A Pure-Jump Transaction-Level Price Model Yielding Cointegration, Leverage, and Nonsynchronous Trading Effects

Hurvich, Clifford and Wang, Yi

Stern School of Business, New York University, Stern School of Business, New York University

5 January 2009

A Pure-Jump Transaction-Level Price Model Yielding Cointegration, Leverage, and Nonsynchronous Trading Effects

Clifford M. Hurvich* Yi Wang *

January 5, 2009

Abstract

We propose a new transaction-level bivariate log-price model, which yields fractional or standard cointegration. The model provides a link between market microstructure and lower-frequency observations. The two ingredients of our model are a Long Memory Stochastic Duration process for the waiting times $\{\tau_k\}$ between trades, and a pair of stationary noise processes ($\{e_k\}$ and $\{\eta_k\}$) which determine the jump sizes in the pure-jump log-price process. Our model includes feedback between the disturbances of the two log-price series at the transaction level, which induces standard or fractional cointegration for any fixed sampling interval Δt. We prove that the cointegrating parameter can be consistently estimated by the ordinary least-squares estimator, and obtain a lower bound on the rate of convergence. We propose transaction-level method-of-moments estimators of the other parameters in our model and discuss the consistency of these estimators. We then use simulations to argue that suitably-modified versions of our model are able to capture a variety of additional properties and stylized facts, including leverage, and portfolio return autocorrelation due to nonsynchronous trading. The ability of the model to capture these effects stems in most cases from the fact that the model treats the (stochastic) intertrade durations in a fully endogenous way.

KEYWORDS: Tick Time; Long Memory Stochastic Duration; Information Share.

*Stern School of Business, New York University.
This is an expanded version of the paper, "A Pure-Jump Transaction-Level Price Model Yielding Cointegration," to appear in *The Journal of Business and Economic Statistics*. The authors thank Yakov Amihud, Torben Andersen, Jushan Bai, Menachem Brenner, Xiaohong Chen, Rohit Deo, Rob Engle, Steve Figlewski, Joel Hasbrouck, Jim Stock, Mark Watson and the participants at the 2006 NBER-NSF Time Series conference in Montreal, and at the Econometrics Seminar, Department of Economics, New York University, for helpful comments and suggestions. The authors also wish to thank two anonymous referees and the associate editor for their help in improving the paper.

I Introduction

In this paper, we propose a transaction-level, pure-jump model for a bivariate price series, in which the intertrade durations are stochastic, and enter into the model in a fully endogenous way. The model is flexible, and able to capture a variety of stylized facts, including standard or fractional cointegration, persistence in durations, volatility clustering, leverage (*i.e.*, a negative correlation between current returns and future volatility), and nonsynchronous trading effects. In our model, all of these features observed at equally-spaced intervals of time are derived from transaction-level properties. Thus, the model provides a link between market microstructure and lower-frequency observations. This paper focuses on the cointegration aspects of the model, presenting theoretical, simulation and empirical analysis.

Cointegration is a well-known phenomenon that has received considerable attention in Economics and Econometrics. Under both standard and fractional cointegration, there is a contemporaneous linear combination of two or more time series which is less persistent than the individual series. Under standard cointegration, the memory parameter is reduced from 1 to 0, while under fractional cointegration the level of reduction need not be an integer. Indeed, in the seminal paper of Engle and Granger (1987), both standard and fractional cointegration were allowed for, although the literature has developed separately for the two cases. Important contributions to the representation, estimation and testing of standard cointegration models include Stock and Watson (1988), Johansen (1988, 1991), and Phillips (1991a). Literature addressing the corresponding problems in fractional cointegration includes Dueker and Startz (1998), Marinucci and Robinson (2001), Robinson and Marinucci (2001), Robinson and Yajima (2002), Robinson and Hualde (2003), Velasco (2003), Velasco and Marmol (2004), Chen and Hurvich (2003a, 2003b, 2006).

A limitation of most existing models for cointegration is that they are based on a particular fixed sampling interval Δt, *e.g.*, one day, one month, *etc.* and therefore do not reflect the dynamics at all levels of aggregation. Indeed, Engle and Granger (1987) assumed a fixed sampling interval. It is also possible to build models for cointegration using diffusion-type continuous-time models such as ordinary or fractional Brownian motion (see Phillips 1991b, Comte 1998, Comte and Renault 1996, 1998), but such models would fail to capture the pure-jump nature of observed asset-price processes.

In this paper, we propose a pure-jump model for a bivariate log-price series such that any discretization of the process to an equally-spaced sampling grid with sampling interval Δt produces fractional or standard cointegration, *i.e.*, there exists a contemporaneous linear combination of the two log-price series which has a smaller memory parameter than the two individual series. A key ingredient in our model is a microstructure noise contribution $\{\eta_k\}$ to the log prices. In the *weak fractional cointegration* case, this noise series is assumed to have memory parameter $d_\eta \in (-\frac{1}{2}, 0)$, in the *strong fractional cointegration* case $d_\eta \in (-1, -\frac{1}{2})$ while in the *standard cointegration* case $d_\eta = -1$. In all three cases, the reduction of the memory parameter is $-d_\eta$. Due to the presence of the microstructure noise term, the discretized log-price series are not Martingales, and the corresponding return series are not linear in either an *i.i.d.* sequence, a Martingale-difference sequence, or a strong-mixing sequence. This is in sharp contrast to existing discrete-time models for cointegration, most of which assume at least that the series has a linear representation with respect to a strong-mixing sequence.

The discretely-sampled returns (*i.e.*, the increments in the log-price series) in our model are not Martingale differences, due to the microstructure noise term. Instead, for small values of Δt they may exhibit noticeable autocorrelations, as observed also in actual returns over short time intervals. Nevertheless, the returns behave asymptotically like Martingale differences as the sampling interval Δt is increased, in the sense that the lag-k autocorrelation tends to zero as Δt tends to ∞ for any fixed k. Again, this is consistent with the near-uncorrelatedness observed in actual returns measured over long time intervals.

The memory parameter of the log prices in our model is 1, in the sense that the variance of the log price increases linearly in t, asymptotically as $t \to \infty$. By contrast, the memory parameter of the appropriate contemporaneous linear combination of the two log-price series is reduced to $(1 + d_\eta) < 1$, thereby establishing the existence of cointegration in our model.

In order to derive the results described above, we will make use of the general theory of point processes, and we will also rely heavily on the theory developed in Deo, Hurvich, Soulier and Wang (2007) for the counting process $N(t)$ induced by a long-memory duration process.

In Section II, we exhibit our pure-jump model for the bivariate log-price series. Since the two series need not have all of their transactions at the same time points (due to nonsynchronous trading), it is not possible to induce cointegration in the traditional way, *i.e.*, by directly imposing in clock time (calendar time) an additive common component for the two series, with a memory parameter equal to 1. Instead, the common component is induced indirectly, and incompletely, by means of a feedback mechanism in transaction time between current log-price disturbances of one asset and past log-price disturbances of the other. This feedback mechanism also induces certain end-effect terms, which we explicitly display and handle in our theoretical derivations using the theory of point processes. Subsection II A provides economic justification of the model, as well as a transaction-level definition of the information share of a market. The subsection also presents some preliminary data analysis results affirming the potential usefulness of certain flexibilities in the model.

In Section III, we give conditions on the microstructure noise process for both fractional and standard cointegration. These conditions are satisfied by a variety of standard time series models. In Section IV, we present the properties of the log-price series implied by our model. In particular, we show that the log price behaves asymptotically like a Martingale as t is increased, and the discretely-sampled returns behave asymptotically like Martingale differences as the sampling interval Δt is increased. In Section V, we establish that our model possesses cointegration, by showing that the cointegrating error has memory parameter $(1 + d_\eta)$. We present three separate theorems for the weak and strong fractional cointegration and standard cointegration cases respectively. In Section VI, we show that the ordinary least squares (OLS) estimator of the cointegrating parameter θ is consistent, and obtain a lower bound on its rate of convergence. In Section VII, we propose an alternative cointegrating parameter estimator based on the tick-level price series. In Section VIII, we propose a method-of-moments estimator for the tick-level model parameters (except the cointegrating parameter θ). The method is based on the observed tick-level returns. In Section IX, we propose a specification test for the transaction-level price model. In Section X, we present simulation results on the OLS estimator of θ, the tick-level cointegrating parameter estimator $\tilde{\theta}$, the method of moments estimator and the proposed specification test. In Section XI, we present a data analysis of buy and sell prices of a single stock (Tiffany, TIF), providing evidence of strong fractional cointegration. The cointegrating parameter is estimated by both OLS regression and the alternative tick-level method proposed in Section VII. The proposed specification test is implemented on the data. Interesting results are observed that are consistent with the existing literature about price discovery process in a market-dealer market. We then consider the information content of buy trades versus sell trades in different market environments. In Section XII, we demonstrate, largely through simulations, that modified versions of our model can reproduce two additional important stylized facts: leverage, and portfolio return autocorrelation due to nonsynchronous trading. We trace these clock-time properties to their tick-time source. In Section XIII, we provide some remarks and discuss possible further generalizations of our model and related future work. In Section XIV we present details on the method-of-moments estimator, establish its consistency, and propose an alternative estimator. Section XV presents proofs.

II A Pure-Jump Model For Log Prices

Before describing our model, we provide some background on transaction-level modeling. Currently, a wealth of transaction-level price data is available, and for such data the (observed) price remains constant between transactions. If there is a diffusion component underlying the price, it is not directly observable. Pure-jump models for prices thus provide a potentially appealing alternative to diffusion-type models. The compound Poisson process proposed in Press (1967) is a pure-jump model for the logarithmic price series, under which innovations to the log price are *i.i.d.*, and these innovations are introduced at random time points, determined by a Poisson process. The model was generalized by Oomen (2006), who introduced an additional innovation term to capture market microstructure.

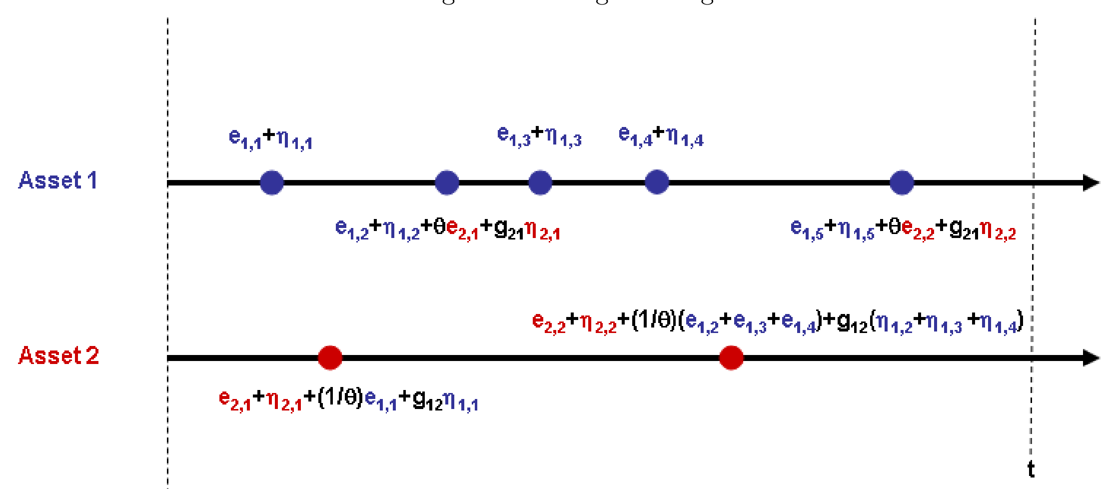
An informative and directly-observable quantity in transaction-level data is the durations $\{\tau_k\}$ between transactions. A seminal paper focusing on durations and, to some extent, on the induced price process, is Engle and Russell (1998). They documented a key empirical fact, i.e., that durations are strongly autocorrelated, quite unlike the *i.i.d.* exponential duration process implied by a Poisson transaction process, and they proposed the Autoregressive Conditional Duration (ACD) model, which is closely related to the GARCH model of Bollerslev (1986). Deo, Hsieh and Hurvich (2006) presented empirical evidence that durations, as well as transaction counts, squared returns and realized volatility have long memory, and introduced the Long Memory Stochastic Duration (LMSD) model, which is closely related to the Long Memory Stochastic Volatility model of Breidt, Crato and de Lima (1998) and Harvey (1998). The LMSD model is $\tau_k = e^{h_k}\epsilon_k$ where $\{h_k\}$ is a Gaussian long-memory series with memory parameter $d_\tau \in (0, \frac{1}{2})$, the $\{\epsilon_k\}$ are *i.i.d.* positive random variables with mean 1, and $\{h_k\}$, $\{\epsilon_k\}$ are mutually independent.

It was shown in Deo, Hurvich, Soulier and Wang (2007) that long memory in durations propagates to long memory in the counting process $N(t)$, where $N(t)$ counts the number of transactions in the time interval $(0, t]$. In particular, if the durations are generated by an LMSD model with memory parameter $d_\tau \in (0, \frac{1}{2})$, then $N(t)$ is long-range count dependent with the same memory parameter, in the sense that $\mathrm{var}N(t) \sim Ct^{2d_\tau+1}$ as $t \to \infty$. This long-range count dependence then propagates to the realized volatility, as studied in Deo, Hurvich, Soulier and Wang (2007).

We now describe the tick-time return interactions that yield cointegration in our model. Suppose that there are two assets, 1 and 2, and that each log price is affected by two types of disturbances when a transaction happens. These disturbances are the value shocks $\{e_{i,k}\}$ and the microstructure noise $\{\eta_{i,k}\}$, for Asset $i = 1, 2$. The subscript i, k pertains to the k'th transaction of asset i. The value shocks are iid and represent permanent contributions to the intrinsic log value of the assets which, in the absence of feedback effects, is a Martingale with respect to full information, both public and private (see Amihud and Mendelson 1987, Glosten 1987). The microstructure shocks represent the remaining contributions to the observed log prices, along similar lines as the noise process considered by Amihud and Mendelson (1987), reflecting transitory price fluctuations due, for example, to liquidity impact of orders. We assume that the m-th tick-time return of Asset 1 incorporates not only its own current disturbances $e_{1,m}$ and $\eta_{1,m}$, but also weighted versions of all intervening disturbances of Asset 2 that were originally introduced between the $(m-1)$-th and m-th transactions of Asset 1. The weight for the value shocks, denoted by θ, may be different from the weight for the microstructure noise, denoted by g_{21} (the impact from Asset 2 to Asset 1). We similarly define the m-th tick-time return of Asset 2, but the weight for the value shocks from Asset 1 to Asset 2 is $(1/\theta)$ and the corresponding weight for the microstructure noise is denoted by g_{12}. The choice of the second impact coefficient $(1/\theta)$ is necessary for the two log-price series to be cointegrated. In general, if the two series are not cointegrated, this constraint is not required.

Figure 1 illustrates the mechanism by which tick-time returns are generated in our model. All disturbances originating from Asset 1 are colored in blue while all disturbances originating from Asset 2 are colored in red. When the first transaction of Asset 1 happens, a value shock $e_{1,1}$ and a microstructure disturbance $\eta_{1,1}$ are introduced. The first transaction of Asset 2 follows in clock time and since the first transaction of Asset 1 occurred before it, the return for this transaction is $(e_{2,1} + \eta_{2,1} + \frac{1}{\theta} e_{1,1} + g_{12}\eta_{1,1})$, *i.e.*, the sum of the first value shock of Asset 2, $e_{2,1}$, the first microstructure disturbance of Asset 2, $\eta_{2,1}$, and a feedback term from the first transaction of Asset 1 whose disturbances are $e_{1,1}$ and $\eta_{1,1}$, weighted by the corresponding feedback impact coefficients $\frac{1}{\theta}$ and g_{12}. In the figure, both log-price processes evolve until time t. Notice that the third return of Asset 1 contains no feedback term from Asset 2 since there is no intervening transaction of Asset 2. The second return of Asset 2 includes its own current disturbances $(e_{2,2}, \eta_{2,2})$ as well as six weighted disturbances $(e_{1,2}, e_{1,3}, e_{1,4}, \eta_{1,2}, \eta_{1,3}$ and $\eta_{1,4})$ from Asset 1 since there are three intervening transactions of Asset 1.

Figure 1: Changes in Log Prices



At a given clock time t, most of the disturbances of Asset 1 are incorporated into the log price of Asset 2 and vice-versa. However, there is an *end effect*. The problem can be easily seen in the figure: since the fifth transaction of Asset 1 happened after the last transaction of Asset 2 before time t, the most recent Asset 1 disturbances $e_{1,5}$ and $\eta_{1,5}$ are not incorporated in the log price of Asset 2 at time t. Eventually, at the next transaction of Asset 2, which will happen after time t, these two disturbances will be incorporated. But this end effect may be present at any given time t. We will handle this end effect explicitly in all derivations in the paper.

Our model for the log prices is then given for all non-negative real t by

$$\begin{aligned}
\log P_{1,t} &= \sum_{k=1}^{N_1(t)} (e_{1,k} + \eta_{1,k}) + \sum_{k=1}^{N_2(t_{1,N_1(t)})} (\theta e_{2,k} + g_{21}\eta_{2,k}) \\
\log P_{2,t} &= \sum_{k=1}^{N_2(t)} (e_{2,k} + \eta_{2,k}) + \sum_{k=1}^{N_1(t_{2,N_2(t)})} (\frac{1}{\theta} e_{1,k} + g_{12}\eta_{1,k}) \ ,
\end{aligned} \tag{1}$$

where $t_{i,k}$ is the clock time for the k-th transaction of Asset i, and $N_i(t)$ $(i = 1, 2)$ are counting processes which count the total number of transactions of Asset i up to time t. Below, we will impose specific conditions on $\{e_{i,k}\}$, $\{\eta_{i,k}\}$ and $N_i(t)$. Note that (1) implies that $\log P_{1,0} = \log P_{2,0} = 0$, the same standardization used in Stock and Watson (1988) and elsewhere. The quantity $N_2(t_{1,N_1(t)})$ represents the total number of transactions of Asset 2 occurring up to the time $(t_{1,N_1(t)})$ of the most recent transaction of Asset 1. An analogous interpretation holds for the quantity $N_1(t_{2,N_2(t)})$.

To exhibit the various components of our model, we rewrite (1) as

$$\log P_{1,t} = \underbrace{\Big(\sum_{k=1}^{N_1(t)} e_{1,k} + \sum_{k=1}^{N_2(t)} \theta e_{2,k} \Big)}_{\text{common component}} + \underbrace{\Big(\sum_{k=1}^{N_1(t)} \eta_{1,k} + \sum_{k=1}^{N_2(t)} g_{21}\eta_{2,k} \Big)}_{\text{microstructure component}} - \underbrace{\sum_{k=N_2(t_{1,N_1(t)})+1}^{N_2(t)} (\theta e_{2,k} + g_{21}\eta_{2,k})}_{\text{end effect}} \tag{2}$$

$$\log P_{2,t} = \underbrace{\Big(\sum_{k=1}^{N_1(t)} \frac{1}{\theta} e_{1,k} + \sum_{k=1}^{N_2(t)} e_{2,k} \Big)}_{\text{common component}} + \underbrace{\Big(\sum_{k=1}^{N_1(t)} g_{12}\eta_{1,k} + \sum_{k=1}^{N_2(t)} \eta_{2,k} \Big)}_{\text{microstructure component}} - \underbrace{\sum_{k=N_1(t_{2,N_2(t)})+1}^{N_1(t)} (\frac{1}{\theta} e_{1,k} + g_{12}\eta_{1,k})}_{\text{end effect}} .$$

The common component is a Martingale, and is therefore $I(1)$. We will show that the microstructure components are $I(1 + d_\eta)$, so these components are less persistent than the common component. The end-effect terms are random sums over time periods that are $O_p(1)$ as $t \to \infty$, (see (39) to (41)), and hence are negligible compared to all other terms. Since both $\log P_{1,t}$ and $\log P_{2,t}$ are $I(1)$ (see Theorem 1) and the linear combination $\log P_{1,t} - \theta \log P_{2,t}$ is $I(1+d_\eta)$ as defined in Section III, the log-price series are cointegrated. (See Theorems 3, 4 and 5.)

Frijns and Schotman (2006) considered a mechanism for generating quotes in tick time which is similar to the mechanism we describe in Figure 1. However, they condition on durations, whereas we endogenize them in our model (1). Furthermore, their model implies standard cointegration, with cointegrating parameter that is known to be 1, and a single value shock component.

Throughout the paper, unless otherwise noted, we will make the following assumptions for our theoretical results. The duration processes $\{\tau_{i,k}\}$ of Asset i, $(i = 1, 2)$, are assumed to possess long memory with memory parameters $d_{\tau_1}, d_{\tau_2} \in (0, \frac{1}{2})$, in order to reflect the empirically-observed persistence in durations and the resulting realized volatility. Specifically, the $\{\tau_{i,k}\}$ are assumed to satisfy the assumptions in Theorem 1 of Deo, Hurvich, Soulier and Wang (2007), which are very general and would allow, for example, the LMSD model of Deo, Hsieh and Hurvich (2006).

We assume that the $\{e_{i,k}\}$ are mutually independent *i.i.d.* processes with mean zero and variance $\sigma^2_{i,e}$, $(i = 1, 2)$. We also assume the $\{\eta_{i,k}\}$ to be mutually independent, with zero mean and memory parameter d_{η_i}. For notational convenience, we set $d_{\eta_1} = d_{\eta_2} = d_\eta$ in our theoretical results. All theorems will continue to hold, however, when d_{η_1} and d_{η_2} are distinct, simply by replacing d_η with $d^*_\eta = \max(d_{\eta_1}, d_{\eta_2})$. For Theorem 6, which establishes the consistency of the OLS estimator of θ, we further assume $\{e_{i,k}\}$ to be $N(0, \sigma^2_{i,e})$.

We assume that $\{\tau_{1,k}\}$ and $\{\tau_{2,k}\}$ are independent of all disturbance series $\{e_{1,k}\}$, $\{e_{2,k}\}$, $\{\eta_{1,k}\}$ and $\{\eta_{2,k}\}$, which are themselves assumed to be mutually independent. We do not require, however, that $N_1(\cdot)$ and $N_2(\cdot)$ be mutually independent, nor do we require that $\{\tau_{1,k}\}$ and $\{\tau_{2,k}\}$ be mutually independent. This is in keeping with recent literature which suggests that there is feedback between the counting processes. See, for example, Nijman, Spierdijk and Soest (2004), Bowsher (2007), and the references therein.

A Economic Justification for the Model

Here, we provide some economic rationale for the transaction-level return interactions leading to Model (1). This supplements our earlier discussion, around Figure 1, on the formal mechanism for price formation. After a brief data analysis affirming the potential usefulness of certain flexibilities of the model, we compare and contrast the model with a clock-time model proposed by Hasbrouck (1995), and then propose a transaction-level generalization of Hasbrouck's definition of the information share of a market.

The model (1) is potentially economically appropriate for pairs of measured prices which are both affected by the same information shocks (i.e., value shocks), possibly in different ways. Examples include: buy prices and sell prices of a single stock; prices of two classified stocks (with different voting rights) from a given company; prices of two different stocks within the same industry; stock and option prices of a given company; option prices on a given stock with different degrees of maturity or moneyness; corporate bond prices at different maturities for a given company; Treasury bond prices at different maturities.

The fundamental (value) prices at time t are an accumulation of information shocks. If we ignore the end effects, these fundamental prices may be thought of as the common components in (2). More precisely, the fundamental prices may be obtained by setting the microstructure shocks in (1) to zero. For definiteness, consider the example of buy prices (Asset 1) and sell prices (Asset 2) of a single stock. Information shocks may be generated on either the buy side or the sell side. According to the model, each buy transaction generates its own information shock, as does each sell transaction. Furthermore, these shocks spill over from the side of the market in which they originated to the other side. Clearly, shocks originating from the sell side of the market cannot be impounded into the buy price until there is a transaction on the buy side. Similarly, in the absence of information arrivals (transactions) on the sell side, any string of intervening information shocks from the buy side will render the most recent sell price stale, until the intervening buy-side shocks are actually impounded into the sell-side price at the next sell transaction.

Shocks spilling over from the buy side to the sell side are weighted by $1/\theta$, while those spilling from the sell side to the buy side are weighted by θ. When $\theta = 1$, shocks spill over from one side to the other in an identical way, and there is just a single fundamental price, shared by both the buy side and sell side. In general, as can be seen from (1), the fundamental (log) price for the buy side is an accumulation of information shocks from both the buy side and the sell side, with the sell-side shocks weighted by θ. Ignoring end effects, as can be seen from (2), the common component on the buy side is proportional to the common component on the sell side, and the constant of proportionality is θ.

Analogous interactions take place on the microstructure shocks, such that a microstructure shock originating on the buy side spills over to the sell side with weight g_{12} and the opposite spillover occurs with weight g_{21}. Even in the absence of spillover of the microstructure shocks ($g_{12} = g_{21} = 0$), the difference between the buy price and θ times the sell price is (except for end effects) an accumulation of microstructure shocks. It seems in accordance with the economic connotation of the term "microstructure" that the microstructure shocks be transitory, i.e., that the aggregate of microstructure shocks be stochastically of smaller order than the aggregate of fundamental shocks, as $t \to \infty$. This will happen if and only if the microstructure shocks have a smaller memory parameter than the fundamental shocks ($d_\eta < 0$), as we assume. Cointegration arises as a consequence of the spillover of the fundamental shocks, together with the assumption $d_\eta < 0$. The spillover of the fundamental shocks induces the common component while the assumption $d_\eta < 0$ ensures that the cointegrating error, arising from microstructure, is less persistent than the common component.

Two questions that might be raised in the context of Model (1) are whether there are situations where the two prices are affected by information in different ways, so that the cointegrating parameter is not equal to 1, and whether it is helpful in practice to allow for fractional cointegration, as opposed to standard cointegration. To address these questions, we briefly present some results of a preliminary data analysis. We considered clock-time option best-available-bid prices and underlying best-available-bid prices for IBM on the NYSE at 390 one-minute intervals from 9:30 AM to 4PM on May 31, 2007. We originally analyzed 74 different options, but removed 5 from consideration since they had either at least one zero bid price during the day or a constant bid price throughout the day. For the remaining 69 options, we regressed the log stock bid price on the log option bid price, and constructed a semiparametric

GPH estimator (Geweke and Porter-Hudak, 1983) of the memory parameter of the residuals. The least-squares regression slopes ranged from -0.21 to 0.39, with a mean of 0.04 and a standard deviation of 0.13. This suggests that information affected the two prices in different ways, for all 69 options. For the GPH estimators, we used $390^{0.5}$ for the number of frequencies. This results in an approximate standard error for the GPH estimator of 0.19. The GPH estimator for the log stock bid price was 1.02. The GPH estimators for the residuals ranged from 0.05 to 1.14 with a mean of 0.55 and a standard deviation of 0.28. Of the 69 sets of residuals, 62 yielded a GPH estimator less than 1, 54 were less than 0.75, and 42 were less than 0.6. Furthermore, 18 were between 0.4 and 0.6. These results suggest the presence of cointegration in most cases, and also that the cointegration in some of these cases may be fractional instead of standard.

It is instructive to compare and contrast our model (1) with the clock-time model of Hasbrouck (1995), in which a single security is traded on several markets and different market prices share an identical random-walk component. Suppose, to facilitate comparisons with the bivariate model (1), that there are two markets. Then for all non-negative integers j, the clock-time log stock prices at time j on the two different markets are given by Hasbrouck's model as

$$\begin{aligned} \log P_{1,j} &= \log P_{1,0} + \sum_{s=1}^{j}(\psi_1 \tilde{e}_{1,s} + \psi_2 \tilde{e}_{2,s}) + v_{1,j} \\ \log P_{2,j} &= \log P_{2,0} + \sum_{s=1}^{j}(\psi_1 \tilde{e}_{1,s} + \psi_2 \tilde{e}_{2,s}) + v_{2,j} \end{aligned} \tag{3}$$

where $\log P_{1,0}$ and $\log P_{2,0}$ are constants, $(\tilde{e}_{1,s}, \tilde{e}_{2,s})'$ is a zero-mean vector of serially uncorrelated disturbances with covariance matrix Ω, $\psi = (\psi_1, \psi_2)$ are the weights for $\tilde{e}_{1,s}, \tilde{e}_{2,s}$, and $\{(v_{1,j}, v_{2,j})'\}$ is a zero-mean stationary bivariate time series. The quantity $\tilde{e}_{i,s}, (i = 1, 2)$ may be regarded as the fundamental shock originating from the i-th market. Hasbrouck (1995) estimated the model on data using a one-second sampling interval.

Both Models (1) and (3) induce a common component, and cointegration. Both have spillover of the fundamental shocks from one market to the other. In Model (3) the spillover is the same in both directions so the common components are identical and the cointegrating parameter is 1, in contrast to Model (1) where the cointegrating parameter need not be equal to 1. In Model (3) the cointegrating error is $I(0)$, while in Model (1) the cointegrating error is allowed to be $I(1+d_\eta)$ for any d_η with $-1 \leq d_\eta < 0$.

In Model (3) the contemporaneous correlation between the fundamental shocks originating from the two markets is allowed to be nonzero, (i.e., Ω is allowed to be non-diagonal), whereas in Model (1) the two fundamental shock series are assumed to be independent. Note, however, that in the transaction-level model (1), the k^{th} transactions of the two assets will (almost surely) occur at different clock times, so any correlation between the two fundamental shocks $e_{1,k}$ and $e_{2,k}$ would fail to be contemporaneous in clock time. This provides one motivation for our assumption that $\{e_{1,k}\}$ and $\{e_{2,k}\}$ are mutually independent. An economic motivation for this assumption stems from the following remarks of Hasbrouck (1995, p. 1183): "In practice, market prices usually change sequentially: a new price is posted in one market, and then the other markets respond. If the observation interval is so long that the sequencing cannot be determined, however, the initial change and the response will appear to be contemporaneous. Therefore, one obvious way of minimizing the correlation is to shorten the interval of observation." Since Model (1) is defined in continuous time, the interval of observation is effectively zero, so at least under the idealized assumptions that there are no truly simultaneous transactions on the two markets and that the time stamps for the transactions are exact, the assumption of mutual independence would be economically reasonable.

In the remainder of this subsection, we discuss the information share, originally defined in Hasbrouck (1995) to measure how market information that drives stock prices is distributed across different exchanges. Hasbrouck (1995) defined the information share of market i based on Model (3) as $S_i = (\psi_i^2 \Omega_{ii})/(\psi \Omega \psi')$, which is the proportional contribution from market i to the total fundamental innovation variance. Only the random-walk component is used in constructing the information share since this is the only permanent component. As discussed in Hasbrouck (1995), the fact that Ω may not be diagonal has the consequence that only a bound for the information share can be estimated. Below, we propose a transaction-level generalization of the concept of information share based on Model (1), which is directly estimable due to our assumption of mutual independence of the transaction-level fundamental disturbance series. The information share proposed in Hasbrouck (1995) measures how the price discovery of one security is fulfilled across difference exchanges. Our information share instead measures how the price-driving information of a security is distributed between buy versus sell trades in a market. The ideas are similar. Indeed, as mentioned in Hasbrouck (1995), his model can be extended to model bid and ask prices dynamics. Nevertheless, as discussed above, our model ultimately is a tick-level model which is different from existing clock-time models, including that in Hasbrouck (1995).

