Econometric Theory, **30**, 2014, 536–579.
doi:10.1017/S0266466613000406

LIMIT LAWS IN TRANSACTION-LEVEL ASSET PRICE MODELS

ALEXANDER AUE
University of California, Davis

LAJOS HORVÁTH
University of Utah

CLIFFORD HURVICH
New York University

PHILIPPE SOULIER
Université Paris X

We consider pure-jump transaction-level models for asset prices in continuous time, driven by point processes. In a bivariate model that admits cointegration, we allow for time deformations to account for such effects as intraday seasonal patterns in volatility and nontrading periods that may be different for the two assets. We also allow for asymmetries (leverage effects). We obtain the asymptotic distribution of the log-price process. For the weak fractional cointegration case, we obtain the asymptotic distribution of the ordinary least squares estimator of the cointegrating parameter based on data sampled from an equally spaced discretization of calendar time, and we justify a feasible method of hypothesis testing for the cointegrating parameter based on the corresponding t-statistic. In the strong fractional cointegration case, we obtain the limiting distribution of a continuously averaged tapered estimator as well as other estimators of the cointegrating parameter, and we find that the rate of convergence can be affected by properties of intertrade durations. In particular, the persistence of durations (hence of volatility) can affect the degree of cointegration. We also obtain the rate of convergence of several estimators of the cointegrating parameter in the standard cointegration case. Finally, we consider the properties of the ordinary least squares estimator of the regression parameter in a spurious regression, i.e., in the absence of cointegration.

1. INTRODUCTION

The increasingly widespread availability of transaction-level financial price data motivates the development of models to describe such data, as well as theory for widely used statistics of interest under the assumption of a given transaction-level generating mechanism. We focus here on a bivariate pure-jump model in

Address correspondence to Clifford Hurvich, Stern School of Business, New York University, 44 West 4th Street, New York NY 10012 USA. e-mail: churvich@stern.nyu.edu

© Cambridge University Press 2013

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

continuous time for log prices proposed by Hurvich and Wang (2009, 2010) that yields fractional or standard cointegration. The motivation for using a pure-jump model is that observed price series are step functions, since no change is possible in observed prices during time periods when there are no transactions. Examples of data sets that would fit into the framework of this model include buy prices and sell prices of a single stock, prices of two different stocks within the same industry, stock and option prices of a given company, option prices on a given stock with different degrees of maturity or moneyness, corporate bond prices at different maturities for a given company, and Treasury bond prices at different maturities.

Though our paper is not entirely focused on the case of fractional cointegration, we present some evidence that this case may arise in practice in financial econometrics. We considered option and underlying best-available bid prices for 69 different options on IBM at 390 one-minute intervals from 9:30 AM to 4 PM on May 31, 2007. Using a log-periodogram estimator based on $390^{0.5}$ frequencies, we found that the logs of the original series had estimated memory parameters close to 1, while the residuals from the ordinary least squares (OLS) regression of the log stock price on the log option price had estimated memory parameters that were typically less than 1. Specifically, of the 69 estimated memory parameters based on these residuals, the values ranged from 0.05 to 1.14, with a mean of 0.55 and a standard deviation of 0.28, with 30 of these estimates lying between 0.5 and 1, while 32 were between 0 and 0.5. Thus, there is evidence for cointegration in most of the series studied, and often the evidence points toward fractional rather than standard cointegration. Furthermore, the OLS estimate of the cointegrating parameter (assuming that cointegration exists) ranged from -0.21 to 0.39, with a mean of 0.04 and a standard deviation of 0.13. This provides evidence that the cointegrating parameter is in general not equal to one in the present context, so it is of interest to study properties of estimates of this parameter.

Two basic questions that we address in this paper are the asymptotic distribution of the log prices as time $t \to \infty$, and of the usual OLS estimator of the cointegrating parameter based on n observations of the log prices at equally spaced time intervals as $n \to \infty$. Most of the existing methods for deriving such limit laws (see Robinson and Marinucci, 2001) cannot be applied here because the continuous-time log-price series are not diffusions and because the discretized log-price series are not linear in either an independent and identically distributed (i.i.d.) sequence, a martingale difference sequence, or a strong mixing sequence. Nevertheless, it is of interest to know whether and under what conditions the existing limit laws, based, say, on linearity assumptions in discrete time, continue to hold under a transaction-level generating mechanism.

In the model of Hurvich and Wang (2009, 2010) the price process in continuous time is specified by a counting process giving the cumulative number of transactions up to time t, together with the process of changes in log price at the transaction times. This structure corresponds to the fact that most transaction-level

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

data consist of a time stamp giving the transaction time as well as a price at that time. Hence the resulting price process can be viewed as a marked point process. Option pricing in this context was studied by Prigent, Renault, and Scaillet (2002, 2004).

In the setting of marked point processes, another observable quantity of interest is the *durations*, i.e., the waiting times between successive transactions of a given asset. There is a growing literature on univariate models for durations, including the seminal paper of Engle and Russell (1998) on the autoregressive conditional duration models (ACD), as well as Bauwens and Veredas (2004) on the stochastic duration model (SCD) and Deo, Hsieh, and Hurvich (2010) on the long-memory stochastic duration model (LMSD).

Deo, Hurvich, Soulier and Wang (2009) showed that, subject to regularity conditions, if partial sums of centered durations, scaled by $n^{-(d+1/2)}$ with $d \in [0, 1/2)$, satisfy a functional central limit theorem then the counting process $N(t)$ has long or short memory (for $d > 0$, $d = 0$, respectively) in the sense that $\mathrm{Var} N(t) \sim Ct^{2d+1}$ as $t \to \infty$ (with $C > 0$), and they gave conditions under which this scaling would lead to long memory in volatility. In particular, LMSD durations with $d > 0$ lead to long memory in volatility. The latter property has been widely observed in the econometrics literature, while evidence for long memory in durations was found in Deo et al. (2010).

Hurvich and Wang (2009, 2010) did not derive limit laws for the log-price series or the OLS estimator of the cointegrating parameter, but focused instead on properties of variances and covariances for log-price series and returns, and on lower bounds on the rate of convergence for the OLS estimator.

In this paper, for a slightly modified version of the model of Hurvich and Wang (2009, 2010), but under assumptions that are more general than theirs, we obtain the limit law for the log prices and for the OLS and tapered estimators of the cointegrating parameter. In our result on the limit law for log prices, Theorem 3.1, we allow for a stochastic time-varying intensity function in the counting processes. This allows for such effects as dynamic intraday seasonality in volatility (as observed, for example, in Deo, Hurvich, and Lu (2006), as well as fixed nontrading intervals such as holidays and overnight periods. We also allow in most of our results for asymmetries (leverage effects), and we show that this opens up the possibility that long memory in durations may affect the rate of convergence of estimators of the cointegrating parameter. This raises some heretofore unrecognized ambiguities in the choice of a definition of standard cointegration. Finally, we consider the properties of the OLS estimator in a spurious regression, i.e., in the absence of cointegration.

The remainder of this paper is organized as follows. In Section 2 we write the model for the log price series and state our assumptions on the counting process, the time-deformation functions, and the return shocks. In Section 3 we provide our main results on the long-run behavior of the log-price process (Section 3.1); the OLS estimator for the cointegrating parameter under weak fractional strong

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

fractional, and standard cointegration (Section 3.2, divided into 3.2.1 on the weak case with leverage, including feasible inference for the cointegrating parameter, and 3.2.2 on the no-leverage strong and no-leverage standard cases), a tapered estimator under weak fractional, strong fractional, and standard cointegration (Section 3.3, allowing for leverage), a continuously-averaged tapered estimator under strong fractional and standard cointegration (Section 3.4, under conditions that are easier to check), and the OLS estimator in the spurious regression case (Section 3.5). Section 4 provides proofs.

Additional lemmas (numbered A1, A2,...) and their proofs are provided in supplementary material to this article, available at Cambridge Journals Online (journals.cambridge.org/ect).

2. TRANSACTION-LEVEL MODEL

As in Hurvich and Wang (2009, 2010), we consider a bivariate pure-jump transaction-level price model that enables cointegration. We define the log-price process $y = \big(y_1, y_2\big) = \big(y(t) \colon t \geq 0\big)$ by

$$y_1(t) = \sum_{k=1}^{N_1(t)} \big(e_{1,k} + \eta_{1,k}\big) + \theta \sum_{k=1}^{N_2(t_{1,N_1(t)})} e_{2,k}, \tag{2.1}$$

$$y_2(t) = \sum_{k=1}^{N_2(t)} \big(e_{2,k} + \eta_{2,k}\big) + \theta^{-1} \sum_{k=1}^{N_1(t_{2,N_2(t)})} e_{1,k}, \tag{2.2}$$

where for $i = 1, 2$, $N_i(\cdot)$ are counting processes on the real line (see Daley and Vere-Jones, 2003, p. 43) such that, for $t \geq 0$, $N_i(t)$ gives the total number of transactions of Asset i in $(0, t]$, and $t_{i,k}$ is the clock time (calendar time) for the kth transaction of Asset i, with $\cdots t_{i,-1} \leq t_{i,0} \leq 0 < t_{i,1} \leq t_{i,2} \cdots$. The quantity $N_2\big(t_{1,N_1(t)}\big)$ denotes the number of transactions of Asset 2 between time 0 and the time $t_{1,N_1(t)}$ of the most recent transaction of Asset 1, with an analogous interpretation for $N_1\big(t_{2,N_2(t)}\big)$. The efficient shock sequences $\{e_{i,k}\}_{k=1}^{\infty}$ model the permanent component, and the microstructure noise sequences $\{\eta_{i,k}\}_{k=1}^{\infty}$ model the transitory component of the log-price process. Efficient shock spillover effects are weighted by θ and θ^{-1}, thus yielding cointegration with cointegrating parameter θ, assumed nonzero.

A detailed economic justification for this model, derivation of a common-components representation, as well as a comparison with certain discrete-time models, is given in Hurvich and Wang (2009, 2010). We summarize some of the salient economic features here. The model is consistent with a variety of stylized facts regarding series of transaction-based prices in financial markets, including the following. Log prices of a given asset are constant between transactions on that asset, so realizations of the log-price series are step functions. The times between transactions on a given asset (the durations) are random and serially dependent. Clusters of short (long) durations lead to periods of high (low)

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

volatility in calendar-time returns. Shocks to the log price consist of both permanent contributions to the intrinsic log value and transitory (microstructure) fluctuations, due, for example, to the liquidity impact of orders and to other effects that produce short-term deviations from a martingale structure of log prices. Leverage-type effects typically observed in actual data (e.g., a negative correlation between the current calendar-time return and the future squared return) are allowed to arise from the transaction-level price dynamics, through a potential dependence between the microstructure noise series and the intertrade durations. The cointegration arises from the fact that each series responds to both sets of efficient shocks, in different ways, producing a common component in the log prices in calendar time (see Hurvich and Wang, 2010, Eq. (2)). The strength of the cointegration is determined (most straightforwardly, in the weak fractional cointegration case) by the rate at which partial sums of the microstructure noise (transitory component) become negligible compared to the efficient components of log price. This rate is of economic importance, as it quantifies how fast deviations from equilibrium will tend to dissipate, so, for example, knowledge of the rate would help pairs traders to assess how long they would need to hold their position in order to have a good chance to profit from a given anomalous configuration of the price series. The transaction-level dynamics for the model are reasonably simple. Whenever a transaction on a given asset takes place, the log price changes by an amount equal to the sum of an efficient and microstructure shock for that asset, in addition to an equally weighted sum of all efficient shocks for the other asset occurring since the previous transaction of the given asset. Through this spillover effect, relevant intervening, economically intrinsic shocks are finally impounded in the price of the given asset, but only when it actually trades.

