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

Clifford M. Hurvich*   Yi Wang *

December 4, 2006

Abstract

We propose a new transaction-level bivariate log-price model, which yields fractional or standard cointegration. Most existing models for cointegration require the choice of a fixed sampling interval Δt. By contrast, our proposed model is constructed at the transaction level, thus determining the properties of returns at all sampling frequencies. 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. The $\{e_k\}$, assumed to be *i.i.d.* Gaussian, produce a Martingale component in log prices. We assume that the microstructure noise $\{\eta_k\}$ obeys a certain model with memory parameter $d_\eta \in (-1/2, 0)$ (fractional cointegration case) or $d_\eta = -1$ (standard cointegration case). Our log-price model includes feedback between the disturbances of the two log-price series. This feedback yields cointegration, in that there exists a linear combination of the two series that reduces the memory parameter from 1 to $1 + d_\eta \in (0.5, 1) \cup \{0\}$. Returns at sampling interval Δt are asymptotically uncorrelated at any fixed lag as Δt increases. 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 several of the other parameters in our model. We present a data analysis, which provides evidence of fractional cointegration. We then consider special cases and generalizations of our model, mostly in simulation studies, to argue that the suitably-modified model is able to capture a variety of additional properties and stylized facts, including leverage, portfolio return autocorrelation due to nonsynchronous trading, Granger causality, and volatility feedback. 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.

*Stern School of Business, New York University.
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.

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

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, and nonsynchronous trading effects. In this introduction, and indeed from here to Section VIII, we focus on cointegration, as this is the area in which we have so far been able to develop theoretical results on our model. Nevertheless, simulations show that a suitably-modified version of our basic model is able to produce the so-called leverage effect (i.e., negative autocorrelation between the current period's return and the next period's absolute return), as well as portfolio return autocorrelation due to nonsynchronous trading, Granger causality, and volatility feedback.

Cointegration 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 (1991). 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 frequency, *e.g.*, one day, one month, *etc.* and therefore do not reflect the dynamics at all levels of aggregation. One could build models for cointegration using diffusion-type continuous-time models such as ordinary or fractional Brownian motion, but such models would fail to capture the pure-jump nature of observed asset-price processes.

Before describing the cointegration aspects of 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 $d_\tau \in (0, 1/2)$, 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 (2006) 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 $d_\tau \in (0, 1/2)$, 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, under the simple return model considered in Deo, Hurvich, Soulier and Wang (2006).

In order to reflect the persistence in durations, we will assume in this paper that durations are generated by an LMSD model with memory parameter $d_\tau \in (0, 1/2)$. Thus, the resulting counting process $N(t)$ will have long-range count dependence with the same memory parameter, d_τ.

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 fractional cointegration case, this noise series obeys a fractional Gaussian noise model, with a corresponding memory parameter $d_\eta \in (-1/2, 0)$, while in the standard cointegration case $\{\eta_k\}$ is the difference of a white noise, and has memory parameter $d_\eta = -1$. In both 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 (2006) for the counting process $N(t)$ induced by LMSD durations.

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 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.

In Section III, 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 Δt is increased. We also present a lemma on the microstructure component of the log-price series. We show that this component, which is a random sum of the microstructure noise, has memory parameter $1 + d_\eta < 1$.

In Section IV, we establish that our model possesses cointegration, by showing that the cointegrating error has memory parameter $1 + d_\eta$. We present two theorems, for the fractional and standard cointegration cases respectively, using a different definition of the memory parameter of the cointegrating error for each case.

In Section V, 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 VI, we present simulation results on the OLS estimator of θ.

In Section VII, we propose a method of moments estimator of the error and mirostructure feedback coefficients and variances. The estimator is based on the observed tick-time returns.

In Section VIII, we present data analyses of prices of classified stocks from a single company, buy and sell prices of a single stock, and transaction prices of stocks of two companies in the same industry, all of which provide evidence of fractional cointegration. We also consider the information share, which can be estimated based on the method of moments estimators from Section VII.

In Section IX, 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 also show that the original model yields volatility feedback, and a modified version of the model can yield Granger causality. We trace all of these clock-time properties to their tick-time source.

In Section X, we provide some remarks and discuss possible further generalizations of our model and related future work.

II A Pure-Jump Model For Log Prices

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 efficient price shocks $\{e_{i,k}\}$ and the microstructure noise $\{\eta_{i,k}\}$, for Asset $i = 1, 2$. We assume that the $\{e_{i,k}\}$ are *i.i.d.* $N(0, \sigma^2_{i,e})$. The fractional cointegration case corresponds to $d_\eta \in (-\frac{1}{2}, 0)$. In this fractional case, we assume that for $i = 1, 2$, the $\{\eta_{i,k}\}$, which are mutually independent, obey a fractional Gaussian noise model, with common memory parameter d_η, i.e.

$$\eta_{i,k} = B_H(k+1) - B_H(k) \tag{1}$$

where $B_H(t)$ is fractional Brownian motion with memory parameter $d_\eta = H - \frac{1}{2} \in (-\frac{1}{2}, 0)$. In this case, we will denote $\sigma^2_{i,\eta} = \text{var}(\eta_{i,k})$. For details on the fractional Gaussian noise, see pages 318–332 of Samorodnitsky and Taqqu (1994). The reason we choose the fractional Gaussian noise model is that it leads to a very simple expression for the variance of a partial sum, which is useful in the proof of Lemma 1.

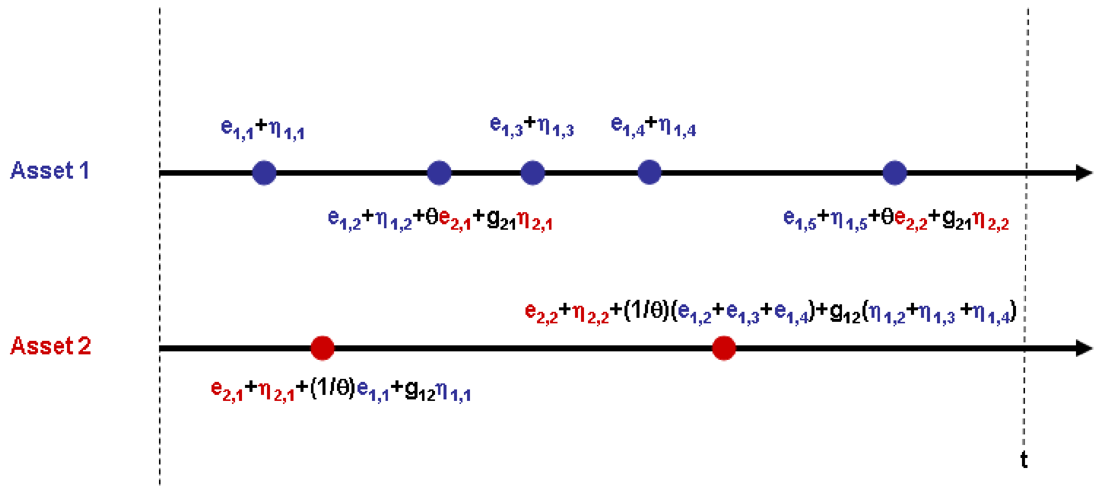
The standard cointegration case corresponds to $d_\eta = -1$, and here we assume that $\eta_{i,k} = \xi_{i,k} - \xi_{i,k-1}$, where $\{\xi_{i,k}\}_{k=1}^{\infty}$ are $i.i.d.$ $(0, \sigma^2_{i,\xi})$ noise series, with the nonrandom initialization $\xi_{i,0} = 0$, $i = 1, 2$. In this case, $\text{var}(\eta_{i,k}) = \sigma^2_{i,\eta} = 2\sigma^2_{i,\xi}$.

The normality assumption on the efficient price shocks $\{e_{i,k}\}$ is only used in Theorem 5. The normality assumption on the microstructure noise $\{\eta_{i,k}\}$ in the fractional case may be relaxed by considering a Fractional Laplace Noise. See Kozubowski, Meerschaert and Podgorski (2006). Note that we do not assume normality of the $\{\eta_{i,k}\}$ in the standard cointegration case.

We now describe the tick-time return interactions that yield cointegration in our model. We will 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 efficient price 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 efficient price 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.

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, an efficient price 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 efficient price 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.

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 follow LMSD models with memory parameters $d_{\tau_1}, d_{\tau_2} \in (0, \frac{1}{2})$. The corresponding counting processes are denoted by $N_i(t)$. Denote the clock time for the k-th transaction of Asset i by $t_{i,k}$. We also assume that $\{\tau_{1,k}\}$ and $\{\tau_{2,k}\}$ are mutually independent and also independent of all disturbance series $\{e_{1,k}\}$, $\{e_{2,k}\}$, $\{\eta_{1,k}\}$ and $\{\eta_{2,k}\}$, which implies that $N_1(\cdot)$ and $N_2(\cdot)$ are mutually independent and independent of all disturbance series. Finally, all disturbance series are assumed to be mutually independent.

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

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

Note that (2) 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 (2) as

$$\log P_{1,t} = \underbrace{\Big(\sum_{k=1}^{N_1(t)} e_{1,k} + \sum_{k=1}^{N_2(t)} \theta e_{2,k} \Big)}_{\text{common component}} + \underbrace{\Big(\sum_{k=1}^{N_1(t)} \eta_{1,k} + \sum_{k=1}^{N_2(t)} g_{21}\eta_{2,k} \Big)}_{\text{microstructure component}} - \underbrace{\sum_{k=N_2(t_{1,N_1(t)})+1}^{N_2(t)} (\theta e_{2,k} + g_{21}\eta_{2,k})}_{\text{end effect}} \qquad (3)$$

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

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. This is accomplished in Lemma 1 and the proof of Theorem 4 for the fractional and standard cointegration cases, respectively. The end-effect terms are random sums over time periods that are $O_p(1)$ as $t \to \infty$, and hence are negligible compared to all other terms. See the discussion around (11) and Lemma 3. 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 $I(1+d_\eta)$ (by Theorems 3 and 4), the log-price series are cointegrated.

From (2) it can be seen that our model for the log price series can be represented in terms of subordinated Brownian motions and fractional Brownian motions, in the spirit of Clark (1973). For example, when $H \in (0, 1/2)$, $\log P_{1,t}$ can be written (up to a constant term) as

$$\Big\{B_1\Big(N_1(t)\Big) + \theta B_2\Big(N_2(t_{1,N_1(t)})\Big)\Big\} + \Big\{B_{1,H}\Big(N_1(t+1)\Big) + g_{21} B_{2,H}\Big(N_2(t_{1,N_1(t)}+1)\Big)\Big\}$$

where B_1 and B_2 are mutually independent Brownian motions, independent of the mutually independent fractional Brownian motions $B_{1,H}$ and $B_{2,H}$. The arguments for $B_{1,H}$ and $B_{2,H}$ would be t rather than $t+1$, and the constant would be zero, if we had defined fractional Gaussian noise as the increment of a fractional Brownian motion at times $k-1$ and k rather than the standard k, $k+1$. Here, the directing processes are the non-decreasing processes $N_1(t)$ and $N_2(t_{1,N_1(t)})$, yielding a pure-jump price process.

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 (2). Furthermore, their model implies standard cointegration, with cointegrating parameter that is known to be 1, and a single efficient shock component.

III Long-Term Martingale-Type Properties Of the Log Prices

Define $\lambda = 1/E^0(\tau_k)$, where E^0 denotes expectation under the Palm distribution, i.e., the stationary distribution of $\{\tau_k\}$. Note that λ is a positive finite constant.

From (3) it can be seen that the microstructure components of the log price are random sums of the microstructure noise. The following lemma shows that such random sums have memory parameter $1 + d_\eta < 1$, where d_η is the memory parameter of the microstructure noise.

Lemma 1 *For durations $\{\tau_k\}$ generated by an LMSD model with memory parameter $d_\tau \in (0, \frac{1}{2})$ and a fractional Gaussian noise series $\{\eta_k\}$ with memory parameter $d_\eta = H - \frac{1}{2} \in (-\frac{1}{2}, 0)$, which is independent of $\{\tau_k\}$,*

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

as $t \to \infty$, where $\sigma^2 = \mathrm{var}[B_H(1)]$ in Equation (1).