For Model (1), we define the information share as follows: for a given clock-time sampling interval Δt, the information share of Asset i is given by

$$S_{1,C} = \frac{\mathrm{var}\Big(\sum_{k=N_1((j-1)\Delta t)+1}^{N_1(j\Delta t)} e_{1,k}\Big)}{\mathrm{var}\Big(\sum_{k=N_1((j-1)\Delta t)+1}^{N_1(j\Delta t)} e_{1,k} + \theta \sum_{k=N_2((j-1)\Delta t)+1}^{N_2(j\Delta t)} e_{2,k}\Big)} = \frac{\lambda_1 \sigma_{1,e}^2}{\lambda_1 \sigma_{1,e}^2 + \theta^2 \lambda_2 \sigma_{2,e}^2} \tag{4}$$

$$S_{2,C} = \frac{\theta^2 \lambda_2 \sigma_{2,e}^2}{\lambda_1 \sigma_{1,e}^2 + \theta^2 \lambda_2 \sigma_{2,e}^2} \ ,$$

where λ_i is the intensity of the counting process $N_i(\cdot)$ (see Daley and Vere-Jones 2003), and represents the intensity of trading (level of market activity) of Asset i. The ultimate expressions for $S_{i,C}$ do not depend on the sampling interval Δt. Note that only the common component in (2) is used to evaluate the information share, as was also done by Hasbrouck (1995). As $\lambda_1/\lambda_2 \to \infty$, $S_{1,C}$ approaches one and $S_{2,C}$ approaches zero. This is consistent with general intuition: an actively-traded security should reveal more information than a thinly-traded one. Indeed, Hasbrouck (1995) found that, for the 30 Dow-Jones stocks, the preponderance of the price discovery takes place at the NYSE and the majority of the transactions occurred on the NYSE. The information share $S_{i,C}$ can be estimated using the transaction-level method of moments as described in section VIII. We exhibit estimates of $S_{i,C}$ computed from transaction-level data in Section XI.

III Conditions on the Microstructure Noise for Fractional and Standard Cointegration

We consider three types of cointegration: weak fractional, strong fractional and standard cointegration. In this section, we describe the conditions assumed for each of these three cases separately.

The weak fractional cointegration case corresponds to $d_\eta \in (-\frac{1}{2}, 0)$. In this case, we will require the following condition, stated for a generic process $\{\eta_k\}$.

Condition A *For $d_\eta \in (-\frac{1}{2}, 0)$, $\{\eta_k\}$ is a weakly stationary zero-mean process with memory parameter d_η in the sense that the spectral density $f(\lambda)$ satisfies*

$$f(\lambda) = \tilde{\sigma}^2 C^* \lambda^{-2d_\eta}(1 + O(\lambda^\beta)) \text{ as } \lambda \to 0^+$$

for some β with $0 < \beta \leq 2$, where $\tilde{\sigma}^2 > 0$ and $C^* = (d_\eta + \frac{1}{2})\Gamma(2d_\eta + 1)\sin((d_\eta + \frac{1}{2})\pi)/\pi > 0$.

Condition A, which was originally used in a semiparametric context by Robinson (1995), is very general, since it only specifies the behavior of the spectral density in a neighborhood of zero frequency. The condition is satisfied by all parametric long-memory models that we have seen in the literature, including the ARFIMA(p, d_η, q) model with $p \geq 0$, $q \geq 0$ and $d_\eta \in (-\frac{1}{2}, 0)$. In the ARFIMA case, $\beta = 2$. Condition A also allows the possibility for seasonal long memory, *i.e.*, poles or zeros of $f(\lambda)$ at nonzero frequencies.

The strong fractional cointegration case corresponds to $d_\eta \in (-1, -\frac{1}{2})$. For this case, we assume

Condition B *For $d_\eta \in (-1, -\frac{1}{2})$, $\eta_k = \varphi_k - \varphi_{k-1}$, $k = 1, 2, \ldots$ where $\varphi_0 = 0$ and $\{\varphi_k\}_{k=1}^\infty$ is a zero-mean weakly-stationary long-memory process with memory parameter $d_\varphi = d_\eta + 1 \in (0, \frac{1}{2})$ in the sense that its autocovariances satisfy*

$$\operatorname{cov}(\varphi_k, \varphi_{k+j}) = K j^{2d_\varphi - 1} + O(j^{2d_\varphi - 3}) \ , \ \ j \geq 1 \tag{5}$$

where $K > 0$.

By Theorem 1 of Lieberman and Phillips (2006), any stationary, invertible ARFIMA(p, d_φ, q) process with $d_\varphi \in (0, \frac{1}{2})$ has autocovariances satisfying (5), with $K = 2f^*(0)\Gamma(1 - 2d_\varphi)\sin(\pi d_\varphi)$, where $f^*(0)$ is the spectral density of the ARMA component of the model at zero frequency.

The standard cointegration case corresponds to $d_\eta = -1$. In this case we assume

Condition C *If $d_\eta = -1$, $\{\eta_k\}_{k=1}^\infty$ is given by $\eta_k = \xi_k - \xi_{k-1}$ with $\xi_0 = 0$. The process $\{\xi_k\}_{k=1}^\infty$ is* *weakly stationary with zero mean and has autocovariance sequence* $\{c_{\xi,r}\}_{r=0}^{\infty}$ *where* $c_{\xi,r} = E(\xi_{k+r}\xi_k)$ *with exponential decay,* $|c_{\xi,r}| \leq A_\xi e^{-K_\xi r}$ *for all* $r \geq 0$*, where* A_ξ *and* K_ξ *are positive constants.*

The assumptions on $\{\xi_k\}$ in Condition C above are satisfied by all stationary, invertible ARMA models.

IV Long-Term Martingale-Type Properties of the Log Prices

In this section, we present the properties of the log-price series generated by Model (1). Define $\lambda_i = 1/E^0(\tau_{i,k})$, where E^0 denotes expectation under the Palm distribution (See Deo, Hurvich, Soulier and Wang 2007 for discussions about the Palm probability measure), i.e., the distribution under which the $\{\tau_{i,k}\}$, $(i = 1, 2)$ are stationary. The following two theorems show that the log-price series in Model (1) have asymptotic variances that scale like t as $t \to \infty$, as would happen for a Martingale, and that their discretized differences are asymptotically uncorrelated as the sampling interval increases, as would happen for a Martingale difference series.

Theorem 1 *For the log-price series in Model (1),*

$$\mathrm{var}(\log P_{i,t}) \sim C_i t, \qquad i = 1, 2$$

as $t \to \infty$*, where* $C_1 = (\sigma_{1,e}^2\lambda_1 + \theta^2\sigma_{2,e}^2\lambda_2)$ *and* $C_2 = (\sigma_{2,e}^2\lambda_2 + \frac{1}{\theta^2}\sigma_{1,e}^2\lambda_1)$.

For a given sampling interval (equally-spaced clock-time period) Δt, the returns (changes in log price) for Asset 1 and 2 corresponding to Model (1) are

$$\begin{aligned} r_{1,j} &= \sum_{k=N_1((j-1)\Delta t)+1}^{N_1(j\Delta t)} (e_{1,k} + \eta_{1,k}) + \sum_{k=N_2(t_{1,N_1((j-1)\Delta t)})+1}^{N_2(t_{1,N_1(j\Delta t)})} (\theta e_{2,k} + g_{21}\eta_{2,k}) \\ r_{2,j} &= \sum_{k=N_2((j-1)\Delta t)+1}^{N_2(j\Delta t)} (e_{2,k} + \eta_{2,k}) + \sum_{k=N_1(t_{2,N_2((j-1)\Delta t)})+1}^{N_1(t_{2,N_2(j\Delta t)})} (\frac{1}{\theta}e_{1,k} + g_{12}\eta_{1,k}) \ . \end{aligned} \tag{6}$$

Theorem 2 *For any fixed integer $k > 0$, the lag-k autocorrelation of $\{r_{i,j}\}_{j=1}^{\infty}$, $i = 1, 2$, tends to 0 as $\Delta t \to \infty$.*

V Properties of the Cointegrating Error

We show that Model (1) implies a cointegrating relationship between the two series, treating the weak and strong fractional as well as standard cointegration cases separately.

Theorem 3 *Under Model (1) with $d_\eta \in (-\frac{1}{2}, 0)$, the memory parameter of the linear combination $(\log P_{1,t} - \theta \log P_{2,t})$ is $(1 + d_\eta) \in (\frac{1}{2}, 1)$, that is,*

$$\mathrm{var}(\log P_{1,t} - \theta \log P_{2,t}) \sim C\, t^{2d_\eta+1}$$

as $t \to \infty$, where $C > 0$. In this sense, $\log P_{1,t}$ and $\log P_{2,t}$ are weakly fractionally cointegrated.

Next, we investigate the standard cointegration case. It is important to note that, unlike in Theorem 3, where we measure the strength of cointegration using the asymptotic behavior of the variance of the cointegrating errors $\mathrm{var}(\log P_{1,t} - \theta \log P_{2,t})$, we need a different measure here since $\log P_{1,t} - \theta \log P_{2,t}$ is stationary and its variance is constant for all t. Instead, we consider the asymptotic covariance of the cointegrating errors

$$\mathrm{cov}(\log P_{1,t} - \theta \log P_{2,t}, \log P_{1,t+j} - \theta \log P_{2,t+j})$$

as $j \to \infty$. We take t and j here to be positive integers, *i.e.*, we sample the log-price series using $\Delta t = 1$, without loss of generality.

Theorem 4 *Under Model (1) with $d_\eta \in (-1, -\frac{1}{2})$, the memory parameter of the cointegrating error $(\log P_{1,t} - \theta \log P_{2,t})$ is $(1 + d_\eta) \in (0, \frac{1}{2})$, that is, for any fixed $t > 0$,*

$$\mathrm{cov}\Big(\log P_{1,t} - \theta \log P_{2,t},\ \log P_{1,t+j} - \theta \log P_{2,t+j}\Big) \sim j^{2(1+d_\eta)-1} \left[C_1 Pr\{N_1(t) > 0\} + C_2 Pr\{N_2(t) > 0\}\right]$$

as $j \to \infty$, where $C_1 > 0$, $C_2 > 0$. In this sense, $\log P_{1,t}$ and $\log P_{2,t}$ are strongly fractionally cointegrated.

We say that a sequence $\{a_j\}$ has *nearly-exponential decay* if $a_j/j^{-\alpha} \to 0$ as $j \to \infty$ for all $\alpha > 0$. We say that a time series has *short memory* if its autocovariances have nearly-exponential decay.

Theorem 5 *Under Model (1), with $d_\eta = -1$, the cointegrating error $(\log P_{1,t} - \theta \log P_{2,t})$ has short memory. In this sense, $\log P_{1,t}$ and $\log P_{2,t}$ are cointegrated.*

VI Least-Squares Estimation of the Cointegrating Parameter

Assume that the log-price series are observed at integer multiples of Δt. The proposed model (1) becomes (with a minor abuse of notation)

$$\begin{aligned} \log P_{1,j} &= \sum_{k=1}^{N_1(j\Delta t)} (e_{1,k} + \eta_{1,k}) + \sum_{k=1}^{N_2(t_{1,N_1(j\Delta t)})} (\theta e_{2,k} + g_{21}\eta_{2,k}) \\ \log P_{2,j} &= \sum_{k=1}^{N_2(j\Delta t)} (e_{2,k} + \eta_{2,k}) + \sum_{k=1}^{N_1(t_{2,N_2(j\Delta t)})} (\frac{1}{\theta} e_{1,k} + g_{12}\eta_{1,k}) \ . \end{aligned} \tag{7}$$

We show that the cointegrating parameter θ can be consistently estimated by OLS regression.

Theorem 6 *For the discretely-sampled log-price series in (7) with normally distributed value shocks $\{e_{1,k}\}$, $\{e_{2,k}\}$, the cointegrating parameter θ can be consistently estimated by $\hat{\theta}$, the ordinary least squares estimator obtained by regressing $\{\log P_{1,j}\}_{j=1}^{n}$ on $\{\log P_{2,j}\}_{j=1}^{n}$ without intercept. For all $\delta > 0$, as $n \to \infty$, we have*

Case I: $d_\eta \in (-\frac{1}{2}, 0)$

$$n^{-d_\eta - \delta}(\hat{\theta} - \theta) \xrightarrow{p} 0,$$

Case II: $d_\eta \in (-1, -\frac{1}{2})$

$$n^{\frac{1}{2} - \delta}(\hat{\theta} - \theta) \xrightarrow{p} 0,$$

Case III: $d_\eta = -1$

$$n^{1 - \delta}(\hat{\theta} - \theta) \xrightarrow{p} 0,$$

In the weak fractional cointegration case $d_\eta \in (-\frac{1}{2}, 0)$, the rate of convergence of $\hat{\theta}$ improves as d_η decreases. In the standard cointegration case where $d_\eta = -1$, the rate is arbitrarily close to n. The n-consistency (super-consistency) of the OLS estimator of the cointegrating parameter in the standard cointegration case has been shown for time series in discrete clock time that are linear with respect to a strong-mixing or *i.i.d.* sequence by Phillips and Durlauf (1986) and Stock (1987). We are currently unable to derive the asymptotic distribution of the OLS estimator of the cointegrating parameter in the standard cointegration case for our model, as we cannot rely on the strong-mixing condition on returns. This condition would not be expected to hold in the case of LMSD durations, since these are not strong-mixing in tick time. In the strong fractional cointegration case $d_\eta \in (-1, -\frac{1}{2})$, though we have established a rate of $n^{1/2-\delta}$, simulations in Section X indicate that the actual rate is faster, at $n^{-d_\eta-\delta}$, in keeping with the rates obtained in the weak fractional and standard cointegration cases.

VII A Tick-level Cointegrating Parameter Estimator

We propose a transaction-level estimator, $\tilde{\theta}$, for the cointegrating parameter θ in this section. It may be argued that the OLS estimator $\hat{\theta}$ discussed in Section VI is not optimal since it is constructed based on discretized log-prices, and therefore only uses partial information. Here we propose a tick-level estimator, $\tilde{\theta}$, that utilizes the full tick-level price series, $\log P_{1,t}$, $\log P_{2,t}$ for $t \in [0, T]$.

Specifically, let $N(T) = N_1(T) + N_2(T)$ be the pooled counting process of transactions for both Asset 1 and Asset 2 in the time interval $(0, T]$, and denote by $\{t_k^\star\}_{k=1}^{N(T)}$ the transaction times for the pooled process. The proposed estimator is

$$\begin{aligned} \tilde{\theta} &= \frac{\int_0^T \log P_{1,t} \cdot \log P_{2,t} dt}{\int_0^T \log P_{2,t}^2 dt} \qquad (8) \\ &= \frac{\sum_{j=1}^{N(T)-1} \left[\log P_{1,N_1(t_j^\star)} \cdot \log P_{2,N_2(t_j^\star)}\right] \cdot (t_{j+1}^\star - t_j^\star) + \left[\log P_{1,N_1(T)} \cdot \log P_{2,N_2(T)}\right] \cdot (T - t_{N(T)}^\star)}{\sum_{j=1}^{N_2(T)-1} \log P_{2,j}^2 \cdot \tau_{2,j+1} + \log P_{2,N_2(T)}^2 \cdot (T - t_{2,N_2(T)})}, \end{aligned}$$

where the numerator is a summation over all transactions, adding up the product of the most recent log prices of Asset 1 and Asset 2 weighted by the corresponding duration for the pooled process. The denominator of $\tilde{\theta}$ has the same structure except that the product is now of Asset 2 log-prices with themselves.

We do not derive asymptotic properties of the estimator $\tilde{\theta}$. Nevertheless, the simulation study in Section X indicates that the tick-level estimator $\tilde{\theta}$ may outperform the OLS estimator $\hat{\theta}$, having smaller bias, variance and Root-Mean-Squared-Error (RMSE), particularly if the sampling interval Δt for $\hat{\theta}$ is large.

VIII Method of Moments Parameter Estimation

In Section XIV, we will propose a transaction-level parameter estimation procedure for model (1) using the method of moments, based on $\log P_{1,t}$, $\log P_{2,t}$ for $t \in [0, T]$. We will make specific assumptions for the sake of definiteness, though most of these assumptions could be relaxed. Specifically, we will assume Gaussian white noise for the value shocks, a Gaussian ARFIMA$(1, d_\eta, 0)$ process for the microstructure noise when $d_\eta \in (-\frac{1}{2}, 0)$, while we will assume that the microstructure noise is the difference of a Gaussian ARFIMA$(1, d_\eta + 1, 0)$ process with the initial value set to zero when $d_\eta \in (-1, -\frac{1}{2})$. In the standard cointegration case $d_\eta = -1$, we will assume that the microstructure noise is the difference of a Gaussian $AR(1)$ process with initial value set to zero.

The method-of-moments estimator $\hat{\Theta} = (\hat{\sigma}^2_{1,e}, \hat{\sigma}^2_{2,e}, \hat{\sigma}^2_{1,\eta}, \hat{\sigma}^2_{2,\eta}, \hat{g}_{21}, \hat{g}_{12}, \hat{d}_{\eta_1}, \hat{d}_{\eta_2}, \hat{\alpha}_1, \hat{\alpha}_2)$ is obtained as the solution to a system of ten equations, as described in Section XIV, where we will also establish the following theorem on consistency of the estimator.

Theorem 7 *The method-of-moments estimator $\hat{\Theta}$ is consistent, i.e.*

$$\hat{\Theta} \xrightarrow{p} \Theta, \quad \text{as } T \to \infty$$

where $\Theta = (\sigma^2_{1,e}, \sigma^2_{2,e}, \sigma^2_{1,\eta}, \sigma^2_{2,\eta}, g_{21}, g_{12}, d_{\eta_1}, d_{\eta_2}, \alpha_1, \alpha_2)$.

Motivated by computational constraints that limit the size of the data set we will be analyzing, we propose an alternative ad hoc estimator $\tilde{\Theta}$ in Section XIV. We evaluate the performance of $\tilde{\Theta}$ in a simulation study in Section X.

IX Model Specification Test

In this section, we propose a specification test for our model (1), based on Theorem 6. The idea is that, according to Theorem 6, if the model (1) is correctly specified, then the OLS estimator is consistent for any particular sampling interval Δt. Suppose we choose two sampling intervals, Δt_1 and Δt_2, and denote the corresponding OLS estimators by $\hat{\theta}^{\Delta t_1}$ and $\hat{\theta}^{\Delta t_2}$. Since both estimators are consistent, their difference must converge in probability to zero. Thus, we propose a specification test to test whether this difference is significantly different from zero. The test is semiparametric, in that model (1) makes no parametric assumptions on either the durations or the microstructure noise.

To implement the test, we divide the entire time span, say one year, into K nonoverlapping subperiods, *e.g.*, divide the data set into months. Within subperiod $k, (k = 1, \ldots, K)$, we sample every Δt_1 to obtain a bivariate log-price series $\{\log P_{1,j,k}^{\Delta t_1}, \log P_{2,j,k}^{\Delta t_1}\}$, where $\log P_{1,j,k}^{\Delta t_1}$ is the j-th sampled Asset 1 log-price in the subperiod k using sampling interval Δt_1 and similarly for $\log P_{2,j,k}^{\Delta t_1}$. Based on these, we obtain an OLS cointegrating parameter estimate $\hat{\theta}_k^{\Delta t_1}$, and similarly we sample every Δt_2 to obtain $\{\log P_{1,j,k}^{\Delta t_2}, \log P_{2,j,k}^{\Delta t_2}\}$ and then obtain $\hat{\theta}_k^{\Delta t_2}$. Repeating the procedure through all K subperiods, we obtain sequences $\{\hat{\theta}_k^{\Delta t_1}\}_{k=1}^K$ and $\{\hat{\theta}_k^{\Delta t_2}\}_{k=1}^K$. The test statistic is proposed to be

$$\hat{\delta}_{12} = \frac{\text{sample mean of } \{\hat{\delta}_{12,k}\}}{\sqrt{\frac{1}{K} \cdot \text{sample variance of } \{\hat{\delta}_{12,k}\}}}$$

where $\hat{\delta}_{12,k} = \hat{\theta}_k^{\Delta t_2} - \hat{\theta}_k^{\Delta t_1}, (k = 1, \ldots, K)$. The distribution of the test statistic under the null hypothesis that all model assumptions are correctly specified is unknown. However, the critical value for the test as well as the corresponding distribution of the test statistic can be simulated under the null hypothesis, based on the estimated parameter values.

The power of the proposed specification test is unknown, since the precise alternative hypothesis is not specified. Similarly as discussed in Hausman (1978), one sufficient requirement is that the two estimators, $\hat{\theta}^{\Delta t_1}$ and $\hat{\theta}^{\Delta t_2}$, have different probability limits under the alternative, in order for the specification test to be consistent.

In Section X, we first investigate the simulation-based distributions of the test statistics for empirically-relevant parameter values, then compute critical values for the specification test on the empirical example, Tiffany (TIF), which are later used in the data analysis in section XI.

X Simulations

A The Estimation of the Cointegrating Parameter: $\hat{\theta}$ and $\tilde{\theta}$

We study the performance of $\hat{\theta}$ as well as $\tilde{\theta}$ in a simulation study carried out as follows.

First, we simulate two mutually independent duration process $\{\tau_{i,k}\}$ for Asset $i = 1, 2$. Note that for simplicity we take the two duration processes to be mutually independent, although this is not required by our theoretical results. Each duration process follows the Long Memory Stochastic Duration (LMSD) model,

$$\tau_{i,k} = e^{h_{i,k}} \epsilon_{i,k}$$

where the $\{\epsilon_{i,k}\}$ are $i.i.d.$ positive random variables with all moments finite, and the $\{h_{i,k}\}$ are a Gaussian long-memory series with zero mean and common memory parameter d_τ. Based on empirical work in Deo, Hsieh and Hurvich (2006), we choose $d_{\tau_1} = d_{\tau_2} = 0.45$. Here, we assume that the $\{\epsilon_{i,k}\}$ follow an exponential distribution with unit mean. We simulated the $\{h_{i,k}\}$ from a Gaussian ARFIMA(0, d_τ, 0) model, with innovation variances chosen so that the mean of the log durations matches those observed in the Tiffany series used in Section XI. Using the simulated durations $\{\tau_{i,k}\}$, $i = 1, 2$, we obtained the corresponding counting processes $\{N_i(t)\}$, using $t_{i,1} = \text{Uniform}[0, \tau_{i,1}]$. This ensures that the counting processes are stationary.

Next, we generate mutually independent disturbance series $\{e_{1,k}\}, \{e_{2,k}\}, \{\eta_{1,k}\}$ and $\{\eta_{1,k}\}$. Here, $\{e_{i,k}\}, i = 1, 2$, are *i.i.d.* Gaussian with zero means. For simplicity, the memory parameters of the microstructure noise series are assumed to be the same: $d_{\eta_1} = d_{\eta_2} = d_\eta$. When $d_\eta \in (-\frac{1}{2}, 0)$, the $\{\eta_{i,k}\}$ are given by ARFIMA(1, d_η, 0). When $d_\eta \in (-1, -\frac{1}{2})$, $\{\eta_{i,k}\}$ are simulated as the differences of ARFIMA(1, $d_\eta + 1$, 0); and when $d_\eta = -1$, $\{\eta_{i,k}\}$ are simulated as the differences of two independent zero-mean Gaussian $AR(1)$ series $\{\xi_{i,k}\}$. The disturbance variances were $\text{var}(e_{i,k}) = 4 \times 10^{-6}$ and $\text{var}(\eta_{i,k}) = 1 \times 10^{-6}$ for $i = 1, 2$. Also, we set $g_{21} = g_{12} = 1$. We selected these particular values as they are close to the corresponding parameter estimates based on several stocks that we analyzed empirically.

We then constructed the log-price series $\{\log P_{i,j}\}_{j=1}^n, i = 1, 2$ from (1), using a fixed sampling interval Δt. We calculated the estimated cointegrating parameter $\hat{\theta}$ by regressing $\{\log P_{1,j}\}_{j=1}^n$ on $\{\log P_{2,j}\}_{j=1}^n$, using ordinary least squares without intercept. The tick-level estimator, $\tilde{\theta}$ is constructed according to (8) using the entire tick-level price series.

In the study, we fixed the cointegrating parameter at $\theta = 1$. We considered various values of the parameters Δt and the sample size n. We think of time as being measured in seconds, so that $\Delta t = 300$ corresponds to observing the price series every 5 minutes, and in this case $n = 390$ would correspond to one week of data. (Each day, there are 6.5 trading hours so sampling every 5 minutes yields 78 observations per day). For each parameter configuration, we generated 1000 realizations of the log-price series. The results are summarized in Table 1.

As the sample size n increases, the bias, the standard deviation and the Root-Mean-Squared Error (RMSE) of $\hat{\theta}$ decrease, as seen in Block A. This is consistent with Theorem 6. We only report results for $d_\eta = -0.75$, however we found similar patterns for $d_\eta = -0.25, -1$.

In A2 together with Block B, we fix the total time span $T = n\Delta t$, while varying the sampling interval Δt and n. For this specific set of empirically-relevant parameter values, the impact of increasing Δt is not obvious until $\Delta t = 9000$, which corresponds to the commonly-used sampling frequency of one day. Both the standard deviation and the RMSE deteriorate as Δt grows. We found the same pattern for $d_\eta = -0.25, -0.75, -1$, although only results for $d_\eta = -0.75$ were reported here. In addition, the bias of

Table 1: Simulation Results on the Estimation of the Cointegrating Parameter

Block	Case	Simulation Parameters				$\hat{\theta}$			$\tilde{\theta}$		
		$n\Delta t$	Δt (sec)	d_η	n	Mean	SD	RMSE	Mean	SD	RMSE
A	A1	39,000	300	–0.75	130	0.9510	0.1192	0.1288	0.9511	0.1192	0.1288
	A2	117,000	300	–0.75	390	0.9769	0.0576	0.0620	0.9768	0.0575	0.0620
	A3	351,000	300	–0.75	1170	0.9875	0.0350	0.0371	0.9876	0.0349	0.0370
	A4	1,053,000	300	–0.75	3510	0.9957	0.0142	0.0148	0.9957	0.0142	0.0148
B	B1	117,000	10	–0.75	11,700	0.9768	0.0575	0.0620	0.9768	0.0575	0.0620
	B2	117,000	60	–0.75	1,950	0.9768	0.0575	0.0620	0.9768	0.0575	0.0620
	B3	117,000	1800	–0.75	65	0.9770	0.0592	0.0635	0.9768	0.0575	0.0620
	B4	117,000	9000	–0.75	13	0.9780	0.0761	0.0792	0.9768	0.0575	0.0620
	B5	117,000	23400	–0.75	5	0.9876	0.1148	0.1154	0.9768	0.0575	0.0620

$\hat{\theta}$ decreases as the sampling interval Δt increases, possibly due to the fact that the end effect is not as important when Δt is large. Finally, in terms of RMSE, $\tilde{\theta}$ performs no worse than $\hat{\theta}$, and performs much better than $\hat{\theta}$ when Δt is large.

We also performed simulations related to the convergence rate of $\hat{\theta}$. In Theorem 6, when $d_\eta \in (-1, -\frac{1}{2})$, the convergence rate is shown to be arbitrarily close to $\sqrt{n}$ and does not depend on d_η. However, simulations indicate a faster rate in this strong fractional cointegration case. For example, when $d_\eta = -0.75$, we simulated the log price series in discrete clock-time using sample sizes n ranging from 1,000 to 20,000 with an equally-spaced increment of 800. The variance of $\hat{\theta}$ for each value of n was obtained based on 1,000 realizations. The estimated convergence rate of $\hat{\theta}$ is $n^{0.78}$, obtained from the estimated slope in a log-log plot of these simulated variances versus the corresponding sample sizes. We ran similar simulations for other values of d_η. Based on these, we conjecture that the actual rate of convergence for $\hat{\theta}$ is $n^{-d_\eta-\delta}$, in keeping with the rates obtained in the weak fractional and standard cointegration cases.

B The performance of $\tilde{\Theta}$

We carried out a simulation study to evaluate the performance of the ad hoc estimator $\tilde{\Theta}$ of Θ discussed in Section XIV. The parameter values were $d_{\tau_1} = d_{\tau_2} = 0.45$, $d_{\eta_1} = d_{\eta_2} = -0.25, -0.75, -1.00$, $\theta = 1$, $\text{var}(e_{i,k}) = 4 \times 10^{-6}, (i = 1, 2)$, $\text{var}(\eta_{i,k}) = 1 \times 10^{-6}, (i = 1, 2)$, $g_{12} = g_{21} = 1$, and $\alpha_1 = \alpha_2 = -0.5$. The $\{h_{i,k}\}$ were simulated as in Section X.

We simulated log prices in model (1) for a clock-time span of $n\Delta t$, with $n = 1716$ and $\Delta t = 300$ seconds (corresponding to a clock-time span of one month). The results, given in Table 2, are averages based on the 1000 realizations, after excluding those realizations that give negative estimates for $\tilde{g}_{21}^2$ or $\tilde{g}_{12}^2$ which happen about 15% of the time. The estimates for both $\sigma_{i,e}^2, (i = 1, 2)$ and $\sigma_{i,\eta}^2, (i = 1, 2)$ are

Table 2: Parameter Estimation using $\tilde{\Theta}$

d_η	$\tilde{\sigma}_{1,e}^2$ ($\times 10^{-6}$) [SD]	$\tilde{\sigma}_{2,e}^2$ ($\times 10^{-6}$) [SD]	$\tilde{\sigma}_{1,\eta}^2$ ($\times 10^{-6}$) [SD]	$\tilde{\sigma}_{2,\eta}^2$ ($\times 10^{-6}$) [SD]	$\tilde{d}_{\eta_1}$ [SD]	$\tilde{\alpha}_1$ [SD]
-0.25	4.01 [0.39]	3.98 [0.43]	0.98 [0.37]	1.00 [0.41]	-0.26 [0.15]	-0.47 [0.20]
-0.75	3.78 [0.47]	3.81 [0.42]	1.20 [0.47]	1.19 [0.42]	-0.79 [0.15]	-0.57 [0.19]
-1.00	3.56 [0.78]	3.56 [0.77]	1.44 [0.79]	1.43 [0.78]	-0.85 [0.28]	-0.69 [0.33]

d_η	$\tilde{g}_{21}$ [SD]	$\tilde{g}_{12}$ [SD]	$\lvert\tilde{g}_{21}\rvert$ [SD]	$\lvert\tilde{g}_{12}\rvert$ [SD]	$\%(\tilde{g}_{21} > 0)$	$\%(\tilde{g}_{12} > 0)$	$\tilde{d}_{\eta_2}$ [SD]	$\tilde{\alpha}_2$ [SD]
-0.25	0.31 [1.05]	0.28 [1.05]	1.04 [0.36]	1.03 [0.36]	64.7%	64.2%	-0.35 [0.15]	-0.48 [0.21]
-0.75	0.38 [1.06]	0.34 [1.06]	1.06 [0.38]	1.05 [0.35]	67.7%	67.6%	-0.66 [0.15]	-0.58 [0.20]
-1.00	0.02 [1.23]	0.03 [1.12]	1.09 [0.56]	1.05 [0.40]	49.7%	51.4%	-0.93 [0.29]	-0.68 [0.34]

reasonably well-behaved. The magnitudes of g_{21} and g_{12} are also well estimated, but the signs are not. This is due to the fact that these signs are determined based on certain five-trade sequences (instead of certain three-trade sequences to estimate the magnitudes) that occur relatively infrequently in the data (see Section XIV in the Appendix for details). Histograms of $\tilde{g}_{21}$ or $\tilde{g}_{12}$ show a bimodal distribution, with peaks centered around +1 and -1, respectively.

