary roles of the temporal error correction component and the endogenous pricing errors in determining the sign, strength, and longevity of the return autocorrelation.

We hasten to add that, unquestionable, the realized volatility approach has drastically improved our ability to monitor and model the return variation process. The point is that *daily* realized volatility measures are largely unaffected by the local biases, as long as the positive and negative autocorrelation regimes are evenly balanced across the trading day, and thus cancel through intertemporal averaging. Nonetheless, it does imply that local realized (spot) volatility estimates are subject to potentially significant biases. Moreover, trading days for which the autocorrelation pattern is highly one-sided will be associated with RV measures, that are subject to a substantial measurement error. A few concurrent papers deal with extreme manifestations of this phenomenon, in the sense that they identify *individual* local episodes of highly significant semimartingale violations in the observed price process, see, e.g., Andersen et al. (2020), or periods of explosive drifts, see, e.g., Christensen et al. (2020). An earlier piece noting the need for filtering episodes featuring unusually strong positive return serial correlation is Barndorff-Nielsen et al. (2008a), who classified certain such episodes as "gradual jumps."

We differ from these studies by documenting that local violations of the semimartingale hypothesis for (observed) intraday return series are common, yet fleeting. We link this finding to the arrival of economic news in the presence of incomplete and asymmetric information, ongoing learning, noise, and trading costs. Specifically, using mid-quote returns from Nasdaq, we find that a flexible price discovery process in line with our information delay model is useful in capturing the high-frequency return dynamics over local intervals. We find that the pricing error is *endogenous* and, consequently, the classic martingale-plus-exogenous-noise model is often inappropriate. Moreover, the incorporation of a sluggish price adjustment is critical for

modeling the dynamics of high-frequency returns. These implications questions the validity of identification assumptions employed in studies, where the statistical properties of MMN are derived via nonparametric asymptotic techniques, assuming the noise is orthogonal to the efficient price. Hence, the evidence for short-run positive return serial dependence in MMN processes, obtained in recent work,[15] might alternatively be viewed through a different lens involving time-varying levels of noise endogeneity.

From an econometric perspective, the primary contributions are as follows. First, compared to the usual high-frequency econometric approach, our model provides a more structural treatment of market microstructure noise capturing sources of endogeneity and temporal dependence. The resulting insights can be exploited to construct conceptually novel model-based estimators of volatility that adapt to this noise structure. Second, we complement long-lasting discussions around the identifiability of parameters in a martingale-plus-noise setting. Within our set-up, we show that the signal-to-noise ratio is only partially identified if the covariance between the MMN and efficient price components is unconstrained, while the fundamental volatility is always fully identified. We further show how full identification may be obtained by exploiting heterogenous variation in the fundamental versus learning-based parameters. Third, our results provide strong evidence for locally shifting return dynamics. These findings have implications for local volatility estimation and should, in periods with unusual return dynamics, help mitigate biases associated with the estimation of volatility over longer daily or weekly horizons. In summary, we anticipate that the combination of market microstructure inspired modeling and explicit recognition of the ever-evolving market dynamics will open up interesting avenues for future work on the construction of alternative local and daily volatility estimators and a deeper understanding of the impact of noise in financial prices.

References

AIT-SAHALIA, Y. AND J. JACOD (2014): *High-Frequency Financial Econometrics*, Princeton University Press.

AIT-SAHALIA, Y., P. MYKLAND, AND L. ZHANG (2005): "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, 351–416.

——— (2006): "Comment," *Journal of Business & Economic Statistics*, 162–167.

AMIHUD, Y. AND H. MENDELSON (1987): "Trading Mechanisms and Stock Returns: An Empirical Investigation," *Journal of Finance*, 42 (3), 533–553.

ANDERSEN, T. G. AND T. BOLLERSLEV (1998): "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885–905.

[15] See, for example, Li and Linton (2020) and Da and Xiu (2020).

——— (2018a): "Introduction," in *Volatility, Volume I*, ed. by T. G. Andersen and T. Bollerslev, Cheltenham, U.K.; Northampton, MA, U.S.: Edward Elgar Publishing, xiii–xli.

——— (2018b): *Volatility, Volume I and II*, Edward Elgar Publishing.

ANDERSEN, T. G., T. BOLLERSLEV, F. DIEBOLD, AND P. LABYS (2000): "Great Realisations," *Risk Magazine*, 18, 105–108.

ANDERSEN, T. G., Y. LI, V. TODOROV, AND B. ZHOU (2020): "Volatility Measurement with Pockets of Extreme Return Persistence," *Journal of Econometrics*, forthcoming.

BANDI, F. AND J. RUSSELL (2006): "Separating Microstructure Noise from Volatility," *Journal of Financial Economics*, 79, 655–692.