The following two theorems show that the log-price series in Model (2) 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 discretization interval increases, as would happen for a Martingale difference series.

Define $\lambda_1 = 1/E^0(\tau_{1,k})$ and $\lambda_2 = 1/E^0(\tau_{2,k})$.

Theorem 1 *For the log-price series in Model (2),*

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

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

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

$$\begin{aligned} r_{1,j} &= \sum_{k=N_1((j-1)\Delta t)+1}^{N_1(j\Delta t)} (e_{1,k} + \eta_{1,k}) + \sum_{k=N_2(t_{1,N_1((j-1)\Delta t)})+1}^{N_2(t_{1,N_1(j\Delta t)})} (\theta e_{2,k} + g_{21}\eta_{2,k}) \\ r_{2,j} &= \sum_{k=N_2((j-1)\Delta t)+1}^{N_2(j\Delta t)} (e_{2,k} + \eta_{2,k}) + \sum_{k=N_1(t_{2,N_2((j-1)\Delta t)})+1}^{N_1(t_{2,N_2(j\Delta t)})} (\frac{1}{\theta} e_{1,k} + g_{12}\eta_{1,k}) \ . \end{aligned} \tag{4}$$

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 $\Delta t \to \infty$.*

IV Properties of the Cointegrating Error

We show that Model (2) implies a cointegrating relationship between the two series, treating the fractional and standard cointegration cases separately.

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

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

as $t \to \infty$, where $C > 0$. Thus, $\log P_{1,t}$ and $\log P_{2,t}$ are 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 $\text{var}(\log P_{1,t} - \theta \log P_{2,t})$, 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

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

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.

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

Theorem 4 *Under Model (2), with $d_\eta = -1$, the cointegrating error $(\log P_{1,t} - \theta \log P_{2,t})$ has short memory. Thus, $\log P_{1,t}$ and $\log P_{2,t}$ are cointegrated.*

V Least-Squares Estimation of the Cointegrating Parameter

Assume that the log-price series are observed at integer multiples of Δt. The proposed model (2) becomes

$$\begin{aligned} \log P_{1,j} &= \sum_{k=1}^{N_1(j\Delta t)} (e_{1,k} + \eta_{1,k}) + \sum_{k=1}^{N_2(t_{1,N_1(j\Delta t)})} (\theta e_{2,k} + g_{21}\eta_{2,k}) \\ \log P_{2,j} &= \sum_{k=1}^{N_2(j\Delta t)} (e_{2,k} + \eta_{2,k}) + \sum_{k=1}^{N_1(t_{2,N_2(j\Delta t)})} (\frac{1}{\theta} e_{1,k} + g_{12}\eta_{1,k}) \ . \end{aligned} \tag{5}$$

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

Theorem 5 *For the discretely-sampled log-price series in (5), 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 $\delta > 0$,*

$$n^{-d_\eta - \delta}(\hat{\theta} - \theta) \xrightarrow{p} 0, \qquad \text{as } n \to \infty,$$

The rate of convergence of $\hat{\theta}$ improves as d_η decreases. In the standard cointegration case $d_\eta = -1$, 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. Even if we consider ACD durations, which are indeed strong-mixing in tick time, there is no guarantee that they lead to returns that are strong-mixing in clock time, even in the standard cointegration case.

VI Simulations on OLS Estimation of the Cointegrating Parameter

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

First, we simulate two mutually independent duration process $\{\tau_{i,k}\}$, $i = 1, 2$, for Asset i. 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 common memory parameter $d_\tau \in (0, 1/2)$. Here, we assume that the $\{\epsilon_{i,k}\}$ follow a Weibull distribution with shape parameter $\kappa = 1$ and scale parameter $\tilde{\lambda} = 1$, so that $E(\epsilon_{i,k}) = 1$. The $\{h_{i,k}\}$ are simulated from a Gaussian ARFIMA(0, d_τ, 0) model with unit innovation variance.

Using the simulated durations $\{\tau_{i,k}\}$, $i = 1, 2$, we obtain 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 $i.i.d.$ $N(0, 1)$. When $d_\eta \in (-\frac{1}{2}, 0)$, the $\{\eta_{i,k}\}$ are given by fractional Gaussian noise as defined in (1) with $\sigma^2_{1,\eta} = \sigma^2_{2,\eta} = 1$, simulated using the algorithm on page 218 of Beran (1994). When $d_\eta = -1$, $\{\eta_{i,k}\}$ are simulated as the differences of two independent $i.i.d.$ zero-mean standard normal series $\{\xi_{i,k}\}$.

We then construct the log-price series $\{\log P_{i,j}\}_{j=1}^n, i = 1, 2$ from (2), using a fixed sampling interval Δt. The estimated cointegrating parameter $\hat{\theta}$ is obtained by regressing $\{\log P_{1,j}\}_{j=1}^n$ on $\{\log P_{2,j}\}_{j=1}^n$.

In the study, we fixed the cointegrating parameter at $\theta = 1$. We considered various values of the parameters $g_{12}, g_{21}, \Delta t, d_\tau, d_\eta$ and the sample size n. For each parameter configuration, 500 replications of the log-price series were generated. The results are summarized in Table 1. The parameter values in block A of the table are varied one by one in the other blocks.

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

Block	Simulation Parameters						Estimated Cointegrating Parameter $\hat{\theta}$		
	g_{12}	g_{21}	Δt	d_τ	d_η	n	Mean	Standard Deviation	Root Mean Squared Error
A	3	3	100	0.25	–0.25	200	0.9689	0.1730	0.1758
B	3	3	100	0.25	–0.25	50	0.9546	0.2504	0.2545
	3	3	100	0.25	–0.25	800	0.9902	0.1315	0.1319
C	3	3	20	0.25	–0.25	200	0.9322	0.2463	0.2554
	3	3	400	0.25	–0.25	200	0.9801	0.1324	0.1339
D	3	3	100	0.05	–0.25	200	0.9668	0.1698	0.1730
	3	3	100	0.15	–0.25	200	0.9883	0.1885	0.1889
	3	3	100	0.35	–0.25	200	0.9766	0.1709	0.1725
	3	3	100	0.45	–0.25	200	0.9759	0.1641	0.1659
E	3	3	100	0.25	–0.05	200	0.7150	0.6479	0.7078
	3	3	100	0.25	–0.15	200	0.9138	0.3591	0.3693
	3	3	100	0.25	–0.35	200	1.0002	0.0731	0.0731
	3	3	100	0.25	–0.45	200	0.9961	0.0538	0.0539
	3	3	100	0.25	–0.75	200	0.9984	0.0242	0.0242
	3	3	100	0.25	–1.00	200	0.9972	0.0235	0.0236
F	1	1	100	0.25	–0.25	200	0.9991	0.0029	0.0030
	9	9	100	0.25	–0.25	200	0.6856	0.5248	0.6117

Not surprisingly, as the sample size n increases, the bias, standard deviation and Root Mean Squared Error (RMSE) of $\hat{\theta}$ decrease, as seen in Blocks A and B. The same pattern holds when the sampling interval Δt is increased (see Blocks A and C), since the end effect is not as important when Δt is large. Similarly, $\hat{\theta}$ performers better when the microstructure noise feedback coefficients g_{12} and g_{21} are small, as shown in Blocks A and F. This is because the cointegrating relationship is obscured less by the microstructure noise when g_{12} and g_{21} are small.

More interestingly, the memory parameter of the duration process d_τ has no discernible impact on the performance of $\hat{\theta}$ (See Blocks A and D). This is in agreement with the theoretical derivation for Theorem 5, in which d_τ plays no role. On the other hand, the variance of $\hat{\theta}$ decreases sharply as the memory parameter of the microstructure noise d_η decreases (see Blocks A and E). This is consistent with the results in Theorem 5, though the case $d_\eta = -0.75$ is not covered by the theorem. (In this case, the microstructure noise $\{\eta_{i,k}\}, i = 1, 2$ were simulated as the difference of the corresponding fractional Gaussian noise with memory parameter $d_\eta + 1 = 0.25$.)

VII Method of Moments Parameter Estimation

We propose a simple (though clearly inefficient) transaction-level parameter estimation procedure for model (2) using the method of moments.

Consider Figure 1. The returns for the first, the third and the fourth transactions of Asset 1 have a simple structure, consisting of the sum of the current Asset 1 efficient and microstructure dusturbances, since for these transactions there was no intervening Asset 2 transaction. In general, we define a Type I transaction of Asset 1 as any Asset 1 transaction with no intervening Asset 2 transaction. Since $\{e_{1,k}\}$ and $\{\eta_{1,k}\}$ are assumed to be mutually independent, the corresponding return variance for Type I Asset 1 transaction is $\sigma^2_{1,e} + \sigma^2_{1,\eta}$.

Consider the third and the fourth transactions of Asset 1 in Figure 1. This is a pair of adjacent Type I transactions. Since $\{e_{1,k}\}$ are assumed to be *i.i.d.*, the covariance of the returns of such a pair is equal to the lag-1 autocovariance of the microstructure noise series $\{\eta_{1,k}\}$, given by

$$f(\sigma^2_{i,\eta}, d_\eta) = \begin{cases} (2^{2d_\eta} - 1)\sigma^2_{i,\eta}, & d_\eta \in (-\frac{1}{2}, 0) \\ (2^{2d_\eta+2} - \frac{1}{2}3^{2d_\eta+1} - \frac{7}{2})\sigma^2_{i,\eta}, & d_\eta \in (-1, -\frac{1}{2}) \\ -\sigma^2_{1,\xi} = -\frac{1}{2}\sigma^2_{i,\eta}, & d_\eta = -1. \end{cases} \tag{6}$$

Next, we define Type-II transactions of Asset 1 to be those with exactly one intervening Asset 2 transaction. An example is the second and the fifth Asset 1 transaction in Figure 1. The corresponding return variance is $\sigma^2_{1,e} + \sigma^2_{1,\eta} + \theta^2\sigma^2_{2,e} + g^2_{21}\sigma^2_{2,\eta}$.

We can define Type-I, adjacent pairs of Type-I and Type-II transactions of Asset 2 in a similar manner.

For both assets, we compute the sample variance of the Type-I and Type-II transactions, as well as the sample covariance between adjacent pairs of Type-I transactions. The method-of-moments estimates $\hat\sigma^2_{1,e}$, $\hat\sigma^2_{1,\eta}$, $\hat\sigma^2_{2,e}$ $\hat\sigma^2_{2,\eta}$, $\hat g_{12}$ and $\hat g_{21}$ are given as the solutions to the following system, consisting of six equations.

$$\begin{aligned}
\widehat{\text{var}}(\text{Type I; Asset } i) &= \hat\sigma^2_{i,e} + \hat\sigma^2_{i,\eta} \qquad (i = 1, 2) \\
\widehat{\text{var}}(\text{Type II; Asset 1}) &= \hat\sigma^2_{1,e} + \hat\sigma^2_{1,\eta} + \hat\theta^2\sigma^2_{2,e} + \hat g^2_{21}\hat\sigma^2_{2,\eta} \\
\widehat{\text{var}}(\text{Type II; Asset 2}) &= \frac{1}{\hat\theta^2}\hat\sigma^2_{1,e} + \hat g^2_{12}\hat\sigma^2_{1,\eta} + \hat\sigma^2_{2,e} + \hat\sigma^2_{2,\eta} \\
\widehat{\text{cov}}(\text{Adjacent pairs of Type I; Asset } i) &= f(\hat\sigma^2_{i,\eta}, \hat d_\eta) \qquad (i = 1, 2)
\end{aligned} \tag{7}$$

where $\hat\theta$ is an OLS estimator of θ as justified in Section V, $\hat d_\eta$ is obtained from the cointegrating residuals using the log-periodogram regression method, and the function $f(\hat\sigma^2_{i,\eta}, \hat d_\eta)$ is defined in (6). Note that since both g_{21} and g_{12} appear as squares in the corresponding variances, we assume both to be positive.

A disadvantage of the method of moments is that the variance estimates $\hat\sigma^2_{1,e}$, $\hat\sigma^2_{1,\eta}$, $\hat\sigma^2_{2,e}$ $\hat\sigma^2_{2,\eta}$ can be negative. The same is true for $\hat g^2_{21}$ and $\hat g^2_{12}$. We set the corresponding estimates to be zero, if negative values are obtained in solving (7).