C Specification Test

We perform a simulation study on the specification test proposed in Section IX. Two sets of empirically-relevant parameter values are used to investigate the simulation-based distribution of the test statistic $\hat{\delta}$.

We choose empirically-relevant parameter values to investigate the simulation-based distribution of the test statistic $\hat{\delta}$. Specifically, we selected four sampling intervals $\Delta t_1 = 60$, $\Delta t_2 = 300$, $\Delta t_3 = 600$ and $\Delta t_4 = 1800$ seconds, respectively. The entire time span is set to be 100 trading days, which is divided into 25 subperiods with 4 trading days each. Other model parameter values are: $d_\eta = d_{\eta_1} = d_{\eta_2} = -0.25, -0.75$, $d_{\tau_1} = d_{\tau_2} = 0.45$. Results are based on 1000 realizations.

Six test-statistic distributions are generated for each pair of sampling intervals. For example, for the pair Δt_1, Δt_2, we obtain a test statistic

$$\hat{\delta}_{12,m} = \frac{\text{sample mean of } \{\hat{\theta}_{k,m}^{\Delta t_2} - \hat{\theta}_{k,m}^{\Delta t_1}\}}{\sqrt{\frac{1}{25} \cdot \text{sample variance of } \{\hat{\theta}_{k,m}^{\Delta t_2} - \hat{\theta}_{k,m}^{\Delta t_1}\}}}$$

for realization m based on $\{\hat{\theta}_{k,m}^{\Delta t_1}\}_{k=1}^{25}$ and $\{\hat{\theta}_{k,m}^{\Delta t_2}\}_{k=1}^{25}$. Overall, we have $\{\hat{\delta}_{12,m}\}_{m=1}^{1000}$, which form the simulation-based empirical distribution of the test statistic $\hat{\delta}_{12}$. This distribution can be used to generate critical values or compute empirical p-values. In Table 3, we summarize the quantiles of these empirical distributions, where Q_q represents the q-th quantile. For each distribution, the null hypothesis of normality is rejected at a nominal size of 1% based on the Kolmogorov-Smirnov Goodness-of-Fit Test.

XI Data Analysis

In this section, we focus on analyzing the buy prices $\{P_{1,t}\}$ and sell prices $\{P_{2,t}\}$ of a single stock. The stock we consider is Tiffany Co. (ticker: TIF). The data were obtained from the TAQ database of WRDS. We considered daily transactions between 9:30 AM and 4:00 PM. Overnight durations and returns are ignored, as was also done, for example, by Hasbrouck (1995). The data span the period from January

Table 3: Summary Statistics of the Simulation-based Empirical Distributions

Case	Test-stat	$Q_{0.005}$	$Q_{0.025}$	$Q_{0.05}$	$Q_{0.5}$	$Q_{0.95}$	$Q_{0.975}$	$Q_{0.995}$	Skewness	Excess Kurtosis
$d_\eta = -0.25$	$\hat{\delta}_{12}$	-2.32	-1.80	-1.55	-0.04	1.47	1.64	2.12	0.05	-0.50
	$\hat{\delta}_{13}$	-2.26	-1.83	-1.54	0.02	1.48	1.70	1.96	0.11	-0.52
	$\hat{\delta}_{14}$	-2.12	-1.80	-1.51	0.00	1.39	1.66	2.16	0.02	-0.45
	$\hat{\delta}_{23}$	-2.43	-1.82	-1.60	0.03	1.57	1.80	2.25	0.08	-0.54
	$\hat{\delta}_{24}$	-2.09	-1.80	-1.50	0.03	1.39	1.65	2.22	0.03	-0.51
	$\hat{\delta}_{34}$	-2.17	-1.81	-1.54	-0.02	1.41	1.74	2.38	0.04	-0.48
$d_\eta = -0.75$	$\hat{\delta}_{12}$	-2.05	-1.66	-1.41	0.00	1.42	1.66	2.10	0.01	-0.72
	$\hat{\delta}_{13}$	-2.45	-1.69	-1.55	-0.02	1.41	1.59	2.02	0.10	-0.51
	$\hat{\delta}_{14}$	-2.14	-1.67	-1.34	-0.05	1.36	1.64	2.06	0.01	-0.17
	$\hat{\delta}_{23}$	-2.47	-1.77	-1.59	-0.02	1.40	1.62	2.17	0.07	-0.63
	$\hat{\delta}_{24}$	-2.15	-1.64	-1.37	-0.03	1.39	1.72	2.06	0.01	-0.24
	$\hat{\delta}_{34}$	-2.17	-1.73	-1.44	-0.01	1.41	1.66	2.11	0.02	-0.35
Standard normal	Z	-2.58	-1.96	-1.64	0.00	1.64	1.96	2.58	0.00	0.00

25, 2000 to July 20, 2000, comprising 124 trading days.

We follow Lee and Ready (1991) to classify individual trades. If the transaction price is higher than the prior bid-ask midpoint, the current trade is labeled as a buy order. If the transaction price is lower, it is labeled as a sell order. If the transaction price is exactly the same as the prior bid-ask midpoint, the tick test (described in Lee and Ready 1991) is used to decide whether it should be classified as a buy or sell order. Lee and Ready (1991) found that the accuracy of their method is at least 85%. Using this method, we found 26,103 buy trades and 32,812 sell trades during the period of study.

We first verify that a strong cointegrating relationship exists between buy and sell prices of TIF. The results are given in Table 4. We estimated the memory parameters of the log buy prices and log sell prices as 1 plus the GPH estimator (see Geweke and Porter-Hudak, 1983) of the memory parameter of the differences. We estimated the memory parameter of the cointegrating error using a GPH estimator based on the levels of the residuals from an OLS regression of $\{\log P_{1,j}\}$ on $\{\log P_{2,j}\}$ for various choices of Δt. Note that the memory parameter of the cointegrating error is $1 + \max(d_{\eta_1}, d_{\eta_2})$. The number of frequencies used in the log periodogram regressions was $n^{0.5}$ As expected, the estimated cointegrating parameter is close to 1. Evidence of strong cointegration is observed. Furthermore, there is some evidence that the cointegration is fractional, not standard.

Table 4: Buy and Sell Prices of TIF

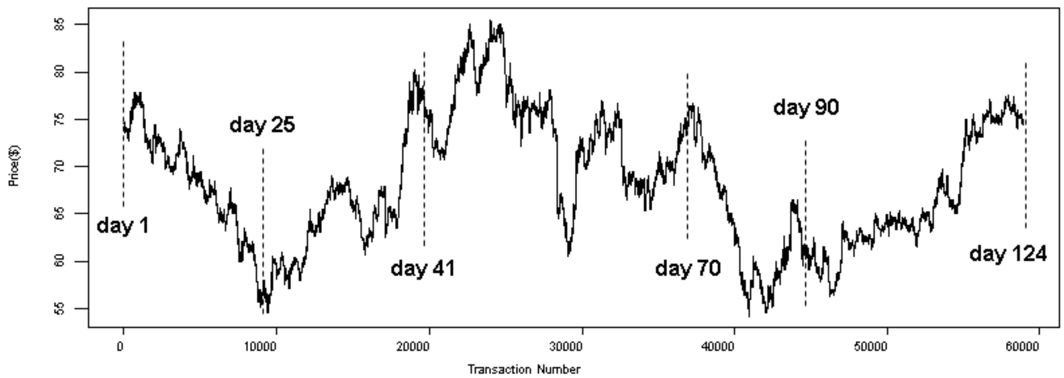
Δt (sec)	n	estimate of θ	$\hat{d}_{\text{buy-price}}[SE]$	$\hat{d}_{\text{sell-price}}[SE]$	$\hat{d}_{\text{coint-error}}[SE]$
1800	1612	$\hat{\theta} = 0.998040$	1.0124 [0.0484]	1.0105 [0.0484]	0.2328 [0.0484]
600	4836	$\hat{\theta} = 0.998046$	1.0223 [0.0330]	1.0208 [0.0330]	0.1312 [0.0330]
300	9672	$\hat{\theta} = 0.998042$	0.9906 [0.0259]	0.9914 [0.0259]	0.1068 [0.0259]
–	–	$\tilde{\theta} = 1.001678$	–	–	–

Next, using the ad hoc estimator $\tilde{\Theta}$, we estimated the model parameters for three clock-time subperiods, as well as the entire period. Period one spans day 1 to day 25, during which the stock price declined by roughly 25%. Period two spans day 41 to day 70, where the price was relatively stable. Period three spans day 90 to day 124, during which the stock price raised by approximately 25%. The results are given in Table 5. The tick-time stock prices are plotted in Figure 2.

Table 5: Method of Moments Parameter Estimates of TIF

Period	Type	# of trades	$\tilde{\sigma}^2_{i,e}(\times 10^{-6})$	$\tilde{\sigma}^2_{i,\eta}(\times 10^{-6})$
1: trading day 1 to 25	Buy	5,852	3.01	3.38
	Sell	6,875	3.05	1.93
2: trading day 41 to 70	Buy	5,360	6.22	0.72
	Sell	7,688	4.08	0.83
3: trading day 90 to 124	Buy	6,896	3.50	1.18
	Sell	8,827	2.00	2.66
entire period	Buy	26,103	4.67	1.26
	Sell	32,812	3.35	1.66

Figure 2: TIF Transaction-level Stock Price



Based on the results in Table 5, we report the following findings:

1) The microstructure noise variance estimates ($\tilde{\sigma}^2_{i,\eta}$) are smaller for subperiod two (during which the stock prices vary substantially but do not show a clear trend), than for subperiods one and three (during which the price showed decreasing and increasing trends, respectively).

2) The value-shock variance estimates ($\tilde{\sigma}^2_{i,e}$) show an opposite pattern, i.e., larger in subperiod two but smaller in subperiods one and three.

3)Comparing $\tilde{\sigma}^2_{i,e}$ and $\tilde{\sigma}^2_{i,\eta}$, the variability of the value shocks usually dominates exceeds that of the microstructure shocks. Indeed, $\tilde{\sigma}^2_{i,e}$ is greater than $\tilde{\sigma}^2_{i,\eta}$ for both buy and sell trades in the entire period.

As for the microstructure noise feedback coefficient estimates, $\tilde{g}_{21}$ and $\tilde{g}_{12}$, their magnitudes are generally around 1, but the signs vary in different subperiods. In some subperiods, the estimates of $\tilde{g}^2_{21}$ or $\tilde{g}^2_{12}$ are negative, thus we set the corresponding $\tilde{g}_{21}$ or $\tilde{g}_{12}$ to be zero. In general, no systematic pattern is observed for $\tilde{g}_{21}$ and $\tilde{g}_{12}$ and their values are not reported in Table 5. We stress, however, that the simulation study in Section X showed that g_{21} and g_{12} are harder to estimate than the other parameters.

The finding 1) is consistent with results from Amihud and Mendelson (1980, 1982), where a market-maker executes buy and sell orders that arrive randomly, with the arrival rate being determined by the quoted bid and ask prices, so as to maximize his expected profit per unit time, under the constraint that his inventory position will not exceed a long and a short positions – L and S respectively. (The analysis applies to traders who act as market makers, that is, quote buying and selling prices and benefit from trading at these prices, rather taking a long-run position in the stock, based on some information.) The market maker sets the pair of bid-ask prices to adjust his inventory, and his policy results in having a preferred inventory position towards which he reverts. The bid-ask spread is minimized at this preferred position while it increases as the inventory diverges from the preferred level. This policy applies when there is no change in information about the security's value, in which case prices show no clear trend, hovering within a range. Amihud and Mendelson (1982, pp. 56-58) analyze a situation of a change in information about the security's value, unknown to the market maker. At first, the market maker maintains the schedule of bid and ask prices that applies to the old valuation, but given the value change, his inventory will deviate from the preferred position and the bid-ask spread will widen. (For example, if the value is lower, the market maker will accumulate a large long position, quoted prices will decline and the bid-ask spread will widen.) After having realized that the value has changed, the market maker shifts his schedule of quoted bid and ask prices and the bid-ask spread reverts to a normal, narrower range. Applying this analysis to the data, subperiods 1 and 3 show a major shift in the security value, reflected in the trend in price. By Amihud and Mendelson (1982), a period of shift in value is associated with wider bid-ask spread. In subperiod 2, when prices vary but do not show a clear trend, the bid-ask spread should be narrower.

A narrower bid-ask spread (a smaller spread magnitude) indicates a smaller microstructure noise variance (see Amihud and Mendelson 1987, *pp.* 536, 547), *i.e.* smaller $\sigma^2_{\eta,i}$ in Model (1). Indeed, we found that the estimated microstructure noise variances are smaller when price fluctuates without a clear trend (subperiod 2) and larger otherwise (subperiods 1 and 3). Unfortunately, it is not possible to test the significance of the change in microstructure noise variances across the three subperiods since the estimates are not independent.

Another interesting question is the price discovery process, which is a popular topic in Finance.

Specifically, here we focus on the price discovery of a single stock, e.g., TIF, during different market environments. To estimate the information share, estimates for the trading intensities λ_1, λ_2, and the value-shock variances $\sigma^2_{1,e}$, $\sigma^2_{2,e}$ are required. To estimate $\lambda_i, (i = 1, 2)$, we use the total number of transactions divided by the total period of observation for asset i. We estimate $\sigma^2_{1,e}$ and $\sigma^2_{2,e}$ by the method of moments as discussed in section VIII. We compute the information share estimates for each of three clock-time subperiods based on results in Table 5. Results are summarized in Table 6.

Table 6: Information Shares (S) of Buy and Sell Price of TIF

Period	$\tilde{S}_{buy}$	$\tilde{S}_{sell}$	$(\tilde{S}_{buy} - \tilde{S}_{sell})$
1: Trading day 1 to 25	45.7%	54.3%	-8.6%
2: Trading day 41 to 70	51.5%	48.5%	3.0%
3: Trading day 90 to 124	57.5%	42.3%	15.2%
Entire Period	52.6%	47.4%	5.2%

For period two, the information shares are approximately equally divided between buys and sells. For period one when the stock price declines dramatically, the sell trades possess more information than buy trades. By contrast, during period three when price is rising, the buy trades have more information. Unfortunately, it is not possible to test the significance of the change in information share across the three subperiods since the estimates are not independent.

As pointed out in Hasbrouck (1995), the information ratios are not related to the microstructure, e.g. spreads, of the markets. This is clear since only the random-walk components of the price series are used in the construction of the information ratios. Therefore, the various results we have presented so far in this section reflect different aspects of the dynamics of the TIF price series.

Finally, we implement of the specification test described in Section IX to test whether model (1) is misspecified. The entire 124 trading days are divided into $K = 62$ subperiods with 2 trading days in each subperiod. We choose four sampling intervals $\Delta t_1 = 1800$, $\Delta t_2 = 600$, $\Delta t_3 = 300$ and $\Delta t_4 = 60$ seconds .

First, we simulation the corresponding empirical distributions of $\hat{\delta}$'s, as in Section X. Results are reported in Table 7.

Table 7: Summary Statistics of the Simulation-based Empirical Distributions for Stock Tiffany (TIF)

Case	Test-stat	$Q_{0.005}$	$Q_{0.025}$	$Q_{0.05}$	$Q_{0.5}$	$Q_{0.95}$	$Q_{0.975}$	$Q_{0.995}$	Skewness	Kurtosis
Tiffany (TIF)	$\hat{\delta}_{12}$	-2.05	-1.68	-1.41	-0.04	1.39	1.69	2.02	0.03	2.45
	$\hat{\delta}_{13}$	-1.98	-1.61	-1.41	-0.03	1.37	1.53	2.04	0.02	2.32
	$\hat{\delta}_{14}$	-2.07	-1.64	-1.43	-0.09	1.28	1.55	2.11	0.08	2.46
	$\hat{\delta}_{23}$	-2.00	-1.70	-1.47	0.00	1.45	1.73	2.27	0.07	2.38
	$\hat{\delta}_{24}$	-1.94	-1.61	-1.44	-0.10	1.34	1.65	2.12	0.13	2.41
	$\hat{\delta}_{34}$	-1.96	-1.68	-1.46	-0.12	1.38	1.69	2.08	0.13	2.39

Based on Table 7, the corresponding simulated test-statistic distributions are used to compute the empirical p-values reported in Table 8 below. For two-sided hypothesis testing with nominal size of 5%,

Table 8: Specification Test for TIF

Δt pair under testing	value of the test statistic [empirical p-value]
Δt_1 vs. Δt_2	1.2356 [0.144]
Δt_1 vs. Δt_3	0.0834 [0.920]
Δt_1 vs. Δt_4	–0.0190 [0.936]
Δt_2 vs. Δt_3	–0.7655 [0.474]
Δt_2 vs. Δt_4	–0.3412 [0.806]
Δt_3 vs. Δt_4	–0.0754 [0.958]

there is no significant evidence to indicate that model (1) is misspecified for the TIF data set, as the null is not rejected in any of the six cases.

XII Modifications of the Model to Capture More Stylized Facts

So far, we have seen that the model (1) yields cointegration, and also captures two stylized facts that have been observed in actual data: volatility clustering, and persistence in durations. In this section, we modify the basic model (1) to capture two additional key stylized facts: the leverage effect, and portfolio autocorrelation due to nonsynchronous trading. Due to limitations in existing theory for point processes, we are currently unable to develop explicit formulas for any of these effects in terms of the model parameters, so we resort primarily to simulations based on the suitably-modified model.

A Portfolio Return Autocorrelation Due to Nonsynchronous Trading

The problem of nonsynchronous trading was first pointed out by Fisher (1966) and the issue has played an important role in the subsequent finance literature. Nonsynchronous trading can adversely affect parameter estimation in the market model, (see, e.g., Scholes and Williams 1977), as well as the estimation of the covariance matrix of the returns (Shanken 1987), and can partially explain the positive autocorrelation of portfolio returns (see, e.g., Atchison, Butler and Simonds 1987, Lo and MacKinlay 1990 a,b, Boudoukh, Richardson and Whitelaw 1994, Kadlec and Patterson 1999).

There are three main approaches to handling nonsynchronous trading in the literature. Scholes and Williams (1977) assumed that, for a given set of equally-spaced time intervals, each asset trades at least once within each time interval. Unfortunately, it is not possible to impose this assumption endogenously, since trading is stochastic. Subsequently, Lo and MacKinlay (1990 a,b) allowed for the possibility of time intervals with no trades, but assumed that the indicator variables for non-trading are serially independent. However, as pointed out by Boudoukh, Whitelaw and Richardson (1994), this is also an unrealistic assumption since the existence of very long durations should be expected to induce positive dependence in the non-trading indicator. In spite of this, Boudoukh, Whitelaw and Richardson (1994) reverted to the even stronger assumption of Scholes and Williams (1977) that there is no nontrading. Nevertheless, in one important respect, the assumptions of Boudoukh, Whitelaw and Richardson (1994) are general, since they allow for cross-sectional dependence of the returns, unlike Lo and MacKinlay (1990 a,b). Recently, Kadlec and Patterson (1999) used a simulation-based approach to assess portfolio autocorrelation due to nonsynchronous trading, in which they use the event times as observed in actual data. Still, Kadlec and Patterson (1999) do not fully endogenize the event times, since if one wanted to run another simulation in their framework, they would have to use the same set of event times.

Up to now, the nontrading mechanism has not been modeled truly endogenously. In this paper, we model the duration process of the price directly, thus endogenize the nontrading mechanism in the price process.

To gain a better picture of the nonsynchronous trading effect implied by our model, we ignore temporarily the microstructure noise. Also, since stock prices may not be cointegrated in general, we change the value shock feedback coefficients in Model (1), $1/\theta$ and θ, to θ_{12} and θ_{21}, respectively. The resulting return series become

$$\begin{aligned} r_{1,j} &= \sum_{k=N_1((j-1)\Delta t)+1}^{N_1(j\Delta t)} e_{1,k} + \sum_{k=N_2(t_{1,N_1((j-1)\Delta t)})+1}^{N_2(t_{1,N_1(j\Delta t)})} \theta_{21} e_{2,k} \\ r_{2,j} &= \sum_{k=N_2((j-1)\Delta t)+1}^{N_2(j\Delta t)} e_{2,k} + \sum_{k=N_1(t_{2,N_2((j-1)\Delta t)})+1}^{N_1(t_{2,N_2(j\Delta t)})} \theta_{12} e_{1,k}. \end{aligned} \tag{9}$$

Lemma 0: *Consider a portfolio consisting of s_1 shares of Asset 1 and s_2 shares of Asset 2, where the returns on the two assets are given by (9). Suppose that $\theta_{12} > 0$ and $\theta_{21} > 0$. Then the lag-1 autocorrelation of the portfolio return is $O(\Delta t^{-1})$ as $\Delta t \to \infty$, and is positive for all values of Δt.*

Table 9 presents simulated averages of the lag-1 autocorrelations of returns of Asset 1, Asset 2 and a portfolio consisting of one share of each asset, i.e., $s_1 = s_2 = 1$, based on 5000 realizations. We also present the minimum and maximum portfolio autocorrelations. The LMSD model implemented here is $\tau_{i,k} = 10e^{h_{i,k}}\epsilon_{i,k}, (i = 1, 2)$. We used $n = 500$, $\theta_{12} = \theta_{21} = 1$, $d_{\tau_1} = d_{\tau_2} = 0.45$ but vary the sampling interval Δt. Other parameter values are the same as described before, unless otherwise listed in the table.

Individual asset returns do not show strong autocorrelation. Nevertheless, the portfolio return has

Table 9: Simulated Lag-1 Autocorrelations to Show Nonsynchronous Trading Effects

	Δt	10	50	200
Asset 1	mean	–0.0018	–0.0014*	–0.0017*
Asset 2	mean	–0.0010	–0.0017*	–0.0014*
Portfolio	mean	0.1077***	0.0894***	0.0376***
	maximum	0.3039	0.3354	0.2369
	minimum	–0.0668	–0.1321	–0.1288

*, ** and *** indicate two-tailed significance at level 5%, 1% and 0.1%, respectively.

significant positive autocorrelation for all sampling intervals Δt considered. The mean autocorrelations range from 0.0376 to 0.1077. The maximum portfolio autocorrelation can be as high as 0.3354. As Δt increases, the portfolio autocorrelation decreases, consistent with the theory described above.

In this paper, we only have two assets. With more assets, it may be possible to obtain far more spurious autocorrelation in the portfolio due to nonsynchronous trading. Empirically, as discussed in Perry (1985), the portfolio lag-1 autocorrelation increases as the number of securities in the portfolio increases. The generalization of our model to the case of $N \geq 3$ assets is beyond the scope of the current paper, but will be the subject of future research.

B The Leverage Effect

The leverage effect is a negative correlation between the current return and future volatility (say, absolute return). We obtain a leverage effect in clock time by introducing a positive lagged cross-correlation between the current value shock e_k and the next-transaction innovation (ν_{k+1}) to the log duration. The moving average representation of the long-memory component $h_{i,k}$ of $\tau_{i,k}$ in the LMSD model for durations can be written as $h_{i,k} = \sum_{j=0}^{\infty} \psi_j \nu_{i,k-j}$ where $\{\psi_j\}$ are constants with $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ and $\{\nu_{i,k}\}$ is an $i.i.d.$ Gaussian series with mean zero and variance $\sigma_{\nu_i}^2$. We will show using simulation that a positive correlation between $\nu_{i,k+1}$ and $e_{i,k}$ in transaction time induces a clock-time leverage effect for the Asset i return.

Specifically, we assume that $e_{i,k} = \sigma_{i,e}(\phi_i \nu_{i,k+1} + w_{i,k})/\sqrt{\phi_i^2 \sigma_{\nu_i}^2 + 1}$, where ψ_i $(i = 1, 2)$ are constants, and the $\{w_{i,k}\}$ are *i.i.d.* standard normal, independent of $\{\nu_{i,k}\}$. Thus, $\mathrm{corr}(e_{i,k}, \nu_{i,k+1}) = \phi_i \sigma_{\nu_i}/\sqrt{\phi_i^2 \sigma_{\nu_i}^2 + 1}$. We assume here that the Asset i durations $\{\tau_{i,k}\}$ follow an LMSD model, $\tau_{i,k} = e^{h_{i,k}} \epsilon_{i,k}$, where $\{h_{i,k}\}$ follow an ARFIMA(0, d_{τ_i}, 0) model and $\{\epsilon_{i,k}\}$, independent of $\{h_{i,k}\}$, are *i.i.d.* exponential with unit mean. A simple calculation yields

$$\mathrm{corr}(e_{i,k}, \tau_{i,k+1}) = \frac{\phi_i \sigma_{\nu_i}^2}{\sqrt{\phi_i^2 \sigma_{\nu_i}^2 + 1}} \cdot \frac{1}{\sqrt{2\, e^{\sigma_{\nu_i}^2 \frac{\Gamma(1-2d_{\tau_i})}{\Gamma^2(1-d_{\tau_i})}} - 1}} \quad .$$

The intuition for why this should produce a leverage effect is that if the current return shock is negative, this induces a below-average shock ν_{k+1} to the log duration, which then persists in the duration series to yield a sequence of below-average durations, i.e., frequent trading in clock time, and above-average volatility.

We verify using simulations that the correlation introduced above yields a leverage effect. For simplicity, we set the microstructure noise to zero. The resulting two-asset return model is given by (9). We simulated $n = 500$ clock-time returns $\{r_{i,j}\}_{j=1}^n$ for each asset, $i = 1, 2$, observed at sampling interval Δt. Sample correlations $\hat{\mathrm{corr}}(r_{i,j}, r_{i,j+1}), \hat{\mathrm{corr}}(|r_{i,j}|, r_{i,j+1})$ and $\hat{\mathrm{corr}}(|r_{i,j}|, r_{i,j-1})$ are calculated for each realization, and the results are averaged, as also done in Andersen, Bollerslev, Frederiksen and Nielsen (2006). We also compared the portfolio return autocorrelations to those simulated under independence of $e_{i,k}$ and $\nu_{i,k+1}$.

Note that $\mathrm{corr}(r_{i,j}, r_{i,j+1})$ is the return lag-1 autocorrelation for Asset $i = 1, 2$, while $\mathrm{corr}(|r_{i,j}|, r_{i,j+1})$ and $\mathrm{corr}(|r_{i,j}|, r_{i,j-1})$ measure the risk-premium effect (RP) and leverage effect (Lev), respectively. Other parameter values used in the simulation are $\theta = 1$, $d_{\tau_1} = d_{\tau_2} = 0.45$, $\sigma_{i,e} = 1$, $\mathrm{var}(\nu_{i,k}) = \frac{\Gamma^2(1-d_\tau)}{\Gamma(1-2d_\tau)}$ so that $\mathrm{var}(h_{i,k}) = 1$ for $i = 1, 2$. Results are based on 5000 realizations, and reported in Table 10.

A positive correlation between $\{e_{i,k}\}$ and $\{\nu_{i,k+1}\}$ induces a significant leverage effect (with the predicted negative sign) for all values of Δt. The magnitude of the leverage effect can be as large as 10%. On the other hand, the magnitude of the simulated risk-premium effect is always much smaller than that of the leverage effect: the corresponding ratio is no larger than 7%. Andersen, et. al. (2006)

Table 10: Risk Premium, Leverage, and Portfolio Autocorrelation from Simulations

			Asset 1		Asset 2		Portfolio
Δt	ϕ_i	$\mathrm{corr}(e_{i,k}, \tau_{i,k+1})$	RP	Lev	RP	Lev	Lag-1 Autocorr
10	0	0	–0.0006	0.0008	–0.0005	–0.0002	0.1077^{***}
	5	0.23	-0.0059^{***}	-0.0924^{***}	-0.0062^{***}	-0.0916^{***}	0.1279^{***}
50	0	0	–0.0005	0.0002	0.0002	–0.0004	0.0894^{***}
	5	0.23	0.0018^{**}	-0.1178^{***}	0.0011	-0.1169^{***}	0.1038^{***}
200	0	0	–0.0008	0.0000	–0.0002	–0.0008	0.0376^{***}
	5	0.23	0.0047^{***}	-0.1097^{***}	0.0043^{***}	-0.1105^{***}	0.0432^{***}

*, ** and *** indicate two-tailed significance at level 5%, 1% and 0.1%, respectively.

concluded from an analysis of 30 blue-chip stocks, there is evidence of a leverage effect, but no convincing evidence of a risk premium effect, so our model is consistent with their findings. The risk premium effect produced by our model, though small, has the interesting property that it is negative for short horizons, but becomes positive for long horizons.

The leverage effect has an impact on the portfolio return autocorrelation, for all sampling frequencies. In each case, the two-sample t-test of equal means for the lag-1 return autocorrelation with and without the leverage effect leads to rejection of the null hypothesis at the 0.1% level. The leverage effect can increase the portfolio return autocorrelation by as much as 2%, as found for $\Delta = 10$. In the Finance literature, it has been concluded that nonsynchronous trading can explain at most part of the portfolio return autocorrelation; see, for example, Lo and MacKinlay (1990 a,b). We feel that this question merits re-investigation, in the light of the model we have proposed in which durations are fully endogenized, and in the light of our current finding of interactions between the leverage effect and nonsynchronous trading effects.

XIII Conclusions

Remark: There is an important caveat regarding the Martingale property in the special case of model (1) in which the microstructure noise components $\{\eta_{1,k}\}$ and $\{\eta_{2,k}\}$ are absent. For each series, as long as the conditioning set involves only returns of the given series up to time t, the log price series (observed at discrete, equally spaced time intervals) is a Martingale. The Martingale property is lost, however, if the conditioning set is augmented to include returns on both assets up to time t. Because of the feedback effect in the model, and the nonsynchronous trading, recent information about Asset 1 can help to predict the Asset 2 return, even though the Asset 2 return is unpredictable based on its own past. Such a situation can occur in actual markets. For example, to predict the (real) return on the sale of a given home, it helps to know the returns on sales of similar homes that have taken place recently, though it may not help at all to know the past returns on sales of the given home, especially if it has not been sold for a long time.