In the mathematical theory presented in this paper, all random variables and stochastic processes are defined on a single probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Expectation with respect to $\mathbb{P}$ will be denoted by $\mathbb{E}$, and var and cov will denote the variance and covariance with respect to $\mathbb{P}$. Convergence in $\mathbb{P}$-probability will be denoted by $\xrightarrow{\mathbb{P}}$, and convergence in distribution under $\mathbb{P}$ of sequences of random variables will be denoted by $\rightarrow$. We use $\Rightarrow$ to denote weak convergence under $\mathbb{P}$ in the space $\mathcal{D}([0,\infty))$ of left-limited, right-continuous (càdlàg) functions, endowed with Skorohod's J_1 topology. See Billingsley (1968) or Whitt (2002) for details about weak convergence in $\mathcal{D}([0,\infty))$. Whenever the limiting process is continuous, this topology can be replaced by the topology of uniform convergence on compact sets.

Following Daley and Vere-Jones (2003, p. 47), a point process is said to be *simple* if the probability is zero that there exists a time t at which more than one event occurs. We do not assume that the counting processes are simple. Thus, we allow for the possibility that several transactions may occur at the same time. The transaction times $t_{i,k}$ are related to the point process by the duality

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

$N_i(t) = k \Leftrightarrow t_{i,k} \leq t < t_{i,k+1}$.

The durations are then defined by

$\tau_{i,k} = t_{i,k} - t_{i,k-1}$.

If the process is simple, then $N_i\left(t_{i,k}\right) = k$. Otherwise, it only holds that $N_i\left(t_{i,k}\right) \geq k$.

There is no requirement that either the point processes N_i or the durations $\left\{\tau_k\right\}$ be stationary under $\mathbb{P}$. Therefore, we make the following ergodicity-type assumptions.

Assumption 2.1. The sequences $\left\{t_{i,k}\right\}$ are nondecreasing and there exists $\lambda_i \in (0, \infty)$ such that

$$t_{i,k}/k \xrightarrow{\mathbb{P}} 1/\lambda_i \ . \tag{2.3}$$

When the counting processes are simple, this is equivalent to $N_i(t)/t \xrightarrow{\mathbb{P}} \lambda_i$. Since we do not assume simplicity, we must introduce an additional assumption.

Assumption 2.2. $N_i(t)/t \xrightarrow{\mathbb{P}} \lambda_i$.

If the counting processes are defined from stationary ergodic durations, then Assumption 2.1 holds. If the counting processes are moreover simple, then Assumption 2.2 also holds. Conversely, if N_i is stationary and ergodic, then Assumption 2.2 holds, and if N_i is moreover simple, then Assumption 2.1 also holds.

It should be stressed that stationarity of durations and of the point process cannot hold simultaneously under the same probability measure (except for the Poisson point process). See, for instance, Baccelli and Brémaud (2003) or Daley and Vere-Jones (2003) for the mathematical theory and Deo et al. (2009) for an econometric interpretation. We now give an example that illustrates this duality.

Example 2.1
For the LMSD model, consider a probability measure (called the Palm measure) P^0 on $(\Omega, \mathcal{F})$ under which $\tau_k = \epsilon_k e^{\sigma Y_k}$, where σ is a positive constant, $\{\epsilon_k\}$ is an i.i.d. sequence of almost surely positive random variables with finite mean, and $\{Y_k\}$ is a stationary standard Gaussian process, independent of $\{\epsilon_k\}$, whose covariance function goes to 0 at infinity. It follows from the latter assumption and Gaussianity that the process $\{Y_k\}$ is ergodic Ibragimov and Rozanov (1978), hence so is $\{\tau_k\}$. Suppose now that the restriction of $\mathbb{P}$ to the sigma-field generated by N is equal to P^0, then Assumption 2.1 holds with $\lambda^{-1} = \mu e^{\sigma^2/2}$, where μ is the expectation of ϵ_0 under P^0, and since the durations are almost surely positive, so does Assumption 2.2. Now, by the Palm duality theory, there exists a probability measure P under which the associated point process N is stationary (in which case, as mentioned above, the durations are no longer stationary). If the restriction of $\mathbb{P}$ to the sigma-field generated by N is equal to P, then Assumptions 2.1 and 2.2 still hold with the same λ as defined above.

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

Example 2.2
The ACD model proposed in Engle and Russell (1998) is

$$\tau_k = \psi_k \epsilon_k, \qquad \psi_k = \omega + \alpha \tau_{k-1} + \beta \psi_{k-1}, \qquad k \in \mathbb{Z}, \tag{2.4}$$

where $\omega > 0$ and $\alpha, \beta \geq 0$, $\{\epsilon_k\}_{k=-\infty}^{\infty}$ is an i.i.d. sequence of almost surely positive random variables with mean 1. If $\alpha + \beta < 1$ then the sequence $\{\tau_k\}_{k=-\infty}^{\infty}$ is strictly stationary and ergodic and has finite mean $\omega/(1-\alpha-\beta)$. As in the above example, we can alternatively assume that the durations form a stationary ACD process under $\mathbb{P}$ or that the associated point process is stationary under $\mathbb{P}$. In both cases, Assumptions 2.1 and 2.2 hold with $\lambda = (1-\alpha-\beta)/\omega$.

We now explain how time deformations can be used to obtain a nonstationary, possibly nonsimple point process from a stationary ergodic simple one, while Assumptions 2.1 and 2.2 are preserved. Let $\tilde{N}(\cdot)$ be a simple, stationary, and ergodic counting process on $\mathbb{R}$ with intensity $\tilde{\lambda} \in (0, \infty)$, and let f be a deterministic or random function such that f is nondecreasing and has càdlàg paths with probability one. Define then

$$N(t) = \tilde{N}(f(t)) \ .$$

If the function f is random, we assume moreover that it is independent of the counting process $\tilde{N}$.

The function f is used to speed up or slow down the trading clock. To incorporate dynamic intraday seasonality in volatility, the same time deformation can be used in each trading period (of length T, say), assuming that $f(t)$ has a periodic derivative (with period T and with probability one), for example, $f(t) = t + .5\sin(2\pi t/T)$. Fixed nontrading intervals, say, $t \in [T_1, T_2)$, could be accommodated by taking $f(t) = f(T_1)$ for $t \in [T_1, T_2)$ so that $f(t)$ remains constant for t in this interval, and then taking $f(T_2) > f(T_1)$ so that $f(t)$ jumps upward when trading resumes at time T_2. The jump allows for the possibility of one or more transactions at time T_2, potentially reflecting information from other markets or assets that did trade in the period $[T_1, T_2)$.

If the values of some series are only recorded at specific time points (e.g., quarterly in the case of certain macroeconomic series), this could be handled by taking the corresponding $f(t)$ to be a pure-jump function. This provides scope for considering two (or more) series, some of which are observed continuously, others at specific times, though not necessarily contemporaneously. In future work, we hope to explore this scenario in detail, and its possible connections with the MIDAS methodology, see Ghysels, Santa-Clara, and Valkanov (2006).

The use of the time-varying intensity function f renders the counting process N nonstationary. Since it is possible that f has (upward) jumps, the N may also not be simple even though the $\tilde{N}$ are simple. We now show, however, that if $\tilde{N}$ satisfies Assumptions 2.1 and 2.2, then so does the time-deformed N under some restrictions on f.

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

LEMMA 2.1. *Assume that f is a nondecreasing (random) function such that* $t^{-1}f(t) \xrightarrow{\mathbb{P}} \gamma \in (0,\infty)$ *and*

$$\sup_{t\geq 0} |f(t) - f(t^-)| \leq C$$

with probability one, where $C \in (0,\infty)$ is a deterministic constant. Let $\tilde{N}$ be a point process such that Assumptions 2.1 and 2.2 hold for some $\tilde{\lambda} \in (0,\infty)$. Let N be the counting process defined by $N(\cdot) = \tilde{N}(f(\cdot))$. Then Assumptions 2.1 and 2.2 hold for N with $\lambda = \tilde{\lambda}\gamma$.

Stationarity of $\tilde{N}$ is not required in this lemma. For example, one could equally well do the time deformation on a counting process that corresponds to a stationary duration sequence (and hence the counting process is not stationary).

In order to show that our results on estimation of the cointegrating parameter (under weak fractional and standard cointegration) hold in this time-deformation framework, we will in Lemma A.2 make further assumptions on the function f. These assumptions mathematically embody natural economic constraints, viz. minimum duration of trading and nontrading periods, maximum duration of nontrading periods, and nonstoppage of trading time during trading periods.

We now state our assumptions on the return shocks.

Assumption 2.3. The efficient shocks $\{e_{i,k}\}$ are mutually independent i.i.d. sequences with zero mean and variance $\sigma_{i,e}^2$.

Although many of our results would continue to hold if the i.i.d. part of Assumption 2.3 were replaced by a weak-dependence assumption, we maintain the i.i.d. assumption here in keeping with the economic motivation for the model as provided by Hurvich and Wang (2010) that in the absence of the microstructure shocks and in the absence of any dependence of the efficient shocks on the counting processes, each of the log-price series would be a martingale with respect to its own past. Since the trades of Asset 1 are not synchronized in calendar time or in transaction time with those of Asset 2, it seems reasonable to assume that the two efficient shock series are mutually independent, as we have done in Assumption 2.3.

The following assumption implies that the microstructure noise does not affect the limiting distribution of the log prices.

Assumption 2.4 (Microstructure noise)**.** The microstructure noise sequences $\{\eta_{i,k}\}$ satisfy $n^{-1/2}\sum_{k=1}^{[n\cdot]} \eta_{i,k} \Rightarrow 0$.

This assumption simply enforces the negligibility of the microstructure noise with respect to the long term behavior of the log price and indeed allows us to establish in Theorem 3.1 that the log price behaves asymptotically as a random

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

walk (and returns at long horizons behave like a martingale difference sequence), consistent with what is assumed in most econometric literature.

Dependence between the counting processes and return shocks allows for leverage effects (for example, a correlation between a return in one time period and a squared return in a subsequent time period). A transaction-level model yielding a leverage effect was proposed (but justified only with simulations) in Hurvich and Wang (2009). Models where the point process need not be independent of the return shocks were discussed in Prigent (2001) in the context of option pricing with marked point processes.

We do not make any assumption of independence between the counting processes and the microstructure shocks, unless explicitly noted otherwise. We will, however, assume that the counting processes are independent of the efficient shocks except in Theorem 3.1 and in Section 3.5.

Assumption 2.4 is all we need to assume about the microstructure noise in order to obtain a limit law for the log-price series (such as Theorem 3.1 below). However, in order to discuss properties of estimators of the cointegrating parameter, it is necessary to make more specific assumptions on the degree of cointegration. In Hurvich and Wang (2009, 2010), three different cases were considered, according to the strength of the memory of the microstructure noise sequences. These cases were labeled as weak fractional cointegration, strong fractional cointegration, and standard cointegration. In the current context, where there may be a dependence between return shocks and counting processes, special care is needed in defining the strong fractional and standard cointegration cases, as long memory in durations may affect the rate of convergence of estimators of the cointegrating parameter in these cases. On the other hand, we will define weak fractional cointegration essentially as in Hurvich and Wang (2010).

Assumption 2.5. The shocks $\{e_{1,k}\}_{k=-\infty}^{\infty}$, $\{e_{2,k}\}_{k=-\infty}^{\infty}$, $\{\eta_{1,k}\}_{k=-\infty}^{\infty}$, and $\{\eta_{2,k}\}_{k=-\infty}^{\infty}$ are mutually independent.

Mutual independence of the efficient and microstructure shock series of a given asset can be justified on economic grounds and is often made in the econometric literature for calendar-time models. See, e.g., Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008). Mutual independence of the two microstructure series is justified by the lack of synchronization of the trading times of the two assets.

We now discuss the weak fractional cointegration case. For $H \in (0,1)$, let B_H denote the standard fractional Brownian motion (FBM) with Hurst index H, i.e., the zero mean Gaussian process with almost surely continuous sample paths and covariance function

$$\operatorname{cov}(B_H(s), B_H(t)) = \frac{1}{2}\left\{s^{2H} - |t-s|^{2H} + t^{2H}\right\}.$$

For $H = 1/2$, $B_{1/2}$ is the standard Brownian motion.