——— (2008): "Microstructure Noise, Realized Variance, and Optimal Sampling," *Review of Economic Studies*, 75, 339–369.

BARNDORFF-NIELSEN, O. E., P. HANSEN, A. LUNDE, AND N. SHEPHARD (2008a): "Realised Kernels in Practice: Trades and Quotes," *Econometrics Journal*, 4, 1–32.

BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2008b): "Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise," *Econometrica*, 76, 1481–1536.

BEVERIDGE, S. AND C. NELSON (1981): "A New Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to the Measurement of the 'Business Cycle'," *Journal of Monetary Economics*, 7, 151–174.

BIBINGER, M., N. HAUTSCH, P. MALEC, AND M. REISS (2014): "Estimating the quadratic covariation matrix from noisy observations: local method of moments and efficiency," *Annals of Statistics*, 42, 1312–1364.

CHAKER, S. (2017): "On high frequency estimation of the frictionless price: The use of observed liquidity variables," *Journal of Econometrics*, 201, 127–143.

CHAN, K. (1993): "Imperfect Information and Cross-Autocorrelation Among Stock Prices," *Journal of Finance*, 48, 1211–1230.

CHRISTENSEN, K., R. OOMEN, AND R. RENÒ (2020): "The drift burst hypothesis," *Journal of Econometrics*, forthcoming.

DA, R. AND D. XIU (2020): "When Moving Average Models Meet High-Frequency Data: Uniform Inference on Volatility," Working Paper, University of Chicago.

DIEBOLD, F. AND G. STRASSER (2013): "On the Correlation Structure of Microstructure Noise: A Financial Economic Approach," *Review of Economic Studies*, 80, 1304–1337.

GLOSTEN, L. AND P. MILGROM (1985): "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Markets Economics*, 14, 71–100.

HANSEN, P. R. AND A. LUNDE (2006): "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, 24, 127–161.

HARVEY, A. C. (1989): *Forecasting. Structural Time Series Models and the Kalman Filter.*, Cambridge University Press.

HASBROUCK, J. (1993): "Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement," *Review of Financial Studies*, 6, 191–212.

HASBROUCK, J. AND T. HO (1987): "Order Arrival, Quote Behavior, and the Return-Generating Process," *Journal of Finance*, 42, 1035–1048.

HOLDEN, C. AND A. SUBRAHMANYAM (1992): "Long-Lived Private Information and Imperfect Competition," *Journal of Finance*, 47, 247–270.

JACOD, J., Y. LI, P. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2009): "Mirostructure noise in the continuous case: the pre-averaging approach," *Stochastic Processes and Their Applications*, 26, 2803–2831.

JACOD, J., Y. LI, AND X. ZHENG (2017): "Statistical Properties of Microstructure Noise," *Econometrica*, 85, 1133–1174.

KALNINA, I. AND O. LINTON (2008): "Estimating Quadratic Variation Consistently in the Presence of Endogenous and Diurnal Measurement Error," *Journal of Econometrics*, 147, 47–59.

KYLE, A. S. (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 53 (6), 1315–1335.

LI, Y., S. XIE, AND X. ZHENG (2016): "Efficient Estimation of Integrated Volatility Incorporating Trading Information," *Journal of Econometrics*, 195, 33–50.

LI, Z. (2020): "Robust Measures of Microstructure Noise," Working Paper, University of Cambridge.

LI, Z., R. LAEVEN, AND M. VELLEKOOP (2020): "Dependent microstructure noise and integrated volatility estimation from high-frequency data," *Journal of Econometrics*, 215, 536–558.

LI, Z. AND O. LINTON (2020): "A ReMeDI for Microstructure Noise," Working Paper, University of Cambridge.

MADHAVAN, A., M. RICHARDSON, AND M. ROOMANS (1997): "Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks," *Review of Financial Studies*, 10, 1035–1064.

MYKLAND, P. AND L. ZHANG (2009): "Inference for Continuous Semimartingales Observed at High Frequency," *Econometrica*, 77, 1403–1445.

RIGOBON, R. (2003): "Identification Through Heteroskedasticity," *Review of Economics and Statistics*, 85, 777–792.

ROLL, R. (1984): "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market," *Journal of Finance*, 1127–1140.

SHEPPARD, K. (2013): "Measuring Market Speed," Working Paper, University of Oxford.

VIVES, X. (1995): "The Speed of Information Revelation in a Financial Market Mechanism," *Journal of Economic Theory*, 67, 187–204.

WATSON, M. W. (1986): "Univariate Detrending Methods with Stochastic Trends," *Journal of Monetary Economics*, 18, 49–75.

ZHANG, L., P. MYKLAND, AND Y. AIT-SAHALIA (2005): "A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.

ZHOU, B. (1996): "High-Frequency Data and Volatility in Foreign-Exchange Rates," *Journal of Business and Economic Statistics*, 45–52.