Next, we list a few possibilities for future work stemming from the current project.

It might be interesting to investigate the interplay between cointegration and option pricing, hedging, asset allocation, pairs trading and index tracking in the current pure-jump context. So far, work has been done for option pricing based on pure-jump processes (Prigent, 2001) and dynamic asset allocation based on jump-diffusion processes (Liu, Longstaff and Pan, 2003), but these papers do not allow for cointegration. Another strand of literature has shown that, in a diffusion context, cointegration may have an impact on option pricing (Duan and Pliska, 2004), and on index tracking (Dunis and Ho, 2005; Alexander and Dimitriu, 2005), but these papers do not allow for a pure-jump process.

Other estimators of the cointegrating parameter could be considered, besides OLS. Though many such estimators have been proposed for both standard and fractional cointegration, none have yet been justified under a transaction-level model such as (1). Semiparametric estimators could be considered, since by the remark above the results of this paper do not require a parametric model for durations.

A possible generalization of Model (1) to the case of $W \geq 2$ price series $P_{1,t}, \cdots, P_{W,t}$ is

$$\begin{aligned}
\log P_{1,t} &= \sum_{k=1}^{N_1(t)} (e_{1,k} + \eta_{1,k}) + \sum_{i=2}^{W} \Big\{ \sum_{k=1}^{N_i(t_{1,N_1(t)})} (\theta_{i1} e_{i,k} + g_{i1} \eta_{i,k}) \Big\} \qquad (10) \\
\log P_{2,t} &= \sum_{k=1}^{N_2(t)} (e_{2,k} + \eta_{2,k}) + \sum_{i=1, i\neq 2}^{W} \Big\{ \sum_{k=1}^{N_i(t_{2,N_2(t)})} (\theta_{i2} e_{i,k} + g_{i2} \eta_{i,k}) \Big\} \\
&\vdots \\
\log P_{W,t} &= \sum_{k=1}^{N_W(t)} (e_{W,k} + \eta_{W,k}) + \sum_{i=1}^{W-1} \Big\{ \sum_{k=1}^{N_i(t_{W,N_W(t)})} (\theta_{iW} e_{i,k} + g_{iW} \eta_{i,k}) \Big\}
\end{aligned}$$

where for $i = 1, \cdots, W$ the $\{e_{i,k}\}$ are mutually independent zero-mean *iid* value shock series, the $\{\eta_{i,k}\}$ are mutually independent microstructure shock series satisfying either Condition A, B or C with memory parameters $d_{\eta_i} \in [-1, 0)$, and for $i \neq j$ the parameters θ_{ij} and η_{ij} represent the impact of the value and microstructure shocks from series i on series j.

In the bivariate case $W = 2$, there are two feedback coefficients for the value shocks, θ_{21} and θ_{12}. When cointegration exists, one coefficient is constrained to be the reciprocal of the other, as in Model (1) where $\theta_{21} = \theta$ and $\theta_{12} = 1/\theta$. In the multivariate model (10) there are $W(W-1)$ such feedback coefficients, and there are at most $(W-1)$ cointegrating vectors, though we do not present here the constraints on the coefficients θ_{ij} that would imply a specific cointegrating rank. It would also be of interest to derive a common-components representation for (10) as was obtained for clock-time multivariate models under standard cointegration by Stock and Watson (1998). Such a representation would generalize the representation (2) to the multivariate case, and would presumably facilitate inference on the cointegrating rank as it did in Stock and Watson (1998). Finally, it would be of interest to derive properties for the OLS and other estimators of the cointegrating vectors in (10), as considered for example for OLS in clock-time multivariate models under standard cointegration by Stock (1987).

References

[1] Alexander, C. (1999), Optimal Hedging Using Cointegration, *Philosophical Transactions of the Royal Society A* **357**, 2039–2058.

[2] Alexander, C. and Dimitriu, A. (2005), Indexing, Cointegration and Equity Market Regimes, *Internatiional Journal of Finance and Economics* **10**, 213–231.

[3] Amihud, Y. and H. Mendelson (1980), Dealership Markets: Market-Making with Inventory, *Journal of Financial Economics* **8**, 31–53.

[4] Amihud, Y. and H. Mendelson (1982), Asset Price Behavior in a Dealershipt Market, *Financial Analysts Journal, May/June.*

[5] Amihud, Y. and H. Mendelson (1987), Trading Mechanisms and Stock Returns: An Empirical Investigation, *Journal of Finance* **XLII**, 533–553.

[6] Andersen, T., Bollerslev, T., Frederiksen, P. and Nielsen M. (2006), Continuous-time Models, Realized Volatilities, and Testable Distributional Implications for Daily Stock Returns. Working Paper.

[7] Atchison, M., K. Butler and R. Simonds (1987), Nonsynchronous Security Trading and Market Index Autocorrelation, *Journal of Finance* **42**, 111–118.

[8] Bertelli, S. and M. Caporin (2002), A Note on Calculating Autocovariances of Long-memory Processes, *Journal of Time Series Analysis* **23**, 503–508.

[9] Bollerslev. T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* **31**, 307–327.

[10] Boudoukh, J., M. Richardson and R. Whitelaw (1994), A Tale of Three Schools: Insights on Autcorrelations of Short-Horizon Returns, *Review of Financial Studies* **7**, 539–573.

[11] Bowsher, C.G. (2007), Modelling Security Market Events in Continuous Time: Intensity Based, Multivariate Point Process Models, *Journal of Econometrics* **141**, 876–912.

[12] Breidt, F., Crato, N. and de Lima, P. (1998), The Detection and Estimation of Long Memory in Stochastic Volatility, *Journal of Econometrics* **83**, 325–348.

[13] Chen, W. and Hurvich, C. (2003a), Estimating Fractional Cointegration in the Presence of Polynomial Trends, *Journal of Econometrics* **117**, 95–121.

[14] Chen, W. and Hurvich, C. (2003b), Semiparametric Estimation of Multivariate Fractional Cointegration, *Journal of the American Statistical Association* **98**, 629–642.

[15] Chen, W. and Hurvich, C. (2006), Semiparametric Estimation of Fractional Cointegrating Subspaces, *Annals of Statistics* **34**, 2939–2979.

[16] Clark, P.K. (1973), A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* **41**, 135–155.

[17] Comte, F. (1998), Discrete and continuous time cointegration - Le cas multidimensionnel, *Journal of Econometrics* **88**, 207–226.

[18] Comte, F. and Renault, E. (1996), Long Memory Continuous Time Models, *Journal of Econometrics* **73**, 101–149.

[19] Comte, F. and Renault, E. (1998), Long Memory in Continuous-time Stochastic Volatility Models, *Mathematical Finance* **8**, 291–323.

[20] Daley, D.J. and Vere-Jones, D. (2003), *An Introduction to the Theory of Point Processes,* 2'nd Ed. New York: Springer.

[21] Deo, R., Hsieh, M. and Hurvich, C. (2006), The Persistence of Memory: From Durations to Realized Volatility. Working Paper, Stern School of Business, New York University.

[22] Deo, R., Hurvich, C., Soulier, P. and Wang, Y. (2007), Conditions for the Propagation of Memory Parameter from Durations to Counts and Realized Volatility, to appear in *Econometric Theory.*

[23] Duan, J.C. and Pliska, S. (2004), Option Valuation with Cointegrated Asset Prices, *Journal of Economic Dynamics and Control* **28**, 727–754.

[24] Dueker, M. and Startz, R. (1998), Maximum-likelihood estimation of fractional cointegration with an application to U.S. and Canadian bond rates, *The Review of Economics and Statistics* **80**, 420–426.

[25] Dunis, C.L. and Ho, R. (2005), Cointegration Portfolios of European Equities for Index Tracking and Market Neutral Strateries, *Journal of Asset Management* **6**, 33-52.

[26] Engle, R. and Granger C.W.J. (1987), Co-integration and Error Correction: Representation, Estimation and Testing, *Econometrica* **66**, 1127–1162.

[27] Engle, R. and Russell J. (1998), Autoregressive Conditional Duration: a New Model for Irregularly Spaced Transaction Data, *Econometrica* **66**, 1127–1162.

[28] Fisher, L. (1966), Some New Stock Market Indices, *Journal of Business* **39**, 191–225.

[29] Frijns, B. and Schotman, P. (2006), Price Discovery in Tick Time, Working Paper.

[30] Geweke, J. and Porter-Hudak, S. (1983), The Estimation and Application of Long Memory Time Series Models, *Journal of Time Series Analysis* **4**, 221–237.

[31] Glosten, L.R. (1987), Components of the Bid-Ask Spread and the Statistical Properties of Transaction Prices, *Journal of Finance* **XLII**, 1293–1307.

[32] Harvey, A.C. (1998), Long Memory in Stochastic Volatility, *in: J. Knight, S. Satchell (Eds.), Forecasting Volatility in Financial Markets*, Butterworth-Heineman, Oxford, 307–320.

[33] Hasbrouck, J. (1995), One Security, Many Markets: Determining the Contributions to Price Discovery, *Journal of Finance* **50**, 1175–1199.

[34] Hausman, J. (1978), Specification tests in econometrics, *Econometrica* **46**, 1251–1271.

[35] Iglehart, D.L. and Whitt, W. (1971) The Equivalence of Functional Central Limit Theorems for Counting Processes and Associated Partial Sums. *Annals of Mathematical Statistics* **42**, 1372–1378.

[36] Isserlis, L. (1918), On a Formula for the Product-Moment Coefficient of any Order of a Normal Frequency Distribution in any Number of Variables, *Biometrika* **12**, 134–139.

[37] Johansen, S. (1988), Statistical Analysis of Cointegration Vectors, *Journal of Economic Dynamics and Control* **12**, 231–254.

[38] Johansen, S. (1991), Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, *Econometrica* **59**, 1551–1580.

[39] Kadlec, G. and Patterson, D. (1999), A Transactions Data Analysis of Nonsynchronous Trading, *Review of Financial Studies* **12**, 609–630.

[40] Lieberman, O. and Phillips, P.C.B. (2006), A Complete Asymptotic Series for the Autocovariance Function of a Long Memory Process. Working Paper, Yale University.

[41] Liu, J., Longstaff, F. and Pan, J. (2003), Dynamic Asset Allocation with Event Risk, *Journal of Finance* **58**, 231–259.

[42] Lee, C. and M. Ready (1991), Inferring Trade Direction from Intraday Data, *Journal of Finance* **46**, 733–746.

[43] Lo, A. and A.C. MacKinlay (1990a), An Econometric Analysis of Nonsynchronous Trading, *Journal of Econometrics* **45**, 181–211.

[44] Lo, A. and A.C. MacKinlay (1990b), When are Constrarian Profits Due to Stock Market Overreaction?, *Review of Financial Studies* **3**, 175–205.

[45] Marinucci, D. and Robinson, P.M. (2001), Semiparametric Fractional Cointegration Analysis. Forecasting and empirical methods in finance and macroeconomics, *Journal of Econometrics* **105**, 225–247.

[46] Nijman, T.E., Spierdijk, L. and Soest, Arthur H.O. (2004), Modeling Co-Movements in Trading Intensities to Distinguish Stock and Sector-Specific News, Working Paper, Tilburg University.

[47] Oomen, R.C.A., (2006), Properties of Realized Variance under Alternative Sampling Schemes, *Journal of Business and Economic Statistics* **24**, 219–237.

[48] Perry, P. (1985), Portfolio Serial Correlation and Nonsynchronous Trading, *Journal of Financial and Quantitative Analysis* **20**, 517–523.

[49] Phillips, P.C.B. (1991a), Optimal Inference in Cointegrated Systems, *Econometrica* **59**, 283–306.

[50] Phillips, P.C.B. (1991b), Error Correction and Long-run Equilibrium in Continuous Time, *Econometrica* **59**, 967–980.

[51] Phillips, P.C.B. and Durlauf, S. (1986), Multiple Time Series Regression with Integrated Processes, *The Review of Economic Studies* **53**, 473–495.

[52] Press, S.J. (1967), A Compound Events Model for Security Prices, *Journal of Business* **40**, 317–335.

[53] Prigent, J-L. (2005), Option Pricing with a General Marked Point Process, *Mathematics of Operations Research* **26**, 50–66.

[54] Robinson, P.M. (1995), Gaussian Semiparametric Estimation of Long Range Dependence, *Annals of Statistics* **23**, 1630–1661.

[55] Robinson, P.M. and Haulde, J. (2003), Cointegration in Fractional Systems with Unknown Integration Orders, *Econometrica* **71**, 1727–1766.

[56] Robinson, P.M. and Marinucci, D. (2001), Narrow-band analysis of nonstationary processes, *Annals of Statistics* **29**, 947–986.

[57] Robinson, P.M. and Yajima, Y. (2002), Determination of cointegrating rank in fractional systems, *Journal of Econometrics* **106**, 217–241.

[58] Scholes M. and J. Williams (1977), Estimating Betas from Nonsynchronous Data, *Journal of Financial Economics* **5**, 309–327.

[59] Shanken J. (1987), Nonsynchronous Data and the Covariance-Factor Structure of Returns, *Journal of Finance* **42**, 221–232.

[60] Stock, J. (1987), Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors, *Econometrica* **55**, 1035–1056.

[61] Stock, J. and Watson M. (1988), Testing for Common Trends, *Journal of American Statistical Association* **83**, 1097–1107.

[62] Velasco, C. (2003), Gaussian Semiparametric Estimation of Fractional Cointegration, *Journal of Time Series Analysis* **24**, 345–378.

[63] Velasco, C. and Marmol, F. (2004), Consistent Testing of Cointegration Relationships, *Econometrica* **72**, 1809–1844.

XIV Appendix: Details on the Method of Moments

A The Method of Moments Estimator $\hat{\Theta}$

Consider an example of a pooled sequence of twelve transactions (ordered according to the occurrence in time):

1 1 1 2 1 2 2 2 1 2 1 1

where "1" denotes an Asset 1 transaction and "2" denotes an Asset 2 transaction. We recognize and utilize eight patterns in the pooled transaction sequence in the method of moments.

Pattern 1 is a doublet of Asset 1 transactions: "1 1", which occurs at the first-second, second-third and eleventh-twelfth positions in the example pooled transaction sequence. The return of the second "1" in the doublet has a simple structure, consisting of only the sum of the current Asset 1 value and microstructure disturbances, since there was no intervening Asset 2 transaction between itself and its previous Asset 1 transaction. In addition, the covariance of the returns of such a doublet is equal to the lag-1 autocovariance of the microstructure noise series $\{\eta_{1,k}\}$. This is because we assume $\{e_{1,k}\}$ to be *i.i.d.*, independent of series $\{\eta_{1,k}\}$, and all disturbances of Asset 1 to be independent of those of Asset 2. Similarly, we define *Pattern 2* as a doublet of Asset 2 transactions: "2 2". In the pooled transaction sequence, the sixth-seventh and seventh-eighth doublets are of this type.

Pattern 3 is defined as a triplet of Asset 1 transactions: "1 1 1" (see the first-second-third positions in the pooled transaction sequence as an example). As discussed for Pattern 1, the covariance of the returns of the first and the third transactions in such a triplet is equal to the lag-2 autocovariance of the microstructure noise series $\{\eta_{1,k}\}$. Similarly, we define *Pattern 4* as a triplet of Asset 2 transactions: "2 2 2", which can be used to compute the lag-2 autocovariance of the microstructure noise series $\{\eta_{2,k}\}$.

An example is the sixth-seventh-eighth positions in the pooled transaction sequence.

Next, we define *Pattern 5*. Consider the triplet "1 2 1" as the third-fourth-fifth positions in the pooled transaction sequence example. The last "1" in the triplet contains exactly four disturbances: two value shocks and two microstructure disturbances because there is exactly one intervening Asset 2 transaction. Similarly, *Pattern 6* is defined as a triplet of "2 1 2", which occurs, for instance, at the forth-fifth-sixth positions in the pooled transaction sequence.

Finally, we define *Pattern 7* as a five-transaction sequence of "1 2 1 2 2" and *Pattern 8* as "2 1 2 1 1". Examples of such patterns in the pooled transactions sequence start from the third and the eighth positions, respectively. Consider the third and the fifth transactions in the "1 2 1 2 2" sequence. We call them a g_{21} *pair* since the returns in a g_{21} pair have covariance equal to the product of g_{21} and the lag-2 autocovariance of the microstructure noise series $\{\eta_{2,k}\}$. Similarly, we define a g_{12} *pair* as a pair of transactions as the third and the fifth transactions in a "2 1 2 1 1" sequence.

In summary, we have defined eight sequence patterns as listed in the following table.

Pattern	Transaction Sequence
1	1 1
2	2 2
3	1 1 1
4	2 2 2
5	1 2 1
6	2 1 2
7	1 2 1 2 2
8	2 1 2 1 1

Denote d_{η_i} and α_i as the memory parameter and the AR coefficient of the ARFIMA$(1, d, 0)$ process for the microstructure noise of Asset $i = 1, 2$. The method-of-moments estimates $\hat{\Theta} = (\hat{\sigma}^2_{1,e}, \hat{\sigma}^2_{2,e}, \hat{\sigma}^2_{1,\eta}, \hat{\sigma}^2_{2,\eta}, \hat{g}_{21}, \hat{g}_{12}, \hat{d}_{\eta_1}, \hat{d}_{\eta_2}, \hat{\alpha}_1, \hat{\alpha}_2)$ are given as the solutions to the following system, consisting of ten equations as in (11). Note that we use the tick-level cointegrating parameter estimator $\tilde{\theta}$

in the equations and treat it as if it is the true parameter θ.

$$
\begin{aligned}
\widehat{\text{var}}(\text{second transaction of Pattern 1}) &= \hat{\sigma}_{1,e}^2 + \hat{\sigma}_{1,\eta}^2 \qquad (11)\\
\widehat{\text{var}}(\text{second transaction of Pattern 2}) &= \hat{\sigma}_{2,e}^2 + \hat{\sigma}_{2,\eta}^2 \\
\widehat{\text{cov}}(\text{first and second transactions of Pattern 1}) &= \hat{\sigma}_{1,\eta}^2 \hat{\rho}_{1,1} \\
\widehat{\text{cov}}(\text{first and second transactions of Pattern 2}) &= \hat{\sigma}_{2,\eta}^2 \hat{\rho}_{2,1} \\
\widehat{\text{cov}}(\text{first and third transactions of Pattern 3}) &= \hat{\sigma}_{1,\eta}^2 \hat{\rho}_{1,2} \\
\widehat{\text{cov}}(\text{first and third transactions of Pattern 4}) &= \hat{\sigma}_{2,\eta}^2 \hat{\rho}_{2,2} \\
\widehat{\text{var}}(\text{third transaction of Pattern 5}) &= \hat{\sigma}_{1,e}^2 + \hat{\sigma}_{1,\eta}^2 + \tilde{\theta}^2 \hat{\sigma}_{2,e}^2 + \hat{g}_{21}^2 \hat{\sigma}_{2,\eta}^2 \\
\widehat{\text{var}}(\text{third transaction of Pattern 6}) &= \hat{\sigma}_{2,e}^2 + \hat{\sigma}_{2,\eta}^2 + \frac{1}{\tilde{\theta}^2} \hat{\sigma}_{1,e}^2 + \hat{g}_{12}^2 \hat{\sigma}_{1,\eta}^2 \\
\widehat{\text{cov}}(g_{21} \text{ pairs in Pattern 7}) &= \hat{g}_{21} \hat{\sigma}_{2,\eta}^2 \hat{\rho}_{2,2} \\
\widehat{\text{cov}}(g_{12} \text{ pairs in Pattern 8}) &= \hat{g}_{12} \hat{\sigma}_{1,\eta}^2 \hat{\rho}_{1,2}
\end{aligned}
$$

where $\widehat{\text{var}}$ and $\widehat{\text{cov}}$ are the sample variance and covariance, $\rho_{i,j}$ is the lag-j autocorrelation of the microstructure disturbances $\{\eta_{i,k}\}$ for Asset $i = 1, 2$, and $\hat{\rho}_{i,j}$ is the resulting estimate of $\rho_{i,j}$.

To solve the ten-equation system (11), we start by taking the ratio of the fourth and the ninth equations which gives $\hat{g}_{21}$. Then using $\hat{g}_{21}$ together with the first, the second and the seventh equations, we solve for $\hat{\sigma}_{2,e}^2$ and $\hat{\sigma}_{2,\eta}^2$. Next, we obtain $\hat{\rho}_{2,1}$ and $\hat{\rho}_{2,2}$ from the fourth and sixth equations given $\hat{\sigma}_{2,\eta}^2$, both of which are functions of $\hat{d}_{\eta_2}$ and $\hat{\alpha}_2$. Finally, we solve for $\hat{d}_{\eta_2}$ and $\hat{\alpha}_2$ (if necessary) using for instance nonlinear minimization method. All other parameters can be solved similarly starting from $\hat{g}_{12}$.

B Proof of Theorem 7 on the Consistency of $\hat{\Theta}$

Let $\{R_j\}$ and $\{A_j\}$ be the pooled return series and the pooled asset label series ("1" for Asset 1 and "2" for Asset 2), respectively. Note that, to construct $\{R_j\}$, we first compute transaction-by-transaction returns of Asset 1 and 2 separately, then pool them together according to the order of occurrence in time.

We define eight groups of transactions, satisfying the following conditions,

$$\begin{aligned}
G_1: &\quad A_j = A_{j+1} = 1 \\
G_2: &\quad A_j = A_{j+1} = 2 \\
G_3: &\quad A_j = A_{j+1} = A_{j+2} = 1 \\
G_4: &\quad A_j = A_{j+1} = A_{j+2} = 2 \\
G_5: &\quad A_j = 1, A_{j+1} = 2, A_{j+2} = 1 \\
G_6: &\quad A_j = 2, A_{j+1} = 1, A_{j+2} = 2 \\
G_7: &\quad A_j = 1, A_{j+1} = 2, A_{j+2} = 1, A_{j+3} = A_{j+4} = 2 \\
G_8: &\quad A_j = 2, A_{j+1} = 1, A_{j+2} = 2, A_{j+3} = A_{j+4} = 1
\end{aligned}$$

with $L_i, (i = 1, \ldots, 8)$ the number of values of j satisfying the condition for group i. Then we have ten moment conditions $m_i, (i = 1, \ldots, 10)$ based on G_1 to G_8, corresponding to the ten equations in (11),

$$\begin{aligned}
m_1 &= \frac{1}{L_1} \sum_{j_1 \in G_1} \left[R_{j_1+1}^2 - (\sigma_{1,e}^2 + \sigma_{1,\eta}^2) \right] \\
m_2 &= \frac{1}{L_2} \sum_{j_2 \in G_2} \left[R_{j_2+1}^2 - (\sigma_{2,e}^2 + \sigma_{2,\eta}^2) \right] \\
m_3 &= \frac{1}{L_1} \sum_{j_1 \in G_1} \left[R_{j_1} R_{j_1+1} - \sigma_{1,\eta}^2 \rho_{1,1} \right] \\
m_4 &= \frac{1}{L_2} \sum_{j_2 \in G_2} \left[R_{j_2} R_{j_2+1} - \sigma_{2,\eta}^2 \rho_{2,1} \right] \\
m_5 &= \frac{1}{L_3} \sum_{j_3 \in G_3} \left[R_{j_3} R_{j_3+2} - \sigma_{1,\eta}^2 \rho_{1,2} \right] \\
m_6 &= \frac{1}{L_4} \sum_{j_4 \in G_4} \left[R_{j_4} R_{j_4+2} - \sigma_{2,\eta}^2 \rho_{2,2} \right] \\
m_7 &= \frac{1}{L_5} \sum_{j_5 \in G_5} \left[R_{j_5+2}^2 - (\sigma_{1,e}^2 + \sigma_{1,\eta}^2 + \theta^2 \sigma_{2,e}^2 + g_{21}^2 \sigma_{2,\eta}^2) \right] \\
m_8 &= \frac{1}{L_6} \sum_{j_6 \in G_6} \left[R_{j_6+2}^2 - (\sigma_{2,e}^2 + \sigma_{2,\eta}^2 + \frac{1}{\theta^2} \sigma_{1,e}^2 + g_{12}^2 \sigma_{1,\eta}^2) \right] \\
m_9 &= \frac{1}{L_7} \sum_{j_7 \in G_7} \left[R_{j_7+2} R_{j_7+4} - g_{21} \sigma_{2,\eta}^2 \rho_{2,2} \right] \\
m_{10} &= \frac{1}{L_8} \sum_{j_8 \in G_8} \left[R_{j_8+2} R_{j_8+4} - g_{12} \sigma_{1,\eta}^2 \rho_{1,2} \right]
\end{aligned}$$

To prove the consistency of GMM estimator $\hat{\Theta}$, it is enough to show that all the moment equations $(m_1, \ldots, m_{10})$ converge in probability to zero hence sample moments are consistent estimators of the corresponding moments. To demonstrate, let us consider the first moment equation

$$m_1 = \frac{1}{L_1} \sum_{j_1 \in G_1} \Big[R^2_{j_1+1} - (\sigma^2_{1,e} + \sigma^2_{1,\eta}) \Big]$$

and show that $m_1 \xrightarrow{p} 0$ as the total time span $T \to \infty$.

Consider a new counting process $N^\star(t)$ that counts only the events that belong to G_1. Denote the corresponding intensity by $\lambda^\star$. By the renewal theorem, $\frac{L_1}{\lambda^\star T} = \frac{N^\star(t)}{\lambda^\star T} \xrightarrow{p} 1$ as $T \to \infty$. Therefore, by Slutsky's Theorem, it is enough to show that

$$\frac{1}{\lambda^\star T} \sum_{j_1 \in G_1} \Big[R^2_{j_1+1} - (\sigma^2_{1,e} + \sigma^2_{1,\eta}) \Big] \xrightarrow{p} 0.$$

Thus by Chebyschev's inequality, it suffices to show that

$$\mathrm{var}\Big\{ \sum_{j_1 \in G_1} \Big[R^2_{j_1+1} - (\sigma^2_{1,e} + \sigma^2_{1,\eta}) \Big] \Big\} = o(T^2).$$

Denote

$$S_1(j) = \sum_{k=1}^{j} I\{A_k = 1\}, \qquad S_2(j) = \sum_{k=1}^{j} I\{A_k = 2\}$$

where $I(\cdot)$ is an indicator function. Note that for $j_1 \in G_1$, we have $R_{j_1+1} = e_{1,S_1(j_1+1)} + \eta_{1,S_1(j_1+1)}$, hence by the total variance formula,

$$\begin{aligned}
& \mathrm{var}\Big\{ \sum_{j_1 \in G_1} \Big[R^2_{j_1+1} - (\sigma^2_{1,e} + \sigma^2_{1,\eta}) \Big] \Big\} \\
= \; & \mathrm{var}\Big\{ \underbrace{E\Big\{ \sum_{j_1 \in G_1} \Big[R^2_{j_1+1} - (\sigma^2_{1,e} + \sigma^2_{1,\eta}) \Big] \Big| N^\star(\cdot) \Big\}}_{0} \Big\} + E\Big\{ \mathrm{var}\Big\{ \sum_{j_1 \in G_1} \Big[R^2_{j_1+1} - (\sigma^2_{1,e} + \sigma^2_{1,\eta}) \Big] \Big| N^\star(\cdot) \Big\} \Big\} \\
= \; & E\Big\{ \mathrm{var}\Big\{ \sum_{j_1 \in G_1} \Big[R^2_{j_1+1} - (\sigma^2_{1,e} + \sigma^2_{1,\eta}) \Big] \Big| N^\star(\cdot) \Big\} \Big\}_. = E\Big\{ \mathrm{var}\Big(\sum_{j_1 \in G_1} R^2_{j_1+1} \Big| N^\star(\cdot) \Big) \Big\}_. \\
= \; & E\Big\{ \sum_{j_1 \in G_1} \mathrm{var}(R^2_{j_1+1}) \Big| N^\star(\cdot) \Big\} + E\Big\{ \sum_{j_1 \in G_1} \sum_{j_1' \in G_1, j_1' \neq j_1} \mathrm{cov}(R^2_{j_1+1}, R^2_{j_1'+1}) \Big| N^\star(\cdot) \Big\} \\
= \; & \underbrace{E(L_1) \cdot \mathrm{var}(R^2_{j_1+1})}_{=O(T)} + E\Big\{ \sum_{j_1 \in G_1} \sum_{j_1' \in G_1, j_1' \neq j_1} \mathrm{cov}\Big(\eta^2_{1,S_1(j_1+1)}, \eta^2_{1,S_1(j_1'+1)} \Big) \Big| N^\star(\cdot) \Big\} \\
= \; & O(T) + E\Big\{ \sum_{j_1 \in G_1} \sum_{j_1' \in G_1, j_1' \neq j_1} \Big[\mathrm{cov}\Big(\eta_{1,S_1(j_1+1)}, \eta_{1,S_1(j_1'+1)} \Big) \Big]^2 \Big| N^\star(\cdot) \Big\}
\end{aligned}$$

where we use the independence of the series $\{e_{1,k}\}$ and the Isserlis (1918) formula for Gaussian $\{\eta_{1,k}\}$ in the last step.

Consider $E\Big\{\sum_{j_1\in G_1}\sum_{j_1'\in G_1, j_1'\neq j_1}\Big[\mathrm{cov}\Big(\eta_{1,S_1(j_1+1)},\eta_{1,S_1(j_1'+1)}\Big)\Big]^2\Big|N^\star(\cdot)\Big\}$, we have

$$\begin{aligned}
& E\Big\{\sum_{j_1\in G_1}\sum_{j_1'\in G_1, j_1'\neq j_1}\Big[\mathrm{cov}\Big(\eta_{1,S_1(j_1+1)},\eta_{1,S_1(j_1'+1)}\Big)\Big]^2\Big|N^\star(\cdot)\Big\} \\
\leq \quad & E\Big\{\mathrm{var}(\eta_{1,k})\cdot\sum_{j_1\in G_1}\underbrace{\sum_{j_1'\in G_1, j_1'\neq j_1}\Big|\mathrm{cov}\Big(\eta_{1,S_1(j_1+1)},\eta_{1,S_1(j_1'+1)}\Big)\Big|}_{\text{summable since } d_{\eta_1}\in[-1,-\frac{1}{2})\bigcup(-\frac{1}{2},0)}\Big|N^\star(\cdot)\Big\} = O(T)
\end{aligned}$$

thus $\mathrm{var}\Big\{\sum_{j_1\in G_1}\Big[R^2_{j_1+1}-(\sigma^2_{1,e}+\sigma^2_{1,\eta})\Big]\Big\} = O(T) = o(T^2)$ and $m_1 \xrightarrow{p} 0$ as $T\to\infty$.

Similarly, it can be shown that

$$m_i \xrightarrow{p} 0, \quad (i = 1,\ldots,10),$$

and therefore $\hat{\Theta} \xrightarrow{p} \Theta$ as $T\to\infty$.