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

Assumption 2.6 (Weak fractional cointegration)**.** There exists $H \in (0, 1/2)$ such that

$$n^{-H} \sum_{k=1}^{[n\cdot]} \eta_{i,k} \Rightarrow c_i B_H^{(i)},$$

where c_1, c_2 are nonnegative constants, not both zero, and $B_H^{(1)}$ and $B_H^{(2)}$ are independent standard fractional Brownian motions with common Hurst index H.

Assumption 2.6 is a strengthening of Assumption 2.4 and is needed to establish the asymptotic distribution of estimators of the cointegrating parameter θ. Whereas Assumption 2.4 allows for cointegration, an assumption such as Assumption 2.6 is a necessary element for defining the degree of cointegration (the rate at which deviations from the long run cointegrating relationship disappear) and is consistent with the finding reported in Section 1 that the cointegrating residuals have weaker memory than a random walk.

Under Assumption 2.5, the independence of all the noise series implies that all the previous convergences hold jointly. The situation where one of the constants c_1 or c_2 is zero could arise naturally if the memory in one of the microstructure noise series is weaker than for the other.

In the case of weak fractional cointegration, we can define the memory parameter of the microstructure noise series as $d_\eta = H - 1/2 \in (-1/2, 0)$, and the degree of fractional cointegration (i.e., the rate of convergence of partial sums of the cointegrating error) is completely determined by d_η. More precisely, in this case the difference between the memory parameters of the series of log prices and the cointegrating error (observed, say, at equally spaced intervals of calendar time) $y_1(j) - \theta y_2(j)$, is $-d_\eta$. This holds regardless of any dependence between the counting processes and the microstructure shocks.

Next we discuss strong fractional and standard cointegration. We start with the assumption that, for $i = 1, 2$, $\eta_{i,k} = \xi_{i,k} - \xi_{i,k-1}$ where $\{\xi_{i,k}\}$ satisfy $\sup_k \mathbb{E}\left[\xi_{i,k}^2\right] < \infty$. It then follows that the cointegrating error at time j is

$$y_1(j) - \theta y_2(j)$$
$$= \sum_{k=N_1\left(t_{2,N_2(j)}\right)+1}^{N_1(j)} e_{1,k} - \theta \sum_{k=N_2\left(t_{1,N_1(j)}\right)+1}^{N_2(j)} e_{2,k} + \xi_{1,N_1(j)} - \xi_{1,0} - \theta\left(\xi_{2,N_2(j)} - \xi_{2,0}\right). \tag{2.5}$$

Under the assumptions we will make in this paper, and also under the assumptions made in Hurvich and Wang (2009, 2010), the first two terms on the right hand side of (2.5) are weakly dependent, so the degree of cointegration is determined by the rate of convergence of partial sums of $\xi_{i,N_i(j)}$.

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

Thus, we will need to study the sequence $\xi_{i,N_i(j)}$. We do not assume that the microstructure shocks are independent of the counting processes. Thus, even if the microstructure shocks have zero mean, it may hold that $\mathbb{E}\left[\xi_{i,N_i(j)}\right] \neq 0$.

In view of the discussion above it is clear that in order to specify the degree of cointegration in the strong fractional and standard cointegration cases, it is necessary to make an assumption on calendar-time aggregates of $\xi_{i,N_i(j)}$. This is in contrast to Assumption 2.6 above, where the degree of fractional cointegration is specified in terms of properties of transaction-level aggregates of the microstructure noise.

Assumption 2.7 (Strong fractional and standard cointegration)**.** The microstucture noise sequences $\{\eta_{i,k}\}$ can be expressed as $\eta_{i,k} = \xi_{i,k} - \xi_{i,k-1}$. There exist $H \in [1/2, 1)$, constants μ_1^*, μ_2^* and nonnegative constants c_1, c_2, not both zero, such that

$$n^{-H} \sum_{j=1}^{[n\cdot]} \left\{\xi_{i,N_i(j)} - \mu_i^*\right\} \Rightarrow c_i B_H^{(i)},$$

where $B_H^{(1)}$ and $B_H^{(2)}$ are independent fractional Brownian motions with Hurst index H.

The case $H > 1/2$ corresponds to strong fractional cointegration, whereas the case $H = 1/2$ corresponds to standard cointegration.

It might be hard to verify Assumption 2.7 unless the durations are integer valued. Since commonly used duration models do not have integer-valued durations, we will introduce a modification of the estimators that involves integrals instead of sums, thus allowing us to avoid this restriction. This change requires a corresponding modification of Assumption 2.7.

Assumption 2.8 (Strong fractional and standard cointegration)**.** The microstucture noise sequences $\{\eta_{i,k}\}$ can be expressed as $\eta_{i,k} = \xi_{i,k} - \xi_{i,k-1}$. There exist $H \in [1/2, 1)$, constants μ_1^*, μ_2^* and nonnegative constants c_1, c_2, not both zero, such that

$$n^{-H} \int_0^{n\cdot} \left\{\xi_{i,N_i(s)} - \mu_i^*\right\} \mathrm{d}s \Rightarrow c_i B_H^{(i)}.$$

In their strong fractional cointegration case, Hurvich and Wang (2010) assumed, for $d_\eta \in (-1, -1/2)$, that $\operatorname{cov}\left(\xi_{i,k}, \xi_{i,k+j}\right) \sim K j^{2d_\xi - 1}$ as $j \to \infty$ where $K > 0$ and $d_\xi = d_\eta + 1 \in (0, 1/2)$. They then showed (in their Lemma 3), under the assumptions made there, that $\operatorname{cov}\left(\xi_{i,N_i(k)}, \xi_{i,N_i(k+j)}\right) \sim K' j^{2d_\xi - 1}$ as $j \to \infty$ where $K' > 0$, so that the degree of fractional cointegration was completely determined by the rate of decay of $\operatorname{cov}\left(\xi_{i,k}, \xi_{i,k+j}\right)$. However, the proof of this result relied on the assumption that the microstructure shocks are independent of the counting processes, an assumption that we do not make here.

We next provide an example showing that under dependence between the microstructure shocks and the counting processes, it is possible for $\left\{\xi_{i,k}\right\}$ to be

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

weakly dependent, and yet the rate of convergence of suitably normalized integrals of the process $\left(\xi_{i,N_i(t)} : t \geq 0\right)$ is determined by the degree of long memory in durations. Suppressing the i subscript, we have the following lemma.

LEMMA 2.2. *Suppose that* $\{\tau_k\}$ *is given by the LMSD model* $\tau_k = \epsilon_k e^{Y_k}$, $\left\{\epsilon_k\right\}$ *are i.i.d. standard exponential, independent of the stationary standard Gaussian series* $\{Y_k\}$, *which satisfies* $\operatorname{cov}\left(Y_0, Y_n\right) \sim cn^{2H_\tau - 2}$ *where* $c > 0$, *and* $H_\tau \in (1/2, 3/4)$. *Define* $\xi_k = Y_{k+1}^2 - 1$. *Then the autocovariance function of* $\left\{\xi_k\right\}$ *is summable, and there exists* $c' > 0$ *such that*

$$n^{-1/2} \sum_{k=1}^{[n\cdot]} \xi_k \Rightarrow c' B \ . \tag{2.6}$$

Nevertheless, the randomly indexed continuous-time process $\xi_{N(t)}$ *has long memory in the sense that there exists a constant* μ^*, *as well as a constant* c'', *such that*

$$n^{-H_\tau} \int_0^{n\cdot} \left\{\xi_{N(s)} - \mu^*\right\} \mathrm{d}s \Rightarrow c'' B_{H_\tau} \ . \tag{2.7}$$

Lemma 2.2 shows that long memory in durations can induce the same degree of long memory in the cointegrating error (2.5) in calendar time, even though the microstructure shocks, which are the source of the cointegration, have short memory as a sequence in transaction time. In Lemma 2.2, this phenomenon was achieved by imposing a particular functional relationship between the (zero mean) microstructure noise and the persistent component of the durations, $\xi_k = Y_{k+1}^2 - 1$. This relationship implies a leverage effect, since $\operatorname{corr}\left(\xi_k, \tau_{k+1}\right) = 1/\sqrt{2(e-1)} \approx .539 > 0$. In other words, a strongly negative microstructure shock to the return leads to a tendency of the next observed duration, as well as subsequent durations, to be shorter than average. Such a string of short durations increases the volatility, e.g., the expectation of squared calendar-time returns, as shown, for example, under a particular return model in Deo et al. (2009). Furthermore, evidence that stock market intertrade durations have long memory was provided in Deo et al. (2010).

In the absence of dependence between the counting processes and microstructure noise series, in both cases of strong fractional and standard cointegration, the memory of durations cannot affect the memory of the cointegrating error. See Lemma A.4 for strong fractional cointegration and Lemma A.5 for standard cointegration.

3. MAIN RESULTS

3.1. The Long-Run Behavior of the Bivariate Log-Price Process

With the assumptions made in Section 2, the long-run behavior of the bivariate process $y = \left(y_1, y_2\right)$ can be determined. The following theorem shows that the

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

log-prices are approximately integrated. Even though independence is assumed between the various shock series, the log-price process $y = \big(y(t)\colon t \geq 0\big)$ exhibits a nontrivial variance-covariance structure that is determined by a complex interplay of the model parameters.

THEOREM 3.1. *Under Assumptions 2.1, 2.2, 2.3, and 2.4,* $n^{-1/2}\big(y_1(n\cdot), y_2(n\cdot)\big) \Rightarrow \mathbb{B}$, *where*

$$\mathbb{B} = \left(\sigma_{1,e}\sqrt{\lambda_1}B_1 + \theta\sigma_{2,e}\sqrt{\lambda_2}B_2\,,\ \theta^{-1}\sigma_{1,e}\sqrt{\lambda_1}B_1 + \sigma_{2,e}\sqrt{\lambda_2}B_2\right)\,, \tag{3.1}$$

and B_1 *and* B_2 *are independent standard Brownian motions.*

In Theorem 3.1, we have not assumed that the counting processes are independent of either the efficient shocks or the microstructure shocks.

Hurvich and Wang (2009, 2010) have in their Theorem 1 computed the long-run variances of $y_1(t)$ and $y_2(t)$, which are given as $(\sigma_{1,e}^2\lambda_1 + \theta^2\sigma_{2,e}^2\lambda_2)t$ and $(\sigma_{1,e}^2\lambda_1/\theta^2 + \sigma_{2,e}^2\lambda_2)t$, respectively. Our theorem yields the variances as well as the covariances in the limiting distribution of $(t^{-1/2}y(t)\colon t \geq 0)$. More importantly, our theorem provides the limiting distribution itself for the (normalized) log-price process y which, in turn, can be used for asymptotic statistical inference. Indeed, most of the subsequent results in this paper use Theorem 3.1 and its proof as a building block. In particular, a slightly generalized version of this theorem directly yields asymptotics for estimators in spurious regressions and therefore can be used to motivate tests for the null hypothesis of no cointegration, as we describe in Section 3.5.