Table 2: Parameter Estimation using the Method of Moments

$g_{21} = g_{12}$	$\hat\sigma^2_{1,e}$	$\hat\sigma^2_{2,e}$	$\hat\sigma^2_{1,\eta}$	$\hat\sigma^2_{2,\eta}$	$\hat g_{21}$	$\hat g_{12}$	$\hat\theta$	$\hat d_\eta$
0.1	1.1858	1.1882	0.8110	0.8123	0.1038	0.0678	0.9896	–0.3400
2.0	1.2427	1.2448	0.7540	0.7557	2.3550	2.3493	1.0011	–0.3639
5.0	1.0069	1.0084	0.9990	0.9999	5.1022	5.0626	0.9479	–0.2739
10.0	0.9591	0.9607	1.0866	1.0900	10.2175	9.4743	0.7322	–0.2643

We carried out a small simulation study to evaluate the performance of the method of moments estimators. The parameter values were $d_{\tau_1} = d_{\tau_2} = 0.25$, $d_\eta = -0.25$, $\theta = 1$, $\mathrm{var}(e_{i,k}) = \mathrm{var}(\eta_{i,k}) = 1, (i = 1, 2)$. We varied g_{12} and g_{21}, which we took to be equal. The $\{h_{i,k}\}$ were simulated as in Section VI. For each of 100 realizations, we simulated log prices in model (2) for a clock-time span of $n\Delta t$, with $n = 100$, $\Delta t = 50$. The estimators $\hat{\theta}$ and $\hat{d}_\eta$ were constructed from the $n = 100$ clock-time log prices, and then these estimates were used together with the tick-time returns to yield the method of moments estimators. The estimator $\hat{d}_\eta$ was based on using the differenced cointegrating residuals in a log-periodogram regression with $n^{0.5}$ frequencies, and then adding 1. The results, given in Table 2, are averages based on the 100 realizations. As g_{12} is increased, all estimates except $\hat{\theta}$ become less biased.

VIII Data Analysis

We analyze three empirical examples, corresponding to three different scenarios: prices of two classified stocks from a given company, buy and sell prices of a single stock, and prices of two different stocks within the same industry. In the first two situations, the cointegrating parameter θ would be expected to be 1, while in the third situation, there is no clear *a priori* value for θ.

Other possible scenarios that we do not pursue here include: (1) stock and option prices of a given company; (2) corporate bond prices at different maturities for a given company.

We obtained our data from the TAQ database of WRDS. We considered daily transactions between 9:30 AM to 4:00 PM. Overnight durations and returns are ignored, as implemented in, for example Hasbrouck (1995).

A Prices of Classified Stocks from a Given Company

Waddell & Reed Financial, Inc.'s initial public offering of Class A common stock (WDR) took place on March 5, 1998. The Class B common stock (WDRB) began trading on November 6, 1998 following the tax-free spin off of Waddell & Reed Financial, Inc. from its former parent company, Torchmark

Corporation. The Class A and Class B common stock were combined as of the close of business on April 30, 2001. Class A stock has one vote per share while Class B stock has five votes per share.

We would expect the cointegrating parameter for the Class A and Class B log-prices to be close to 1. This is because Class A and B stocks have the same expected future cash flow. The only difference is the voting right, which only changes infrequently.

Our data spans the time period January 3, 2000 to April 30, 2001. Overall, there are 55,255 transactions of WDR and 10,689 transactions of WDRB. The average durations are 131.78 and 653.16 seconds for WDR and WDRB, respectively.

Based on the log-prices of WDR and WDRB observed every Δt seconds, we computed the ordinary least squares estimates $\hat{\theta}$ of θ, as well as log-periodogram regression estimates $\hat{d}$ of the memory parameters for both log-price series as well as the cointegrating residuals (RES). The log-periodogram regression estimators were based on differences (with 1 subsequently added to the result), and used a number of frequencies equal to $n^{0.7}$ for $\hat{d}_{WDR}$ and $\hat{d}_{WDRB}$, and $n^{0.6}$ for $\hat{d}_{RES}$, chosen by visual inspection. The asymptotic estimated standard errors are also reported. The three choices of Δt correspond to 1 minute, 5 minutes and 30 minutes. The results are reported in Table 3.

Table 3: Prices of Classified Stocks from a Given Company: WDR and WDRB

Δt (sec)	n	$\hat{\theta}$	$\hat{d}_{WDR}$	$\hat{d}_{WDRB}$	$\hat{d}_{RES}$	$SE(\hat{d}_{WDR})$	$SE(\hat{d}_{WDRB})$	$SE(\hat{d}_{RES})$
60	118,063	0.98428	1.0024	1.0343	0.7236	0.0108	0.0108	0.0193
300	23,718	0.98429	1.0198	1.0354	0.7964	0.0189	0.0189	0.0312
1800	4,008	0.98432	1.0388	1.0371	0.7642	0.0352	0.0352	0.0532

It can be seen that, for all choices of Δt, the estimated cointegrating parameter is very close to 1, the log-price series both have memory parameters that are insignificantly different from 1, while the residuals have a memory parameter that is significantly greater than 0 and significantly less than 1. Thus, there is evidence of fractional, but not standard, cointegration. Although the evidence of fractional cointegration is strong, the degree of this cointegration seems rather weak, as the memory parameter is only reduced from (roughly) 1 to 0.75.

B Buy and Sell Prices of a Single Stock

We consider buy and sell prices for a single stock. We analyze two different buy-sell data sets, one for a heavily-traded stock, Coca Cola (KO), and the other for a thinly-traded stock, Commercial Federal Bank (CFB). The data span the period from June 1, 2000 to August 31, 2000. Within this three-month period, there were 65 trading days (The market is closed on July 4, 2000) and 144,606 transactions of KO, 6,397 transactions of CFB.

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 sell order. If the transaction price is lower, it is labeled as a buy 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.

We study the buy and sell prices because they are closely related so that a strong cointegrating relationship is expected. Separating the buy and sell prices makes two series free of bid-ask bounce.

Table 4: Buy and Sell Prices of a Single Stock

Δt (sec)	n	$\hat{\theta}$	$\hat{d}_{KO}^{buy}$	$\hat{d}_{KO}^{sell}$	$\hat{d}_{RES}$	$SE(\hat{d}_{KO}^{buy})$	$SE(\hat{d}_{KO}^{sell})$	$SE(\hat{d}_{RES})$
60	25,227	0.9997468	1.0320	1.0368	0.1074	0.0185	0.0185	0.0509
300	5,062	0.9997473	1.0432	1.0458	0.0825	0.0324	0.0324	0.0496
1800	845	0.9997408	1.1451	1.1145	0.0350	0.0433	0.0433	0.0606

Δt (sec)	n	$\hat{\theta}$	$\hat{d}_{CFB}^{buy}$	$\hat{d}_{CFB}^{sell}$	$\hat{d}_{RES}$	$SE(\hat{d}_{CFB}^{buy})$	$SE(\hat{d}_{CFB}^{sell})$	$SE(\hat{d}_{RES})$
60	24,704	0.9988320	1.0241	1.0236	0.0918	0.0186	0.0186	0.0512
300	4,965	0.9988289	1.0153	1.0462	0.1139	0.0213	0.0213	0.0499
1800	840	0.9988086	0.8859	0.9121	0.0844	0.0434	0.0434	0.1191

The results are given in Table 4. The number of frequencies used in the log periodogram regressions vary from $n^{0.5}$ to $n^{0.8}$, chosen by visual inspection of log-log periodogram plots. As expected, the estimated cointegrating parameter is close to 1. Evidence of strong cointegration is found for both stocks. Furthermore, there is some evidence that the cointegration is fractional, not standard.

C Transaction Prices of Two Company Stocks within an Industry

We consider prices for the stocks of two companies within the same industry. Unlike in the previous examples, here there is no *a priori* value for the cointegrating parameter θ.

The two companies we study are GM (GM) and Ford (F), within a one month period from June 1 to June 30, 2000. The results are given in Table 5. The cointegrating relationship between GM and Ford

Table 5: Transaction Prices of Two Company Stocks within an Industry

Δt (sec)	n	$\hat{\theta}$	$\hat{d}_{GM}$	$\hat{d}_F$	$\hat{d}_{RES}$	$SE(\hat{d}_{GM})$	$SE(\hat{d}_F)$	$SE(\hat{d}_{RES})$
60	8,542	0.918506	0.9774	0.9848	0.8914	0.0172	0.0172	0.0270
300	1,715	0.918512	0.9659	0.9859	0.9468	0.0326	0.0326	0.0473
1800	286	0.918533	0.8968	1.1208	1.0338	0.0668	0.0668	0.1175

prices is much weaker than for the previous two examples, and because of this it is only significant for the smallest choice of the sampling interval Δt.

D Information Share

In Hasbrouck (1995), a single security is traded on several markets and different market prices share a common random-walk component. Suppose there are two markets. Then the clock-time log stock prices at time j on two different markets can be written as

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

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. We regard $\tilde{e}_{i,s}, (i = 1, 2)$ as the innovation originating from the i-th market. The model in Hasbrouck (1995) is defined in clock time and is estimated using a one-second sampling interval. There, the information share of market i is defined as

$$S_i = \frac{\psi_i^2 \Omega_{ii}}{\psi \Omega \psi'},$$

which is the proportional contribution from market i to the total random walk innovation variance. Only the random-walk component is used in constructing the information share since this is the only permanent component.

In our price model (3), we can also evaluate the information share, as described in words above. We consider two series, not necessarily the price of a given security on two different markets. For a given clock-time sampling interval Δt, the information share of Asset 1, denoted by $S_{1,C}$, is given by

$$S_{1,C} = \frac{\operatorname{var}\left(\sum_{k=N_1((j-1)\Delta t)+1}^{N_1(j\Delta t)} e_{1,k}\right)}{\operatorname{var}\left(\sum_{k=N_1((j-1)\Delta t)+1}^{N_1(j\Delta t)} e_{1,k} + \theta \sum_{k=N_2((j-1)\Delta t)+1}^{N_2(j\Delta t)} e_{2,k}\right)} = \frac{\lambda_1 \sigma_{1,e}^2}{\lambda_1 \sigma_{1,e}^2 + \theta^2 \lambda_2 \sigma_{2,e}^2}.$$

Similarly, the information share of Asset 2 is given by

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

Note that only the common component in (3) is used to evaluate the information share, as was also done by Hasbrouck (1995).

In Hasbrouck (1995), since the model is built in clock time, the trading intensities λ_1, λ_2 do not appear explicitly in the information share formulas, but instead the impact of these intensities is reflected in $\psi \Omega \psi'$. By contrast, λ_1, λ_2 appear explicitly in our formulas for $S_{1,C}$ and $S_{2,C}$.

As $\lambda_1/\lambda_2 \to \infty$, $S_{1,C}$ approaches one and $S_{2,C}$ approaches zero. This is consistent with the 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.

To estimate the information share, estimates for the trading intensities λ_1, λ_2, and the efficient innovation variances $\sigma^2_{1,e}$, $\sigma^2_{2,e}$ are needed. 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 VII. We estimate θ using OLS, with $\Delta t = 60$ seconds.

We consider the information shares of the buy and sell prices of a single stock: Coca Cola (KO). A question of interest is whether buy trades contain more information and therefore are more important for the price discovery process than sell trades. The data spans a 65 trading-day period from June 1 to August 31, 2000. The tick-time stock prices are plotted in Figure 2.