C An Alternative Ad Hoc Estimator $\tilde{\Theta}$

Although the method of moments estimator $\hat{\Theta}$ is consistent, simulations not shown here indicate that a very large number of trades must be observed in order for $\hat{\Theta}$ to yield physically meaningful and accurate estimates. The difficulty is mainly due to the final two equations in (11). They are based on subsequences of five trades, which occur far less frequently than the two-trade or three-trade subsequences used in the other equations. In practice, then, if the number of trades is not very large, g_{21} and g_{12} may be poorly estimated by the method of moments, resulting in negative variance estimates in $\hat{\Theta}$. For the same set of parameter values and time spans used in the simulation study in Section X, more than 90% of the realizations lead to negative variance estimates in $\hat{\Theta}$. Furthermore, a dramatic increase in the time span of the data set was not feasible for us due to computational constraints.

We therefore propose an alternative estimator $\tilde{\Theta}$ which, in part, avoids using the final two equations in (11). The method is ad hoc, but simulations reported in Section X indicate that it gives reasonable estimates.

We start by taking the ratio of the third and the fifth equation in (11), which gives us a numerical estimate of $\rho_{1,1}/\rho_{1,2}$, which we denote by $\tilde{\rho}_{1,1}/\tilde{\rho}_{1,2}$. Then, on a grid of values of (d, α), we computed the corresponding ratio $\rho_{1,1}/\rho_{1,2}$ for the ARFIMA$(1, d, 0)$ model with parameters (d, α). We used the algoritm of Bertelli and Caporin (2002) to compute $\rho_{1,1}$ and $\rho_{1,2}$, since there is no attractive closed form for the autocovariances of an ARFIMA$(1, d, 0)$ process. The supports of (d, α) are $(-1, -\frac{1}{2}) \cup (-\frac{1}{2}, 0)$ and $(-1, 1)$, respectively. Next, we construct $\tilde{d}_{\eta_1}$ and $\tilde{\alpha}_1$ such that $\left|\frac{\tilde{\rho}_{1,1}}{\tilde{\rho}_{1,2}} - \frac{\rho_{1,1}}{\rho_{1,2}}\right|$ is minimized, *i.e.*

$$(\tilde{d}_{\eta_1}, \tilde{\alpha}_1) = \min_{d,\alpha} \left|\frac{\tilde{\rho}_{1,1}}{\tilde{\rho}_{1,2}} - \frac{\rho_{1,1}}{\rho_{1,2}}\right|.$$

In addition, $\tilde{\rho}_{1,1}$ and $\tilde{\rho}_{1,2}$ are obtained. Similarly, we obtain $(\tilde{d}_{\eta_2}, \tilde{\alpha}_2)$, as well as $\tilde{\rho}_{2,1}$ and $\tilde{\rho}_{2,2}$. We then obtain the remaining parameter estimates in $\tilde{\Theta}$ from (11). Using $\tilde{\rho}_{1,1}$ in the third equation of (11), we get $\tilde{\sigma}^2_{1,\eta}$ which then is used in the first equation to get $\tilde{\sigma}^2_{1,e}$. Similarly, we obtain $\tilde{\sigma}^2_{2,\eta}$ and $\tilde{\sigma}^2_{2,e}$. Next, we obtain $\tilde{g}^2_{21}$ and $\tilde{g}^2_{12}$ based on the seventh and the eighth equations in (11). At this point, we have obtained $\tilde{g}^2_{21}$ and $\tilde{g}^2_{12}$ as well as all entries of $\tilde{\Theta}$ except for $\tilde{g}_{12}$ and $\tilde{g}_{21}$ using only the first eight equations of (11). Finally, we use the last two equations in (11) (which are inherently less accurate than the others since they are based on five-trade sequences) to determine the signs of $\tilde{g}_{21}$ and $\tilde{g}_{12}$.

XV Appendix: Proofs

A Lemmas

From (2) it can be seen that the microstructure components of the log price are random sums of the microstructure noise. Lemmas 1 and 3 below show for the weak and strong fractional cointegration cases, respectively, that such random sums have memory parameter $1 + d_\eta < 1$, where d_η is the memory parameter of the microstructure noise.

Lemma 1 *Suppose that* $\{\eta_k\}$ *has memory parameter* $d_\eta \in (-\frac{1}{2}, 0)$*, is independent of* $\{\tau_k\}$*, and satisfies Condition A. Then*

$$\operatorname{var}(\sum_{k=1}^{N(t)} \eta_k) \sim (\tilde{\sigma}^2 \lambda^{2d_\eta+1}) t^{2d_\eta+1}$$

as $t \to \infty$.

The following lemma is used for proving Lemma 1

Lemma 2 *For* $d_\eta \in (-\frac{1}{2}, 0)$*, suppose that* $\{\eta_k\}$ *satisfies Condition A. Then there exists a positive constant* C *such that for all nonnegative integers* s*,* $\operatorname{var}(\sum_{k=1}^{s} \eta_k) = \tilde{\sigma}^2 s^{2d_\eta+1} + R(s)$*, where* $|R(s)| \leq C s^{\max(2d_\eta+1-\beta,0)}$.

Lemma 3 *For* $d_\eta \in (-1, -\frac{1}{2})$*, suppose that* $\{\eta_k\}$ *satisfies Condition B, and is independent of* $N(\cdot)$*. Then for any fixed* $t > 0$*,*

$$\operatorname{cov}(\sum_{k=1}^{N(t)} \eta_k, \sum_{k=1}^{N(t+j)} \eta_k) \sim C j^{2d_\eta+1} Pr\{N(t) > 0\} \tag{12}$$

as $j \to \infty$ *where* $C > 0$ *is a constant not depending on* t.

The following two lemmas are used in the proofs of Theorems 3, 4 and 5.

Lemma 4 *If the durations* $\{\tau_k\}$ *are generated by a Long Memory Stochastic Duration (LMSD) model with memory parameter* $d_\tau \in (0, \frac{1}{2})$ *and all moments of the durations* $\{\tau_k\}$ *are finite, then all moments of the backward recurrence time* (BRT_t)*, as defined in (23), are also finite.*

Lemma 5 *For durations* $\{\tau_k\}$ *satisfying the assumptions in Lemma 4,* $E[N(s)^m] \leq K_m(s^m + 1)$ *for all* $s > 0$*, where* $K_m < \infty$*,* $m = 1, 2, \cdots$.

B Proofs of Lemmas

Proof of Lemma 1: Since $N(\cdot)$ is independent of $\{\eta_k\}$, conditioning on $N(\cdot)$, we obtain

$$\begin{aligned}\mathrm{var}(\sum_{k=1}^{N(t)} \eta_k) &= E(\sum_{k=1}^{N(t)} \eta_k)^2 = E\Big[E\Big((\sum_{k=1}^{N(t)} \eta_k)^2|N(\cdot)\Big)\Big] = E\Big[\mathrm{var}\Big(\sum_{k=1}^{N(t)} \eta_k|N(\cdot)\Big)\Big] \\ &= E\{\tilde{V}[N(t)]\},\end{aligned}$$

where $\tilde{V}(s) = \mathrm{var}(\sum_{k=1}^{s} \eta_k)$. In Lemma 2, we show that there exists a positive constant C such that for all nonnegative integers s, $\tilde{V}(s) = \tilde{\sigma}^2 s^{2d_\eta+1} + R(s)$, where $|R(s)| \leq C s^{\max(2d_\eta+1-\beta,0)}$. Thus,

$$\mathrm{var}(\sum_{k=1}^{N(t)} \eta_k) = E\{\tilde{V}[N(t)]\} = \tilde{\sigma}^2 E\Big\{[N(t)]^{2d_\eta+1}\Big\} + E[R(N(t))], \tag{13}$$

where

$$|E[R(N(t))]| \leq E|R(N(t))| \leq CE\Big\{[N(t)]^{\max(2d_\eta+1-\beta,0)}\Big\}. \tag{14}$$

We evaluate $E\Big\{[N(t)]^{2d_\eta+1}\Big\}$ in (13) as follows. Denote $Z(t) = \frac{N(t)-\lambda t}{t^{\frac{1}{2}+d_\tau}}$. As shown by Deo, Hurvich, Soulier and Wang (2007, in the proof of Theorem 1) using Iglehart and Whitt (1971, Theorem 1), $Z(t) \xrightarrow{D} CB_{d_\tau+\frac{1}{2}}(1)$ as $t \to \infty$, where $\xrightarrow{D}$ denotes converge in distribution and C is a positive constant. Since $d_\tau < \frac{1}{2}$, as $t \to \infty$,

$$\frac{N(t)}{\lambda t} = 1 + \frac{1}{\lambda} t^{d_\tau - \frac{1}{2}} Z(t) \xrightarrow{p} 1$$

and thus $\Big[\frac{N(t)}{\lambda t}\Big]^{2d_\eta+1} \xrightarrow{p} 1$.

Next, we will prove that

$$E\Big[\frac{N(t)}{\lambda t}\Big]^{2d_\eta+1} \to 1 \tag{15}$$

by showing that $\limsup_t E\Big[\frac{N(t)}{\lambda t}\Big]^{2d_\eta+1+\delta} < \infty$ for some positive δ. Since $d_\eta < 0$, we choose $\delta = 3 - 2d_\eta > 0$. Using the fact that for all real x

$$\begin{aligned}(1+x)^4 &= [(1+x)^2]^2 \leq (2+2x^2)^2 = 4(x^4+2x^2+1) \\ &\leq 4[x^4 + (x^4+1) + 1] = 8(x^4+1)\end{aligned}$$

we obtain that, for $t \geq 1$

$$\Big[\frac{N(t)}{\lambda t}\Big]^{2d_\eta+1+\delta} = \Big[1 + \frac{1}{\lambda} t^{d_\tau-\frac{1}{2}} Z(t)\Big]^4 \leq \Big[1 + \Big|\frac{1}{\lambda} Z(t)\Big|\Big]^4 \leq 8 + \frac{8}{\lambda^4} Z^4(t) \qquad .$$

By Lemma 2 in Deo, Hurvich, Soulier and Wang (2007), $\limsup_t E[Z^4(t)] < \infty$. Therefore, $\limsup_t E\Big[\frac{N(t)}{\lambda t}\Big]^{2d_\eta+1+\delta} < \infty$ and we obtain (15). Similarly,

$$E\Big[\frac{N(t)}{\lambda t}\Big]^{\max(2d_\eta+1-\beta,0)} \to 1.$$

From (13) and (14), we obtain

$$\frac{\mathrm{var}(\sum_{k=1}^{N(t)}\eta_k)}{(\lambda t)^{2d_\eta+1}} = \tilde{\sigma}^2 E\Big[\frac{N(t)}{\lambda t}\Big]^{2d_\eta+1} + E\Big[\frac{R(N(t))}{(\lambda t)^{2d_\eta+1}}\Big] \quad \to \quad \tilde{\sigma}^2 > 0. \qquad \square$$

Proof of Lemma 2: Without any loss of generality, we can approximate the behavior of the spectral density of $\{\eta_k\}$ around zero frequency by that of a fractional Gaussian noise with variance $\tilde{\sigma}^2$, and Hurst exponent $d_\eta + \frac{1}{2}$. If $\{\eta_k^*\}$ obeys this fully-parametric model (see Samorodnitsky and Taqqu 1994, pp. 332–339), then $\mathrm{var}(\sum_{k=1}^{s}\eta_k^*) = \tilde{\sigma}^2 s^{2d_\eta+1}$. Denoting the spectral density of $\{\eta_k^*\}$ by $f_{FGN}(\lambda)$, then $f(\lambda)/f_{FGN}(\lambda) \to 1$ as $\lambda \to 0+$, and since $\lambda^{-2d_\eta}/f_{FGN}(\lambda)$ is twice-differentiable at zero, we can without loss of generality write the spectral density of $\{\eta_k\}$ as

$$f(\lambda) = f_{FGN}(\lambda) + g(\lambda)$$

where $g(\lambda) = O(\lambda^{-2d_\eta+\beta})$ as $\lambda \to 0^+$ and $\int_{-\pi}^{\pi}|g(\lambda)|d\lambda < \infty$, the latter bound following from weak stationarity of $\{\eta_k\}$. We therefore have, for all non-negative integers s,

$$\mathrm{var}(\sum_{k=1}^{s}\eta_k) = 2\pi s\int_{-\pi}^{\pi}F_s(\lambda)f(\lambda)d\lambda = \tilde{\sigma}^2 s^{2d_\eta+1} + R(s)$$

where $R(0) = 0$ and for $s \geq 1$, $R(s) = \int_{-\pi}^{\pi}F_s(\lambda)g(\lambda)d\lambda$, with

$$F_s(\lambda) = \frac{1}{2\pi s}\frac{\sin^2(s\lambda/2)}{\sin^2(\lambda/2)}.$$

There exist positive constants $\epsilon \in (0,\pi)$, C_1 and C_2 such that $|g(\lambda)| \leq C_1|\lambda|^{-2d_\eta+\beta}$ for $\lambda \in [0,\epsilon]$ and such that $sF_s(\lambda) \leq C_2$ for $\lambda \in [\epsilon,\pi]$. Thus,

$$|R(s)| \leq 4\pi s\int_0^{\epsilon}F_s(\lambda)|g(\lambda)|d\lambda + 4\pi s\int_{\epsilon}^{\pi}F_s(\lambda)|g(\lambda)|d\lambda$$

$$\leq 4\pi s C_1\int_0^{\epsilon}F_s(\lambda)|\lambda|^{-2d_\eta+\beta}d\lambda + 4\pi C_2\int_{\epsilon}^{\pi}|g(\lambda)|d\lambda$$

$$\leq 4\pi s C_1 \int_0^{\pi} F_s(\lambda)|\lambda|^{-2d_\eta+\beta} d\lambda + 4\pi C_2 \int_0^{\pi} |g(\lambda)| d\lambda.$$

The second term on the righthand side is a finite positive constant. The first term is $C_1 \text{var}(\sum_{k=1}^{s} \tilde{\eta}_k)$ where $\{\tilde{\eta}_k\}$ is an $I(d_\eta - \beta/2)$ process. This variance is $O(s^{2d_\eta-\beta+1})$ if $d_\eta - \beta/2 \in (-\frac{1}{2}, 0)$, and is $O(1)$ if $d_\eta - \beta/2 \in (-3/2, -\frac{1}{2})$. Thus, overall, there exists a positive constant C such that $|R(s)| \leq Cs^{\max(2d_\eta+1-\beta,0)}$. $\square$

Proof of Lemma 3: For $j, k \geq 1$, define $K_0 = \text{var}(\varphi_k)$, and write $\text{cov}(\varphi_k, \varphi_{k+j}) = Kj^{2d_\varphi-1} + R(j)$, where $|R(j)| \leq K_1 j^{2d_\varphi-3}$, with $K_1 > 0$. We have

$$\begin{aligned}
&\text{cov}(\sum_{k=1}^{N(t)} \eta_k, \sum_{k=1}^{N(t+j)} \eta_k) = E\Big[E\Big\{\sum_{k=1}^{N(t)} \eta_k \sum_{k=1}^{N(t+j)} \eta_k \Big| N(\cdot)\Big\}\Big] \\
= \;& E\Big[E\big\{\varphi_{N(t)}\varphi_{N(t+j)} I\{N(t)>0\}\Big| N(\cdot)\big\}\Big] \\
= \;& E\Big[\Big(K\Delta N_{t,t+j}^{2d_\varphi-1} \cdot I\{\Delta N_{t,t+j}>0\} + R(\Delta N_{t,t+j}) \cdot I\{\Delta N_{t,t+j}>0\} + K_0 \cdot I\{\Delta N_{t,t+j}=0\}\Big) I\{N(t)>0\}\Big]
\end{aligned}$$

where $\Delta N_{t,t+j} = N(t+j) - N(t)$ and I is an indicator function.

As shown by (48), $E[I\{\Delta N_{t,t+j} = 0\}] = P(\Delta N_{t,t+j} = 0)$ has *nearly-exponential decay* (defined just before Theorem 5) as $j \to \infty$. Furthermore, since

$$\begin{aligned}
\Big|E\Big[R(\Delta N_{t,t+j}) \cdot I\{\Delta N_{t,t+j}>0\} \cdot I\{N(t)>0\}\Big]\Big| \;&\leq\; E\Big[|R(\Delta N_{t,t+j})| \cdot I\{\Delta N_{t,t+j}>0\}\Big] \\
&\leq\; E\Big[K_1 \Delta N_{t,t+j}^{2d_\varphi-3} \cdot I\{\Delta N_{t,t+j}>0\}\Big] \quad ,
\end{aligned}$$

it is sufficient to show that

$$E\Big[\Delta N_{t,t+j}^{2d_\varphi-1} \cdot I\{\Delta N_{t,t+j}>0\} \cdot I\{N(t)>0\}\Big] \sim Cj^{2d_\varphi-1} Pr\{N(t)>0\} \tag{16}$$

and

$$E\Big[\Delta N_{t,t+j}^{2d_\varphi-3} \cdot I\{\Delta N_{t,t+j}>0\}\Big] \sim Cj^{2d_\varphi-3} \quad . \tag{17}$$

First, we show (16). From the proof of Lemma 1, we know $\frac{\Delta N_{t,t+j}}{\lambda j} \xrightarrow{p} 1$ as $j \to \infty$. Since $P\{\Delta N_{t,t+j} = 0\}$ has nearly exponential decay, $I\{\Delta N_{t,t+j} > 0\} \xrightarrow{p} 1$ as $j \to \infty$. Hence,

$$\Big(\frac{\Delta N_{t,t+j}}{\lambda j}\Big)^{2d_\varphi-1} \cdot I\{\Delta N_{t,t+j}>0\} \xrightarrow{p} 1, \qquad \text{as } j \to \infty$$

and (16) will follow from uniform integrability if it can be shown that

$$\sup_{j\geq 4} E\Big[\Big(\frac{\Delta N_{t,t+j}}{\lambda j}\Big)^{-\alpha}\cdot I\{\Delta N_{t,t+j}>0\}\Big]<\infty$$

where $\alpha=(1-2d_{\varphi}+\delta)\in(0,1)$ for some $\delta>0$, with the convention $\infty\cdot 0=0$. Since by Chung (1974) Theorem 3.2.1, p. 42,

$$E\Big[\Big(\frac{\Delta N_{t,t+j}}{\lambda j}\Big)^{-\alpha}\cdot I\{\Delta N_{t,t+j}>0\}\Big]\leq 1+\sum_{s=1}^{\infty}P\Big[\Big(\frac{\Delta N_{t,t+j}}{\lambda j}\Big)^{-\alpha}\cdot I\{\Delta N_{t,t+j}>0\}\geq s\Big]$$

it suffices to show that

$$\sup_{j\geq 4}\sum_{s=1}^{\infty}P\Big[\Big(\frac{\Delta N_{t,t+j}}{\lambda j}\Big)^{-\alpha}\cdot I\{\Delta N_{t,t+j}>0\}\geq s\Big]<\infty\ . \tag{18}$$

Next, we establish (18). As in the proof of Lemma 2 in Deo, Hurvich, Soulier and Wang (2007), we set $\lambda=1$ without loss of generality. Furthermore, in view of the stationarity of the counting process, we replace the integer variable j by the continuous variable t, replace $\Delta N_{t,t+j}$ by $N(t)$, and prove

$$\sup_{t\geq 4}\sum_{s=1}^{\infty}P\Big[\Big(\frac{N(t)}{t}\Big)^{-\alpha}\cdot I\{N(t)>0\}\geq s\Big]<\infty. \tag{19}$$

Consider a fixed value of $t\geq 4$. As $s\to\infty$, $s^{-1/\alpha}\to 0$. Thus the infinite sum on s reduces to a finite sum, up to the greatest integer such that $ts^{-1/\alpha}\geq 1$. Note that

$$\Big[\Big(\frac{N(t)}{t}\Big)^{-\alpha}\cdot I\{N(t)>0\}\Big]\geq s\quad\Longleftrightarrow\quad\frac{N(t)}{t}\leq s^{-1/\alpha}\qquad\text{and}\qquad N(t)>0\ .$$

Thus, (19) becomes

$$\sup_{t\geq 4}\sum_{s=1}^{\lfloor t^{\alpha}\rfloor}P\Big[N(t)\leq ts^{-1/\alpha}\text{ and }N(t)>0\Big]<\infty\ .$$

It suffices to show that

$$\sup_{t\geq 4}\sum_{s=1}^{\lfloor t^{\alpha}\rfloor}P\Big[N(t)<2ts^{-1/\alpha}\Big]<\infty\ . \tag{20}$$

To prove (20), we split the sum into two parts,

$$\sup_{t\geq 4}\sum_{s=1}^{\lfloor t^{\alpha}\rfloor}P\Big[N(t)<2ts^{-1/\alpha}\Big]\leq\underbrace{\sup_{t\geq 4}\sum_{s=1}^{\lfloor s^{\star}\rfloor}P\Big[N(t)<2ts^{-1/\alpha}\Big]}_{G_1}+\underbrace{\sup_{t\geq 4}\sum_{s=\lfloor s^{\star}\rfloor+1}^{\lfloor t^{\alpha}\rfloor}P\Big[N(t)<2ts^{-1/\alpha}\Big]}_{G_2}$$

where $s^\star = \max\Big(\big(\frac{2t}{t-1}\big)^\alpha, 8^\alpha\Big)$. The term G_1 is finite since $\sum_{s=1}^{\lfloor s^\star\rfloor} P\Big[N(t) < 2ts^{-1/\alpha}\Big] \leq s^\star$. Next, we consider G_2. Since for all real positive k, $N(t) < k \Longleftrightarrow \sum_{i=1}^{\lceil k\rceil} u_i > t$, we have

$$P\Big[N(t) < 2ts^{-1/\alpha}\Big] = P\Big[\sum_{i=1}^{\lceil 2ts^{-\frac{1}{\alpha}}\rceil} u_i > t\Big] = P\Big[\sum_{i=1}^{\lceil v(t,s)\rceil} u_i > t\Big]$$

where $v(t,s) = 2ts^{-\frac{1}{\alpha}}$. For s in the range $(\lfloor s^\star\rfloor + 1)$ to $\lfloor t^\alpha\rfloor$, $(t - 2ts^{-1/\alpha} - 1) > 0$, hence by Chebyshev's inequality,

$$\begin{aligned} P\Big[\sum_{i=1}^{\lceil v(t,s)\rceil} u_i > t\Big] &= P\Big(W > \frac{t - \lceil v(t,s)\rceil}{\lceil v(t,s)\rceil^{\frac{1}{2}+d_\tau}}\Big) \leq P\Big(|W| > \frac{t - 2ts^{-1/\alpha} - 1}{(2ts^{-1/\alpha}+1)^{\frac{1}{2}+d_\tau}}\Big) \\ &\leq E(|W|^{2+\epsilon}) \cdot \frac{(2ts^{-1/\alpha}+1)^{(\frac{1}{2}+d_\tau)(2+\epsilon)}}{(t - 2ts^{-1/\alpha} - 1)^{(2+\epsilon)}} \end{aligned}$$

where $\epsilon > 0$ is a positive constant, $d_\tau \in (0, \frac{1}{2})$ is the memory parameter of the durations and

$$W = \frac{\sum_{i=1}^{\lceil v(t,s)\rceil} u_i - \lceil v(t,s)\rceil}{\lceil v(t,s)\rceil^{\frac{1}{2}+d_\tau}}\ .$$

For s in the range $(\lfloor s^\star\rfloor + 1)$ to $\lfloor t^\alpha\rfloor$, the smallest value $2ts^{-1/\alpha}$ can achieve is $\min(t-1, 2)$, thus $2ts^{-1/\alpha} > 1$, and $(1 - 4s^{-1/\alpha}) > \frac{1}{2}$ (since $s \geq 8^\alpha$). We obtain

$$\begin{aligned} \sum_{s=\lfloor s^\star\rfloor+1}^{\lfloor t^\alpha\rfloor} P\Big[\sum_{i=1}^{\lceil v(t,s)\rceil} u_i > t\Big] &\leq \sum_{s=\lfloor s^\star\rfloor+1}^{\lfloor t^\alpha\rfloor} E(|W|^{2+\epsilon}) \cdot \frac{(2ts^{-1/\alpha}+1)^{(\frac{1}{2}+d_\tau)(2+\epsilon)}}{(t - 2ts^{-1/\alpha} - 1)^{(2+\epsilon)}} \\ &\leq E(|W|^{2+\epsilon}) \cdot \sum_{s=\lfloor s^\star\rfloor+1}^{\lfloor t^\alpha\rfloor} \frac{(4ts^{-1/\alpha})^{(\frac{1}{2}+d_\tau)(2+\epsilon)}}{(\frac{1}{2}t - 2ts^{-1/\alpha})^{(2+\epsilon)}} \\ &= CE(|W|^{2+\epsilon}) \cdot t^{(2+\epsilon)(-\frac{1}{2}+d_\tau)} \cdot \sum_{s=\lfloor s^\star\rfloor+1}^{\lfloor t^\alpha\rfloor} \frac{s^{-\frac{1}{\alpha}(\frac{1}{2}+d_\tau)(2+\epsilon)}}{(1 - 4s^{-1/\alpha})^{(2+\epsilon)}} \\ &\leq CE(|W|^{2+\epsilon}) \cdot t^{(2+\epsilon)(-\frac{1}{2}+d_\tau)} \sum_{s=\lfloor s^\star\rfloor+1}^{\lfloor t^\alpha\rfloor} s^{-\frac{1}{\alpha}(\frac{1}{2}+d_\tau)(2+\epsilon)} \end{aligned}$$

with

$$-\frac{1}{\alpha}(\frac{1}{2} + d_\tau)(2+\epsilon) < -1 \tag{21}$$

and $(2+\epsilon)(-\frac{1}{2} + d_\tau) < 0$. Thus,

$$\sum_{s=\lfloor s^\star\rfloor+1}^{\lfloor t^\alpha\rfloor} P\Big[\sum_{i=1}^{\lceil v(t,s)\rceil} u_i > t\Big]$$

is uniformly bounded for $t \geq 4$, provided that $\sup_{t\geq 4, s\geq 1} E(|W|^{2+\epsilon}) < \infty$. But this condition is already proved in equation (25) of Deo, Hurvich, Soulier and Wang (2007).

Overall, we obtain that $G_2 < \infty$, so that (20) and (16) are proved.

To establish (17), we follow the same steps as above, but using $\alpha = (3 - 2d_\varphi + \delta) \in (2, 3)$. The proof goes through without change, except that for (21) to hold, we need to select $\epsilon > 4$. $\square$

Proof of Lemma 4: First, by exercise 3.4.1 on page 59 of Daley and Vere-Jones (2002),

$$BRT_t \overset{d}{\equiv} u_1 \tag{22}$$

where $\overset{d}{\equiv}$ denotes equivalence in distribution and u_1 is the time of occurrence of the first transaction following time zero. Since $0 < u_1 \leq \tau_1$, and we have assumed that all moments of τ_1 are finite, $E(BRT_t^m) = E(u_1^m) \leq E(\tau_1^m) = C < \infty$ for all $m > 0$. $\square$

Proof of Lemma 5: By Proposition 1 in Deo, Hurvich, Soulier and Wang (2007), and the fact that for $a > 0, b > 0$ and positive integer m, $(a+b)^m \leq 2^{m-1}(a^m + b^m)$ (which can be shown using Jensen's inequality and the convexity of the function $x^m, x > 0$), we obtain that, for $s > 0$,

$$\begin{aligned} E[N(s)^m] &= E\{[\lambda s + Z(s)s^{\frac{1}{2}+d_\tau}]^m\} \leq E\{[\lambda s + |Z(s)|s^{\frac{1}{2}+d_\tau}]^m\} \\ &\leq 2^{m-1}\left[\lambda^m s^m + E|Z(s)|^m s^{m(\frac{1}{2}+d_\tau)}\right] \leq K_m(s^m + 1), \end{aligned}$$

where K_m is a finite constant, $Z(s) = \frac{N(s)-\lambda s}{s^{\frac{1}{2}+d_\tau}}$ and λ as defined before. $\square$

C Proof of Theorem 1

We first consider the fractional cointegration case, $d_\eta \in (-\frac{1}{2}, 0)$. We focus on $\log P_{1,t}$, since the proof for $\log P_{2,t}$ follows along similar lines.

The log price of Asset 1 is

$$\log P_{1,t} = \sum_{k=1}^{N_1(t)} (e_{1,k} + \eta_{1,k}) + \sum_{k=1}^{N_2(t_{1,N_1(t)})} (\theta e_{2,k} + g_{21}\eta_{2,k}) \quad .$$

Note that the two terms on the righthand side are uncorrelated. By Lemma 1, since $d_\eta < 0$, we obtain

$$\begin{aligned} \mathrm{var}[\sum_{k=1}^{N_1(t)}(e_{1,k}+\eta_{1,k})] &= \sigma_{1,e}^2 E[N_1(t)] + \mathrm{var}[\sum_{k=1}^{N_1(t)}\eta_{1,k}] \\ &\sim (\sigma_{1,e}^2\lambda_1)t + (\sigma_{1,\eta}^2\lambda_1^{2d_{\eta_1}+1})t^{2d_{\eta_1}+1} = (\sigma_{1,e}^2\lambda_1)t + o(t). \end{aligned}$$

Next, consider $E\{N_1(t) - N_1(t_{2,N_2(t)})\}$, which is the expected number of transactions of Asset 1 after the most recent transaction of Asset 2 up to time t. Define the *backward recurrence time* for Asset 2 at time t as

$$BRT_{2,t} = \inf\{s > 0 : N_2(t) - N_2(t-s) > 0\}. \tag{23}$$

Clearly, $BRT_{2,t} = t - t_{2,N_2(t)}$. By stationarity of $N_2(\cdot)$, and using (3.1.7), page 43 of Daley and Vere-Jones (2003), we obtain $E\{N_1(t) - N_1(t_{2,N_2(t)})\} = E[-N_1(-BRT_{2,t})] = E[-N_1(-BRT_{2,0})]$. In the righthand equality, we used the fact that, since $N_2(\cdot)$ is a stationary point process, $BRT_{2,t}$ has the same distribution as $BRT_{2,0}$, which does not depend on t. (See Daley and Vere-Jones (2002), page 58–59 for a detailed discussion.) Thus

$$E\{N_1(t) - N_1(t_{2,N_2(t)})\} = \tilde{C}_1, \tag{24}$$

a finite constant, independent of t. Similarly

$$E\{N_2(t) - N_2(t_{1,N_1(t)})\} = \tilde{C}_2 \tag{25}$$

is also a finite constant, independent of t as well.