3.2. OLS Estimator of the Cointegrating Parameter

In this section, we derive the asymptotic behavior of the OLS estimator of the cointegrating parameter θ. To do so, we assume that the log-price series are observed at integer multiples of Δt. We will work here, without loss of generality, with $\Delta t = 1$ in order to keep the notation simple. Then (2.1) and (2.2) become

$$y_1(j) = \sum_{k=1}^{N_1(j)} \big(e_{1,k} + \eta_{1,k}\big) + \theta \sum_{k=1}^{N_2\left(t_{1,N_1(j)}\right)} e_{2,k},$$

$$y_2(j) = \sum_{k=1}^{N_2(j)} \big(e_{2,k} + \eta_{2,k}\big) + \theta^{-1} \sum_{k=1}^{N_1\left(t_{2,N_2(j)}\right)} e_{1,k}.$$

Regressing $y_1(1), \ldots, y_1(n)$ on $y_2(1), \ldots, y_2(n)$ without intercept, we obtain the OLS estimator of θ as

$$\hat{\theta}_n^{OLS} = \frac{\sum_{j=1}^n y_2(j)y_1(j)}{\sum_{j=1}^n y_2^2(j)}. \tag{3.2}$$

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

Hurvich and Wang (2009, 2010) have shown in their Theorem 6 (under conditions that are for the most part stronger than the ones we assume here) that $\hat{\theta}_n^{OLS}$ is weakly consistent for θ and obtained a lower bound on the rate of convergence in the case of weak fractional, strong fractional, and standard cointegration. The exact limit distributions, however, were not given. We fill in this gap next for weak fractional cointegration.

3.2.1. OLS and Inference for the Cointegrating Parameter Under Weak Fractional Cointegration

THEOREM 3.2. *Let Assumptions 2.1, 2.2, 2.3, 2.5, and 2.6 hold. Assume in addition that the counting processes N_1 and N_2 are mutually independent and independent of the efficient shocks and there exists a constant C such that*

$$\sup_{s\geq 0} \mathbb{E}\left[t_{i,N_i(s)+1} - s\right] \leq C \,. \tag{3.3}$$

Then

$$n^{1/2-H}(\hat{\theta}_n^{OLS} - \theta) \to \Sigma_1 \frac{\int_0^1 B(t) B_H(t)\,\mathrm{d}t}{\int_0^1 B^2(t)\,\mathrm{d}t},$$

where B is a standard Brownian motion, B_H is a fractional Brownian motion, independent of B, and

$$\Sigma_1^2 = \frac{c_1^2 \lambda_1^{2H} + \theta^2 c_2^2 \lambda_2^{2H}}{\theta^{-2}\lambda_1 \sigma_{1,e}^2 + \lambda_2 \sigma_{2,e}^2} \,.$$

The result in Theorem 3.2 is similar to that obtained in Robinson and Marinucci (2001, Prop. 6.5, eq. (6.8)), under their Assumption 6.1, for which a sufficient condition (their formula (6.5)) was verified in Marinucci and Robinson (2000) to hold for weak (but not strong) fractional cointegration in the case where the process is linear with respect to i.i.d. innovations.

Next, we provide sufficient conditions for condition (3.3) to hold for the LMSD and ACD models under both frameworks for stationarity considered in Examples 2.1 and 2.2.

- Note first that as long as the point processes N_i are stationary under $\mathbb{P}$, then condition (3.3) holds provided that $\mathbb{E}[t_1] < \infty$ (equivalently $E^0\left[\tau_0^2\right] < \infty$, where E^0 represents expectation with respect to the Palm measure), since the forward recurrence time $\left\{t_{i,N_i(s)} - s\right\}$ is then stationary.
- If the durations form a stationary LMSD sequence under $\mathbb{P}$, then Lemma A.3 shows that if $\mathbb{E}\left[\epsilon_0^q\right] < \infty$ for all $q \geq 1$, then (3.3) holds.
- If the durations form a stationary ACD sequence under $\mathbb{P}$, then Lemma A.1 shows that (3.3) holds as long as $\mathbb{E}\left[\tau_0^3\right] < \infty$. By Carrasco and Chen (2002, Cor. 6), this holds true if $\mathbb{E}\left[\left(\beta + \alpha\epsilon_t\right)^3\right] < 1$.

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

- If $\tilde{N}$ is a point process that satisfies condition (3.3) under $\mathbb{P}$, and if f is a time-deformation function that satisfies the economically justified assumptions of Lemma A.2, then the time-deformed point process N (defined by $N(t) = \tilde{N}(f(t))$ still satisfies (3.3).

Next, we consider inference for the cointegrating parameter θ under weak fractional cointegration. It is seen from Theorem 3.2 that the asymptotic distribution of $n^{1/2-H}(\hat{\theta}_n^{OLS} - \theta)$ depends on a variety of unknown quantities. To alleviate the dependence on nuisance parameters and to thereby facilitate inference on θ, we consider the t-statistic for testing the null hypothesis $H_0 : \theta = \theta_0$. If the null hypothesis is true, then the t-statistic is given by

$$t_n = \frac{\hat{\theta}_n^{OLS} - \theta}{\hat{\sigma}_{\hat{\theta}_n^{OLS}}},$$

where $\hat{\sigma}_{\hat{\theta}_n^{OLS}}$ is the traditional estimated standard error for $\hat{\theta}_n^{OLS}$, with

$$\hat{\sigma}^2_{\hat{\theta}_n^{OLS}} = \frac{n^{-1}\sum_{j=1}^n \left[y_1(j) - \hat{\theta}_n^{OLS} y_2(j)\right]^2}{\sum_{j=1}^n y_2^2(j)}.$$

We have the following theorem.

THEOREM 3.3. *Under the assumptions of Theorem 3.2,*

$$n^{-1/2} t_n \to \frac{\int_0^1 B(t) B_H(t) dt}{\sqrt{\int_0^1 B^2(t)dt \int_0^1 B_H^2(t)dt - \left[\int_0^1 B(t) B_H(t) dt\right]^2}},$$

where B is a standard Brownian motion and B_H is a standard fractional Brownian motion, independent of B.

Note that by the Cauchy-Schwarz inequality,

$$\left|\int_0^1 B(t) B_H(t) dt\right|^2 \le \int_0^1 B^2(t)dt \int_0^1 B_H^2(t)dt,$$

and in fact the inequality above is strict since B and B_H are mutually independent, so that the limiting distribution in Theorem 3.3 is well defined.

Note that the limiting distribution in Theorem 3.3 depends only on H. Thus, if H can be consistently estimated, one can conduct asymptotically valid hypothesis tests for θ based on the test statistic $n^{-1/2}t_n$. The asymptotic null distribution of the test statistic is given by Theorem 3.3, and it is easily seen that if $\theta \neq \theta_0$ then $n^{-1/2}t_n$ diverges.

We now show that H can indeed be consistently estimated, using an aggregation method considered in Giraitis, Robinson, and Surgailis (1999). Let m be an

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

integer sequence such that $1/m + m/n \to 0$. We work with the differences of the cointegrating residuals, $x_j = \Delta\big(y_1(j) - \hat{\theta}_n^{OLS} y_2(j)\big)$, where Δ is the differencing operator. Divide the data set into contiguous, nonoverlapping blocks of size m. The kth block average is

$$X_k^{(m)} = \frac{1}{m}\sum_{t=1}^{m} x_{t+(k-1)m} \ .$$

Denote the sample variance of these block averages by

$$s_m^2 = \lfloor n/m \rfloor^{-1} \sum_{k=1}^{\lfloor n/m \rfloor} \left(X_k^{(m)}\right)^2 \ .$$

Now, define $\hat{H} = 1 + \frac{\log s_m^2}{2\log m}$. The next theorem states that $\hat{H}$ is a consistent estimator of H under assumptions that imply those of Theorem 3.2, but are more restrictive. We make these restrictions to facilitate a reasonably simple proof.

THEOREM 3.4. *Let Assumptions 2.3 and 2.5 hold. Assume that* $\eta_{i,t} = \sigma_i\big\{B_{i,H}(t) - B_{i,H}(t-1)\big\}$, $i = 1, 2$ *where* $B_{i,H}$ *are mutually independent standard fractional Brownian motions with* $H \in (0, 1/2)$. *Assume that* N_1 *and* N_2 *are stationary mutually independent ergodic point processes, independent of the processes* $B_{i,H}$, *such that* $\mathbb{E}\big[N_i^4(1)\big] < \infty$ *and*

$$\sup_{t\geq 2} \mathbb{E}\left[\left(\frac{N_i(t)}{t}\right)^{4H-8} \mathbf{1}_{\{N_i(t)>0\}}\right] < \infty \ . \tag{3.4}$$

Then, $\hat{H} \xrightarrow{\mathbb{P}} H$.

We now continue the discussion of the LMSD and ACD models (see Examples 2.1 and 2.2 and the discussion immediately after Theorem 3.2). The proof of the following claims is in the online supplement following Lemma A.6, which it uses.

- Assume that the point process N is stationary under the probability $\mathbb{P}$ and that under the Palm probability P^0 the durations form an LMSD sequence $\{\tau_k\}$ with memory parameter $H_\tau \in (1/2, 1)$ as in Example 2.1. If $E^0\big[\epsilon_0^{9-4H}\big] < \infty$, condition (3.4) holds. If $E^0\big[\epsilon_0^q\big] < \infty$ with $q > 4/(1 - H_\tau)$, then $\mathbb{E}\big[N^4(1)\big] < \infty$.
- Assume that point process N is stationary under the probability $\mathbb{P}$ and that under the Palm probability P^0 the durations form an ACD sequence $\{\tau_k\}$ as in Example 2.2. If moreover ϵ_0 admits a strictly positive density on $[0, \infty)$ and $E^0\big[\tau_1^{9-4H+\eta}\big] < \infty$ for some $\eta > 0$, then $\mathbb{E}\big[N^4(1)\big] < \infty$ and condition (3.4) holds. A condition for $E^0\big[\tau_1^{9-4H+\eta}\big] < \infty$ is $E^0\big[\big(\alpha\epsilon_0 + \beta\big)^{9-4H+\eta}\big] < 1$.

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

We next present some results from a simulation study on statistical inference for θ in the case of weak fractional cointegration based on Theorems 3.3 and 3.4. For $H = 0.1, 0.2, 0.3, 0.4$, we first tabulated the asymptotic null distribution of $n^{-1/2}t_n$ based on Theorem 3.3 using 1,000 replications. From this, we obtained (approximations to) various relevant percentiles of the asymptotic distribution. Next, for each of two sample sizes, $n = 500$ and $n = 2,000$, we simulated $1,000$ replications of $(y_1(t), y_2(t))$ from the model (2.1), (2.2), from which we constructed the equally-spaced observations $y_1(1), \ldots, y_1(n)$ and $y_2(1), \ldots, y_2(n)$.

Our specific choices in simulating the model (2.1), (2.2) were as follows. We used $d_\eta = -0.3$, corresponding to $H = 0.2$. This gives weak fractional cointegration of a degree that we feel is often observed in practice. Other choices could be tried, for this and for other aspects of the model, but the main point of this study is to examine the feasibility of inference based on Theorems 3.3 and 3.4. The durations for both series were taken to be mutually independent LMSD processes (see Example 2.1), both with memory parameter $d_\tau = 0.2$, which is consistent with the stylized facts for the persistence in volatility of financial assets. Specifically, in each LMSD model we took ϵ_k to be unit exponential and $\{Y_k\}$ to obey the $ARFIMA(0, d_\tau, 0)$ model $Y_k = (1 - B)^{-d_\tau} e_k$, where B is the backshift operator, $\{e_k\}$ are i.i.d. standard normal, independent of the $\{\epsilon_k\}$, and σ was selected to yield intensities $\lambda_1 = \lambda_2 = 20$. So one unit of time corresponds to the amount of time in which (on average) 20 transactions take place. We deliberately took a reasonably small number of transactions per unit of time to examine the effect (if any) of the pure-jump nature of the realizations on the distributions of various statistics of interest. For the cointegrating parameter, we used $\theta = 1$. The efficient shocks $e_{1,k}$ and $e_{2,k}$ were mutually independent i.i.d. standard normal sequences. The microstructure shocks $\{\eta_{1,k}\}$ and $\{\eta_{2,k}\}$ were mutually independent Gaussian $ARFIMA(0, -0.3, 0)$ with unit innovation variance.