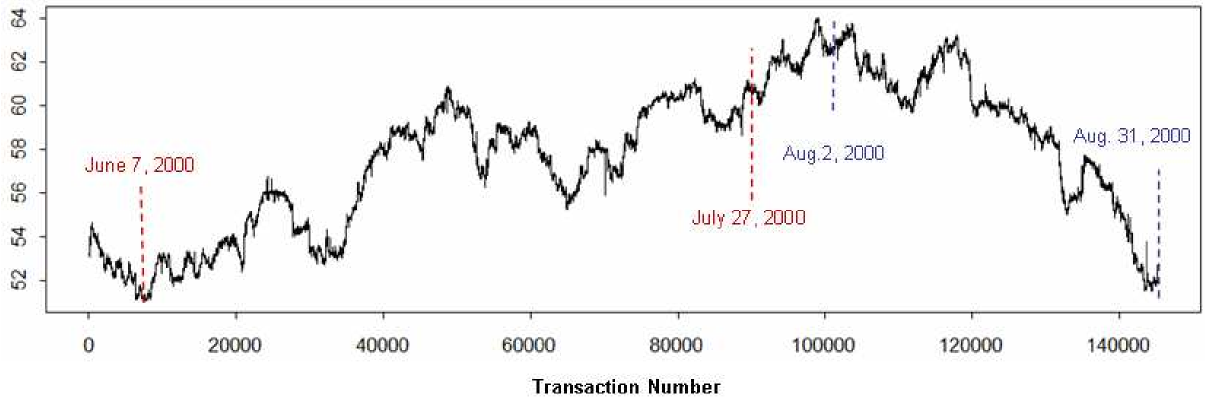
We estimate the information share for each of three clock-time periods. Period one is the entire three-month interval comprising 65 trading days. Period two spans 36 trading days in which the stock price rose by roughly 20%. Period three comprises 22 trading days in which the stock price dropped by approximately 20%. The results are given in Table 6.

Table 6: Information Shares of Buy and Sell Price of KO

Period	Type	# of trades	$\hat{\lambda}_i$(per day)	$\hat{\sigma}^2_{i,e}$	$\hat{\sigma}^2_{i,\eta}$	$\hat{S}_{i,C}$
1: 06/01 – 08/31/2000	Buy	74,856	1151.63	5.38e-07	4.45e-08	0.5144
	Sell	69,750	1073.08	5.46e-07	4.75e-08	0.4856
2: 06/07 – 07/27/2000	Buy	42,804	1189.00	6.96e-07	4.39e-08	0.5813
	Sell	39,437	1095.47	5.44e-07	4.37e-08	0.4187
3: 08/02 – 08/31/2000	Buy	23,800	1081.82	3.15e-07	4.26e-08	0.4120
	Sell	21,626	983.00	4.95e-07	6.74e-08	0.5880

For the entire three-month period, the information shares are almost equally divided between buys and sells. For period two when the stock price increases dramatically, the buy trades possess much more information than sell trades. By contrast, during period three when price has a significant drop, the sell trades have more information.

Figure 2: KO Stock Price in June to August, 2000



IX Modifications of the Model to Capture More Stylized Facts

So far, we have seen that the model (2) 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 (2) to capture two additional key stylized facts: the leverage effect, and portfolio autocorrelation due to nonsynchronous trading. We also show that the original model yields volatility feedback, and a modified version of the model can yield Granger causality. 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 Volatility Feedback

In Model (2) we have assumed that $N_1(\cdot)$ and $N_2(\cdot)$ are mutually independent. Thus, for any fixed sampling interval Δt, the resulting series of counts, $\{\Delta N_{1,j}\}$, $\{\Delta N_{2,j}\}$ are mutually independent, where $\Delta N_{1,j} = N_1(j\Delta t) - N_1([j-1]\Delta t)$ and $\Delta N_{2,j} = N_2(j\Delta t) - N_2([j-1]\Delta t)$. From Clark (1973) (see also Deo, Hurvich, Soulier and Wang (2006)), it is known for univariate series that the autocorrelation properties of realized volatility are related to those of counts. Thus, it may appear that for the bivariate returns (4) corresponding to model (2), the realized volatilities of the two series should be mutually independent. However, inspection of (4) reveals that both $N_1(\cdot)$ and $N_2(\cdot)$ appear in the equations for both return series, $\{r_{1,j}\}$ and $\{r_{2,j}\}$. Therefore, there is reason to suspect that in fact the realized volatilities for the two return series will be mutually dependent. For example, if in a given time period the durations of Asset 1 are shorter than average (yielding a large contribution to the realized volatility of Asset 1), then although this will have no effect on the durations of Asset 2 it will still tend to produce a large number of shocks in the Asset 2 return, due to the return feedback mechanism shown in Figure 1, leading to a large contribution to realized volatility for Asset 2 from this time period.

We performed a small simulation study to confirm the volatility feedback effect. The parameter values were $d_{\tau_1} = d_{\tau_2} = 0.35$, $d_\eta = -0.25$, $\theta = 1$, $\mathrm{var}(e_{i,k}) = \mathrm{var}(\eta_{i,k}) = 1, (i = 1, 2)$. We varied g_{12} and g_{21} (which we took to be equal), to include or exclude the microstructure noise. We also varied the sampling frequency Δt and the expected values of the durations of the two assets. The $\{h_{i,k}\}$ were simulated as in Section VI with unit innovation variance. The results, presented in Table 7, are based on 100 realizations of length $n = 500$. We denote by r_{RV} the contemporaneous cross correlation of the realized volatilities of the two assets. Here, the realized volatilities were computed by summing the tick-time squared returns within each time period of width Δt. This version of realized volatility was also considered in Andersen, Bollerslev, Frederiksen and Nielsen (2006).

It is seen from Table 7 that, as the average duration decreases or the sampling interval increases, r_{RV} increases. This correlation decreases when microstructure noise is introduced.

Table 7: Simulation for the Volatility Feedback Effect

Δt	$g_{12} = g_{21}$	$E(\tau_{1,k})$	$E(\tau_{2,k})$	mean(r_{RV})	SE(r_{RV})
10	0	2.18	2.18	0.5914	0.0084
10	0	2.18	21.8	0.1975	0.0079
10	0	21.8	21.8	0.1582	0.0090
100	0	2.18	2.18	0.9609	0.0016
100	0	2.18	21.8	0.5763	0.0100
100	0	21.8	21.8	0.6157	0.0082
10	5	2.18	2.18	0.1140	0.0048
10	5	2.18	21.8	0.1591	0.0060
10	5	21.8	21.8	0.0508	0.0037

B 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, Kedlac 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 efficient shock feedback coefficients in Model (2), $1/\theta$ and θ, to θ_{12} and θ_{21}, respectively. The resulting return series become

$$\begin{aligned} r_{1,j} &= \sum_{k=N_1((j-1)\Delta t)+1}^{N_1(j\Delta t)} e_{1,k} + \sum_{k=N_2(t_{1,N_1((j-1)\Delta t)})+1}^{N_2(t_{1,N_1(j\Delta t)})} \theta_{21} e_{2,k} \\ r_{2,j} &= \sum_{k=N_2((j-1)\Delta t)+1}^{N_2(j\Delta t)} e_{2,k} + \sum_{k=N_1(t_{2,N_2((j-1)\Delta t)})+1}^{N_1(t_{2,N_2(j\Delta t)})} \theta_{12} e_{1,k}. \end{aligned} \tag{8}$$

Lemma 2 *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 (8). 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 8 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.

Table 8: 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.

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

C Granger Causality

Consider the return model (8). Suppose that $\theta_{12} \neq 0$ but $\theta_{21} = 0$. Then the clock-time returns for Asset 2, $\{r_{2,j}\}$, will contain contributions from both tick-time shock series $\{e_{1,k}\}$ and $\{e_{2,k}\}$, whereas the returns from Asset 1 will only contain contributions from $\{e_{1,k}\}$. Roughly speaking, new information flows from Asset 1 to Asset 2, but not from Asset 2 to Asset 1. It seems plausible that the directionality of the tick-time interactions in prices should induce some form of causality in clock time, with the same directionality. If, for example, we were to fit a (misspecified) bivariate $AR(1)$ model to the return data, we might expect to find that $\{r_{1,j}\}$ Granger-causes $\{r_{2,j}\}$ but that $\{r_{2,j}\}$ does not Granger-cause $\{r_{1,j}\}$. To get a clearer idea of why this might happen, note that although the individual return series are serially uncorrelated, there is a cross-correlation between the two returns when Asset 1 leads Asset 2 but not when Asset 2 leads Asset 1. This follows from the proof of Lemma 2 and is also in accord with intuition. For example, if Asset 1 was the last asset to trade in time period $j-1$ then the corresponding Asset 1 shock will be incorporated into the Asset 2 return at a time period after $j-1$. However, no Asset 2 shock will ever be incorporated into the Asset 1 return.

To study the causality properties of Model (8) under various restrictions, we simulated returns from the model using the same parameter values as in Table 7 (unless otherwise indicated). For each pair of simulated returns, we ran two OLS regressions: (1) Current returns of Asset 1 on lagged returns of both assets; (2) Current returns of Asset 2 on lagged returns of both assets. Table 9 reports means (over the 100 replications), and corresponding standard errors, for the estimated coefficient of lagged returns of Asset 2 in regression (1), and lagged returns of Asset 1 in regression (2). Denoting the population versions of these two regression coefficients as π_{12} and π_{21}, it is seen that there is strong evidence that $\pi_{21} > 0$ but we cannot reject the hypothesis that $\pi_{12} = 0$. Thus, at least in the context of the misspecified bivariate $AR(1)$ model, it seems that the above-conjectured patterns in Granger causality indeed hold. The strength of π_{21} diminishes as Δt increases, since it is the nonsynchronous trading effect that induces the causality.

Table 9: Simulation Results for Granger Causality

Δt	θ_{12}	θ_{21}	mean($\hat{\pi}_{12}$)	SE($\hat{\pi}_{12}$)	t-stat($\hat{\pi}_{12}$)	mean($\hat{\pi}_{21}$)	SE($\hat{\pi}_{21}$)	t-stat($\hat{\pi}_{21}$)
10	10	0	–0.0001	0.0005	–0.20	4.6588	0.0565	82.40
20	10	0	–0.0006	0.0007	–0.84	4.8621	0.0728	66.82
50	10	0	0.0011	0.0010	1.14	4.6951	0.0911	51.56
20	5	0	–0.0013	0.0014	–0.90	2.2726	0.0378	60.14
20	1	0	–0.0041	0.0033	–1.25	0.2402	0.0088	27.42

D 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 efficient 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, $\text{corr}(e_{i,k}, \nu_{i,k+1}) = \phi_i \sigma_{\nu_i}/\sqrt{\phi_i^2 \sigma_{\nu_i}^2 + 1}$. As described in Section VI, 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.* Weibull with shape parameter κ_i and scale parameter $\tilde{\lambda}_i$ such that $E(\epsilon_{i,k}) = 1$. A simple calculation yields

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

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 (8). 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 $\hat{\text{corr}}(r_{i,j}, r_{i,j+1}), \hat{\text{corr}}(|r_{i,j}|, r_{i,j+1})$ and $\hat{\text{corr}}(|r_{i,j}|, r_{i,j-1})$ 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 $i = 1, 2$, $\kappa_i = \tilde{\lambda}_i = 1, (i = 1, 2)$. Results are based on 5000 realizations, and reported in Table 10.

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.

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) 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.

X Remarks and Suggestions for Future Work

Remark 1: Although we have assumed that the durations are generated by the LMSD model, the theoretical results of Sections III, IV and V on cointegration continue to hold under the more general conditions given in Theorem 1 of Deo, Hurvich, Soulier and Wang (2006), which would allow, for example, the Autoregressive Conditional Duration model of Engle and Russell (1998).

Remark 2: In the fractional cointegration case, we have assumed that the memory parameter d_η of the microstructure noise components $\{\eta_{i,k}\}$ $(i = 1, 2)$ lies in the range $(-0.5, 0)$. Two generalizations may be of interest. First is the case $d_\eta = 0$, which implies that there is no cointegration, as would be the case for a factor model such as CAPM, see Sharpe (1964). Second is the case $d_\eta \in (-1, -1/2)$, which should presumably result in stronger fractional cointegration than we have allowed with our previous restrictions. We have only a partial understanding of what would happen in this case. To obtain $d_\eta \in (-1, -1/2)$, we could define $\{\eta_{i,k}\}$ as the difference of a fractional Gaussian noise. The random partial sum of such a series is stationary, so its variance does not grow with t. Thus in order to gauge the long memory parameter of the partial sum we would need to consider something other than its variance. Perhaps the autocorrelation of the partial sum could be derived, but this is so far intractable, because it is related to the autocorrelation of fractional Gaussian noise, at a random lag. It is easily seen that we would get fractional cointegration in this case, but we are currently unable to establish the exact degree of cointegration; we can merely show that it is at least is strong as what would be obtained for $d_\eta \in (-1/2, 0)$. The OLS estimator of θ would still be consistent, with a rate that is at least as fast as $\sqrt{n}$.