Using (13),

$$\begin{aligned} \mathrm{var}\Big[\sum_{k=1}^{N_2(t_{1,N_1(t)})}(\theta e_{2,k} + g_{21}\eta_{2,k})\Big] &= \theta^2\sigma_{2,e}^2 \underbrace{E\{N_2(t_{1,N_1(t)})\}}_{T_1} + g_{21}^2\sigma_{2,\eta}^2 \underbrace{E\Big\{[N_2(t_{1,N_1(t)})]^{2d_\eta+1}\Big\}}_{T_2} \\ &\quad + g_{21}^2\sigma_{2,\eta}^2 \underbrace{E\{R(N_2(t_{1,N_1(t)}))\}}_{T_3}. \end{aligned}$$

By (25), the first term equals

$$T_1 = E\{N_2(t)\} - \tilde{C}_2 = \lambda_2 t - \tilde{C}_2 \sim \lambda_2 t,$$

as $t \to \infty$.

As for the second term, since when $x > 0$ and $0 < p = (2d_\eta + 1) < 1$, the function x^p is concave, we can apply Jensen's inequality to obtain

$$T_2 \leq \Big\{E[N_2(t_{1,N_1(t)})]\Big\}^{2d_\eta+1} = (\lambda_2 t - \tilde{C}_2)^{2d_\eta+1} = o(t).$$

Using (14) and arguing as above, we obtain

$$|T_3| \leq CE\Big\{[N_2(t_{1,N_1(t)})]^{\max(2d_\eta+1-\beta,0)}\Big\} = o(t).$$

Therefore,

$$\text{var}\Big[\sum_{k=1}^{N_2(t_{1,N_1(t)})} (\theta e_{2,k} + g_{21}\eta_{2,k})\Big] \sim (\theta^2\sigma^2_{2,e}\lambda_2)t$$

as $t \to \infty$.

Overall,

$$\text{var}[\log P_{1,t}] \sim (\sigma^2_{1,e}\lambda_1)t + (\theta^2\sigma^2_{2,e}\lambda_2)t = C_1 t$$

where $C_1 = (\sigma^2_{1,e}\lambda_1 + \theta^2\sigma^2_{2,e}\lambda_2)$.

Similarly,

$$\text{var}[\log P_{2,t}] \sim (\sigma^2_{2,e}\lambda_2)t + (\frac{1}{\theta^2}\sigma^2_{1,e}\lambda_1)t = C_2 t$$

where $C_2 = (\sigma^2_{2,e}\lambda_2 + \frac{1}{\theta^2}\sigma^2_{1,e}\lambda_1)$.

Next, for both the strong fractional cointegration case ($d_\eta \in (-1, -\frac{1}{2})$) and the standard cointegration case ($d_\eta = -1$), the proof is identical to that for the weak fractional cointegration case, except that here we have $\text{var}(\sum_{k=1}^{N_i(t)} \eta_{i,k}), (i = 1, 2)$, equal to some finite constants, which does not increase with t. $\square$

D Proof of Theorem 2

We first consider the fractional cointegration case, $d_\eta \in (-\frac{1}{2}, 0)$. We focus on the returns $\{r_{1,j}\}$ of Asset 1, which corresponds to the first equation in (6) since the proof for $\{r_{2,j}\}$ follows along similar lines.

Consider the lag-1 autocorrelation of

$$r_{1,j} = \underbrace{\sum_{k=N_1[(j-1)\Delta t]+1}^{N_1[j\Delta t]} e_{1,k}}_{T_1} + \underbrace{\sum_{k=N_1[(j-1)\Delta t]+1}^{N_1[j\Delta t]} \eta_{1,k}}_{T_2} + \underbrace{\sum_{k=N_2(t_{1,N_1((j-1)\Delta t)})+1}^{N_2(t_{1,N_1(j\Delta t)})} \theta e_{2,k}}_{T_3} + \underbrace{\sum_{k=N_2(t_{1,N_1((j-1)\Delta t)})+1}^{N_2(t_{1,N_1(j\Delta t)})} g_{21}\eta_{2,k}}_{T_4} .$$

Denote $\Delta N_{1,j} = N_1(j\Delta t) - N_1((j-1)\Delta t)$ and $\Delta N_{2,j} = N_2(j\Delta t) - N_2((j-1)\Delta t)$. We know that $E(\Delta N_{1,j}) = \lambda_1 \Delta t$ and $E(\Delta N_{2,j}) = \lambda_2 \Delta t$. Thus,

$$\begin{aligned} \mathrm{var}(T_1) &= E\Big\{\Big[\sum_{k=N_1((j-1)\Delta t)+1}^{N_1(j\Delta t)} e_{1,k}\Big]^2\Big\} = E\Big[E\Big\{\Big[\sum_{k=N_1((j-1)\Delta t)+1}^{N_1(j\Delta t)} e_{1,k}\Big]^2\Big|N_1(\cdot)\Big\}\Big] \\ &= \sigma_{1,e}^2 E\{N_1(j\Delta t) - N_1((j-1)\Delta t)\} \\ &= \sigma_{1,e}^2 E(\Delta N_{1,j}) = \sigma_{1,e}^2 \lambda_1 \Delta t \quad . \end{aligned} \tag{26}$$

By Lemma 2, we have

$$\mathrm{var}(T_2) = \sigma_{2,\eta}^2 E[\{\Delta N_{1,j}\}^{2d_\eta+1}] + \sigma_{2,\eta}^2 E\{R(\Delta N_{1,j})\}.$$

Since the function x^p is concave when $x > 0$ and $0 < p < 1$, by Jensen's inequality for $d_\eta \in (-0.5, 0)$, the first part satisfies

$$\sigma_{2,\eta}^2 E[\{\Delta N_{1,j}\}^{2d_\eta+1}] \le \sigma_{2,\eta}^2 \{E[\Delta N_{1,j}]\}^{2d_\eta+1} = \sigma_{2,\eta}^2 \{\lambda_1 \Delta t\}^{2d_\eta+1} = o(\Delta t), \tag{27}$$

as $\Delta t \to \infty$. As for the second part,

$$\Big|E\{R(\Delta N_{1,j})\}\Big| \le CE\Big\{(\Delta N_{1,j})^{\max(2d_\eta+1-\beta,0)}\Big\} \le CE\Big\{(\Delta N_{1,j})^{2d_\eta+1}\Big\} = o(\Delta t).$$

Hence, $\mathrm{var}(T_2) = o(\Delta t)$ as $\Delta t \to \infty$.

Next, by Lemma 2 and equations (24) and (25),

$$\begin{aligned} \mathrm{var}(T_3) &= \theta^2 \sigma_{2,e}^2 E\{N_2(t_{1,N_1(j\Delta t)}) - N_2(t_{1,N_1((j-1)\Delta t)})\} \\ &= \theta^2 \sigma_{2,e}^2 E[N_2(j\Delta t) - N_2((j-1)\Delta t)] \\ &= \theta^2 \sigma_{2,e}^2 E[\Delta N_{2,j}] = \theta^2 \sigma_{2,e}^2 \lambda_2 \Delta t \end{aligned} \tag{28}$$

and

$$\begin{aligned}
\mathrm{var}(T_4) &= g_{21}^2\sigma_{2,\eta}^2 E\Big[\{\underbrace{N_2(t_{1,N_1(j\Delta t)})-N_2(t_{1,N_1((j-1)\Delta t)})}_{J}\}^{2d_\eta+1}\Big] + g_{21}^2\sigma_{2,\eta}^2 E\Big\{R(J)\Big\} \\
&\le g_{21}^2\sigma_{2,\eta}^2\Big[E\{N_2(t_{1,N_1(j\Delta t)})-N_2(t_{1,N_1((j-1)\Delta t)})\}\Big]^{2d_\eta+1} + g_{21}^2\sigma_{2,\eta}^2 E\Big\{R(J)\Big\} \\
&= g_{21}^2\sigma_{2,\eta}^2\Big\{E[\Delta N_{2,j}]\Big\}^{2d_\eta+1} + g_{21}^2\sigma_{2,\eta}^2 E\Big\{R(J)\Big\} \\
&= g_{21}^2\sigma_{2,\eta}^2(\lambda_2\Delta t)^{2d_\eta+1} + g_{21}^2\sigma_{2,\eta}^2 E\Big\{R(J)\Big\} = o(\Delta t). \qquad (29)
\end{aligned}$$

since by Lemma 2, $E\Big\{R(J)\Big\} \le C\Big\{E[\Delta N_{2,j}]\Big\}^{2d_\eta+1} = o(\Delta t)$.

As for the covariance terms, by the Cauchy-Schwartz inequality and Equations (26) to (29),

$$\begin{aligned}
|\mathrm{cov}(T_1,T_2)| &\le \sqrt{\mathrm{var}(T_1)\mathrm{var}(T_2)} \le \sqrt{\sigma_{1,e}^2\sigma_{2,\eta}^2(\lambda_1\Delta t)^{2d_\eta+2}} = o(\Delta t) \qquad (30)\\
|\mathrm{cov}(T_1,T_4)| &\le \sqrt{\mathrm{var}(T_1)\mathrm{var}(T_4)} \le \sqrt{g_{21}^2\sigma_{1,e}^2\sigma_{2,\eta}^2(\lambda_1\Delta t)(\lambda_2\Delta t)^{2d_\eta+1}} = o(\Delta t) \qquad (31)\\
|\mathrm{cov}(T_2,T_3)| &\le \sqrt{\mathrm{var}(T_2)\mathrm{var}(T_3)} \le \sqrt{\theta^2\sigma_{2,e}^2\sigma_{2,\eta}^2(\lambda_2\Delta t)(\lambda_1\Delta t)^{2d_\eta+1}} = o(\Delta t) \qquad (32)\\
|\mathrm{cov}(T_2,T_4)| &\le \sqrt{\mathrm{var}(T_2)\mathrm{var}(T_4)} \le \sqrt{g_{21}^2\sigma_2^4(\lambda_1\Delta t)^{2d_\eta+1}(\lambda_2\Delta t)^{2d_\eta+1}} = o(\Delta t) \qquad (33)\\
|\mathrm{cov}(T_3,T_4)| &\le \sqrt{\mathrm{var}(T_3)\mathrm{var}(T_4)} \le \sqrt{\theta^2 g_{21}^2\sigma_{2,e}^2\sigma_{2,\eta}^2(\lambda_2\Delta t)^{2d_\eta+2}} = o(\Delta t) \qquad (34)
\end{aligned}$$

since $d_\eta < 0$. Also,

$$\mathrm{cov}(T_1,T_3) = 0 \qquad (35)$$

since $\{e_{1,k}\}$ and $\{e_{2,k}\}$ are mutually independent *i.i.d.* series.

Overall, by (26) to (35), we obtain $\mathrm{var}(r_{1,j}) \sim (\sigma_{1,e}^2\lambda_1 + \theta^2\sigma_{2,e}^2\lambda_2)\Delta t$, as $\Delta t \to \infty$, i.e.

$$\lim_{\Delta t\to\infty}\frac{\mathrm{var}(r_{1,j})}{\Delta t} = (\sigma_{1,e}^2\lambda_1 + \theta^2\sigma_{2,e}^2\lambda_2).$$

Similarly, for

$$(r_{1,j}+r_{1,j+1}) = \sum_{k=N_1((j-1)\Delta t)+1}^{N_1((j+1)\Delta t)}(e_{1,k}+\eta_{1,k}) + \sum_{k=N_2(t_{1,N_1((j-1)\Delta t)})+1}^{N_2(t_{1,N_1((j+1)\Delta t)})}(\theta e_{2,k}+g_{21}\eta_{2,k})$$

we obtain

$$\mathrm{var}(r_{1,j}+r_{1,j+1}) \sim 2(\sigma_{1,e}^2\lambda_1 + \theta^2\sigma_{2,e}^2\lambda_2)\Delta t$$

i.e.

$$\lim_{\Delta t \to \infty} \frac{\mathrm{var}(r_{1,j} + r_{1,j+1})}{2\Delta t} = (\sigma_{1,e}^2 \lambda_1 + \theta^2 \sigma_{2,e}^2 \lambda_2).$$

Therefore,

$$\begin{aligned} \mathrm{corr}(r_{1,j}, r_{1,j+1}) = \frac{\mathrm{cov}(r_{1,j}, r_{1,j+1})}{\mathrm{var}(r_{1,j})} &= \frac{\frac{1}{2}\mathrm{var}(r_{1,j} + r_{1,j+1}) - \mathrm{var}(r_{1,j})}{\mathrm{var}(r_{1,j})} = \frac{\frac{1}{2}\mathrm{var}(r_{1,j} + r_{1,j+1})}{\mathrm{var}(r_{1,j})} - 1 \\ &= \frac{\frac{\mathrm{var}(r_{1,j}+r_{1,j+1})}{2\Delta t}}{\frac{\mathrm{var}(r_{1,j})}{\Delta t}} - 1 \to 0, \end{aligned}$$

as $\Delta t \to \infty$.

The fact that the lag-2 autocorrelation also converges to zero can be shown by recognizing that

$$\mathrm{corr}(r_{1,j}, r_{1,j+2}) = \frac{1}{2}\Big[\frac{\mathrm{var}(r_{1,j} + r_{1,j+1} + r_{1,j+2})}{\mathrm{var}(r_{1,j})} - 3 - 4\mathrm{corr}(r_{1,j}, r_{1,j+1})\Big]$$

and using the lag-1 autocorrelation results proved above as well as

$$\lim_{\Delta t \to \infty} \frac{\mathrm{var}(r_{1,j} + r_{1,j+1} + r_{1,j+2})}{3\Delta t} = (\sigma_{1,e}^2 \lambda_1 + \theta^2 \sigma_{2,e}^2 \lambda_2).$$

The result follows for any fixed lag k by induction.

Next, for both the strong fractional cointegration case ($d_\eta \in (-1, -\frac{1}{2})$) and the standard cointegration case ($d_\eta = -1$), the proof is identical to that for the weak fractional cointegration case, except that here we have $\mathrm{var}(\sum_{k=N_i((j-1)\Delta t)+1}^{N_i(j\Delta t)} \eta_{i,k})$, $(i = 1, 2)$ (as well as other similar terms) equal to some finite constants, which does not increase with Δt. $\square$

E Proof of Theorem 3

Consider a linear combination of $\log P_{1,t}$ and $\log P_{2,t}$ using vector $(1, -\theta)$,

$$
\begin{aligned}
& \log P_{1,t} - \theta \log P_{2,t} \\
= & \sum_{k=N_1(t_{2,N_2(t)})+1}^{N_1(t)} e_{1,k} - \theta \sum_{k=N_2(t_{1,N_1(t)})+1}^{N_2(t)} e_{2,k} \\
& + \sum_{k=1}^{N_1(t)} \eta_{1,k} - \theta g_{12} \sum_{k=1}^{N_1(t_{2,N_2(t)})} \eta_{1,k} - \theta \sum_{k=1}^{N_2(t)} \eta_{2,k} + g_{21} \sum_{k=1}^{N_2(t_{1,N_1(t)})} \eta_{2,k} \\
= & \underbrace{\sum_{k=N_1(t_{2,N_2(t)})+1}^{N_1(t)} e_{1,k}}_{T_1} - \theta \underbrace{\sum_{k=N_2(t_{1,N_1(t)})+1}^{N_2(t)} e_{2,k}}_{T_2} + (1 - \theta g_{12}) \underbrace{\sum_{k=1}^{N_1(t)} \eta_{1,k}}_{T_3} \\
& + \theta g_{12} \underbrace{\sum_{k=N_1(t_{2,N_2(t)})+1}^{N_1(t)} \eta_{1,k}}_{T_4} - (\theta - g_{21}) \underbrace{\sum_{k=1}^{N_2(t)} \eta_{2,k}}_{T_5} - g_{21} \underbrace{\sum_{k=N_2(t_{1,N_1(t)})+1}^{N_2(t)} \eta_{2,k}}_{T_6}.
\end{aligned} \tag{36}
$$

Since all shock series are mutually independent and also independent of the counting processes $N_1(t)$ and $N_2(t)$, we obtain

$$
\begin{aligned}
\mathrm{var}\Big[\log P_{1,t} - \theta \log P_{2,t} \Big] \quad = \quad & \mathrm{var}(T_1) + \theta^2 \mathrm{var}(T_2) + (1 - \theta g_{12})^2 \mathrm{var}(T_3) + \theta^2 g_{12}^2 \mathrm{var}(T_4) \\
& + 2\theta g_{12}(1 - \theta g_{12}) \mathrm{cov}(T_3, T_4) + (\theta - g_{21})^2 \mathrm{var}(T_5) + g_{21}^2 \mathrm{var}(T_6) \\
& + 2 g_{21}(\theta - g_{21}) \mathrm{cov}(T_5, T_6).
\end{aligned} \tag{37}
$$

First, by Lemma 1

$$
\begin{aligned}
\mathrm{var}(T_3) \quad & \sim \quad (\sigma_{1,\eta}^2 \lambda_1^{2d_\eta+1}) t^{2d_\eta+1} \\
\mathrm{var}(T_5) \quad & \sim \quad (\sigma_{2,\eta}^2 \lambda_2^{2d_\eta+1}) t^{2d_\eta+1} \qquad .
\end{aligned} \tag{38}
$$

Using (24) and Lemma 2, we obtain

$$\begin{aligned}
\mathrm{var}(T_4) &= \sigma_{1,\eta}^2 E\Big[\{N_1(t) - N_1(t_{2,N_2(t)})\}^{2d_\eta+1}\Big] + \sigma_{1,\eta}^2 E\Big\{R\Big[N_1(t) - N_1(t_{2,N_2(t)})\Big]\Big\} \\
&\leq \sigma_{1,\eta}^2 \Big[E\{N_1(t) - N_1(t_{2,N_2(t)})\}\Big]^{2d_\eta+1} + \sigma_{1,\eta}^2 C\Big[E\{N_1(t) - N_1(t_{2,N_2(t)})\}\Big]^{2d_\eta+1} \\
&= (1+C)\sigma_{1,\eta}^2 \tilde{C}_1^{2d_\eta+1} \qquad (39)
\end{aligned}$$

where we apply Jensen's inequality in the last inequality, noting that for $x > 0$ and $0 < p = (2d_\eta + 1) < 1$, the function x^p is concave. Similarly,

$$\mathrm{var}(T_6) \leq (1+C)\sigma_{2,\eta}^2 \tilde{C}_2^{2d_\eta+1}. \qquad (40)$$

Also, by (24) and (25)

$$\begin{aligned}
\mathrm{var}(T_1) &= \mathrm{var}(e_{1,k}) E\{N_1(t) - N_1(t_{2,N_2(t)})\} = \sigma_{1,e}^2 \tilde{C}_1 \qquad (41) \\
\mathrm{var}(T_2) &= \mathrm{var}(e_{2,k}) E\{N_2(t) - N_2(t_{1,N_1(t)})\} = \sigma_{2,e}^2 \tilde{C}_2. \qquad (42)
\end{aligned}$$

Next, we consider the covariance terms in (37) using Cauchy-Schwartz inequality. By (38) and (39)

$$|\mathrm{cov}(T_3, T_4)| \leq \sqrt{\mathrm{var}(T_3)\mathrm{var}(T_4)} \leq \sqrt{(1+C)\sigma_{1,\eta}^2 \tilde{C}_1^{2d_\eta+1} \mathrm{var}(T_3)} = o(t^{2d_\eta+1}) \qquad (43)$$

and similarly by (38) and (40)

$$|\mathrm{cov}(T_5, T_6)| \leq \sqrt{\mathrm{var}(T_5)\mathrm{var}(T_6)} \leq \sqrt{(1+C)\sigma_{2,\eta}^2 \tilde{C}_2^{2d_\eta+1} \mathrm{var}(T_5)} = o(t^{2d_\eta+1}). \qquad (44)$$

Overall, using (38) to (44) for (37), we obtain

$$\mathrm{var}\Big(\log P_{1,t} - \theta \log P_{2,t}\Big) \sim C t^{2d_\eta+1} \qquad (45)$$

where $C = (1 - \theta g_{12})^2 (\sigma_{1,\eta}^2 \lambda_1^{2d_{\eta_1}+1}) + (\theta - g_{21})^2 (\sigma_{2,\eta}^2 \lambda_2^{2d_{\eta_2}+1})$.

The cointegrating vector is $(1, -\theta)$ and the memory parameter decreases from 1 for both log prices to $1 + d_\eta$. $\square$

F Proof of Theorem 4

The proof follows along the same lines as the proof of Theorem 3, except that we now use Lemma 3 to obtain the asymptotic behavior of the autocovariances of the partial sums of the microstructure noise.

As in the proof of Theorem 3 we have

$$\begin{aligned}
& \log P_{1,t} - \theta \log P_{2,t} \\
= & \sum_{k=N_1(t_{2,N_2(t)})+1}^{N_1(t)} e_{1,k} - \theta \sum_{k=N_2(t_{1,N_1(t)})+1}^{N_2(t)} e_{2,k} \\
& + \sum_{k=1}^{N_1(t)} \eta_{1,k} - \theta g_{12} \sum_{k=1}^{N_1(t_{2,N_2(t)})} \eta_{1,k} - \theta \sum_{k=1}^{N_2(t)} \eta_{2,k} + g_{21} \sum_{k=1}^{N_2(t_{1,N_1(t)})} \eta_{2,k} \\
= & \underbrace{\sum_{k=N_1(t_{2,N_2(t)})+1}^{N_1(t)} e_{1,k}}_{T_{1,t}} - \theta \underbrace{\sum_{k=N_2(t_{1,N_1(t)})+1}^{N_2(t)} e_{2,k}}_{T_{2,t}} + (1-\theta g_{12}) \underbrace{\sum_{k=1}^{N_1(t)} \eta_{1,k}}_{T_{3,t}} \\
& + \theta g_{12} \underbrace{\sum_{k=N_1(t_{2,N_2(t)})+1}^{N_1(t)} \eta_{1,k}}_{T_{4,t}} - (\theta - g_{21}) \underbrace{\sum_{k=1}^{N_2(t)} \eta_{2,k}}_{T_{5,t}} - g_{21} \underbrace{\sum_{k=N_2(t_{1,N_1(t)})+1}^{N_2(t)} \eta_{2,k}}_{T_{6,t}} .
\end{aligned} \tag{46}$$

The dominant term in $\text{cov}(\log P_{1,t} - \theta \log P_{2,t}, \log P_{1,t+j} - \theta \log P_{2,t+j})$, based on Lemma 3, is

$$(1-\theta g_{12})^2 \text{cov}(T_{3,t}, T_{3,t+j}) + (\theta - g_{21})^2 \text{cov}(T_{5,t}, T_{5,t+j})$$

$$\sim C j^{2d_\eta+1} [(1-\theta g_{12})^2 Pr\{N_1(t) > 0\} + (\theta - g_{21})^2 Pr\{N_2(t) > 0\}] \ .$$

It suffices to show that $\text{cov}(T_{L,t}, T_{M,t+j}) = o(j^{2d_\eta+1})$ for $L, M \in 1, \cdots, 6$, with $(L, M) \neq (3,3)$ and $(L, M) \neq (5,5)$. We will explicitly consider the cases with $L \leq M$, as the proofs for the remaining cases are similar.

It will be shown in the proof of Theorem 5 that $\text{cov}(T_{1,t}, T_{1,t+j})$ and $\text{cov}(T_{2,t}, T_{2,t+j})$ both have nearly exponential decay. Independence assumptions made previously imply that $\text{cov}(T_{L,t}, T_{M,t+j}) = 0$ for $(L, M) = (1, M)$ with $M \neq 1$, for $(L, M) = (2, M)$ with $M \neq 2$, and also for $(L, M) = (3,5), (3,6), (4,5), (4,6)$.

Let $\tilde{\Delta}_{t,t+j} = N_1(t_{2,N_2(t+j)}) - N_1(t)$. Note that $\tilde{\Delta}_{t,t+j}$ is a possibly negative integer. First, we show that for fixed t, $\tilde{\Delta}_{t,t+j}/j \xrightarrow{p} \lambda_1$ as $j \to \infty$. We have

$$\frac{\tilde{\Delta}_{t,t+j}}{j} \stackrel{D}{=} \frac{\tilde{\Delta}_{0,j}}{j} = \frac{N_1(t_{2,N_2(j)})}{j} = \frac{N_1(t_{2,N_2(j)})}{t_{2,N_2(j)}} \frac{t_{2,N_2(j)}}{N_2(j)} \frac{N_2(j)}{j} \xrightarrow{p} \lambda_1 \frac{1}{\lambda_2} \lambda_2 = \lambda_1$$

as $j \to \infty$, since $t_{2,N_2(j)} \to \infty$ and $N_2(t_{2,N_2(j)}) = N_2(j)$.

We show next that $Pr\{\tilde{\Delta}N_{t,t+j} \leq 0\}$ has nearly exponential decay for fixed t as $j \to \infty$. Define $Z_1(j) = \frac{N_1(j) - \lambda_1 j}{j^{1/2+d_\tau}}$, and $R(j) = N_1(t_{2,N_2(j)}) - N_1(j)$. It was shown in Deo, Hurvich, Soulier and Wang (2007) that $E|Z_1(j)|^m$ is bounded uniformly in j for all m. So is $E|R(j)|^m$ by Lemmas 4 and 5. (See also the proof of Theorem 5, part 2.ii). Thus, so is $E|\tilde{R}(j)|^m$ where $\tilde{R}(j) = R(j)/j^{1/2+d_\tau}$. By stationarity of $(N_1(\cdot), N_2(\cdot))$,

$$\begin{aligned} &Pr\{\tilde{\Delta}N_{t,t+j} \leq 0\} = Pr\{\tilde{\Delta}N_{0,j} \leq 0\} = Pr\{N_1(t_{2,N_2(j)}) \leq 0\} = Pr\{N_1(j) + R(j) \leq 0\} \\ &\leq P\Big[|Z_1(j) + \tilde{R}(j)| \geq \lambda_1 j^{\frac{1}{2}-d_\tau}\Big] \leq \frac{2^{m-1}E|Z_1(j)|^m + 2^{m-1}E|\tilde{R}(j)|^m}{\lambda_1^m j^{m(\frac{1}{2}-d_\tau)}} = O(j^{m(d_\tau - \frac{1}{2})}). \end{aligned}$$

Consider $\text{cov}(T_{4,t}, T_{4,t+j})$. Note that

$$T_{4,t} = \varphi_{1,N_1(t)} \cdot I\{N_1(t) > 0\} - \varphi_{1,N_1(t_{2,N_2(t)})} \cdot I\{N_1(t_{2,N_2(t)}) > 0\}.$$

Thus, $\text{cov}(T_{4,t}, T_{4,t+j})$ will consist of four terms. We will consider two of these in detail. The first term is

$$E\left[E[\varphi_{1,N_1(t)} \cdot I\{N_1(t) > 0\}\, \varphi_{1,N_1(t+j)} \big| N_1(\cdot)]\right] \sim C j^{2d_\eta+1} Pr\{N_1(t) > 0\}$$

as $j \to \infty$, where C is the constant on the righthand side of (16) in the proof of Lemma 3. The next term is

$$-E\left[E[\varphi_{1,N_1(t)} \cdot I\{N_1(t) > 0\}\, \varphi_{1,N_1(t_{2,N_2(t+j)})} \cdot I\{N_1(t_{2,N_2(t+j)}) > 0\} \big| N_1(\cdot), N_2(\cdot)]\right]. \tag{47}$$

As in the proof of Lemma 3, the inside conditional mean in (47) is

$$\begin{aligned} &\Big(K\tilde{\Delta}N_{t,t+j}^{2d_\varphi-1} \cdot I\{\tilde{\Delta}N_{t,t+j} > 0\} + R(\tilde{\Delta}N_{t,t+j}) \cdot I\{\tilde{\Delta}N_{t,t+j} > 0\}\Big) I\{N_1(t) > 0\} I\{N_1(t_{2,N_2(t+j)}) > 0\} \\ &+ \Big(K(-\tilde{\Delta}N_{t,t+j})^{2d_\varphi-1} \cdot I\{\tilde{\Delta}N_{t,t+j} < 0\} + R(-\tilde{\Delta}N_{t,t+j}) \cdot I\{\tilde{\Delta}N_{t,t+j} < 0\}\Big) I\{N_1(t) > 0\} I\{N_1(t_{2,N_2(t+j)}) > 0\} \end{aligned}$$

$$+\Big(K_0 \cdot I\{\tilde{\Delta} N_{t,t+j} = 0\}\Big)\, I\{N_1(t) > 0\} I\{N_1(t_{2,N_2(t+j)}) > 0\}$$

Since $Pr\{\tilde{\Delta} N_{t,t+j} \leq 0\}$ has nearly exponential decay for fixed t as $j \to \infty$, so do the expectations of the second and third terms above. Arguing as in the proof of Lemma 3, we find that (47) is

$$-E\Big[E[\varphi_{1,N_1(t)} \cdot I\{N_1(t) > 0\}\, \varphi_{1,N_1(t_{2,N_2(t+j)})} \cdot I\{N_1(t_{2,N_2(t+j)}) > 0\} \big| N_1(\cdot), N_2(\cdot)]\Big] \sim -C j^{2d_\eta + 1} Pr\{N_1(t) > 0\}$$

with the same constant C as above, in view of the facts that $I\{\tilde{\Delta} N_{t,t+j} > 0\} \xrightarrow{p} 1$ and $I\{N_1(t_{2,N_2(t+j)}) > 0\} \xrightarrow{p} 1$. By a similar argument, the other two terms in $\text{cov}(T_{4,t}, T_{4,t+j})$ also cancel each other, so that overall, $\text{cov}(T_{4,t}, T_{4,t+j}) = o(j^{2d_\eta+1})$.

Similarly, $\text{cov}(T_{6,t}, T_{6,t+j}) = o(j^{2d_\eta+1})$.

Next, consider $\text{cov}(T_{3,t}, T_{4,t+j}) =$

$$\text{cov}\Big(\varphi_{1,N_1(t)} \cdot I\{N_1(t) > 0\}\ ,\ \varphi_{1,N_1(t+j)} \cdot I\{N_1(t+j) > 0\} - \varphi_{1,N_1(t_{2,N_2(t+j)})} \cdot I\{N_1(t_{2,N_2(t+j)}) > 0\}\Big) = o(j^{2d_\eta+1})$$

due to the cancelation of the two previously considered contributions to $\text{cov}(T_{4,t}, T_{4,t+j})$.

Similarly, $\text{cov}(T_{5,t}, T_{6,t+j}) = o(j^{2d_\eta+1})$. $\square$

G Proof of Theorem 5

As in the proof of Theorem 3, we denote

$$\begin{aligned} S_t = \log P_{1,t} - \theta \log P_{2,t} &= \underbrace{\sum_{k=N_1(t_{2,N_2(t)})+1}^{N_1(t)} e_{1,k}}_{S_{1,t}} - \theta \underbrace{\sum_{k=N_2(t_{1,N_1(t)})+1}^{N_2(t)} e_{2,k}}_{S_{2,t}} \\ &\quad + \underbrace{\sum_{k=1}^{N_1(t)} \eta_{1,k}}_{S_{3,t}} - \theta g_{12} \underbrace{\sum_{k=1}^{N_1(t_{2,N_2(t)})} \eta_{1,k}}_{S_{4,t}} - \theta \underbrace{\sum_{k=1}^{N_2(t)} \eta_{2,k}}_{S_{5,t}} + g_{21} \underbrace{\sum_{k=1}^{N_2(t_{1,N_1(t)})} \eta_{2,k}}_{S_{6,t}} \\ &= S_{1,t} - \theta S_{2,t} + S_{3,t} - \theta g_{12} S_{4,t} - \theta S_{5,t} + g_{12} S_{6,t} \qquad , \end{aligned}$$

and evaluate the terms in $\text{cov}(S_t, S_{t+j})$.