For the null hypothesis $H_0 : \theta = 1$, which would often be the hypothesis of econometric interest in practice, we constructed, for each realization, the t-statistic as well as the estimated degree of cointegration as measured by $\hat{H}$. We used the aggregation-based estimator $\hat{H}$ considered in Theorem 3.4, but we first standardized the data set of differenced cointegrating residuals $x_1, ..., x_n$ by dividing by the sample standard deviation. This makes the estimator invariant to the multiplicative scale of the input series, and it can be easily shown that $\hat{H}$ is still consistent under this modification. We then tested the null hypothesis versus a two-tailed alternative $H_A : \theta \neq 1$. The test statistic used was $|n^{-1/2}t_n|$. The critical values were obtained from the tabulations of the asymptotic distribution of $n^{-1/2}t_n$, which is symmetric about zero. This distribution does depend on H, however, and for H we used the value in $\{0.1, 0.2, 0.3, 0.4\}$ that came closest to $\hat{H}$. We used two values for the significance level, $\alpha = 0.05$ and $\alpha = 0.01$. The critical values for the two corresponding tests were, respectively, the 97.5th percentile and 99.5th percentile of the distribution for $n^{-1/2}t_n$, using the value of H as described above. For the purpose of calibration, we also considered an infeasible version of these tests where H is assumed known to be 0.2.

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

For the nominal significance level $\alpha = 0.05$, based on the 1,000 simulated realizations, the rate at which the null hypothesis was (incorrectly) rejected (with rates for the infeasible test in parentheses) was 0.065 (0.049) for $n = 500$ and 0.078 (0.054) for $n = 2,000$. For the nominal significance level $\alpha = 0.01$, the rejection rates were 0.017 (0.007) for $n = 500$ and 0.024 (0.015) for $n = 2,000$.

For the infeasible method, the observed rates are not statistically significantly different from the nominal 0.05 and 0.01, given that we have only 1,000 realizations. For the feasible method, the tests are somewhat oversized, apparently due to the variability arising from the estimation of H. It should also be noted that for both the infeasible and feasible tests, the differences between the observed rejection rates for the two sample sizes are not statistically significant.

3.2.2. OLS Under Strong Fractional and Standard Cointegration. We now consider the case where the microstructure noise series $\{\eta_{i,k}\}$ are differences of strongly or weakly dependent processes $\{\xi_{i,k}\}$.

THEOREM 3.5. *Let Assumptions 2.1, 2.2, 2.3, 2.5, and 2.7 hold. Assume moreover that*

- *the efficient shocks are i.i.d. Gaussian;*
- *the counting processes N_1 and N_2 are independent of each other and independent of the microstructure noise sequences and of the efficient shocks, and there exists a constant C such that*

$$\sup_{t\geq 0}\mathbb{E}\left[\left(t_{i,N_i(t)+1}-t\right)^2\right]\leq C\,; \tag{3.5}$$

- $\mathbb{E}[\xi_{i,k}]=0$, $\sup_k \mathbb{E}[\xi_{i,k}^2]<\infty$, *and* $\xi_{i,0}=0$.

Then,

- *if* $1/2 < H < 1$,

$$n^{3/2-H}\left(\hat{\theta}_n^{OLS}-\theta\right)\to\sqrt{\frac{c_1^2+\theta^2c_2^2}{\theta^{-2}\lambda_1\sigma_{1,e}^2+\lambda_2\sigma_{2,e}^2}}\,\frac{\int_0^1 B(s)\,\mathrm{d}B_H(s)}{\int_0^1 B^2(s)\,\mathrm{d}s}\,, \tag{3.6}$$

  where B is standard Brownian motion independent of the standard fractional Brownian motion B_H;

- *if* $H = 1/2$, $n\left(\hat{\theta}_n^{OLS}-\theta\right)=O_P(1)$.

The rate of convergence obtained in the standard cointegration case improves on the one obtained by Hurvich and Wang (2010). The sufficient conditions for (3.3) discussed after Theorem 3.2 become sufficient conditions for (3.5) after augmenting by 1 the exponent in the moment conditions appearing there. The assumptions in Theorem 3.5 are quite strong, ruling out leverage effects and providing one motivation for our subsequent consideration of tapered estimators.

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

3.3. A Tapered Estimator of the Cointegrating Parameter

Even in existing discrete-time models for cointegration, the OLS estimator lacks any particular optimality properties. Here we consider an estimator based on discrete Fourier transforms of the tapered differences of $y_1(j)$, $y_2(j)$, $1 \le j \le n$. It was shown in Chen and Hurvich (2003a) that this estimator can have a faster rate of convergence than OLS in certain cases of fractional cointegration. In the weak fractional cointegration case, our limit results for the tapered estimator (Theorem 3.6) are obtained under identical conditions to those assumed in Theorem 3.2 for OLS. However, under strong fractional and standard cointegration, the conditions for our results on the tapered estimator (Theorem 3.7) allow for leverage, unlike the corresponding theorem for OLS.

We introduce all relevant notation using a generic time series $\{x_j\}_{j=-\infty}^{\infty}$. Let $h\colon I \mapsto \mathbb{R}$ be a general continuous taper function on an open interval I containing $[0,1]$ such that $h(0) = h(1) = 0$. For $\ell = 1, 2, \ldots$, denote by $\omega_\ell = 2\pi\ell/n$ the Fourier frequencies. The tapered DFT of $\{x_j\}_{j=-\infty}^{\infty}$ with taper function h is defined by

$$d_{x,\ell} = \sum_{j=1}^{n} h\left(\frac{j}{n}\right) x_j \, \mathrm{e}^{\mathrm{i}j\omega_\ell} = \sum_{j=1}^{n} h_\ell\left(\frac{j}{n}\right) x_j,$$

where $h_\ell(t) = h(t)\mathrm{e}^{2\pi\,\mathrm{i}\ell t}$. Denote by $\{\Delta x_j\}_{j=-\infty}^{\infty}$ the first difference of the series $\{x_j\}$, where $\Delta x_j = x_j - x_{j-1}$. We define the tapered DFT of the first difference $\{\Delta x_j\}_{j=-\infty}^{\infty}$ with taper function h by

$$d_{\Delta x,\ell} = \sum_{j=1}^{n} h\left(\frac{j}{n}\right) \Delta x_j \, \mathrm{e}^{\mathrm{i}j\omega_\ell} = \sum_{j=1}^{n} h_\ell\left(\frac{j}{n}\right) \Delta x_j. \tag{3.7}$$

In our setting, we observe the cointegrated component processes y_1 and y_2 at equidistant sample points. Defining the cointegrating error $z_j = y_{1,j} - \theta y_{2,j}$ and following Chen and Hurvich (2003b), we can now introduce the estimator

$$\hat{\theta}_n^{Tap} = \mathrm{Re}\left(\tilde{\theta}_n\right),$$

where $\mathrm{Re}(z)$ signifies the real part of a complex number $z = a + \mathrm{i}b$ and, letting $\bar{z} = a - \mathrm{i}b$ be the complex conjugate of z,

$$\tilde{\theta}_n = \frac{\sum_{\ell=1}^{m} d_{\Delta y_1,\ell}\, \bar{d}_{\Delta y_2,\ell}}{\sum_{\ell=1}^{m} |d_{\Delta y_2,\ell}|^2}.$$

Therein, any tapered DFT of differenced sequences is defined according to (3.7). Note that $\hat{\theta}_n^{Tap}$ is the real part of the ratio of the averaged tapered cross-periodogram between the series y_1 and y_2 and the averaged tapered periodogram of the series y_2.

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

3.3.1. Discrete Tapered Estimator Under Weak Fractional Cointegration

THEOREM 3.6. *Let Assumptions 2.1, 2.2, 2.3, 2.5, and 2.6 hold. Assume moreover that the counting processes are mutually independent and independent of the efficient shocks and (3.3) holds. Then,*

$$n^{1/2-H}\left(\hat{\theta}_n^{Tap}-\theta\right)\to\sqrt{\frac{c_1^2\lambda_1^{2H}+\theta^2c_2^2\lambda_2^{2H}}{\theta^{-2}\lambda_1\sigma_{1,e}^2\lambda_1+\lambda_2\sigma_{2,e}^2}}\,\frac{\sum_{\ell=1}^m \mathrm{Re}\left(\int_0^1 h_\ell(s)\,\mathrm{d}B(s)\int_0^1 h_\ell(t)\,\mathrm{d}B_H(t)\right)}{\sum_{\ell=1}^m\left|\int_0^1 h_\ell(s)\,\mathrm{d}B(s)\right|^2},$$

where B is a standard Brownian motion, B_H is a standard fractional Brownian motion, and B and B_H are independent.

Since the assumptions of Theorem 3.6 are the same as in Theorem 3.2, the comments immediately following that theorem also apply here.

3.3.2. Discrete Tapered Estimator Under Strong Fractional and Standard Cointegration

THEOREM 3.7. *Let Assumptions 2.1, 2.2, 2.3, 2.5, and 2.7 hold. Assume moreover that the counting processes are mutually independent and independent of the efficient shocks and (3.5) holds.*

- *If $1/2 < H < 1$, then*

$$n^{3/2-H}(\hat{\theta}_n^{Tap}-\theta)\to\sqrt{\frac{c_1^2+\theta^2c_2^2}{\theta^{-2}\lambda_1\sigma_{1,e}^2+\lambda_2\sigma_{2,e}^2}}\times\frac{\sum_{\ell=1}^m \mathrm{Re}\left(\int_0^1 h_\ell(s)\,\mathrm{d}B(s)\int_0^1 h'_\ell(s)\,\mathrm{d}B_H(s)\right)}{\sum_{\ell=1}^m|\int_0^1 h_\ell(s)\,\mathrm{d}B(s)|^2},$$

  where B_H is a standard fractional Brownian motion independent of the standard Brownian motion B.

- *If $H=1/2$, $n\left(\hat{\theta}_n^{Tap}-\theta\right)=O_P(1)$.*

The assumptions of this theorem are weaker than those of Theorem 3.5 on the OLS estimator. The microstructure shocks are not assumed to be independent of the counting processes, and the efficient shocks are not assumed to be Gaussian. Theorem 3.5 can presumably be proved without the Gaussian assumption. It might be much more difficult in the proof of Theorem 3.5 to avoid the assumption of independence between the microstructure shocks and the counting processes.

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

3.4. A Continuous-Time Tapered Estimator

The estimators of θ we have considered so far are based on equally spaced observations of the log-price series. However, under the model (2.1), (2.2), a continuous-time record is available, and it is of interest to consider using all of the available data to estimate θ. Here, for the sake of theoretical tractability, and in view of the discussion just above Assumption 2.8, we consider a tapered estimator $\tilde{\theta}$ based on continuously averaged log prices on adjacent nonoverlapping time intervals. Since the problems with discretization appear only in the strong fractional and standard cointegration cases, we only consider them in this section. There is no difference in the case of weak fractional cointegration.

We first establish some notation. Let $\{X(t)\}$ be any time series defined for all $t \geq 0$, and suppose that we have data on $\{X_t\}$ for $t \in [0, T]$. Let $\delta > 0$ be fixed. In practice, we might take δ to be 5 minutes, but the choice of δ does not affect the asymptotic distribution we derive below. Define $n = \lfloor T/\delta \rfloor$, $\tilde{X}(0) = 0$, and

$$\tilde{X}(k) = \int_{u=(k-1)\delta}^{k\delta} X(u)\,\mathrm{d}u \quad k = 1, \ldots, n\,.$$

Then we can define an estimator $\tilde{\theta}_\delta$ based on these averaged observations by

$$\hat{\theta}_{n,\delta}^{Tap} = \mathrm{Re}\left(\tilde{\theta}_{n,\delta}\right)\,,$$

with

$$\tilde{\theta}_{n,\delta} = \frac{\sum_{\ell=1}^{m} d_{\Delta\tilde{y}_1,\ell}\bar{d}_{\Delta\tilde{y}_2,\ell}}{\sum_{\ell=1}^{m} |d_{\Delta\tilde{y}_2,\ell}|^2}\,.$$

THEOREM 3.8. *Let Assumptions 2.1, 2.2, 2.3, 2.5, and 2.8 hold. Assume moreover that the counting processes are mutually independent and independent of the efficient shocks and (3.5) holds.*

- *If* $1/2 < H < 1$, *then*

$$n^{3/2-H}(\hat{\theta}_{n,\delta}^{Tap} - \theta) \to \sqrt{\frac{\delta^{2H}(c_1^2 + \theta^2 c_2^2)}{\theta^{-2}\lambda_1\sigma_{1,e}^2 + \lambda_2\sigma_{2,e}^2}}\,\frac{\sum_{\ell=1}^{m}\mathrm{Re}\left(\int_0^1 h_\ell(s)\,\mathrm{d}B(s)\int_0^1 h'_\ell(s)\mathrm{d}B_H(s)\right)}{\sum_{\ell=1}^{m}|\int_0^1 h_\ell(s)\,\mathrm{d}B(s)|^2},$$

  where B_H *is a standard FBM independent of the standard Brownian motion* B.