Remark 3: There is an important caveat regarding the Martingale property in the special case of model (2) 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.

Remark 4: In certain situations it might be useful to allow for different additive constants in the model (2) for $\log P_{1,t}$ and $\log P_{2,t}$. In particular, we could consider adding a positive constant C to $\log P_{1,t}$ and subtracting C from $\log P_{2,t}$ ($cf.$ Roll, 1984). This would have no effect on any of the theoretical results on cointegration or the OLS estimation of the cointegrating parameter. In the example we considered in Section VIII of buy and sell prices of a given stock, the constant C could represent transaction costs. Note that it would still not necessarily be the case that $\log P_{1,t}$ exceeds $\log P_{2,t}$, but such a constraint is not needed here since the buy and sell markets trade nonsynchronously. The constant C would not be estimable by OLS (which would still be run without an intercept), but could be estimable from the cointegrating residuals if there is strong cointegration, with $d_\eta < -1/2$.

Finally, 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 (2). Semiparametric estimators could be considered, since by the remark above the results of this paper do not require a parametric model for durations.

Generalizations of our model (2) could also be studied. For example, we could relax the assumptions that $N_1(\cdot)$ and $N_2(\cdot)$ are independent. For a related model, see Hsieh and Hurvich (2006). Finally, the possibility of more than two assets as well as deterministic linear trends should be considered.

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] 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.

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

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

[6] 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.

[7] 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.

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

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

[10] Chen, W. and Hurvich, C. (2006), Semiparametric Estimation of Fractional Cointegrating Subspaces, Forthcoming in *Annals of Statistics*.

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

[12] 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.

[13] Deo, R., Hurvich, C., Soulier, P. and Wang Y. (2006), Conditions for the Propagation of Memory Parameter from Durations to Counts and Realized Volatility. Working Paper, Stern School of Business New York University.

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

[15] 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.

[16] 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.

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

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

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

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

[21] 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.

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

[23] Hsieh, M. and Hurvich, C.M. (2006), Long Memory Models for Durations of Bivariate Series, Working Paper.

[24] 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.

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

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

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

[28] Kozubowski, T., Meerschaert, M. and Podgorski, K. (2006), Fractional Laplace Motion, *Advances in Applied Probability* (in press).

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

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

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

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

[33] 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.

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

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

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

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

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

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

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

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

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

[43] Roll, R. (1984), A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market, *Journal of Finance* **39**, 1127–1139.

[44] Samorodnitsky, G. and Taqqu, M.S. (1994), *Stable Non-Gaussian Random Processes*, New York: Chapman and Hall.

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

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

[47] Sharpe, W. (1964), Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk, *Journal of Finance* **19**, 425–442.

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

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

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

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

XI Appendix

A Proof of Lemma 1

Proof: Because $N(\cdot)$ 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 $\tilde{V}(s) = \mathrm{var}(\sum_{k=1}^{s}\eta_k)$. Because $\{\eta_k\}$ is fractional Gaussian noise, $\sum_{k=1}^{s}\eta_k = B_H(s+1) - B_H(1)$. Using the definition of the fractional Brownian motion,

$$\mathrm{cov}[B_H(t_1), B_H(t_2)] = \frac{\sigma^2}{2}[|t_1|^{2H} + |t_2|^{2H} - |t_1 - t_2|^{2H}].$$

We obtain

$$\begin{aligned}\tilde{V}(s) = \mathrm{var}(\sum_{k=1}^{s}\eta_k) &= \mathrm{var}[B_H(s+1) - B_H(1)]\\ &= \mathrm{var}[B_H(s+1)] + \mathrm{var}[B_H(1)] - 2\mathrm{cov}[B_H(s+1), B_H(1)]\\ &= \sigma^2|s+1|^{2H} + \sigma^2 - \sigma^2\{|s+1|^{2H} + 1 - |s|^{2H}\}\\ &= \sigma^2|s|^{2H} = \sigma^2|s|^{2(d_\eta+1/2)} = \sigma^2|s|^{2d_\eta+1}.\end{aligned}$$

Therefore,

$$\mathrm{var}(\sum_{k=1}^{N(t)}\eta_k) = \sigma^2 E\Big\{[N(t)]^{2d_\eta+1}\Big\}. \tag{9}$$

We evaluate $E\Big\{[N(t)]^{2d_\eta+1}\Big\}$ in (9) as follows. Denote $Z(t) = \frac{N(t)-\lambda t}{t^{1/2+d_\tau}}$. As shown by Deo, Hurvich, Soulier and Wang (2006, 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 $d_\tau < 1/2$, 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 $\left[\frac{N(t)}{\lambda t}\right]^{2d_\eta+1} \xrightarrow{p} 1$.

Next, we will prove that

$$E\Big[\frac{N(t)}{\lambda t}\Big]^{2d_\eta+1} \to 1 \tag{10}$$

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

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

we obtain that, for $t \ge 1$

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

By Lemma 2 in Deo, Hurvich, Soulier and Wang (2006), $\limsup_t E[Z^4(t)] < \infty$. Therefore, $\limsup_t E\Big[\frac{N(t)}{\lambda t}\Big]^{2d_\eta+1+\delta} < \infty$ and we obtain (10).

From (9) and (10), we obtain

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

B Proof of Theorem 1

Proof: 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

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

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

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

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

$$BRT_{2,t} = \inf\{s > 0 : N_2(t) - N_2(t-s) > 0\}. \tag{11}$$

Clearly, $BRT_{2,t} = t - t_{2,N_2(t)}$ and thus $E\{N_1(t) - N_1(t_{2,N_2(t)})\} = \lambda_1 E[BRT_{2,t}]$. Because $N_2(t)$ is a stationary point process, $BRT_{2,t}$ has the same distribution as $BRT_{2,0}$ and $E[BRT_{2,t}] = E[BRT_{2,0}] < \infty$ does not depend on t (Daley and Vere-Jones (2002), page 58–59 for a detailed discussion, and our Lemma 3). Thus

$$E\{N_1(t) - N_1(t_{2,N_2(t)})\} = \lambda_1 E[BRT_{2,t}] = \lambda_1 E[BRT_{2,0}] = \tilde{C}_1, \tag{12}$$

a finite constant, independent of t. Similarly

$$E\{N_2(t) - N_2(t_{1,N_1(t)})\} = \lambda_2 E[BRT_{1,t}] = \lambda_2 E[BRT_{1,0}] = \tilde{C}_2 \tag{13}$$

is also a finite constant, independent of t as well. Intuitively, both (12) and (13) make sense. For example, (12) says that the expected number of transactions of Asset 1 after the most recent transaction of Asset 2 up to time t increases as the expected duration of Asset 1 decreases (λ_1 increases) and/or as the expected backward recurrence time of Asset 2, $E[BRT_{2,0}]$, increases.

Using (9),

$$\text{var}\Big[\sum_{k=1}^{N_2(t_{1,N_1(t)})}(\theta e_{2,k} + g_{21}\eta_{2,k})\Big] = \theta^2 \underbrace{\sigma_{2,e}^2 E\{N_2(t_{1,N_1(t)})\}}_{T_1} + g_{21}^2\sigma_{2,\eta}^2 \underbrace{E\Big\{[N_2(t_{1,N_1(t)})]^{2d_\eta+1}\Big\}}_{T_2}$$

By (13), the first term equals

$$\theta^2 T_1 = \theta^2\sigma_{2,e}^2(E\{N_2(t)\} - \tilde{C}_2) = \theta^2\sigma_{2,e}^2(\lambda_2 t - \tilde{C}_2) \sim (\theta^2\sigma_{2,e}^2\lambda_2)t,$$

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

$$g_{21}^2 \sigma_{2,\eta}^2 T_2 \leq g_{21}^2 \sigma_{2,\eta}^2 \Big\{ E[N_2(t_{1,N_1(t)})] \Big\}^{2d_\eta+1} = g_{21}^2 \sigma_{2,\eta}^2 (\lambda_2 t - \tilde{C}_2)^{2d_\eta+1} = o(t).$$

Therefore,

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

as $t \to \infty$. Overall,

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

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

Similarly,

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

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

Next, we consider the standard cointegration case, $d_\eta = -1$. The proof is identical to that for the fractional cointegration case, except that here we have $\mathrm{var}(\sum_{k=1}^{N_i(t)} \eta_{i,k}) = 2\sigma_{i,\xi}^2$, $i = 1, 2$, which does not increase with t. $\square$

C Proof of Theorem 2

Proof: 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 (4) 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,

$$\begin{aligned}
\text{var}(T_1) &= E\Big\{\Big[\sum_{k=N_1((j-1)\Delta t)+1}^{N_1(j\Delta t)} e_{1,k}\Big]^2\Big\} = E\Big[E\Big\{\Big[\sum_{k=N_1((j-1)\Delta t)+1}^{N_1(j\Delta t)} e_{1,k}\Big]^2\Big|N_1(\cdot)\Big\}\Big] \\
&= \sigma_{1,e}^2 E\{N_1(j\Delta t) - N_1((j-1)\Delta t)\} \\
&= \sigma_{1,e}^2 E(\Delta N_{1,j}) = \sigma_{1,e}^2 \lambda_1 \Delta t \quad . \qquad (14)
\end{aligned}$$

By the proof of (9), $\text{var}(T_2) = \sigma_{2,\eta}^2 E[\{\Delta N_{1,j}\}^{2d_\eta+1}]$. Since the function x^p is concave when $x > 0$ and $0 < p < 1$, by Jensen's inequality for $d_\eta \in (-0.5, 0)$,

$$\text{var}(T_2) = \sigma_{2,\eta}^2 E[\{\Delta N_{1,j}\}^{2d_\eta+1}] \le \sigma_{2,\eta}^2 \{E[\Delta N_{1,j}]\}^{2d_\eta+1} = \sigma_{2,\eta}^2 \{\lambda_1 \Delta t\}^{2d_\eta+1} = o(\Delta t), \qquad (15)$$

as $\Delta t \to \infty$.

Next, by the proof of (9) and equations (12) and (13),

$$\begin{aligned}
\text{var}(T_3) &= \theta^2 \sigma_{2,e}^2 E\{N_2(t_{1,N_1(j\Delta t)}) - N_2(t_{1,N_1((j-1)\Delta t)})\} \\
&= \theta^2 \sigma_{2,e}^2 E[N_2(j\Delta t) - N_2((j-1)\Delta t)] \\
&= \theta^2 \sigma_{2,e}^2 E[\Delta N_{2,j}] = \theta^2 \sigma_{2,e}^2 \lambda_2 \Delta t \qquad (16)
\end{aligned}$$

and

$$\begin{aligned}
\text{var}(T_4) &= g_{21}^2 \sigma_{2,\eta}^2 E\Big[\{N_2(t_{1,N_1(j\Delta t)}) - N_2(t_{1,N_1((j-1)\Delta t)})\}^{2d_\eta+1}\Big] \\
&\le g_{21}^2 \sigma_{2,\eta}^2 \Big[E\{N_2(t_{1,N_1(j\Delta t)}) - N_2(t_{1,N_1((j-1)\Delta t)})\}\Big]^{2d_\eta+1} \\
&= g_{21}^2 \sigma_{2,\eta}^2 \Big\{E[\Delta N_{2,j}]\Big\}^{2d_\eta+1} = g_{21}^2 \sigma_{2,\eta}^2 (\lambda_2 \Delta t) = o(\Delta t). \qquad (17)
\end{aligned}$$

As for the covariance terms, by the Cauchy-Schwartz inequality and Equations (14) to (17),

$$\begin{aligned}
|\mathrm{cov}(T_1,T_2)| &\le \sqrt{\mathrm{var}(T_1)\mathrm{var}(T_2)} \le \sqrt{\sigma_{1,e}^2\sigma_{2,\eta}^2(\lambda_1\Delta t)^{2d_\eta+2}} = o(\Delta t) && (18)\\
|\mathrm{cov}(T_1,T_4)| &\le \sqrt{\mathrm{var}(T_1)\mathrm{var}(T_4)} \le \sqrt{g_{21}^2\sigma_{1,e}^2\sigma_{2,\eta}^2(\lambda_1\Delta t)(\lambda_2\Delta t)^{2d_\eta+1}} = o(\Delta t) && (19)\\
|\mathrm{cov}(T_2,T_3)| &\le \sqrt{\mathrm{var}(T_2)\mathrm{var}(T_3)} \le \sqrt{\theta^2\sigma_{2,e}^2\sigma_{2,\eta}^2(\lambda_2\Delta t)(\lambda_1\Delta t)^{2d_\eta+1}} = o(\Delta t) && (20)\\
|\mathrm{cov}(T_2,T_4)| &\le \sqrt{\mathrm{var}(T_2)\mathrm{var}(T_4)} \le \sqrt{g_{21}^2\sigma_2^4(\lambda_1\Delta t)^{2d_\eta+1}(\lambda_2\Delta t)^{2d_\eta+1}} = o(\Delta t) && (21)\\
|\mathrm{cov}(T_3,T_4)| &\le \sqrt{\mathrm{var}(T_3)\mathrm{var}(T_4)} \le \sqrt{\theta^2 g_{21}^2\sigma_{2,e}^2\sigma_{2,\eta}^2(\lambda_2\Delta t)^{2d_\eta+2}} = o(\Delta t) && (22)
\end{aligned}$$

since $d_\eta < 0$. Also,

$$\mathrm{cov}(T_1,T_3) = 0 \tag{23}$$

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

Overall, by (14) to (23), 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

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

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

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

The result follows for any fixed lag k by induction.