1) Consider $\mathrm{cov}(S_{1,t}, S_{1,t+j}) = E(S_{1,t}S_{1,t+j})$. The term $S_{1,t}$ is a sum of shocks occurring in the time interval between the last transaction of Asset 2 before time t and time t. Similarly, $S_{1,t+j}$ is a sum of shocks occurring between the last transaction of Asset 2 before time $t+j$ and time $t+j$. Clearly, if at least one transaction of Asset 2 occurs in $(t, t+j]$, we must have $t_{2,N_2(t+j)} > t$ so that $E[S_{1,t}S_{1,t+j}|N_1(\cdot), N_2(\cdot)] = 0$ because $\{e_{1,k}\}$ is *i.i.d.*. Otherwise, $t_{2,N_2(t+j)} = t_{2,N_2(t)}$ and $E[S_{1,t}S_{1,t+j}|N_1(\cdot), N_2(\cdot)] = \sigma_{1,e}^2[N_1(t) - N_1(t_{2,N_2(t)})]$. Therefore, by the Cauchy-Schwartz inequality,

$$\begin{aligned}
\mathrm{cov}(S_{1,t}, S_{1,t+j}) &= E(S_{1,t}S_{1,t+j}) = E\Big\{E[S_{1,t}S_{1,t+j}|N_1(\cdot), N_2(\cdot)]\Big\} \\
&= E\Big\{\sigma_{1,e}^2[N_1(t) - N_1(t_{2,N_2(t)})] \cdot I\{N_2(t+j) - N_2(t) = 0\}\Big\} \\
&\leq \sigma_{1,e}^2\{E[N_1(t) - N_1(t_{2,N_2(t)})]^2\}^{\frac{1}{2}} \cdot \{P[N_2(t+j) - N_2(t) = 0]\}^{\frac{1}{2}} .
\end{aligned}$$

By Lemma 5 and the stationarity of $N_1(\cdot)$, we obtain

$$\begin{aligned}
E\{[N_1(t) - N_1(t_{2,N_2(t)})]^2\} &= E\{[N_1(t - t_{2,N_2(t)})]^2\} = E\{[N_1(BRT_{2,t})]^2\} \\
&= E\Big(E\{[N_1(BRT_{2,t})]^2\Big|N_1(\cdot), N_2(\cdot)\}\Big) \\
&\leq E\Big[K_2(BRT_{2,t}^2 + 1)\Big]
\end{aligned}$$

which is bounded uniformly in t using Lemma 4.

Next, since $N_2(\cdot)$ is stationary, for any positive integer m, we obtain

$$\begin{aligned}
&P\Big[N_2(t+j) - N_2(t) = 0\Big] = P\Big[N_2(j) \leq 0\Big] \\
\leq \;& P\Big[|Z_2(j)| \geq \lambda_2 j^{\frac{1}{2}-d_\tau}\Big] \leq \frac{E|Z_2(j)|^m}{\lambda_2^m j^{m(\frac{1}{2}-d_\tau)}} = O(j^{m(d_\tau - \frac{1}{2})}),
\end{aligned} \tag{48}$$

where $Z_2(j) = \frac{N_2(j) - \lambda_2 j}{j^{\frac{1}{2}+d_\tau}}$. This is true since it follows from the proof of Proposition 1 in Deo, Hurvich, Soulier and Wang (2007) that $E|Z_2(j)|^m$ is bounded uniformly in j for all m. Therefore, $P\Big[N_2(t+j) - N_2(t) = 0\Big]$ has nearly-exponential decay, because (48) holds for all m. Thus, $\mathrm{cov}(S_{1,t}, S_{1,t+j})$ has nearly-exponential decay.

Similarly, $\mathrm{cov}(S_{2,t}, S_{2,t+j})$ has nearly-exponential decay.

2) Next, we consider $\mathrm{cov}(S_{3,t}, S_{3,t+j})$, $\mathrm{cov}(S_{3,t}, S_{4,t+j})$, $\mathrm{cov}(S_{4,t}, S_{3,t+j})$ and $\mathrm{cov}(S_{4,t}, S_{4,t+j})$.

2.i) First, we have

$$\begin{aligned}\Big|\mathrm{cov}(S_{3,t},S_{3,t+j})\Big| &= \Big|\mathrm{cov}\Big(\xi_{1,N_1(t)}I\{N_1(t)>0\}\,,\,\xi_{1,N_1(t+j)}I\{N_1(t+j)>0\}\Big)\Big| \\ &\le \sigma_{1,\xi}^2\sum_{r=0}^{\infty}|c_{\xi_{1,r}}|\,P\Big[N_1(t+j)-N_1(t)=r\Big]. \end{aligned} \tag{49}$$

If $Z_1(j)=\frac{N_1(j)-\lambda_1 j}{j^{\frac{1}{2}+d_\tau}}$, since $E|Z_1(j)|^m$ is bounded uniformly in j for all m, we get

$$\begin{aligned} &\sum_{r=0}^{\infty}|c_{\xi_{1,r}}|\,P\Big[N_1(t+j)-N_1(t)=r\Big]\le\sum_{r=0}^{\infty}|c_{\xi_{1,r}}|\,P\Big[N_1(j)\le r\Big] \\ \le\;& \sum_{r=0}^{\infty}|c_{\xi_{1,r}}|\,P\Big[|Z_1(j)-rj^{-\frac{1}{2}-d_\tau}|\ge\lambda_1 j^{\frac{1}{2}-d_\tau}\Big] \\ \le\;& \sum_{r=0}^{\infty}|c_{\xi_{1,r}}|\,\frac{E|Z_1(j)-rj^{-\frac{1}{2}-d_\tau}|^m}{\lambda_1^m j^{m(\frac{1}{2}-d_\tau)}} \\ \le\;& j^{m(d_\tau-\frac{1}{2})}C_m\sum_{r=0}^{\infty}e^{-K_{\xi_1}r}[1+r^m j^{m(-\frac{1}{2}-d_\tau)}]=O(j^{m(d_\tau-\frac{1}{2})}), \end{aligned}$$

where C_m are finite positive constants. Note that Minkowski's inequality is used in the last inequality.

2.ii) Next, we consider $\mathrm{cov}(S_{3,t},S_{4,t+j})$. We have

$$\begin{aligned} &\Big|\mathrm{cov}(S_{3,t},S_{4,t+j})\Big|=\Big|\mathrm{cov}\Big(\xi_{1,N_1(t)}I\{N_1(t)>0\}\,,\,\xi_{1,N_1(t_{2,N_2(t+j)})}I\{N_1(t_{2,N_2(t+j)})>0\}\Big)\Big| \\ \le\;& \sigma_{1,\xi}^2\sum_{r=0}^{\infty}|c_{\xi_{1,r}}|\,P\Big[N_1(t_{2,N_2(t+j)})-N_1(t)=r\Big] \\ \le\;& \sigma_{1,\xi}^2\sum_{r=0}^{\infty}|c_{\xi_{1,r}}|\,P\Big[N_1(t_{2,N_2(t+j)})-N_1(t)\le r\Big] \\ \le\;& \sigma_{1,\xi}^2\sum_{r=0}^{\infty}|c_{\xi_{1,r}}|\,P\Big\{[N_1(t+j)-N_1(t)]-\underbrace{[N_1(t+j)-N_1(t_{2,N_2(t+j)})]}_{X_{t,j}}\le r\Big\} \\ =\;& \sigma_{1,\xi}^2\sum_{r=0}^{\infty}|c_{\xi_{1,r}}|\,P\Big\{\frac{[N_1(t+j)-N_1(t)]-\lambda_1 j}{j^{\frac{1}{2}+d_\tau}}-\frac{X_{t,j}}{j^{\frac{1}{2}+d_\tau}}-\frac{r}{j^{\frac{1}{2}+d_\tau}}\le\frac{-\lambda_1 j}{j^{\frac{1}{2}+d_\tau}}\Big\} \\ \le\;& \sigma_{1,\xi}^2\sum_{r=0}^{\infty}|c_{\xi_{1,r}}|\,P\Big\{\Big|\frac{[N_1(t+j)-N_1(t)]-\lambda_1 j}{j^{\frac{1}{2}+d_\tau}}-\frac{X_{t,j}}{j^{\frac{1}{2}+d_\tau}}-\frac{r}{j^{\frac{1}{2}+d_\tau}}\Big|\ge\lambda_1 j^{\frac{1}{2}-d_\tau}\Big\} \\ \le\;& \sigma_{1,\xi}^2\sum_{r=0}^{\infty}|c_{\xi_{1,r}}|\,E\Big|\frac{[N_1(t+j)-N_1(t)]-\lambda_1 j}{j^{\frac{1}{2}+d_\tau}}-\frac{X_{t,j}}{j^{\frac{1}{2}+d_\tau}}-\frac{r}{j^{\frac{1}{2}+d_\tau}}\Big|^m\lambda_1^{-m}j^{m(d_\tau-\frac{1}{2})} \end{aligned} \tag{50}$$

for any positive integer m. If we could show that both $E\Big|\frac{[N_1(t+j)-N_1(t)]-\lambda_1 j}{j^{\frac{1}{2}+d_\tau}}\Big|^m$ and $E|X_{t,j}|^m$ are uniformly bounded, then by Minkowski's inequality and the fact that $|c_{\xi_1,r}|\le A_{\xi_1}e^{-K_{\xi_1}r}$ for all $r\ge 0$, we could obtain $\sum_{r=0}^{\infty} |c_{\xi_{1,r}}| \, E\Big|\frac{[N_1(t+j)-N_1(t)]-\lambda_1 j}{j^{\frac{1}{2}+d_\tau}} - \frac{X_{t,j}}{j^{\frac{1}{2}+d_\tau}} - \frac{r}{j^{\frac{1}{2}+d_\tau}}\Big|^m$ summable in r, hence $\Big|\text{cov}(S_{3,t}, S_{4,t+j})\Big| = O(j^{m(d_\tau - \frac{1}{2})})$ for any positive integer m, i.e., $\text{cov}(S_{3,t}, S_{4,t+j})$ has nearly-exponential decay.

Using the stationarity of $N_1(\cdot)$,

$$E\Big|\frac{[N_1(t+j)-N_1(t)]-\lambda_1 j}{j^{\frac{1}{2}+d_\tau}}\Big|^m = E\Big|\frac{N_1(j)-\lambda_1 j}{j^{\frac{1}{2}+d_\tau}}\Big|^m = E|Z_1(j)|^m,$$

which is bounded uniformly in j, by the proof of Proposition 1 in Deo, Hurvich, Soulier and Wang (2007).

By Lemma 5, we obtain

$$\begin{aligned} E|X_{t,j}|^m &= E\Big\{E\Big[|X_{t,j}|^m \Big| N_2(\cdot)\Big]\Big\} = E\Big\{E\Big[\Big(N_1(t+j) - N_1(t_{2,N_2(t+j)})\Big)^m \Big| N_2(\cdot)\Big]\Big\} \\ &\leq K_m E(BRT^m_{2,t+j} + 1), \end{aligned}$$

which is uniformly bounded in t and j by Lemma 4. Thus, $\text{cov}(S_{3,t}, S_{4,t+j})$ has nearly-exponential decay.

2.iii) Next, we consider $\text{cov}(S_{4,t}, S_{3,t+j})$. Since

$$\begin{aligned} \Big|\text{cov}(S_{4,t}, S_{3,t+j})\Big| &= \Big|\text{cov}\Big(\xi_{1,N_1(t_{2,N_2(t)})} I\{N_1(t_{2,N_2(t)}) > 0\}\, ,\, \xi_{1,N_1(t+j)} I\{N_1(t+j) > 0\}\Big)\Big| \\ &\leq \sigma^2_{1,\xi} \sum_{r=0}^{\infty} |c_{\xi_{1,r}}| P\Big[N_1(t+j) - N_1(t_{2,N_2(t)}) = r\Big] \\ &\leq \sigma^2_{1,\xi} \sum_{r=0}^{\infty} |c_{\xi_{1,r}}| P\Big[N_1(t+j) - N_1(t) = r\Big] \end{aligned}$$

Thus, $\text{cov}(S_{4,t}, S_{3,t+j})$ has nearly-exponential decay, by the proof of (49).

Finally, since

$$\begin{aligned} \Big|\text{cov}(S_{4,t}, S_{4,t+j})\Big| &= \Big|\text{cov}\Big(\xi_{1,N_1(t_{2,N_2(t)})} I\{N_1(t_{2,N_2(t)}) > 0\}\, ,\, \xi_{1,N_1(t_{2,N_2(t+j)})} I\{N_1(t_{2,N_2(t+j)}) > 0\}\Big)\Big| \\ &\leq \sigma^2_{1,\xi} \sum_{r=0}^{\infty} |c_{\xi_{1,r}}| P\Big[N_1(t_{2,N_2(t+j)}) - N_1(t_{2,N_2(t)}) = r\Big] \\ &\leq \sigma^2_{1,\xi} \sum_{r=0}^{\infty} |c_{\xi_{1,r}}| P\Big[N_1(t_{2,N_2(t+j)}) - N_1(t_{2,N_2(t)}) \leq r\Big] \\ &\leq \sigma^2_{1,\xi} \sum_{r=0}^{\infty} |c_{\xi_{1,r}}| P\Big[N_1(t_{2,N_2(t+j)}) - N_1(t) \leq r\Big] \qquad (51) \end{aligned}$$

which as we have shown in (50) has nearly-exponential decay. The last inequality in (51) holds since $N_1(t_{2,N_2(t)}) \leq N_1(t)$.

2.iv) Similarly to the above proofs, we can show that $\mathrm{cov}(S_{5,t}, S_{5,t+j})$, $\mathrm{cov}(S_{5,t}, S_{6,t+j})$, $\mathrm{cov}(S_{6,t}, S_{5,t+j})$ and $\mathrm{cov}(S_{6,t}, S_{6,t+j})$ have nearly-exponential decay.

3) So far, we have shown that the following terms have nearly-exponential decay as $j \to \infty$: $\mathrm{cov}(S_{1,t}, S_{1,t+j})$, $\mathrm{cov}(S_{2,t}, S_{2,t+j})$, $\mathrm{cov}(S_{3,t}, S_{3,t+j})$, $\mathrm{cov}(S_{3,t}, S_{4,t+j})$, $\mathrm{cov}(S_{4,t}, S_{3,t+j})$, $\mathrm{cov}(S_{4,t}, S_{4,t+j})$, $\mathrm{cov}(S_{5,t}, S_{5,t+j})$, $\mathrm{cov}(S_{5,t}, S_{6,t+j})$, $\mathrm{cov}(S_{6,t}, S_{5,t+j})$ and $\mathrm{cov}(S_{6,t}, S_{6,t+j})$. Since $\{e_{1,k}\}$, $\{e_{2,k}\}$, $\{\eta_{1,k}\}$ and $\{\eta_{2,k}\}$ are mutually independent, the remaining covariances are all zero. $\square$

H Proof of Theorem 6

We will treat the weak fractional cointegration case (Case 1), the standard cointegration case (Case 3) and the strong fractional cointegration case (Case 2) separately.

Case 1: weak fractional cointegration, $d_\eta \in (-\frac{1}{2}, 0)$.

The log prices given by (7) can be written as

$$\begin{aligned}
A_j &\equiv \log P_{1,j} = \sum_{k=1}^{N_1(j\Delta t)} (e_{1,k} + \eta_{1,k}) + \sum_{k=1}^{N_2(t_{1,N_1(j\Delta t)})} (\theta e_{2,k} + g_{21}\eta_{2,k}) \\
B_j &\equiv \log P_{2,j} = \sum_{k=1}^{N_2(j\Delta t)} (e_{2,k} + \eta_{2,k}) + \sum_{k=1}^{N_1(t_{2,N_2(j\Delta t)})} (\frac{1}{\theta} e_{1,k} + g_{12}\eta_{1,k}) \\
&= \sum_{k=1}^{N_2(j\Delta t)} (e_{2,k} + \eta_{2,k}) + \sum_{k=1}^{N_1(j\Delta t)} (\frac{1}{\theta} e_{1,k} + g_{12}\eta_{1,k}) - \sum_{k=N_1(t_{2,N_2(j\Delta t)})+1}^{N_1(j\Delta t)} (\frac{1}{\theta} e_{1,k} + g_{12}\eta_{1,k}) \\
&= \underbrace{\sum_{k=1}^{N_2(j\Delta t)} e_{2,k} + \frac{1}{\theta} \sum_{k=1}^{N_1(j\Delta t)} e_{1,k}}_{B_{1,j}} + \underbrace{\sum_{k=1}^{N_2(j\Delta t)} \eta_{2,k}}_{B_{2,j}} + g_{12} \underbrace{\sum_{k=1}^{N_1(j\Delta t)} \eta_{1,k}}_{B_{3,j}} \\
&- \frac{1}{\theta} \underbrace{\sum_{k=N_1(t_{2,N_2(j\Delta t)})+1}^{N_1(j\Delta t)} e_{1,k}}_{B_{4,j}} - g_{12} \underbrace{\sum_{k=N_1(t_{2,N_2(j\Delta t)})+1}^{N_1(j\Delta t)} \eta_{1,k}}_{B_{5,j}}
\end{aligned}$$

and

$$
\begin{aligned}
T_j \equiv A_j - \theta B_j \quad = \quad & \underbrace{\sum_{k=N_1(t_{2,N_2(j\Delta t)})+1}^{N_1(j\Delta t)} e_{1,k}}_{T_{1,j}=B_{4,j}} - \theta \underbrace{\sum_{k=N_2(t_{1,N_1(j\Delta t)})+1}^{N_2(j\Delta t)} e_{2,k}}_{T_{2,j}} + (1-\theta g_{12}) \underbrace{\sum_{k=1}^{N_1(j\Delta t)} \eta_{1,k}}_{T_{3,j}=B_{3,j}} \\
& + \theta g_{12} \underbrace{\sum_{k=N_1(t_{2,N_2(j\Delta t)})+1}^{N_1(j\Delta t)} \eta_{1,k}}_{T_{4,j}=B_{5,j}} - (\theta - g_{21}) \underbrace{\sum_{k=1}^{N_2(j\Delta t)} \eta_{2,k}}_{T_{5,j}=B_{2,j}} - g_{21} \underbrace{\sum_{k=N_2(t_{1,N_1(j\Delta t)})+1}^{N_2(j\Delta t)} \eta_{2,k}}_{T_{6,j}} \quad .
\end{aligned}
$$

The OLS slope estimator $\hat{\theta}$ obtained from regressing $\{\log P_{1,j}\}_{j=1}^n$ on $\{\log P_{2,j}\}_{j=1}^n$ is

$$
\hat{\theta} = \frac{\sum_{j=1}^n A_j B_j}{\sum_{j=1}^n B_j^2} = \frac{\sum_{j=1}^n (\theta B_j + T_j) B_j}{\sum_{j=1}^n B_j^2} = \theta + \frac{\sum_{j=1}^n T_j B_j}{\sum_{j=1}^n B_j^2} \quad . \tag{52}
$$

First, we show that $n^{-r} \sum_{j=1}^n T_j B_j \xrightarrow{p} 0$, where $r = 2 + d_\eta + \delta$ for $\forall\ \delta > 0$. By the Cauchy-Schwartz inequality,

$$
\frac{1}{n^r} \sum_{j=1}^n T_{i,j} B_{k,j} \leq \sqrt{\left(\frac{1}{n^{2r-2}} \sum_{j=1}^n T_{i,j}^2\right)\left(\frac{1}{n^2} \sum_{j=1}^n B_{k,j}^2\right)} \; . \tag{53}
$$

It is therefore sufficient to show that the righthand side of (53) converges in probability to zero, for all $i = 1, \ldots, 6$ and $k = 1, \ldots, 5$.

By (24), (25), Lemma 1, Lemma 2 and Jensen's inequality $E(X^{2d_\eta+1}) \leq (EX)^{2d_\eta+1}$ for $x \geq 0$,

$d_\eta \in (-\frac{1}{2}, 0)$, we obtain that, for any $\epsilon > 0$,

$$\begin{aligned}
\frac{1}{n^{2r-2}} P(\sum_{j=1}^{n} T_{1,j}^2 > \epsilon) &\le \frac{E(\sum_{j=1}^{n} T_{1,j}^2)}{n^{2r-2}\epsilon} = \frac{\sum_{j=1}^{n} \operatorname{var}(T_{1,j})}{n^{2r-2}\epsilon} = \frac{\sum_{j=1}^{n} \sigma_{1,e}^2 \tilde{C}_1}{n^{2r-2}\epsilon} = \frac{\sigma_{1,e}^2 \tilde{C}_1}{n^{2r-3}\epsilon} \to 0, \\
\frac{1}{n^{2r-2}} P(\sum_{j=1}^{n} T_{2,j}^2 > \epsilon) &\le \frac{E(\sum_{j=1}^{n} T_{2,j}^2)}{n^{2r-2}\epsilon} = \frac{\sum_{j=1}^{n} \operatorname{var}(T_{2,j})}{n^{2r-2}\epsilon} = \frac{\sum_{j=1}^{n} \sigma_{2,e}^2 \tilde{C}_2}{n^{2r-2}\epsilon} = \frac{\sigma_{2,e}^2 \tilde{C}_2}{n^{2r-3}\epsilon} \to 0, \\
\frac{1}{n^{2r-2}} P(\sum_{j=1}^{n} T_{3,j}^2 > \epsilon) &\le \frac{E(\sum_{j=1}^{n} T_{3,j}^2)}{n^{2r-2}\epsilon} = \frac{\sum_{j=1}^{n} \operatorname{var}(T_{3,j})}{n^{2r-2}\epsilon} \\
&\le \frac{\sum_{j=1}^{n} (1+C)\sigma_{1,\eta}^2 \{\lambda_1 j \Delta t\}^{2d_\eta+1}}{n^{2r-2}\epsilon} = \frac{O(n^{2d_\eta+2})}{n^{2r-2}\epsilon} \to 0, \\
\frac{1}{n^{2r-2}} P(\sum_{j=1}^{n} T_{4,j}^2 > \epsilon) &\le \frac{\sum_{j=1}^{n} \operatorname{var}(T_{4,j})}{n^{2r-2}\epsilon} \le \frac{\sum_{j=1}^{n} (1+C)\sigma_{1,\eta}^2 E\{[N_1(j\Delta t) - N_1(t_{2,N_2(j\Delta t)})]^{2d_\eta+1}\}}{n^{2r-2}\epsilon} \\
&\le \frac{\sum_{j=1}^{n} (1+C)\sigma_{1,\eta}^2 \{E[N_1(j\Delta t) - N_1(t_{2,N_2(j\Delta t)})]\}^{2d_\eta+1}}{n^{2r-2}\epsilon} \\
&= \frac{\sum_{j=1}^{n} (1+C)\sigma_{1,\eta}^2 (\tilde{C}_1)^{2d_\eta+1}}{n^{2r-2}\epsilon} \to 0, \\
\frac{1}{n^{2r-2}} P(\sum_{j=1}^{n} T_{5,j}^2 > \epsilon) &\to 0 \quad (\text{similar as for } \frac{1}{n^{2r-2}} P(\sum_{j=1}^{n} T_{3,j}^2 > \epsilon)), \\
\frac{1}{n^{2r-2}} P(\sum_{j=1}^{n} T_{6,j}^2 > \epsilon) &\to 0 \quad (\text{similar as for } \frac{1}{n^{2r-2}} P(\sum_{j=1}^{n} T_{4,j}^2 > \epsilon)),
\end{aligned}$$

as $n \to \infty$, since $d_\eta \in (-\frac{1}{2}, 0)$, $(2r-2) = 2d_\eta + 2 + \delta$ and $(2r-3) > 1$.

Therefore,

$$\frac{1}{n^{2r-2}} \sum_{j=1}^{n} T_{i,j}^2 \xrightarrow{p} 0 \tag{54}$$

for $i = 1, \ldots, 6$.

Next, since

$$\begin{aligned}
P\Big[\frac{1}{n^2} \sum_{j=1}^{n} B_{1,j}^2 > \mu\Big] &\le \frac{E(\sum_{j=1}^{n} B_{1,j}^2)}{n^2\mu} = \frac{\sum_{j=1}^{n} (\sigma_{2,e}^2 \lambda_2 j \Delta t + \frac{1}{\theta^2}\sigma_{1,e}^2 \lambda_1 j \Delta t)}{n^2\mu} \\
&= \frac{1}{2}(\sigma_{2,e}^2 \lambda_2 \Delta t + \frac{1}{\theta^2}\sigma_{1,e}^2 \lambda_1 \Delta t)(1 + \frac{1}{n})\frac{1}{\mu}
\end{aligned}$$

and for any $\epsilon > 0$ and all $n > 1$, we can choose $\mu > \frac{1}{\epsilon}(\sigma_{2,e}^2 \lambda_2 \Delta t + \frac{1}{\theta^2}\sigma_{1,e}^2 \lambda_1 \Delta t)$, so that

$$P\Big[\frac{1}{n^2} \sum_{j=1}^{n} B_{1,j}^2 > \mu\Big] < \epsilon,$$

we obtain

$$\frac{1}{n^2}\sum_{j=1}^{n} B_{1,j}^2 = O_p(1). \tag{55}$$

Since $B_{2,j} = T_{5,j}$, $B_{3,j} = T_{3,j}$, $B_{4,j} = T_{1,j}$ and $B_{5,j} = T_{4,j}$, it follows from (54) that

$$\frac{1}{n^2}\sum_{j=1}^{n} B_{i,j}^2 \xrightarrow{p} 0, \tag{56}$$

for $i = 2, \ldots, 5$.

Applying (54), (55), (56) in (53), we obtain

$$\frac{1}{n^r}\sum_{j=1}^{n} T_j B_j \xrightarrow{p} 0, \tag{57}$$

where $r = 2 + d_\eta + \delta$ for any $\delta > 0$.

Next, we show that $\frac{1}{\frac{1}{n^2}\sum_{j=1}^{n} B_j^2}$ is $O_p(1)$ by bounding it by a random variable that converges in distribution.

Since $n\sum_{j=1}^{n} a_j^2 \geq (\sum_{j=1}^{n} a_j)^2$ for any sequence $\{a_j\}$, we have,

$$\frac{1}{\frac{1}{n^2}\sum_{j=1}^{n} B_j^2} \leq \frac{1}{\frac{1}{n^3}(\sum_{j=1}^{n} B_j)^2}.$$

Note that

$$\frac{1}{n^3}(\sum_{j=1}^{n} B_j)^2 = \frac{1}{n^3}(\sum_{j=1}^{n} B_{1,j})^2 + \frac{1}{n^3}\sum_{i=2}^{5}(\sum_{j=1}^{n} B_{i,j})^2 + \frac{1}{n^3}\sum_{i=1}^{5}\sum_{s\neq i, s=1}^{5}\Big[(\sum_{j=1}^{n} B_{i,j})(\sum_{j=1}^{n} B_{s,j})\Big].$$

We will show that

$$\frac{1}{n^{3/2}}\sum_{j=1}^{n} B_{1,j} \xrightarrow{d} \sqrt{\frac{1}{3}\sigma_{2,e}^2\lambda_2\Delta t + \frac{1}{3\theta^2}\sigma_{1,e}^2\lambda_1\Delta t}\; Z \qquad , \tag{58}$$

where Z is standard normal and

$$\frac{1}{n^{3/2}}\sum_{j=1}^{n} B_{i,j} \xrightarrow{p} 0 \tag{59}$$

for $i = 2, \ldots, 5$, so that

$$\frac{1}{\frac{1}{n^3}(\sum_{j=1}^n B_j)^2} \xrightarrow{d} \Big(\frac{3\theta^2}{\theta^2\sigma_{2,e}^2\lambda_2\Delta t + \sigma_{1,e}^2\lambda_1\Delta t}\Big)\frac{1}{Z^2} \qquad ,$$

and

$$\frac{1}{\frac{1}{n^2}\sum_{j=1}^n B_j^2} = O_p(1). \tag{60}$$

To show (58), we write

$$\frac{1}{n^{3/2}}\sum_{j=1}^{n} B_{1,j} = \underbrace{\frac{1}{n^{3/2}}\sum_{j=1}^{n}\sum_{k=1}^{N_2(j\Delta t)} e_{2,k}}_{G_1} + \underbrace{\frac{1}{n^{3/2}}\frac{1}{\theta}\sum_{j=1}^{n}\sum_{k=1}^{N_1(j\Delta t)} e_{1,k}}_{G_2} \qquad ,$$

where G_1 and G_2 are independent.

Since $\{e_{2,k}\}$ is serially independent,

$$\begin{aligned} G_1 &= \frac{1}{n^{3/2}}\Big(n\sum_{k=1}^{N_2(\Delta t)} e_{2,k} + (n-1)\sum_{k=N_2(\Delta t)+1}^{N_2(2\Delta t)} e_{2,k} + \ldots + \sum_{k=N_2((n-1)\Delta t)+1}^{N_2(n\Delta t)} e_{2,k}\Big) \\ &\overset{d}{\equiv} \frac{\sigma_{2,e}}{n^{3/2}}\Big(n\sqrt{\Delta N_{2,1}}Z_1 + (n-1)\sqrt{\Delta N_{2,2}}Z_2 + \ldots + \sqrt{\Delta N_{2,n}}Z_n\Big) \\ &\overset{d}{\equiv} \frac{\sigma_{2,e}}{n^{3/2}}\Big(\sqrt{n^2\Delta N_{2,1} + (n-1)^2\Delta N_{2,2} \ldots + \Delta N_{2,n}}\ Z\Big) \\ &= \sigma_{2,e}\sqrt{\underbrace{\frac{1}{n^3}\sum_{k=1}^{n}(n-k+1)^2\Delta N_{2,k}}_{D}}\ Z \end{aligned} \tag{61}$$

where $\overset{d}{\equiv}$ denotes equivalence in distribution, $\Delta N_{2,j} = N_2(j\Delta t) - N_2((j-1)\Delta t)$, $\{Z_k\}_{k=1}^n$ are *i.i.d.* standard normal and Z is a standard normal random variable.