- *If* $H = 1/2$, $n\left(\hat{\theta}_{n,\delta}^{Tap} - \theta\right) = O_P(1)$.

Because Assumption 2.8 involves an integral rather than a sum, we are able to verify that it holds for certain models with noninteger durations such as ACD and LMSD under certain relationships with the microstructure shocks.

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

In Theorem 3.8, we allow for leverage effects, and therefore care is required in defining standard cointegration. As demonstrated in Lemma 2.2 (which assumes LMSD durations), if there is a leverage effect, even when the microstructure shocks are the differences of a weakly dependent sequence, the cointegrating error need not be $I(0)$. In such a case we have strong fractional cointegration rather than the standard cointegration that might have been anticipated.

It is also possible that even though a leverage effect exists, the memory of durations has no effect on the degree of cointegration. Specifically, if in Lemma 2.2 we replace $\xi_k = Y_{k+1}^2 - 1$ by $\xi_k = H_2(Y_{k+1}) - .75H_3(Y_{k+1})$, where $H_2(y) = y^2 - 1$ and $H_3(y) = y^3 - 3y$ (the second and third Hermite polynomials, respectively), then there is a leverage effect with $\mathrm{corr}(\tau_{k+1}, \xi_k) = .082$. Nevertheless it follows from an argument similar to the proof of Lemma 2.2 that Assumption 2.8 holds in this example with $H = 1/2$, so that we have standard cointegration and Theorem 3.8 holds with $H = 1/2$.

Lemma A.5 provides an example of standard cointegration allowing for both time deformation and dependence between the counting processes and microstructure shocks. Theorem 3.8 would hold for this example with $H = 1/2$.

3.5. Spurious Regressions

In this section only, we consider a noncointegrated version of the model defined by (2.1) and (2.2),

$$y_1(t) = \sum_{k=1}^{N_1(t)} \left(e_{1,k} + \eta_{1,k}\right) + \theta_{21} \sum_{k=1}^{N_2(t_{1,N_1(t)})} e_{2,k}, \tag{3.8}$$

$$y_2(t) = \sum_{k=1}^{N_2(t)} \left(e_{2,k} + \eta_{2,k}\right) + \theta_{12} \sum_{k=1}^{N_1(t_{2,N_2(t)})} e_{1,k}, \tag{3.9}$$

where $\theta_{12} \neq \theta_{21}^{-1}$. We examine here the properties of the OLS estimator in the (spurious) regression of y_1 on y_2 in discrete time and then briefly discuss corresponding tests for the null hypothesis of cointegration. Corollary 3.1 below follows directly from the proof of Theorem 3.1.

COROLLARY 3.1. *If Assumptions 2.1, 2.2, 2.3, and 2.4 are satisfied and* $y = (y_1, y_2)$ *is given by (3.8) and (3.9) with* $\theta_{12} \neq \theta_{21}^{-1}$, *then as* $n \to \infty$,

$$\left(\frac{1}{\sqrt{n}} y(nu) \colon u \in [0,1]\right) \Rightarrow B_y = (B_y(u) \colon u \in [0,1]),$$

where B_y *is a bivariate Brownian motion with* 2×2 *covariance matrix* $\Sigma = (\Sigma_{i,j} : i, j = 1, 2)$ *given by the entries*

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

$$\Sigma_{1,1} = \lambda_1\sigma_{1,e}^2 + \theta_{21}^2\lambda_2\sigma_{2,e}^2\,, \quad \Sigma_{2,2} = \theta_{12}^2\lambda_1\sigma_{1,e}^2 + \lambda_2\sigma_{2,e}^2\,,$$
$$\Sigma_{1,2} = \theta_{12}\lambda_1\,\sigma_{1,e}^2 + \theta_{21}\lambda_2\,\sigma_{2,e}^2 = \Sigma_{2,1}.$$

Next, we consider the discretization of $y_1(t)$ and $y_2(t)$ given by (3.8) and (3.9) at integer time values,

$$y_{1,j} = \sum_{k=1}^{N_1(j)} (e_{1,k} + \eta_{1,k}) + \theta_{21} \sum_{k=1}^{N_2(t_{1,N_1(j)})} e_{2,k}, \tag{3.10}$$

$$y_{2,j} = \sum_{k=1}^{N_2(j)} (e_{2,k} + \eta_{2,k}) + \theta_{12} \sum_{k=1}^{N_1(t_{2,N_2(j)})} e_{1,k}. \tag{3.11}$$

Regressing $y_{1,1}, \ldots, y_{1,n}$ on $y_{2,1}, \ldots, y_{2,n}$ without intercept, we obtain the OLS estimator

$$\hat{\delta}_n = \frac{\sum_{j=1}^n y_{2,j}\, y_{1,j}}{\sum_{j=1}^n y_{2,j}^2}. \tag{3.12}$$

Corollary 3.2 below follows directly from Corollary 3.1 and the continuous mapping theorem.

COROLLARY 3.2. *If Assumptions 2.1, 2.2, 2.3, and 2.4 are satisfied and* $y = (y_1, y_2)$ *is given by (3.8) and (3.9) with* $\theta_{12} \neq \theta_{21}^{-1}$, *then as* $n \to \infty$,

$$\hat{\delta}_n \to \frac{\int_0^1 B_{2,y}(u) B_{1,y}(u)\, du}{\int_0^1 B_{1,y}^2(u)\, du}\,,$$

where $B_y = \left(B_{1,y}, B_{2,y}\right)$ *is the bivariate Brownian motion given in Corollary 3.1.*

Corollary 3.2 together with Corollary 3.1 can be used to motivate tests for the null hypothesis of no cointegration. We do not pursue the details here, but it seems clear that the null distribution for unit root tests based on the residuals $\{y_{1,j} - \hat{\delta}_n y_{2,j}\}_{j=1}^n$ can be derived from Corollaries 3.1 and 3.2, and that these null distributions will have form similar to the distributions listed, for example, in Hamilton (1994, Prop. 19.4).

4. PROOFS

Proof of Lemma 2.1. Write

$$\frac{N(t)}{t} = \frac{\tilde{N}(f(t))}{t} = \frac{\tilde{N}(f(t))}{f(t)}\,\frac{f(t)}{t}\,.$$

By assumption, $f(t) \mapsto \infty$ (in probability if f is random), thus $\tilde{N}(f(t))/f(t)$ converges in probability to $\tilde{\lambda}$. By assumption, it also holds that $t^{-1}f(t) \xrightarrow{\mathbb{P}} \gamma$.

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

Thus, $\frac{N(t)}{t} \xrightarrow{\mathbb{P}} \tilde{\lambda}\gamma$, so that Assumption 2.2 holds for N with $\lambda = \tilde{\lambda}\gamma$. Next, we note that $N(t_k^-) \le k$, thus

$$1 \le \frac{N(t_k)}{k} = 1 + \frac{N(t_k) - k}{k} \le 1 + \frac{N(t_k) - N(t_k^-)}{k}$$

$$= 1 + \frac{\tilde{N}(f(t_k^-), f(t_k))}{k} \le 1 + \frac{\tilde{N}(f(t_k) - C, f(t_k))}{f(t_k)} \frac{f(t_k)}{k}$$

using the definition of N and the boundedness requirement on the jumps of f. Since $f(t_k)$ tends to infinity, it suffices to prove that if Assumptions 2.1 and 2.2 hold for $\tilde{N}$, then for any fixed positive C, it holds that

$$\frac{\tilde{N}(t - C, t]}{t} \xrightarrow{\mathbb{P}} 0 .$$

Without loss of generality, set $\tilde{\lambda} = 1$. Fix some $\epsilon \in (0, 1/2)$. Since $\tilde{t}_k / k \xrightarrow{\mathbb{P}} 1$, with probability arbitrarily close to 1 (say, bigger than $1 - \epsilon$), there exists k_0 such that $\tilde{t}_k / k \in [1 - \epsilon, 1 + \epsilon]$ for all $k \ge k_0$. For $k \ge k_0$, if $t_k \in (t - C, t]$, it necessarily holds that $t - C \le k(1 + \epsilon)$ and $k(1 - \epsilon) \le t$. Hence, $(t - C)/(1 + \epsilon) \le k \le t/(1 - \epsilon)$. This implies that for t large enough ($t < C + k_0(1 + \epsilon)$), $\tilde{N}(t - C, t)$ is less than the number of integers between $(t - C)/(1 + \epsilon)$ and $t/(1 - \epsilon)$, i.e., $\tilde{N}(t - C, t] \le c\epsilon t$, for some constant $c > 0$. Thus, $\tilde{N}(t)/t \le c\epsilon$ with probability tending to 1, and this proves that $\tilde{N}(t - C, t]/t \xrightarrow{\mathbb{P}} 0$. It follows that $N(t_k)/k = 1 + o_P(1)$, i.e., $N(t_k)/t_k$ converges in probability to 1. Thus,

$$\frac{t_k}{k} = \frac{t_k}{N(t_k)} \frac{N(t_k)}{k} \xrightarrow{\mathbb{P}} \frac{1}{\lambda} . \qquad \blacksquare$$

Proof of Lemma 2.2. Denote H_τ by H to simplify the notation. Define $\xi_k = Y_{k+1}^2 - 1$. Then ξ_k is centered, has finite variance summable autocovariance function, since $\operatorname{cov}(\xi_0, \xi_k) = 2\operatorname{cov}^2(Y_0, Y_{k+1})$. Thus, $\{\xi_k\}$ has a summable autocovariance function because $H \in (1/2, 3/4)$. By Arcones (1994, Thm. 4), this implies that $\{\xi_k\}$ is in the domain of attraction of the standard Brownian motion, i.e.,

$$n^{-1/2} \sum_{k=1}^{[n\cdot]} \xi_k \Rightarrow c' B ,$$

with $c'^2 = \operatorname{var}(\xi_0) + 2\sum_{k=1}^{\infty} \operatorname{cov}(\xi_0, \xi_k)$. This proves (2.6).