Next, we consider the standard cointegration case, $d_\eta = -1$. The proof is identical to that for the fractional cointegration case, except that here we have $\text{var}(\sum_{k=N_i((j-1)\Delta t)+1}^{N_i(j\Delta t)} \eta_{i,k}) = 2\sigma_{i,\xi}^2$, $i = 1, 2$, and other similar terms which do not increase with Δt. $\square$

D Proof of Theorem 3

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

$$\begin{aligned}
&\log P_{1,t} - \theta \log P_{2,t} \\
= &\sum_{k=N_1(t_{2,N_2(t)})+1}^{N_1(t)} e_{1,k} - \theta \sum_{k=N_2(t_{1,N_1(t)})+1}^{N_2(t)} e_{2,k} \\
&+ \sum_{k=1}^{N_1(t)} \eta_{1,k} - \theta g_{12} \sum_{k=1}^{N_1(t_{2,N_2(t)})} \eta_{1,k} - \theta \sum_{k=1}^{N_2(t)} \eta_{2,k} + g_{21} \sum_{k=1}^{N_2(t_{1,N_1(t)})} \eta_{2,k} \\
= &\underbrace{\sum_{k=N_1(t_{2,N_2(t)})+1}^{N_1(t)} e_{1,k}}_{T_1} - \theta \underbrace{\sum_{k=N_2(t_{1,N_1(t)})+1}^{N_2(t)} e_{2,k}}_{T_2} + (1 - \theta g_{12}) \underbrace{\sum_{k=1}^{N_1(t)} \eta_{1,k}}_{T_3} \\
&+ \theta g_{12} \underbrace{\sum_{k=N_1(t_{2,N_2(t)})+1}^{N_1(t)} \eta_{1,k}}_{T_4} - (\theta - g_{21}) \underbrace{\sum_{k=1}^{N_2(t)} \eta_{2,k}}_{T_5} - g_{21} \underbrace{\sum_{k=N_2(t_{1,N_1(t)})+1}^{N_2(t)} \eta_{2,k}}_{T_6}.
\end{aligned} \tag{24}$$

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

and $N_2(t)$, we obtain

$$\begin{aligned}\text{var}\Big[\log P_{1,t} - \theta \log P_{2,t}\Big] &= \text{var}(T_1) + \theta^2 \text{var}(T_2) + (1-\theta g_{12})^2 \text{var}(T_3) + \theta^2 g_{12}^2 \text{var}(T_4) \\ &\quad +2\theta g_{12}(1-\theta g_{12})\text{cov}(T_3, T_4) + (\theta - g_{21})^2 \text{var}(T_5) + g_{21}^2 \text{var}(T_6) \\ &\quad +2g_{21}(\theta - g_{21})\text{cov}(T_5, T_6). \end{aligned} \tag{25}$$

First, by Lemma 1

$$\begin{aligned}\text{var}(T_3) &\sim (\sigma_{1,\eta}^2 \lambda_1^{2d_\eta+1}) t^{2d_\eta+1} \\ \text{var}(T_5) &\sim (\sigma_{2,\eta}^2 \lambda_2^{2d_\eta+1}) t^{2d_\eta+1} \end{aligned} \quad . \tag{26}$$

Using (12) and the proof of (9), we obtain

$$\begin{aligned}\text{var}(T_4) &= \text{var}[B_H(N_1(t)+1) - B_H(N_1(t_{2,N_2(t)})+1)] \\ &= \sigma_{1,\eta}^2 E\Big[\{N_1(t) - N_1(t_{2,N_2(t)})\}^{2d_\eta+1}\Big] \\ &\leq \sigma_{1,\eta}^2 \Big[E\{N_1(t) - N_1(t_{2,N_2(t)})\}\Big]^{2d_\eta+1} = \sigma_{1,\eta}^2 \tilde{C}_1^{2d_\eta+1} \end{aligned} \tag{27}$$

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

$$\text{var}(T_6) \leq \sigma_{2,\eta}^2 \tilde{C}_2^{2d_\eta+1}. \tag{28}$$

Also, by (12) and (13)

$$\text{var}(T_1) = \text{var}(e_{1,k}) E\{N_1(t) - N_1(t_{2,N_2(t)})\} = \sigma_{1,e}^2 \tilde{C}_1 \tag{29}$$

$$\text{var}(T_2) = \text{var}(e_{2,k}) E\{N_2(t) - N_2(t_{1,N_1(t)})\} = \sigma_{2,e}^2 \tilde{C}_2. \tag{30}$$

Next, we consider the covariance terms in (25) using Cauchy-Schwartz inequality. By (26) and (12)

$$|\text{cov}(T_3, T_4)| \leq \sqrt{\text{var}(T_3)\text{var}(T_4)} \leq \sqrt{\sigma_{1,\eta}^2 \tilde{C}_1^{2d_\eta+1} \text{var}(T_3)} \sim (\sigma_{1,\eta}^2 \tilde{C}_1^{d_\eta+\frac{1}{2}} \lambda_1^{d_\eta+\frac{1}{2}}) t^{d_\eta+\frac{1}{2}} = o(t^{2d_\eta+1}) \tag{31}$$

and similarly by (26) and (13)

$$|\text{cov}(T_5, T_6)| \leq \sqrt{\text{var}(T_5)\text{var}(T_6)} \leq \sqrt{\sigma_{2,\eta}^2 \tilde{C}_2^{2d_\eta+1} \text{var}(T_5)} \sim (\sigma_{2,\eta}^2 \tilde{C}_2^{d_\eta+\frac{1}{2}} \lambda_2^{d_\eta+\frac{1}{2}}) t^{d_\eta+\frac{1}{2}} = o(t^{2d_\eta+1}). \tag{32}$$

Overall, using (26) to (32) for (25), we obtain

$$\mathrm{var}\Big(\log P_{1,t} - \theta \log P_{2,t}\Big) \sim Ct^{2d_\eta+1} \tag{33}$$

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

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

E Proof of Theorem 4

We will need the following lemmas.

Lemma 3 *If the durations $\{\tau_k\}$ are generated by a Long Memory Stochastic Duration (LMSD) model with memory parameter $d_\tau \in (0, 1/2)$ and all moments of the durations $\{\tau_k\}$ are finite, then all moments of the backward recurrence time (BRT_t), as defined in (11), are also finite.*

Proof of Lemma 3: First, by exercise 3.4.1 on page 59 of Daley and Vere-Jones (2002),

$$BRT_t \overset{d}{\equiv} u_1 \tag{34}$$

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 $m > 0$. $\square$

Lemma 4 *For durations $\{\tau_k\}$ satisfying the assumptions in Lemma 3, $E[N(s)^m] \leq K_m(s^m + 1)$ for all $s > 0$, where $K_m < \infty$, $m = 1, 2, \cdots$.*

Proof of Lemma 4: By Proposition 1 in Deo, Hurvich, Soulier and Wang (2006), 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 $s > 0$,

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

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

We now present the proof of Theorem 4. 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 $\mathrm{cov}(S_t, S_{t+j})$.

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 $E[S_{1,t}S_{1,t+j}|N_1(\cdot), N_2(\cdot)] = \sigma^2_{1,e}[N_1(t) - N_1(t_{2,N_2(t)})]$. Therefore, by the Cauchy-Schwartz inequality,

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

By Lemma 4 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 3.

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

$$\begin{aligned} &P\Big[N_2(t+j) - N_2(t) = 0\Big] = P\Big[N_2(j) \leq 0\Big] \\ \leq\ & P\Big[|Z_2(j)| \geq \lambda_2 j^{1/2-d_\tau}\Big] \leq \frac{E|Z_2(j)|^m}{\lambda_2^m j^{m(1/2-d_\tau)}} = O(j^{m(d_\tau - 1/2)}), \end{aligned} \tag{35}$$

where $Z_2(j) = \frac{N_2(j)-\lambda_2 j}{j^{1/2+d_\tau}}$. This is true since it follows from the proof of Proposition 1 in Deo, Hurvich, Soulier and Wang (2006) 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 (35) 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,

$$\begin{aligned} \mathrm{cov}(S_{3,t}, S_{3,t+j}) &= \mathrm{cov}\Big(\xi_{1,N_1(t)} - \xi_0, \xi_{1,N_1(t+j)} - \xi_0\Big) = \mathrm{cov}\Big(\xi_{1,N_1(t)}, \xi_{1,N_1(t+j)}\Big) \\ &= \mathrm{cov}\Big(\xi_{1,N_1(t)}, \xi_{1,N_1(t+j)}\Big) = \sigma_{1,\xi}^2 P\Big[N_1(t+j) - N_1(t) = 0\Big] \end{aligned}$$

which has nearly-exponential decay, as shown above for $N_2(\cdot)$ in (35).

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

$$
\begin{aligned}
0 \le \text{cov}(S_{3,t}, S_{4,t+j}) &= \text{cov}\Big(\xi_{1,N_1(t)} - \xi_0,\ \xi_{1,N_1(t_{2,N_2(t+j)})} - \xi_0\Big) = \text{cov}\Big(\xi_{1,N_1(t)}, \xi_{1,N_1(t_{2,N_2(t+j)})}\Big) \\
&= \sigma_{1,\xi}^2 P\Big[N_1(t_{2,N_2(t+j)}) - N_1(t) = 0\Big] \\
&\le \sigma_{1,\xi}^2 P\Big[N_1(t_{2,N_2(t+j)}) - N_1(t) \le 0\Big] \\
&\le \sigma_{1,\xi}^2 P\Big\{[N_1(t+j) - N_1(t)] - \underbrace{[N_1(t+j) - N_1(t_{2,N_2(t+j)})]}_{X_{t,j}} \le 0\Big\} \\
&= \sigma_{1,\xi}^2 P\Big\{\frac{[N_1(t+j) - N_1(t)] - \lambda_1 j}{j^{1/2+d_\tau}} - \frac{X_{t,j}}{j^{1/2+d_\tau}} \le \frac{-\lambda_1 j}{j^{1/2+d_\tau}}\Big\} \\
&\le \sigma_{1,\xi}^2 P\Big\{\Big|\frac{[N_1(t+j) - N_1(t)] - \lambda_1 j}{j^{1/2+d_\tau}} - \frac{X_{t,j}}{j^{1/2+d_\tau}}\Big| \ge \lambda_1 j^{1/2-d_\tau}\Big\} \\
&\le \sigma_{1,\xi}^2 E\Big|\frac{[N_1(t+j) - N_1(t)] - \lambda_1 j}{j^{1/2+d_\tau}} - \frac{X_{t,j}}{j^{1/2+d_\tau}}\Big|^m \lambda_1^{-m} j^{m(d_\tau - 1/2)} \qquad (36)
\end{aligned}
$$

for any positive integer m. Thus, $\text{cov}(S_{3,t}, S_{4,t+j})$ has nearly-exponential decay, provided that $E\Big|\frac{[N_1(t+j)-N_1(t)]-\lambda_1 j}{j^{1/2+d_\tau}} - \frac{X_{t,j}}{j^{1/2+d_\tau}}\Big|^m$ is bounded uniformly in t and j. By Minkowski's inequality, it is sufficient to show that both $E\Big|\frac{[N_1(t+j)-N_1(t)]-\lambda_1 j}{j^{1/2+d_\tau}}\Big|^m$ and $E|X_{t,j}|^m$ are uniformly bounded.