Consider D defined in (61). Applying the summation by parts formula for two sequences $\{f_k\}$ and $\{g_k\}$,

$$\sum_{k=m}^{n} f_k(g_{k+1} - g_k) = (f_{n+1}g_{n+1} - f_m g_m) - \sum_{k=m}^{n} g_{k+1}(f_{k+1} - f_k) \qquad ,$$

we obtain

$$\begin{aligned}
D &= \sum_{k=1}^{n}(n-k+1)^2\Delta N_{2,k} = \sum_{k=1}^{n}(n-k+1)^2\Big[N_2(k\Delta t)-N_2((k-1)\Delta t)\Big] \\
&= \sum_{k=0}^{n-1}\underbrace{(n-k)^2}_{f_k}\Big[\underbrace{N_2((k+1)\Delta t)}_{g_{k+1}}-\underbrace{N_2(k\Delta t)}_{g_k}\Big] \\
&= (f_n g_n - f_0 g_0) - \sum_{k=0}^{n-1} N_2((k+1)\Delta t)\Big[(2n-2k-1)(-1)\Big] \\
&= \sum_{k=0}^{n-1}(2n-2k-1)N_2((k+1)\Delta t) \qquad \text{(since } f_n = 0 \text{ and } g_0 = 0) \\
&= \sum_{k=1}^{n}(2n-2k+1)N_2(k\Delta t)
\end{aligned}$$

thus

$$\begin{aligned}
E(\frac{1}{n^3}D) &= \frac{1}{n^3}\sum_{k=1}^{n}(2n-2k+1)E[N_2(k\Delta t)] = \frac{\lambda_2\Delta t}{n^3}\sum_{k=1}^{n}(2n-2k+1)k \\
&= \frac{\lambda_2\Delta t}{n^3}\Big[2n\frac{n(n+1)}{2} - 2\frac{n(n+1)(2n+1)}{6} + \frac{n(n+1)}{2}\Big] \to \frac{1}{3}\lambda_2\Delta t, \qquad (62)
\end{aligned}$$

and

$$\begin{aligned}
\operatorname{var}(\frac{1}{n^3}D) &\le \frac{1}{n^6}\sum_{j=1}^{n}\sum_{s=1}^{n}(2n-2j+1)(2n-2s+1)\Big|\operatorname{cov}(N_2(j\Delta t),N_2(s\Delta t))\Big| \\
&\le \frac{1}{n^6}\sum_{j=1}^{n}\sum_{s=1}^{n}(2n-2j+1)(2n-2s+1)\sqrt{\operatorname{var}(N_2(j\Delta t))\operatorname{var}(N_2(s\Delta t))} \\
&\le \frac{4n^2}{n^6}\Big(\sum_{j=1}^{n}\sqrt{\operatorname{var}(N_2(j\Delta t))}\Big)\Big(\sum_{s=1}^{n}\sqrt{\operatorname{var}(N_2(s\Delta t))}\Big) = O(n^{2d_\tau-1}) \to 0 \qquad (63)
\end{aligned}$$

as $n \to \infty$ since $d_\tau \in (0, \frac{1}{2})$ and by Theorem 1 of Deo, Hurvich, Soulier and Wang (2007),

$$\operatorname{var}(N_2(n\Delta t)) \sim C(n\Delta t)^{2d_\tau+1} \qquad \text{as } n \to \infty.$$

By (62) and (63), $\left(\frac{1}{n^3}D - \frac{1}{3}\lambda_2\Delta t\right)$ converges in mean-square to zero, which implies that

$$\frac{1}{n^3}D = \frac{1}{n^3}\sum_{k=1}^{n}(n-k+1)^2\Delta N_{2,k} \xrightarrow{p} \frac{1}{3}\lambda_2\Delta t \qquad . \qquad (64)$$

Using (64) in (61), by Slutsky's theorem

$$G_1 \xrightarrow{d} \sigma_{2,e}\sqrt{\frac{1}{3}\lambda_2\Delta t}\; Z_1 \qquad (65)$$

and similarly

$$G_2 \xrightarrow{d} \frac{\sigma_{1,e}}{\theta}\sqrt{\frac{1}{3}\lambda_1\Delta t}\; Z_2 \tag{66}$$

where Z_1 and Z_2 are independent standard normals.

Overall, by (65), (66) and the independence between G_1 and G_2, (58) is obtained.

To show (59), since for any $\epsilon > 0$, and $i = 2, \ldots, 5$, by Chebyshev's inequality,

$$P\Big(\Big|\frac{1}{n^{3/2}}\sum_{j=1}^{n} B_{i,j}\Big| > \epsilon\Big) \le \frac{\mathrm{var}(\sum_{j=1}^{n} B_{i,j})}{n^3\epsilon^2}$$

it is enough to show that,

$$\frac{\mathrm{var}(\sum_{j=1}^{n} B_{i,j})}{n^3} \to 0, \qquad i = 2, \ldots, 5. \tag{67}$$

Since

$$\begin{aligned}
\mathrm{var}(B_{2,j}) &= \sigma_{2,\eta}^2 E[(N_2(j\Delta t))^{2d_\eta+1}] + \sigma_{2,\eta}^2 E\{R(N_2(j\Delta t))\} \\
&\le (1+C)\sigma_{2,\eta}^2\{E[N_2(j\Delta t)]\}^{2d_\eta+1} = (1+C)\sigma_{2,\eta}^2(\lambda_2 j\Delta t)^{2d_\eta+1} \\
&\le (1+C)\sigma_{2,\eta}^2(\lambda_2\Delta t)^{2d_\eta+1}n^{2d_\eta+1} \qquad \text{(since } j \le n \text{ and } 2d_\eta + 1 > 0\text{)} \\
\mathrm{var}(B_{3,j}) &\le \sigma_{1,\eta}^2(\lambda_1\Delta t)^{2d_\eta+1}n^{2d_\eta+1} \qquad \text{(similar as for } \mathrm{var}(\sum_{j=1}^{n} B_{2,j})\text{)} \\
\mathrm{var}(B_{4,j}) &= \sigma_{1,e}^2\tilde{C}_1 \\
\mathrm{var}(B_{5,j}) &= \sigma_{1,\eta}^2 E\Big[(N_1(j\Delta t) - N_1(t_{2,N_2(j\Delta t)}))^{2d_\eta+1}\Big] + \sigma_{1,\eta}^2 E\Big\{R\Big(N_1(j\Delta t) - N_1(t_{2,N_2(j\Delta t)})\Big)\Big\} \\
&\le (1+C)\sigma_{1,\eta}^2\Big\{E\Big[(N_1(j\Delta t) - N_1(t_{2,N_2(j\Delta t)})\Big]^{2d_\eta+1}\Big\} = (1+C)\sigma_{1,\eta}^2(\tilde{C}_1)^{2d_\eta+1} \qquad ,
\end{aligned}$$

we obtain

$$\begin{aligned}
\mathrm{var}(\sum_{j=1}^{n} B_{2,j}) \le \sum_{j=1}^{n}\sum_{s=1}^{n}|\mathrm{cov}(B_{2,j}, B_{2,s})| &\le \sum_{j=1}^{n}\sum_{s=1}^{n}(1+C)\sigma_{2,\eta}^2(\lambda_2 n\Delta t)^{2d_\eta+1} = O(n^{2d_\eta+3}) \qquad (68) \\
\mathrm{var}(\sum_{j=1}^{n} B_{3,j}) \le \sum_{j=1}^{n}\sum_{s=1}^{n}|\mathrm{cov}(B_{3,j}, B_{3,s})| &= O(n^{2d_\eta+3}) \qquad \text{(similar as above)} \\
\mathrm{var}(\sum_{j=1}^{n} B_{4,j}) \le \sum_{j=1}^{n}\sum_{s=1}^{n}|\mathrm{cov}(B_{4,j}, B_{4,s})| &= \sum_{j=1}^{n}\sum_{s=1}^{n}\sigma_{1,e}^2\tilde{C}_1 = O(n^2) \\
\mathrm{var}(\sum_{j=1}^{n} B_{5,j}) \le \sum_{j=1}^{n}\sum_{s=1}^{n}|\mathrm{cov}(B_{5,j}, B_{5,s})| &= \sum_{j=1}^{n}\sum_{s=1}^{n}(1+C)\sigma_{1,\eta}^2(\tilde{C}_1)^{2d_\eta+1} = O(n^2) \qquad .
\end{aligned}$$

This implies (67) and (59), since $d_\eta < 0$.

Overall, since (58), (59) are proved, we obtain (60). Thus, by (52), (57) and (60),

$$n^{2-r}(\hat{\theta} - \theta) = \frac{\frac{1}{n^r}\sum_{j=1}^{n} T_j B_j}{\frac{1}{n^2}\sum_{j=1}^{n} B_j^2} \xrightarrow{p} 0.$$

<u>Case 3:</u> standard cointegration, $d_\eta = -1$.

When $d_\eta = -1$, $\eta_{1,k} = \xi_{1,k} - \xi_{1,k-1}$ and $\eta_{2,k} = \xi_{2,k} - \xi_{2,k-1}$. Denote

$$\begin{aligned} B_j \quad &\equiv \quad \underbrace{\sum_{k=1}^{N_2(t_{1,N_1(j\Delta t)})} e_{2,k}}_{B^*_{1,j}} + \underbrace{\sum_{N_2(t_{1,N_1(j\Delta t)})+1}^{N_2(j\Delta t)} e_{2,k}}_{B^*_{2,j}} + \frac{1}{\theta} \underbrace{\sum_{k=1}^{N_1(t_{2,N_2(j\Delta t)})} e_{1,k}}_{B^*_{3,j}} \\ &+ \quad g_{12} \cdot \underbrace{\xi_{1,N_1(t_{2,N_2(j\Delta t)})} I\{N_1(t_{2,N_2(j\Delta t)}) > 0\}}_{B^*_{4,j}} + \underbrace{\xi_{2,N_2(j\Delta t)} I\{N_2(j\Delta t) > 0\}}_{B^*_{5,j}} \end{aligned}$$

and

$$\begin{aligned} T_j \equiv A_j - \theta B_j \quad &= \quad \underbrace{\sum_{k=N_1(t_{2,N_2(j\Delta t)})+1}^{N_1(j\Delta t)} e_{1,k}}_{T^*_{1,j}} - \theta \underbrace{\sum_{k=N_2(t_{1,N_1(j\Delta t)})+1}^{N_2(j\Delta t)} e_{2,k}}_{T^*_{2,j} = B^*_{2,j}} \\ &+ \quad \underbrace{\xi_{1,N_1(j\Delta t)} I\{N_1(j\Delta t) > 0\}}_{T^*_{3,j}} - \theta g_{12} \cdot \underbrace{\xi_{1,N_1(t_{2,N_2(j\Delta t)})} I\{N_1(t_{2,N_2(j\Delta t)}) > 0\}}_{T^*_{4,j} = B^*_{4,j}} \\ &- \quad \theta \cdot \underbrace{\xi_{2,N_2(j\Delta t)} I\{N_2(j\Delta t) > 0\}}_{T^*_{5,j} = B^*_{5,j}} + g_{21} \cdot \underbrace{\xi_{2,N_2(t_{1,N_1(j\Delta t)})} I\{N_1(j\Delta t) > 0\}}_{T^*_{6,j}} \end{aligned}$$

1) Consider $\sum_{j=1}^{n} B^*_{1,j}T^*_{1,j}$. Since $E(B^*_{1,j}T^*_{1,j}) = E[E(B^*_{1,j}T^*_{1,j}|N_1(\cdot), N_2(\cdot))] = 0$, we obtain

$$
\begin{aligned}
& \operatorname{var}(\sum_{j=1}^{n} B^*_{1,j}T^*_{1,j}) \\
= & E\Big[\sum_{j=1}^{n}\sum_{s=1}^{n} B^*_{1,j}B^*_{1,s}T^*_{1,j}T^*_{1,s}\Big] - \Big\{E(\sum_{j=1}^{n} B^*_{1,j}T^*_{1,j})\Big\}^2 \\
= & \sum_{j=1}^{n} E({B^*_{1,j}}^2 {T^*_{1,j}}^2) + 2\sum_{j=1}^{n}\sum_{s=j+1}^{n} E(B^*_{1,j}B^*_{1,s}T^*_{1,j}T^*_{1,s}) \\
= & \sum_{j=1}^{n} E[E({B^*_{1,j}}^2 {T^*_{1,j}}^2|N_1(\cdot), N_2(\cdot))] + 2\sum_{j=1}^{n}\sum_{s=j+1}^{n} E[E(B^*_{1,j}B^*_{1,s}T^*_{1,j}T^*_{1,s}|N_1(\cdot), N_2(\cdot))] \\
= & \sum_{j=1}^{n} E[E({B^*_{1,j}}^2|N_1(\cdot), N_2(\cdot)) \cdot E({T^*_{1,j}}^2|N_1(\cdot), N_2(\cdot))] \\
& + 2\sum_{j=1}^{n}\sum_{s=j+1}^{n} E[E(B^*_{1,j}B^*_{1,s}|N_1(\cdot), N_2(\cdot)) \cdot E(T^*_{1,j}T^*_{1,s}|N_1(\cdot), N_2(\cdot))] \\
= & \underbrace{\sigma^2_{1,e}\sigma^2_{2,e}\sum_{j=1}^{n} E\{N_2(t_{1,N_1(j\Delta t)}) \cdot [N_1(j\Delta t) - N_1(t_{2,N_2(j\Delta t)})]\}}_{O(n^2),\text{ as shown in below}} \qquad (69) \\
& + 2\sigma^2_{1,e}\sigma^2_{2,e}\sum_{j=1}^{n}\sum_{s=j+1}^{n} E\Big\{N_2(t_{1,N_1(j\Delta t)}) \cdot [N_1(j\Delta t) - N_1(t_{2,N_2(j\Delta t)})] \cdot I\{N_2(s\Delta t) - N_2(j\Delta t) = 0\}\Big\} \quad .
\end{aligned}
$$

By the Cauchy-Schwartz inequality,

$$
\begin{aligned}
& E\{N_2(t_{1,N_1(j\Delta t)}) \cdot [N_1(j\Delta t) - N_1(t_{2,N_2(j\Delta t)})]\} \\
\leq & \sqrt{E\{[N_2(t_{1,N_1(j\Delta t)})]^2\} \cdot E\{[N_1(j\Delta t) - N_1(t_{2,N_2(j\Delta t)})]^2\}} = O(j)
\end{aligned}
$$

because by Lemma 4 and Lemma 5, $E\{[N_1(j\Delta t) - N_1(t_{2,N_2(j\Delta t)})]^2\}$ is bounded uniformly in j and by Theorem 1 in Deo, Hurvich, Soulier and Wang (2007),

$$E\{[N_2(t_{1,N_1(j\Delta t)})]^2\} \leq E\{[N_2(j\Delta t)]^2 = \{E[N_2(j\Delta t)]\}^2 + \operatorname{var}[N_2(j\Delta t)] = (\lambda_2 j\Delta t)^2 + O(j^{2d_\tau+1}) = O(j^2)$$

hence

$$\sigma^2_{1,e}\sigma^2_{2,e}\sum_{j=1}^{n} E\{N_2(t_{1,N_1(j\Delta t)}) \cdot [N_1(j\Delta t) - N_1(t_{2,N_2(j\Delta t)})]\} = O(n^2),$$

as indicated in (69).

Similarly, since

$$\begin{aligned}
& \sum_{j=1}^{n}\sum_{s=j+1}^{n} E\Big\{N_2(t_{1,N_1(j\Delta t)})\cdot[N_1(j\Delta t)-N_1(t_{2,N_2(j\Delta t)})]\cdot I\{N_2(s\Delta t)-N_2(j\Delta t)=0\}\Big\} \\
\leq\ & \sum_{j=1}^{n}\sum_{s=j+1}^{n} \sqrt{E\{[N_2(t_{1,N_1(j\Delta t)})]^2\}}\cdot\{E[N_1(j\Delta t)-N_1(t_{2,N_2(j\Delta t)})]^4\}^{1/4}\cdot\{P[N_2(s\Delta t)-N_2(j\Delta t)=0]\}^{1/4} \\
\leq\ & \underbrace{\sqrt{E\{[N_2(n\Delta t)]^2\}}}_{O(n)}\cdot\sum_{j=1}^{n}\sum_{s=j+1}^{n}\{\underbrace{E[N_1(j\Delta t)-N_1(t_{2,N_2(j\Delta t)})]^4}_{\text{bounded uniformly in } j}\}^{1/4}\cdot\{\underbrace{P[N_2(s\Delta t)-N_2(j\Delta t)=0]}_{\leq K(s-j)^{m(d_\tau-\frac{1}{2})},\ \forall m\geq 1}\}^{1/4}
\end{aligned}$$

and by Lemma 4 and Lemma 5, $E\{[N_1(j\Delta t)-N_1(t_{2,N_2(j\Delta t)})]^4\}$ is bounded uniformly in j, while by (48), $P[N_2(s\Delta t)-N_2(j\Delta t)=0]\leq K(s-j)^{m(d_\tau-\frac{1}{2})}$ for all $m\geq 1$. We obtain that,

$$\operatorname{var}(\sum_{j=1}^{n} B^*_{1,j}T^*_{1,j}) \leq O(n^2)+Kn\sum_{j=1}^{n}\sum_{s=j+1}^{n}(s-j)^{m(d_\tau-\frac{1}{2})/4}. \tag{70}$$

Consider $\sum_{j=1}^{n}\sum_{s=j+1}^{n}(s-j)^{m(d_\tau-\frac{1}{2})/4}$. For any fixed integer $1\leq j\leq n$, we choose $m>\frac{8}{1-2d_\tau}$ so that $\sum_{s=j+1}^{n}(s-j)^{m(d_\tau-\frac{1}{2})/4}$ is summable in s, hence $\sum_{j=1}^{n}\sum_{s=j+1}^{n}(s-j)^{m(d_\tau-\frac{1}{2})/4}=O(n)$. Therefore, $\operatorname{var}(\sum_{j=1}^{n}B^*_{1,j}T^*_{1,j})=O(n^2)$ and by Chebyshev's inequality, we obtain that for any $\delta>0$,

$$\frac{1}{n^{1+\delta}}\sum_{j=1}^{n}B^*_{1,j}T^*_{1,j}\xrightarrow{p}0 \qquad .$$

2) Next, we consider $\sum_{j=1}^{n}B^*_{1,j}T^*_{2,j}$. Since $E(B^*_{1,j}T^*_{2,j})=E[E(B^*_{1,j}T^*_{2,j}|N_1(\cdot),N_2(\cdot))]=0$, we have

$$\begin{aligned}
& \operatorname{var}(\sum_{j=1}^{n}B^*_{1,j}T^*_{2,j}) \\
=\ & E\Big[\sum_{j=1}^{n}\sum_{s=1}^{n}B^*_{1,j}B^*_{1,s}T^*_{2,j}T^*_{2,s}\Big]-\Big\{E(\sum_{j=1}^{n}B^*_{1,j}T^*_{2,j})\Big\}^2 \\
=\ & \sum_{j=1}^{n}E({B^*_{1,j}}^2{T^*_{2,j}}^2)+2\sum_{j=1}^{n}\sum_{s=j+1}^{n}E(B^*_{1,j}B^*_{1,s}T^*_{2,j}T^*_{2,s}) \\
=\ & \sum_{j=1}^{n}E[E({B^*_{1,j}}^2{T^*_{2,j}}^2|N_1(\cdot),N_2(\cdot))]+2\sum_{j=1}^{n}\sum_{s=j+1}^{n}E[E(B^*_{1,j}B^*_{1,s}T^*_{2,j}T^*_{2,s}|N_1(\cdot),N_2(\cdot))]
\end{aligned}$$

Since conditionally on $N_1(\cdot)$ and $N_2(\cdot)$, $B^*_{1,j},B^*_{1,s},T^*_{2,j}$ and $T^*_{2,j}$ are zero-mean normals, using Isserlis'

Formula (Isserlis, 1918), we obtain

$$
\begin{aligned}
& \operatorname{var}(\sum_{j=1}^{n} B_{1,j}^{*} T_{2,j}^{*}) \\
= & \sum_{j=1}^{n} E[E({B_{1,j}^{*}}^{2}|N_1(\cdot),N_2(\cdot)) \cdot E({T_{2,j}^{*}}^{2}|N_1(\cdot),N_2(\cdot))] \\
& + 2\sum_{j=1}^{n}\sum_{s=j+1}^{n} E\Big[\ E(B_{1,j}^{*}B_{1,s}^{*}|N_1(\cdot),N_2(\cdot)) \cdot E(T_{2,j}^{*}T_{2,s}^{*}|N_1(\cdot),N_2(\cdot)) \\
& + \underbrace{E(B_{1,j}^{*}T_{2,j}^{*}|N_1(\cdot),N_2(\cdot))}_{0} \cdot E(B_{1,s}^{*}T_{2,s}^{*}|N_1(\cdot),N_2(\cdot)) + \underbrace{E(B_{1,j}^{*}T_{2,s}^{*}|N_1(\cdot),N_2(\cdot))}_{0} \cdot E(B_{1,s}^{*}T_{2,j}^{*}|N_1(\cdot),N_2(\cdot)) \quad \Big] \\
= & \sum_{j=1}^{n} E[E({B_{1,j}^{*}}^{2}|N_1(\cdot),N_2(\cdot)) \cdot E({T_{2,j}^{*}}^{2}|N_1(\cdot),N_2(\cdot))] \\
& + 2\sum_{j=1}^{n}\sum_{s=j+1}^{n} E\Big[E(B_{1,j}^{*}B_{1,s}^{*}|N_1(\cdot),N_2(\cdot)) \cdot E(T_{2,j}^{*}T_{2,s}^{*}|N_1(\cdot),N_2(\cdot))\Big] \\
= & \sigma_{2,e}^{4} \sum_{j=1}^{n} E\{N_2(t_{1,N_1(j\Delta t)}) \cdot [N_2(j\Delta t) - N_2(t_{1,N_1(j\Delta t)})]\} \\
& + 2\sigma_{2,e}^{4} \sum_{j=1}^{n}\sum_{s=j+1}^{n} E\Big\{N_2(t_{1,N_1(j\Delta t)}) \cdot [N_2(j\Delta t) - N_2(t_{1,N_1(j\Delta t)})] \cdot I\{N_1(s\Delta t) - N_1(j\Delta t) = 0\}\Big\} \qquad .
\end{aligned}
$$

which is similar to (69). Following along similar lines as for (69), we obtain

$$
\frac{1}{n^{1+\delta}} \sum_{j=1}^{n} B_{1,j}^{*} T_{2,j}^{*} \xrightarrow{p} 0, \qquad \forall\ \delta > 0
$$

3) Similarly to 1), for $\sum_{j=1}^{n} B_{1,j}^{*} T_{3,j}^{*} = \sum_{j=1}^{n} B_{1,j}^{*} \xi_{1,N_1(j\Delta t)} I\{N_1(j\Delta t) > 0\}$, we have

$$
\begin{aligned}
& \operatorname{var}(\sum_{j=1}^{n} B_{1,j}^{*} \xi_{1,N_1(j\Delta t)} I\{N_1(j\Delta t) > 0\}) \\
\leq & \ \sigma_{2,e}^{2}\sigma_{1,\xi}^{2} \sum_{j=1}^{n} E[N_2(t_{1,N_1(j\Delta t)})] + 2\sigma_{2,e}^{2}\sigma_{1,\xi}^{2} \sum_{j=1}^{n}\sum_{s=j+1}^{n}\sum_{r=0}^{\infty} E\Big[N_2(t_{1,N_1(j\Delta t)}) \cdot I\{N_1(s\Delta t) - N_1(j\Delta t) = r\}\Big] \cdot |c_{\xi_{1,r}}| \\
\leq & \ \sigma_{2,e}^{2}\sigma_{1,\xi}^{2} \sum_{j=1}^{n} E[N_2(j\Delta t)] + 2\sigma_{2,e}^{2}\sigma_{1,\xi}^{2} \sum_{j=1}^{n}\sum_{s=j+1}^{n}\sum_{r=0}^{\infty} E\Big[N_2(j\Delta t) \cdot I\{N_1(s\Delta t) - N_1(j\Delta t) = r\}\Big] \cdot |c_{\xi_{1,r}}| \\
\leq & \ \sigma_{2,e}^{2}\sigma_{1,\xi}^{2}\lambda_2\Delta t \frac{n(n+1)}{2} + 2\sigma_{2,e}^{2}\sigma_{1,\xi}^{2} \sum_{j=1}^{n}\sum_{s=j+1}^{n}\sum_{r=0}^{\infty} \sqrt{E\{[N_2(j\Delta t)]^2\} \cdot P[N_1(s\Delta t) - N_1(j\Delta t) = r]} \cdot |c_{\xi_{1,r}}| \\
\leq & \ \sigma_{2,e}^{2}\sigma_{1,\xi}^{2}\lambda_2\Delta t \frac{n(n+1)}{2} + 2\sigma_{2,e}^{2}\sigma_{1,\xi}^{2} \underbrace{\sqrt{E\{[N_2(n\Delta t)]^2\}}}_{O(n)} \cdot \sum_{j=1}^{n}\sum_{s=j+1}^{n}\sum_{r=0}^{\infty} \sqrt{P[N_1(s\Delta t) - N_1(j\Delta t) \leq r]} \cdot |c_{\xi_{1,r}}|.
\end{aligned}
$$

Since

$$
\begin{aligned}
& \sum_{r=0}^{\infty} \sqrt{P[N_1(s\Delta t) - N_1(j\Delta t) \le r]} \cdot |c_{\xi_{1,r}}| \\
\le \quad & \sum_{r=0}^{\infty} |c_{\xi_{1,r}}| \sqrt{\frac{E|Z_1(s-j) - r(s-j)^{-\frac{1}{2}-d_\tau}|^m}{\lambda_1^m (s-j)^{m(\frac{1}{2}-d_\tau)}}} \\
\le \quad & (s-j)^{\frac{m}{2}(d_\tau - \frac{1}{2})} \cdot C_m \sum_{r=0}^{\infty} e^{-K_{\xi_1} r} \Big[1 + r^{\frac{m}{2}} (s-j)^{\frac{m}{2}(-\frac{1}{2}-d_\tau)} \Big] \\
= \quad & O\Big((s-j)^{\frac{m}{2}(d_\tau - \frac{1}{2})} \Big)
\end{aligned}
$$

we can choose m sufficiently large so that $\sum_{s=j+1}^{n} \sum_{r=0}^{\infty} \sqrt{P[N_1(s\Delta t) - N_1(j\Delta t) \le r]} \cdot |c_{\xi_{1,r}}|$ is summable in s. Hence

$$
\underbrace{\sqrt{E\{[N_2(n\Delta t)]^2\}}}_{O(n)} \cdot \sum_{j=1}^{n} \underbrace{\sum_{s=j+1}^{n} \sum_{r=0}^{\infty} \sqrt{P[N_1(s\Delta t) - N_1(j\Delta t) \le r]} \cdot |c_{\xi_{1,r}}|}_{\text{summable in s}} = O(n^2) \qquad .
$$

Therefore, $\mathrm{var}(\sum_{j=1}^{n} B^*_{1,j} \xi_{1,N_1(j\Delta t)}) = O(n^2)$, and

$$
\frac{1}{n^{1+\delta}} \sum_{j=1}^{n} B^*_{1,j} T^*_{3,j} \xrightarrow{p} 0, \qquad \forall\ \delta > 0
$$

using Chebyshev's inequality.

By similar arguments for $\sum_{j=1}^{n} B^*_{1,j} T^*_{3,j}$, we obtain that $\forall\ \delta > 0$

$$
\frac{1}{n^{1+\delta}} \sum_{j=1}^{n} B^*_{1,j} T_{i,j} \xrightarrow{p} 0, \qquad i = 4, 5, 6.
$$

4) The proof for $\sum_{j=1}^{n} B^*_{3,j} T_{i,j}, (i = 1, \ldots, 6)$ follows along similar lines as for $\sum_{j=1}^{n} B^*_{1,j} T_{i,j}, (i = 1, \ldots, 6)$, since $B^*_{3,j}$ and $B^*_{1,j}$ are essentially the same since one is for Asset 1 and the other is for Asset 2. Thus, $\forall\ \delta > 0$

$$
\frac{1}{n^{1+\delta}} \sum_{j=1}^{n} B^*_{3,j} T^*_{i,j} \xrightarrow{p} 0, \qquad i = 1, \ldots, 6.
$$

5) The remaining terms $\sum_{j=1}^{n} B^*_{i,j} T^*_{k,j}, (i = 2, 4, 5)$ and $(k = 1, \ldots, 6)$ are all $O_p(n)$, as can easily be shown by using the Cauchy-Schwartz inequality and Chebyshev's inequality. For example:

5.1) We have

$$\sum_{j=1}^{n} B_{2,j}^{*} T_{1,j}^{*} \leq \sqrt{\sum_{j=1}^{n} {B_{2,j}^{*}}^{2} \cdot \sum_{j=1}^{n} {T_{1,j}^{*}}^{2}} = O_p(n)$$

since by Chebyshev's inequality, for any $\epsilon > 0$, we can choose $M > \frac{\sigma_{2,e}^2 \tilde{C}_2}{\epsilon}$, so that

$$P\Big(\frac{1}{n}\sum_{j=1}^{n} {B_{2,j}^{*}}^{2} > M\Big) \leq \frac{E(\sum_{j=1}^{n} {B_{2,j}^{*}}^{2})}{nM} = \frac{\sum_{j=1}^{n} \mathrm{var}(B_{2,j}^{*})}{nM} = \frac{\sigma_{2,e}^2 \tilde{C}_2}{M} < \epsilon$$

and similarly $\sum_{j=1}^{n} {T_{1,j}^{*}}^{2} = O_p(n)$.

5.2) We have

$$\sum_{j=1}^{n} B_{2,j}^{*} T_{2,j}^{*} = \sum_{j=1}^{n} {B_{2,j}^{*}}^{2} = O_p(n).$$

Therefore, $\forall\ \delta > 0$

$$\frac{1}{n^{1+\delta}} \sum_{j=1}^{n} B_{i,j}^{*} T_{k,j}^{*} \xrightarrow{p} 0, \qquad i = 2, 4, 5 \text{ and } k = 1, \ldots, 6.$$

6) Overall, when $d_\eta = -1$

$$\frac{1}{n^{1+\delta}} \sum_{j=1}^{n} B_j T_j \xrightarrow{p} 0 \tag{71}$$

for any $\delta > 0$.

Furthermore, the proof for (60) in the standard cointegration case is identical to that for the fractional cointegration case, except that here we have $\mathrm{var}(\sum_{k=1}^{N_i(t)} \eta_{i,k}) \leq 2\sigma_{i,\xi}^2$, $(i = 1, 2)$, which does not increase with t. (We still have the telescope sum even if ξ_i is not *i.i.d.* and the variance of the partial sum is still some constant.) This, together with (71), gives that

$$n^{1-\delta}(\hat{\theta} - \theta) \xrightarrow{p} 0.$$

Case 2: strong fractional cointegration, $d_\eta \in (-1, -\frac{1}{2})$.

Following along the same lines as the proof of Case 1, we can show that, the convergence rate of $\hat{\theta}$ is arbitrarily close to $\sqrt{n}$, using the fact that the variance of the partial sums of the microstructure noise is a constant and not increasing with time. □