Assume now that $\tau_k = \epsilon_k e^{\sigma Y_k}$ (with $\sigma = 1$ in the statement of the lemma). The properties of Hermite polynomials yield that $\mathbb{E}[e^{\sigma Y_0} H_j(Y_0)] = \sigma^j e^{\sigma^2/2}$ for all $j \ge 1$. Denote now $\lambda^{-1} = \mathbb{E}[\tau_k] = \mathbb{E}[e^{\sigma Y_k}] = e^{\sigma^2/2}$, $m = \mathbb{E}[\xi_{k-1}\tau_k] = \mathbb{E}[(Y_k^2 - 1)e^{Y_k}] = \sigma^2 e^{\sigma^2/2}$, and $G(y) = (y^2 - 1)e^{\sigma y} - m$. We now prove that (2.7) holds

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

with $\mu^* = \lambda m$. Write

$$\begin{aligned}
&\int_0^T \left(\xi_{N(s)} - \lambda m\right) \mathrm{d}s \\
&\quad = t_1\xi_0 + \sum_{k=1}^{N(T)} \tau_{k+1}\xi_k - \lambda m T + \left(t_{N(T)+1} - T\right)\xi_{N(T)+1} \\
&\quad = t_1\xi_0 + \sum_{k=1}^{N(T)} (\epsilon_{k+1} - 1)\,\xi_k \mathrm{e}^{\sigma Y_{k+1}} + \sum_{k=0}^{N(T)} G\left(Y_{k+1}\right) \qquad \textbf{(4.1)} \\
&\quad + m(N(T) - \lambda T) - (t_{N(T)+1} - T)\xi_{N(T)+1}\,.
\end{aligned}$$

By Lemma A.3 and applying Hölder's inequality, it can be shown that $\left(t_{N(T)+1} - T\right)(\xi_{N(T)+1} - \rho) = O_P(1)$. Since the sequence $\{\epsilon_k\}$ is independent of the Gaussian process $\{Y_k\}$, the second term in the right-hand side of (4.1) is in the domain of attraction of the standard Brownian motion, and the normalizing sequence is $\sqrt{n}$. Thus, we must obtain the joint asymptotic behavior of $\sum_{k=1}^{N(Tt)} G(Y_k)$ and $N(Tt) - \lambda T t$.

The durations are in the domain of attraction of the fractional Brownian motion with Hurst index H, since

$$\sum_{k=1}^{n} \left(\tau_k - \lambda^{-1}\right) = \sum_{k=1}^{n} (\epsilon_k - 1)\,\mathrm{e}^{\sigma Y_k} + \sum_{k=1}^{n} \left(\mathrm{e}^{\sigma Y_k} - \lambda^{-1}\right)\,.$$

The first term in the right-hand side is $O_P(\sqrt{n})$, and the second sum, suitably normalized, converges to the fractional Brownian motion with Hurst index H because the function $x \mapsto \mathrm{e}^{\sigma x} - \lambda^{-1}$ has Hermite rank 1. See, e.g., Arcones (1994). More precisely, let $c_1 = \mathbb{E}[Y_1 \mathrm{e}^{\sigma Y_1}] = \sigma \mathrm{e}^{\sigma^2/2}$ and define $g(y) = \mathrm{e}^{\sigma y} - \lambda^{-1} - c_1 y$. The function g has Hermite rank 2, and since $H \in (1/2, 3/4)$, this implies that

$$\mathrm{var}\left(\sum_{k=1}^{n} g(Y_k)\right) = O(n)\,.$$

Thus, $\sum_{k=1}^{n} \left(\tau_k - \lambda^{-1}\right)$ is asymptotically equivalent to $c_1 \sum_{k=1}^{n} Y_k$. Let B_H denote the standard fractional Brownian motion with Hurst index H. The assumption on the covariance of the Gaussian process $\{y_k\}$ implies that

$$n^{-H} \sum_{k=1}^{[n\cdot]} Y_k \Rightarrow \varphi B_H$$

with $\varphi^2 = c/\{H(2H-1)\}$. Denote now $c_2 = \mathbb{E}[Y_1 G(Y_1)] = \sigma(\sigma^2 + 2)\mathrm{e}^{\sigma^2/2}$ and define $h(y) = G(y) - c_2 y$. Then h has Hermite rank 2 and thus by similar arguments as above, $\sum_{k=1}^{n} G(Y_k)$ is asymptotically equivalent to $c_2 \sum_{k=1}^{n} Y_k$. Thus we obtain

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

$$n^{-H}\left(\sum_{k=1}^{[nt]}(\tau_k-\lambda^{-1}),\sum_{k=1}^{[nt]}G(Y_k)\right)\Rightarrow(c_1\varphi B_H(t),c_2\varphi B_H(t))\ .$$

By Vervaat's lemma (see Vervaat, 1972; or Resnick, 2007, Prop. 3.3), the previous convergence implies that

$$n^{-H}\left(N(nt)-\lambda nt,\sum_{k=1}^{[nt]}G(Y_k)\right)\Rightarrow(-\lambda c_1\varphi B_H(\lambda t),c_2\varphi B_H(t))\ .$$

By the continuity of the composition map, this yields

$$n^{-H}\left(N(nt)-\lambda nt,\sum_{k=1}^{N(nt)}G(Y_k)\right)\Rightarrow(-\lambda c_1\varphi B_H(\lambda t),c_2\varphi B_H(\lambda t))\ .$$

Next we obtain that

$$n^{-H}\left\{\sum_{k=1}^{N(nt)}G(Y_k)+m(N(nt)-\lambda nt)\right\}\Rightarrow\varphi\,(c_2-\lambda mc_1)\,B_H(\lambda t),$$

with $c_2-\lambda mc_1=2\sigma\,\mathrm{e}^{\sigma^2/2}>0$. We conclude that $n^{-H}\int_0^{n\cdot}\{\xi_{N(s)}-\lambda m\}\,\mathrm{d}s\Rightarrow\varphi(c_2-\lambda mc_1)B_H$. ■

4.1. Proof of Theorem 3.1 and Corollary 3.1

We first need the following lemma.

LEMMA 4.1. *Under Assumptions 2.1 and 2.2, $N_i(t_{j,N_j(nt)})/n$ converges in probability uniformly on compact sets to $\lambda_i t$, where $\{i,j\}=\{1,2\}$.*

Proof of Lemma 4.1. The sequence of (random) functions $N_i(n\cdot)/n$ is nondecreasing and converges pointwise in probability to $\lambda_i t$ by ergodicity. A sequence of nondecreasing function converging to a continuous function converges uniformly on compact sets. This results is known as Dini's theorem; cf. Resnick (1987, p. 3). Thus the convergence of $N_i(n\cdot)/n$ is uniform on compact sets. Assumptions 2.1 and 2.2 imply that $N_i(t)\xrightarrow{\mathbb{P}}\infty$ and $t_{i,n}\xrightarrow{\mathbb{P}}\infty$. Thus,

$$\frac{N_i\left(t_{j,N_j(nu)}\right)}{n}=\frac{N_i\left(t_{j,N_j(nu)}\right)}{t_{j,N_j(nu)}}\times\frac{t_{j,N_j(nu)}}{N_j(nu)}\times\frac{N_j(nu)}{n}\xrightarrow{\mathbb{P}}\lambda_i\times\frac{1}{\lambda_j}\times\lambda_j u=\lambda_i u\ .$$

Applying again Dini's lemma, we also have that $N_i(t_{j,N_j(nu)})/n$ converges uniformly on compact sets to $\lambda_i u$. ■

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

Proof of Theorem 3.1. Denote $S^e_{i,n}(t) = \sum_{k=1}^{[nt]} e_{i,k}$ and $S^\eta_{i,n}(t) = \sum_{k=1}^{[nt]} \eta_{i,k}$, $i = 1, 2$. Under Assumptions 2.3 and 2.4, $n^{-1/2}\left(S^e_{1,n}, S^e_{2,n}, S^\eta_{1,n}, S^\eta_{1,n}\right)$ converges weakly to $(\sigma_{1,e}B_1, \sigma_{2,e}B_2, 0, 0)$, where B_1 and B_2 are independent standard Brownian motions. This follows from the independence of e_1 and e_2 and the local uniform convergence to 0 in probability of $n^{-1/2}S^\eta_{i,n}$. With the previous notation, (3.8) and (3.9) become

$$y_1(nt) = S^e_{1,n}\left(N_1(nt)\right) + \theta_{21} S^e_{2,n}\left(N_2(t_{1,N_1(nt)})\right) + S^\eta_{1,n}\left(N_1(nt)\right) ,$$

$$y_2(nt) = S^e_{2,n}\left(N_2(nt)\right) + \theta_{12} S^e_{1,n}\left(N_1(t_{2,N_2(nt)})\right) + S^\eta_{2,n}\left(N_2(nt)\right) .$$

By Lemma 4.1 and the continuity of the composition map on $\mathcal{C} \times \mathcal{C}$ endowed with the metric of uniform convergence on compact sets (see, e.g., Billingsley, 1968, Ch. 3, Sec. 17), we obtain the joint convergence of

$$n^{-1/2}\Big(S^e_{1,n}\left(N_1(n\cdot)\right), S^e_{1,n}\left(N_1\left(t_{2,N_2(n\cdot)}\right)\right), S^e_{2,n}\left(N_2(n\cdot)\right), S^e_{2,n}\left(N_2(t_{1,N_1(n\cdot)})\right), S^\eta_{1,n}\left(N_1(n\cdot)\right), S^\eta_{2,n}\left(N_2(n\cdot)\right)\Big)$$

to $\left(\sigma_{1,e}\sqrt{\lambda_1}B_1, \sigma_{1,e}\sqrt{\lambda_2}B_1, \sigma_{2,e}\sqrt{\lambda_2}B_2, \sigma_{2,e}\sqrt{\lambda_2}B_2, 0, 0\right)$. This yields Corollary 3.1 and Theorem 3.1 by setting $\theta_{21} = \theta$ and $\theta_{12} = \theta^{-1}$. ■

4.2. Proof of Theorems 3.2, 3.3, 3.4, and 3.5

Proof of Theorem 3.2. Write

$$\hat{\theta}_n^{OLS} = \theta + \frac{\sum_{j=1}^n \{y_1(j) - \theta y_2(j)\} y_2(j)}{\sum_{j=1}^n y_2^2(j)} .$$

Assumptions 2.1, 2.2, 2.3, 2.5, and 2.6 imply those of Theorem 3.1. Thus, we can apply the continuous mapping theorem and obtain

$$n^{-2}\sum_{j=1}^n y_2^2(j) \to \left\{\theta^{-2}\lambda_1\sigma_{1,e}^2 + \lambda_2\sigma_{2,e}^2\right\}\int_0^1 B^2(s)\,\mathrm{d}s , \tag{4.2}$$

where B is a standard Brownian motion. Thus, in order to study the convergence of $\hat{\theta}_n^{OLS} - \theta$ suitably renormalized, it suffices to study the sum

$$\sum_{j=1}^n \{y_1(j) - \theta y_2(j)\}\, y_2(j) .$$

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

We further decompose the cointegrating error. Denote

$$y_1^e(j) = \sum_{k=1}^{N_1(j)} e_{1,k} + \theta \sum_{k=1}^{N_2\left(t_{1,N_1(j)}\right)} e_{2,k}\,, \quad y_1^\eta(j) = \sum_{k=1}^{N_1(j)} \eta_{1,k}\,,$$

$$y_2^e(j) = \sum_{k=1}^{N_2(j)} e_{2,k} + \theta^{-1} \sum_{k=1}^{N_1\left(t_{2,N_2(j)}\right)} e_{1,k}\,, \quad y_2^\eta(j) = \sum_{k=1}^{N_2(j)} \eta_{2,k}\,,$$

$$r_{1,j} = \sum_{k=N_1\left(t_{2,N_2(j)}\right)+1}^{N_1(j)} e_{1,k}\,, \quad r_{2,j} = \sum_{k=N_2\left(t_{1,N_1(j)}\right)+1}^{N_2(j)} e_{2,k}\,.$$

With this notation, we can write

$$\sum_{j=1}^{n} \{y_1(j) - \theta y_2(j)\}\, y_2(j) = \sum_{j=1}^{n} \left\{r_{1,j} - \theta r_{2,j}\right\} y_2(j) + \sum_{j=1}^{n} \left\{y_1^\eta(j) - \theta y_2^\eta(j)\right\} y_2(j)\,. \tag{4.3}$$