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

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

which is bounded uniformly in j, by the proof of Proposition 1 in Deo, Hurvich, Soulier and Wang (2006).

By Lemma 4, we obtain

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

which is uniformly bounded in t and j by Lemma 3. 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

$$
\begin{aligned}
0 \le \text{cov}(S_{4,t}, S_{3,t+j}) &= \text{cov}\Big(\xi_{1,N_1(t_{2,N_2(t)})} - \xi_0, \xi_{1,N_1(t+j)} - \xi_0\Big) \\
&= \sigma_{1,\xi}^2 P\Big[N_1(t+j) - N_1(t_{2,N_2(t)}) = 0\Big] \le \sigma_{1,\xi}^2 P\Big[N_1(t+j) - N_1(t) = 0\Big] .
\end{aligned}
$$

Thus, $\mathrm{cov}(S_{4,t}, S_{3,t+j})$ has nearly-exponential decay, by the proof of (35).

Finally, since

$$\begin{aligned} 0 \le \mathrm{cov}(S_{4,t}, S_{4,t+j}) &= \mathrm{cov}\Big(\xi_{1,N_1(t_{2,N_2(t)})} - \xi_0, \xi_{1,N_1(t_{2,N_2(t+j)})} - \xi_0\Big) \\ &= \sigma^2_{1,\xi} P\Big[N_1(t_{2,N_2(t+j)}) - N_1(t_{2,N_2(t)}) = 0\Big] \\ &\le \sigma^2_{1,\xi} P\Big[N_1(t_{2,N_2(t+j)}) - N_1(t_{2,N_2(t)}) \le 0\Big] \\ &\le \sigma^2_{1,\xi} P\Big[N_1(t_{2,N_2(t+j)}) - N_1(t) \le 0\Big] \qquad (37) \end{aligned}$$

which as we have shown in (36) has nearly-exponential decay. The last inequality in (37) holds since $N_1(t_{2,N_2(t)}) \le N_1(t)$.

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. $\square$

F Proof of Theorem 5

Proof: We will treat the fractional cointegration case and standard cointegration case separately.

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

The log prices given by (5) 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

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

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

$$
\hat{\theta} = \frac{\sum_{j=1}^n A_j B_j}{\sum_{j=1}^n B_j^2} = \frac{\sum_{j=1}^n (\theta B_j + T_j) B_j}{\sum_{j=1}^n B_j^2} = \theta + \frac{\sum_{j=1}^n T_j B_j}{\sum_{j=1}^n B_j^2} \quad . \tag{38}
$$

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,

$$
\frac{1}{n^r} \sum_{j=1}^n T_{i,j} B_{k,j} \leq \sqrt{(\frac{1}{n^{2r-2}} \sum_{j=1}^n T_{i,j}^2)(\frac{1}{n^2} \sum_{j=1}^n B_{k,j}^2)} \quad . \tag{39}
$$

It is therefore sufficient to show that the righthand side of (39) converges in probability to zero, for all $i = 1, \ldots, 6$ and $k = 1, \ldots, 5$.

By (9), (12), (13), Lemma 1 and Jensen's inequality $E(X^{2d_\eta+1}) \leq (EX)^{2d_\eta+1}$ for $x \geq 0$, $d_\eta \in (-1/2, 0)$, we obtain that, for any $\epsilon > 0$,

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

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

Therefore,

$$\frac{1}{n^{2r-2}} \sum_{j=1}^{n} T_{i,j}^2 \xrightarrow{p} 0 \tag{40}$$

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

Next, since

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

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

$$\frac{1}{n^2} \sum_{j=1}^{n} B_{1,j}^2 = O_p(1). \tag{41}$$

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 (40) that

$$\frac{1}{n^2}\sum_{j=1}^{n} B_{i,j}^2 \xrightarrow{p} 0, \tag{42}$$

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

Applying (40), (41), (42) in (39), we obtain

$$\frac{1}{n^r}\sum_{j=1}^{n} T_j B_j \xrightarrow{p} 0, \tag{43}$$

where $r = 2 + \max(d_\eta, d_\eta) + \delta$ 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

$$\frac{1}{n^{3/2}}\sum_{j=1}^{n} B_{1,j} \xrightarrow{d} \sqrt{\frac{1}{3}\sigma_{2,e}^2\lambda_2\Delta t + \frac{1}{3\theta^2}\sigma_{1,e}^2\lambda_1\Delta t}\; Z \qquad , \tag{44}$$

where Z is standard normal and

$$\frac{1}{n^{3/2}}\sum_{j=1}^{n} B_{i,j} \xrightarrow{p} 0 \tag{45}$$

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

$$\frac{1}{\frac{1}{n^2}\sum_{j=1}^{n} B_j^2} = O_p(1). \tag{46}$$

To show (44), 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,

$$\begin{aligned}G_1 &= \frac{1}{n^{3/2}}\Big(n\sum_{k=1}^{N_2(\Delta t)}e_{2,k}+(n-1)\sum_{k=N_2(\Delta t)+1}^{N_2(2\Delta t)}e_{2,k}+\ldots+\sum_{k=N_2((n-1)\Delta t)+1}^{N_2(n\Delta t)}e_{2,k}\Big)\\ &\overset{d}{\equiv} \frac{\sigma_{2,e}}{n^{3/2}}\Big(n\sqrt{\Delta N_{2,1}}Z_1+(n-1)\sqrt{\Delta N_{2,2}}Z_2+\ldots+\sqrt{\Delta N_{2,n}}Z_n\Big)\\ &\overset{d}{\equiv} \frac{\sigma_{2,e}}{n^{3/2}}\Big(\sqrt{n^2\Delta N_{2,1}+(n-1)^2\Delta N_{2,2}\ldots+\Delta N_{2,n}}\;Z\Big)\\ &= \sigma_{2,e}\sqrt{\frac{1}{n^3}\underbrace{\sum_{k=1}^{n}(n-k+1)^2\Delta N_{2,k}}_{D}}\;Z\end{aligned}\tag{47}$$

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 (47). Applying the summation by parts formula for two sequences $\{f_k\}$ and $\{g_k\}$,

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

we obtain

$$\begin{aligned}D &= \sum_{k=1}^{n}(n-k+1)^2\Delta N_{2,k}=\sum_{k=1}^{n}(n-k+1)^2\Big[N_2(k\Delta t)-N_2((k-1)\Delta t)\Big]\\ &= \sum_{k=0}^{n-1}\underbrace{(n-k)^2}_{f_k}\Big[\underbrace{N_2((k+1)\Delta t)}_{g_{k+1}}-\underbrace{N_2(k\Delta t)}_{g_k}\Big]\\ &= (f_ng_n-f_0g_0)-\sum_{k=0}^{n-1}N_2((k+1)\Delta t)\Big[(2n-2k-1)(-1)\Big]\\ &= \sum_{k=0}^{n-1}(2n-2k-1)N_2((k+1)\Delta t)\qquad(\text{since } f_n=0 \text{ and } g_0=0)\\ &= \sum_{k=1}^{n}(2n-2k+1)N_2(k\Delta t)\end{aligned}$$

thus

$$
\begin{aligned}
E(\frac{1}{n^3}D) &= \frac{1}{n^3}\sum_{k=1}^{n}(2n-2k+1)E[N_2(k\Delta t)] = \frac{\lambda_2\Delta t}{n^3}\sum_{k=1}^{n}(2n-2k+1)k \\
&= \frac{\lambda_2\Delta t}{n^3}\Big[2n\frac{n(n+1)}{2} - 2\frac{n(n+1)(2n+1)}{6} + \frac{n(n+1)}{2}\Big] \to \frac{1}{3}\lambda_2\Delta t, \qquad (48)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{var}(\frac{1}{n^3}D) &\leq \frac{1}{n^6}\sum_{j=1}^{n}\sum_{s=1}^{n}(2n-2j+1)(2n-2s+1)\Big|\mathrm{cov}(N_2(j\Delta t), N_2(s\Delta t))\Big| \\
&\leq \frac{1}{n^6}\sum_{j=1}^{n}\sum_{s=1}^{n}(2n-2j+1)(2n-2s+1)\sqrt{\mathrm{var}(N_2(j\Delta t))\mathrm{var}(N_2(s\Delta t))} \\
&\leq \frac{4n^2}{n^6}\Big(\sum_{j=1}^{n}\sqrt{\mathrm{var}(N_2(j\Delta t))}\Big)\Big(\sum_{s=1}^{n}\sqrt{\mathrm{var}(N_2(s\Delta t))}\Big) = O(n^{2d_\tau - 1}) \to 0 \qquad (49)
\end{aligned}
$$

as $n \to \infty$ since $d_\tau \in (0, \frac{1}{2})$ and by Theorem 1 of Deo, Hurvich, Soulier and Wang (2006),

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

By (48) and (49), $\left(\frac{1}{n^3}D - \frac{1}{3}\lambda_2\Delta t\right)$ converges in mean-square to zero, which implies that

$$
\frac{1}{n^3}D = \frac{1}{n^3}\sum_{k=1}^{n}(n-k+1)^2\Delta N_{2,k} \xrightarrow{p} \frac{1}{3}\lambda_2\Delta t \quad . \qquad (50)
$$

Using (50) in (47), by Slutsky's theorem

$$
G_1 \xrightarrow{d} \sigma_{2,e}\sqrt{\frac{1}{3}\lambda_2\Delta t}\; Z_1 \qquad (51)
$$

and similarly

$$
G_2 \xrightarrow{d} \frac{\sigma_{1,e}}{\theta}\sqrt{\frac{1}{3}\lambda_1\Delta t}\; Z_2 \qquad (52)
$$

where Z_1 and Z_2 are independent standard normals.

Overall, by (51), (52) and the independence between G_1 and G_2, (44) is obtained.

To show (45), since for any $\epsilon > 0$, and $i = 2, \ldots, 5$, by Chebyshev's inequality,

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

it is enough to show that,

$$\frac{\mathrm{var}(\sum_{j=1}^{n} B_{i,j})}{n^3} \to 0, \qquad i = 2, \ldots, 5. \tag{53}$$

Since

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

we obtain

$$\begin{aligned}
\mathrm{var}(\sum_{j=1}^{n} B_{2,j}) \le \sum_{j=1}^{n}\sum_{s=1}^{n} |\mathrm{cov}(B_{2,j}, B_{2,s})| &\le \sum_{j=1}^{n}\sum_{s=1}^{n} \sigma_{2,\eta}^2 (\lambda_2 n\Delta t)^{2d_\eta+1} = O(n^{2d_\eta+3}) \qquad (54) \\
\mathrm{var}(\sum_{j=1}^{n} B_{3,j}) \le \sum_{j=1}^{n}\sum_{s=1}^{n} |\mathrm{cov}(B_{3,j}, B_{3,s})| &= O(n^{2d_\eta+3}) \qquad (\text{similar as above}) \\
\mathrm{var}(\sum_{j=1}^{n} B_{4,j}) \le \sum_{j=1}^{n}\sum_{s=1}^{n} |\mathrm{cov}(B_{4,j}, B_{4,s})| &= \sum_{j=1}^{n}\sum_{s=1}^{n} \sigma_{1,e}^2 \tilde{C}_1 = O(n^2) \\
\mathrm{var}(\sum_{j=1}^{n} B_{5,j}) \le \sum_{j=1}^{n}\sum_{s=1}^{n} |\mathrm{cov}(B_{5,j}, B_{5,s})| &= \sum_{j=1}^{n}\sum_{s=1}^{n} \sigma_{1,e}^2 (\tilde{C}_1)^{2d_\eta+1} = O(n^2) \qquad .
\end{aligned}$$

This implies (53) and (45), since $d_\eta < 0$.

Overall, since (44), (45) are proved, we obtain (46). Thus, by (38), (43) and (46),

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

<u>Case 2:</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

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

and

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

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

$$\begin{aligned} & \mathrm{var}(\sum_{j=1}^n B^*_{1,j}T^*_{1,j}) \\ = & E\Big[\sum_{j=1}^n\sum_{s=1}^n B^*_{1,j}B^*_{1,s}T^*_{1,j}T^*_{1,s}\Big] - \Big\{E(\sum_{j=1}^n B^*_{1,j}T^*_{1,j})\Big\}^2 \\ = & \sum_{j=1}^n E({B^*_{1,j}}^2 {T^*_{1,j}}^2) + 2\sum_{j=1}^n\sum_{s=j+1}^n E(B^*_{1,j}B^*_{1,s}T^*_{1,j}T^*_{1,s}) \\ = & \sum_{j=1}^n E[E({B^*_{1,j}}^2{T^*_{1,j}}^2|N_1(\cdot),N_2(\cdot))] + 2\sum_{j=1}^n\sum_{s=j+1}^n E[E(B^*_{1,j}B^*_{1,s}T^*_{1,j}T^*_{1,s}|N_1(\cdot),N_2(\cdot))] \\ = & \sum_{j=1}^n E[E({B^*_{1,j}}^2|N_1(\cdot),N_2(\cdot))\cdot E({T^*_{1,j}}^2|N_1(\cdot),N_2(\cdot))] \\ & + 2\sum_{j=1}^n\sum_{s=j+1}^n E[E(B^*_{1,j}B^*_{1,s}|N_1(\cdot),N_2(\cdot))\cdot E(T^*_{1,j}T^*_{1,s}|N_1(\cdot),N_2(\cdot))] \\ = & \underbrace{\sigma^2_{1,e}\sigma^2_{2,e}\sum_{j=1}^n E\{N_2(t_{1,N_1(j\Delta t)})\cdot[N_1(j\Delta t) - N_1(t_{2,N_2(j\Delta t)})]\}}_{O(n^2),\text{ as shown in below}} \qquad (55) \\ & + 2\sigma^2_{1,e}\sigma^2_{2,e}\sum_{j=1}^n\sum_{s=j+1}^n E\Big\{N_2(t_{1,N_1(j\Delta t)})\cdot[N_1(j\Delta t) - N_1(t_{2,N_2(j\Delta t)})]\cdot I\{N_2(s\Delta t) - N_2(j\Delta t) = 0\}\Big\} \quad . \end{aligned}$$

By the Cauchy-Schwartz inequality,

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

because by Lemma 3 and Lemma 4, $E\{[N_1(j\Delta t) - N_1(t_{2,N_2(j\Delta t)})]^2\}$ is bounded uniformly in j and by Theorem 1 in Deo, Hurvich, Soulier and Wang (2006),

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

hence

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

as indicated in (55).

Similarly, since

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

and by Lemma 3 and Lemma 4, $E\{[N_1(j\Delta t) - N_1(t_{2,N_2(j\Delta t)})]^4\}$ is bounded uniformly in j, while by (35), $P[N_2(s\Delta t) - N_2(j\Delta t) = 0] \leq K(s-j)^{m(d_\tau-1/2)}$ for all $m \geq 1$. We obtain that,

$$\text{var}(\sum_{j=1}^{n} B_{1,j}^* T_{1,j}^*) \leq O(n^2) + Kn \sum_{j=1}^{n}\sum_{s=j+1}^{n} (s-j)^{m(d_\tau-1/2)/4}. \tag{56}$$

Consider $\sum_{j=1}^{n}\sum_{s=j+1}^{n}(s-j)^{m(d_\tau-1/2)/4}$. For any fixed integer $1 \leq j \leq n$, we choose $m > \frac{8}{1-2d_\tau}$ so that $\sum_{s=j+1}^{n}(s-j)^{m(d_\tau-1/2)/4}$ is summable in s, hence $\sum_{j=1}^{n}\sum_{s=j+1}^{n}(s-j)^{m(d_\tau-1/2)/4} = O(n)$. Therefore, $\text{var}(\sum_{j=1}^{n} B_{1,j}^* T_{1,j}^*) = O(n^2)$ and by Chebyshev's inequality, we obtain that for any $\delta > 0$,

$$\frac{1}{n^{1+\delta}} \sum_{j=1}^{n} B_{1,j}^* T_{1,j}^* \xrightarrow{p} 0 \quad .$$

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

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

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

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

which is similar to (55). Following along similar lines as for (55), 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 $\sum_{j=1}^n B^*_{1,j}T^*_{3,j} = \sum_{j=1}^n B^*_{1,j}\xi_{1,N_1(j\Delta t)}$, we consider $\mathrm{var}(\sum_{j=1}^n B^*_{1,j}\xi_{1,N_1(j\Delta t)})$ and

obtain

$$
\begin{aligned}
& \operatorname{var}(\sum_{j=1}^{n} B^*_{1,j}\xi_{1,N_1(j\Delta t)}) \\
= \;& \sigma^2_{2,e}\sigma^2_{1,\xi}\sum_{j=1}^{n} E[N_2(t_{1,N_1(j\Delta t)})] + 2\sigma^2_{2,e}\sigma^2_{1,\xi}\sum_{j=1}^{n}\sum_{s=j+1}^{n} E\Big[N_2(t_{1,N_1(j\Delta t)})\cdot I\{N_1(s\Delta t)-N_1(j\Delta t)=0\}\Big] \\
\leq \;& \sigma^2_{2,e}\sigma^2_{1,\xi}\sum_{j=1}^{n} E[N_2(j\Delta t)] + 2\sigma^2_{2,e}\sigma^2_{1,\xi}\sum_{j=1}^{n}\sum_{s=j+1}^{n} E\Big[N_2(j\Delta t)\cdot I\{N_1(s\Delta t)-N_1(j\Delta t)=0\}\Big] \\
\leq \;& \sigma^2_{2,e}\sigma^2_{1,\xi}\lambda_2\Delta t\frac{n(n+1)}{2} + 2\sigma^2_{2,e}\sigma^2_{1,\xi}\sum_{j=1}^{n}\sum_{s=j+1}^{n}\sqrt{E\{[N_2(j\Delta t)]^2\}\cdot P[N_1(s\Delta t)-N_1(j\Delta t)=0]} \\
\leq \;& \sigma^2_{2,e}\sigma^2_{1,\xi}\lambda_2\Delta t\frac{n(n+1)}{2} + 2\sigma^2_{2,e}\sigma^2_{1,\xi}\underbrace{\sqrt{E\{[N_2(n\Delta t)]^2\}}}_{O(n)}\cdot\sum_{j=1}^{n}\sum_{s=j+1}^{n}\underbrace{\sqrt{P[N_1(s\Delta t)-N_1(j\Delta t)=0]}}_{\leq K(s-j)^{m(d_\tau-1/2)}} \quad .
\end{aligned}
$$

Since $E\{[N_2(j\Delta t)]^2\} = O(j^2)$ and $P[N_1(s\Delta t)-N_1(j\Delta t)=0] \leq K(s-j)^{m(d_\tau-1/2)}$ for all $m \geq 1$, we can choose m large enough so that $\sum_{j=1}^{n}\sum_{s=j+1}^{n}\sqrt{E\{[N_2(j\Delta t)]^2\}\cdot P[N_1(s\Delta t)-N_1(j\Delta t)=0]} = O(n^2)$, following similar lines as for the double summation in the second term on the righthand side of (56). Therefore, $\operatorname{var}(\sum_{j=1}^{n} B^*_{1,j}\xi_{1,N_1(j\Delta t)}) = O(n^2)$, 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

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

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 $d_\eta = -1$

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

for any $\delta > 0$.

Furthermore, the proof for (46) in the standard cointegration case is identical to that for the fractional cointegration case, except that here we have $\text{var}(\sum_{k=1}^{N_i(t)} \eta_{i,k}) = 2\sigma_{i,\xi}^2$, $i = 1, 2$, which does not increase with t. This, together with (57), gives that

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

□

G Proof of Lemma 2

By the serial and mutual independence of $\{e_{1,k}\}$ and $\{e_{2,k}\}$, we have

$$\text{cov}(r_{1,j}, r_{1,j+1}) = \text{cov}(r_{2,j}, r_{2,j+1}) = 0.$$

On the other hand,

$$\begin{aligned}
0 < \text{cov}(r_{1,j}, r_{2,j+1}) &= \theta_{12}\sigma_{1,e}^2 E\Big\{[N_1(j\Delta t) - N_1(t_{2,N_2(j\Delta t)})] \cdot I\{N_2((j+1)\Delta t) - N_2(j\Delta t) > 0\}\Big\} \\
&\leq \theta_{12}\sigma_{1,e}^2 \lambda_1 E(BRT_{2,0}) \\
0 < \text{cov}(r_{2,j}, r_{1,j+1}) &= \theta_{21}\sigma_{2,e}^2 \lambda_2 E\Big\{[N_2(j\Delta t) - N_2(t_{1,N_1(j\Delta t)})] \cdot I\{N_1((j+1)\Delta t) - N_1(j\Delta t) > 0\}\Big\} \\
&\leq \theta_{21}\sigma_{2,e}^2 \lambda_2 E(BRT_{1,0}).
\end{aligned}$$

For a portfolio consisting of s_1 shares of Asset 1 and s_2 shares of Asset 2, its return at time $j\Delta t$ is $r_j = s_1 r_{1,j} + s_2 r_{2,j}$. Thus,

$$0 < \text{cov}(r_j, r_{j+1}) \leq s_1 s_2 \Big[\theta_{12}\sigma_{1,e}^2 \lambda_1 E(BRT_{2,0}) + \theta_{21}\sigma_{2,e}^2 \lambda_2 E(BRT_{1,0})\Big].$$

Meanwhile,

$$\begin{aligned}
\text{var}(r_{1,j}) &= \sigma_{1,e}^2 \lambda_1 \Delta t + \theta_{21}^2 \sigma_{2,e}^2 \lambda_2 \Delta t \\
\text{var}(r_{2,j}) &= \sigma_{2,e}^2 \lambda_1 \Delta t + \theta_{12}^2 \sigma_{1,e}^2 \lambda_1 \Delta t \\
0 \leq \text{cov}(r_{1,j}, r_{2,j}) &= \theta_{12}\sigma_{1,e}^2 E\Big\{[N_1(t_{2,N_2(j\Delta t)}) - N_1((j-1)\Delta t)] \cdot I\{N_2(j\Delta t) - N_2((j-1)\Delta t) > 0\}\Big\} \\
&+ \theta_{21}\sigma_{2,e}^2 E\Big\{[N_2(t_{1,N_1(j\Delta t)}) - N_2((j-1)\Delta t)] \cdot I\{N_1(j\Delta t) - N_1((j-1)\Delta t) > 0\}\Big\}.
\end{aligned}$$

Thus, $\text{var}(r_j) = s_1^2 \text{var}(r_{1,j}) + s_2^2 \text{var}(r_{2,j}) + 2 s_1 s_2 \text{cov}(r_{1,j}, r_{2,j})$ is between

$$\Big[\sigma_{1,e}^2 \lambda_1 (s_1^2 + s_2^2 \theta_{12}^2) + \sigma_{2,e}^2 \lambda_2 (s_2^2 + \theta_{21}^2 s_1^2)\Big]\Delta t \quad \text{and} \quad \Big[\sigma_{1,e}^2 \lambda_1 (s_1 + s_2 \theta_{12})^2 + \sigma_{2,e}^2 \lambda_2 (s_2 + \theta_{21} s_1)^2\Big]\Delta t.$$

Thus, $\text{corr}(r_j, r_{j+1}) = O(\Delta t^{-1})$.

Overall, the lag-1 autocorrelation of a portfolio consisting of s_1 shares of Asset 1 and s_2 shares of Asset 2 is positive for a given Δt, but as Δt increases, the nonsynchronous-trading-induced portfolio autocorrelation converges to zero.

□