Applying Theorem 3.1, Assumption 2.6, and the continuous mapping theorem, we obtain

$$\begin{aligned} n^{-3/2-H} &\sum_{j=1}^{n} \left\{y_1^\eta(j) - \theta y_2^\eta(j)\right\} y_2(j) \\ &\to \int_0^1 \left\{\theta^{-1}\sqrt{\lambda_1}\sigma_{1,e} B_1(t) + \sqrt{\lambda_2}\sigma_{2,e} B_2(t)\right\} \left\{c_1 B_{1,H}(\lambda_1 t) \right. \\ &\left. -\theta c_2 B_{2,H}(\lambda_2 t)\right\} \mathrm{d}t \\ &\overset{law}{=} \Sigma \int_0^1 B(t) B_H(t)\, \mathrm{d}t, \end{aligned}$$

where B is a standard Brownian motion, B_H is a fractional Brownian motion, independent of B, and

$$\Sigma^2 = \left(\theta^{-2}\lambda_1\sigma_{1,e}^2 + \lambda_2\sigma_{2,e}^2\right)\left(c_1^2\lambda_1^{2H} + \theta^2 c_2^2 \lambda_2^{2H}\right)\,. \tag{4.4}$$

There only remains to prove that, for $i = 1, 2$,

$$n^{-3/2} \sum_{j=1}^{n} r_{i,j}\, y_2(j) = O_P(1)\,. \tag{4.5}$$

The convergence of $n^{-1/2} y_2$ is uniform on $[0, 1]$, so $n^{-1/2} \max_{1 \le j \le n} |y_2(j)| = O_P(1)$. Therefore, it suffices to prove that

$$n^{-1} \sum_{j=1}^{n} |r_{i,j}| = O_P(1)\,. \tag{4.6}$$

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

Recall that $N_i(s) < k \Leftrightarrow t_{i,k} > s$. Thus, for $k \le N_1(n)$,

$$N_1\left(t_{2,N_2(j)}\right) < k \le N_1(j) \Leftrightarrow t_{2,N_2(j)} < t_{1,k} \le j\ .$$

The first inequality on the right-hand side means that there is no point of N_2 between $t_{1,k}$ and j, i.e., $j \le t_{2,N_2(t_{1,k})+1}$. Let $A_2(t) = t_{2,N_2(t)+1} - t$ denote the forward recurrence time of N_2, i.e., the time between t and the next event of N_2 after t. Thus,

$$\sum_{j=1}^{n} |r_{1,j}| \le \sum_{j=1}^{n} \sum_{k=N_1(t_{2,N_2(j)})+1}^{N_1(j)} |e_{1,k}| = \sum_{k=1}^{N_1(n)} |e_{1,k}|\{A_2(t_{1,k}) + 1\}\ .$$

We thus get the bound for the conditional expectation given the sigma-field $\mathcal{N}$ generated by the counting processes N_1 and N_2,

$$\mathbb{E}\left[\sum_{j=1}^{n} |r_{1,j}|\,|\,\mathcal{N}\right] \le C \sum_{k=1}^{N_1(n)} A_2(t_{1,k})\ .$$

Conditioning on N_1 and applying (3.3) yields

$$\mathbb{E}\left[\sum_{j=1}^{n} |r_{1,j}|\,|\,N_1\right] \le C N_1(n) = O_P(n)\ .$$

This proves (4.6) and concludes the proof of Theorem 3.2. ■

Proof of Theorem 3.3. We have

$$n^{-1/2} t_n = \frac{n^{1/2-H}\left(\hat{\theta}_n^{OLS} - \theta\right)}{n^{1/2-H}\sqrt{n\hat{\sigma}^2_{\hat{\theta}_n^{OLS}}}}\ .$$

Furthermore,

$$\begin{aligned} n\hat{\sigma}^2_{\hat{\theta}_n^{OLS}} &= \frac{\sum_{j=1}^{n}\left[y_1(j) - \hat{\theta}_n^{OLS} y_2(j)\right]^2}{\sum_{j=1}^{n} y_2^2(j)} \\ &= \frac{\sum_{j=1}^{n}\left[y_1(j) - \theta y_2(j) + \left(\theta - \hat{\theta}_n^{OLS}\right) y_2(j)\right]^2}{\sum_{j=1}^{n} y_2^2(j)} \\ &= \frac{\sum_{j=1}^{n}[y_1(j) - \theta y_2(j)]^2}{\sum_{j=1}^{n} y_2^2(j)} - \left(\theta - \hat{\theta}_n^{OLS}\right)^2\ . \end{aligned}$$

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

Note that

$$n^{-2H-1}\sum_{j=1}^{n}[y_1(j)-\theta y_2(j)]^2$$
$$=n^{-2H-1}\sum_{j=1}^{n}\left[y_1^{\eta}(j)-\theta y_2^{\eta}(j)\right]^2$$
$$+n^{-2H-1}\sum_{j=1}^{n}\left(r_{1,j}-\theta r_{2,j}\right)^2+2n^{-2H-1}\sum_{j=1}^{n}\left(r_{1,j}-\theta r_{2,j}\right)\left(y_1^{\eta}(j)-\theta y_2^{\eta}(j)\right).$$

By Theorem 3.1 and the continuous mapping theorem, the first term is $O_P(1)$. We will prove below that $n^{-1}\sum_{j=1}^{n}\left(r_{1,j}-\theta r_{2,j}\right)^2=O_P(1)$, implying that the second and last terms are $o_P(1)$. By the proof of Theorem 3.2, we also have that

$$n^{1/2-H}\left(\hat{\theta}_n^{OLS}-\theta\right)=\frac{n^{-3/2-H}\sum_{j=1}^{n}\left[y_1^{\eta}(j)-\theta y_2^{\eta}(j)\right]y_2(j)+o_P(1)}{n^{-2}\sum_{j=1}^{n}y_2^2(j)}.$$

Thus, we can write

$$n^{-1/2}t_n=\frac{n^{-3/2-H}\sum_{j=1}^{n}\left[y_1^{\eta}(j)-\theta y_2^{\eta}(j)\right]y_2(j)+o_P(1)}{n^{-2}\sum_{j=1}^{n}y_2^2(j)}$$
$$\cdot\left\{\frac{n^{-2H-1}\sum_{j=1}^{n}\left[y_1^{\eta}(j)-\theta y_2^{\eta}(j)\right]^2+o_P(1)}{n^{-2}\sum_{j=1}^{n}y_2^2(j)}\right.$$
$$\left.-\left[\frac{n^{-3/2-H}\sum_{j=1}^{n}\left[y_1^{\eta}(j)-\theta y_2^{\eta}(j)\right]y_2(j)+o_P(1)}{n^{-2}\sum_{j=1}^{n}y_2^2(j)}\right]^2\right\}^{-1/2}.$$

By Theorem 3.1 and Assumption 2.6, we know that $(n^{-1/2}y_2(n\cdot), n^{-H}[y_1^{\eta}(n\cdot)-\theta y_2^{\eta}(n\cdot)])$ converge jointly to $(\varsigma B, \varsigma_H B_H)$, where B is a standard Brownian motion, B_H is a standard fractional Brownian motion, mutually independent, and ς and ς_H are positive constants. Thus, by the continuous mapping theorem, we have

$$n^{-1/2}t_n\to\frac{\varsigma\varsigma_H\int_0^1 B(t)B_H(t)dt}{\varsigma^2\int_0^1 B^2(t)dt}\left\{\frac{\varsigma_H^2\int_0^1 B_H^2(t)dt}{\varsigma^2\int_0^1 B^2(t)dt}-\frac{\varsigma^2\varsigma_H^2\left[\int_0^1 B(t)B_H(t)dt\right]^2}{\varsigma^4\left[\int_0^1 B^2(t)dt\right]^2}\right\}^{-1/2}$$
$$=\frac{\int_0^1 B(t)B_H(t)dt}{\sqrt{\int_0^1 B^2(t)dt\int_0^1 B_H^2(t)dt-\left[\int_0^1 B(t)B_H(t)dt\right]^2}}.$$

We now deal with the remainder term $n^{-1}\sum_{j=1}^{n}\left(r_{1,j}-\theta r_{2,j}\right)^2$. We only prove that $n^{-1}\sum_{j=1}^{n}r_{1,j}^2=O_P(1)$, the proof for the term involving $r_{2,j}$ being similar.

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

Since the counting processes and the efficient shocks are independent, taking conditional expectations, we obtain

$$\mathbb{E}\left[\sum_{j=1}^{n} r_{1,j}^2 \mid \mathcal{N}\right] = \sum_{j=1}^{n} \sum_{k=N_1(t_{2,N_2(j)})+1}^{N_1(j)} \sigma_{1,e}^2 = \sigma_{1,e}^2 \sum_{k=1}^{N(n)} \left\{A_2(t_{1,k})+1\right\} .$$

Since the counting processes are mutually independent, we can apply Condition (3.3) to see that the expectation of the sum in the right-hand side is $O(n)$. Thus, $\sum_{j=1}^{n} r_{1,j}^2 = O_P(n)$, and this concludes the proof. ■

Proof of Theorem 3.4. We will prove below that there exists a positive constant C such that

$$m^{2-2H} s_m^2 \xrightarrow{\mathbb{P}} C . \tag{4.7}$$

Thus,

$$\log\left(m^{2-2H} s_m^2\right) \xrightarrow{\mathbb{P}} \log C ,$$

and

$$\hat{H} = 1 + \frac{\log\left(s_m^2\right)}{2\log m} = H + \frac{\log\left(m^{2-2H} s_m^2\right)}{2\log m} \xrightarrow{\mathbb{P}} H .$$

■

Proof of (4.7). Elementary algebra yields

$$m X_k^{(m)} = A_{k,m} + V_{k,m} - R_{k,m},$$

where

$$\begin{aligned} A_{k,m} = {} & \sigma_1 \left\{B_{1,H}\left(N_1(km)\right) - B_{1,H}\left(N_1((k-1)m)\right)\right\} \\ & -\theta\sigma_2 \left\{B_{2,H}\left(N_2(km)\right) - B_{2,H}\left(N_2((k-1)m)\right)\right\} , \end{aligned}$$

$$R_{k,m} = \left(\hat{\theta}_n^{OLS} - \theta\right)\left[y_2(km) - y_2((k-1)m)\right] ,$$

$$\begin{aligned} V_{k,m} = {} & \sum_{\ell=N_1(t_{2,N_2(km)})+1}^{N_1(km)} e_{1,\ell} - \theta \sum_{\ell=N_2(t_{1,N_1(km)})+1}^{N_2(km)} e_{2,\ell} \\ & - \sum_{\ell=N_1(t_{2,N_2((k-1)m)})+1}^{N_1((k-1)m)} e_{1,\ell} + \theta \sum_{\ell=N_2(t_{1,N_1((k-1)m)})+1}^{N_2((k-1)m)} e_{2,\ell} . \end{aligned}$$

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

The convergence (4.7) is a consequence of three convergences,

$$m^{1-2H}n^{-1}\sum_{k=1}^{\lfloor n/m\rfloor} R_{k,m}^2 \xrightarrow{\mathbb{P}} 0\,, \tag{4.8}$$

$$m^{1-2H}n^{-1}\sum_{k=1}^{\lfloor n/m\rfloor} V_{k,m}^2 \xrightarrow{\mathbb{P}} 0\,, \tag{4.9}$$

$$m^{1-2H}n^{-1}\sum_{k=1}^{\lfloor n/m\rfloor} A_{k,m}^2 \xrightarrow{\mathbb{P}} C\,. \tag{4.10}$$

We will only prove (4.10), the other convergences being similarly and more easily obtained. Let $T_n = m^{1-2H}n^{-1}\sum_{k=1}^{\lfloor n/m\rfloor} A_{k,m}^2$ and let $\mathcal{N}$ denote the sigma-field generated by the point processes N_1 and N_2. We will prove that there exists a positive constant C such that

$$\lim_{n\to\infty} \mathbb{E}[T_n \mid \mathcal{N}] = C\,, \tag{4.11}$$

$$\lim_{n\to\infty} \operatorname{var}(T_n \mid \mathcal{N}) = 0\,. \tag{4.12}$$

By the Bienaymé-Chebyshev inequality, (4.11) and (4.12), we have

$$\begin{aligned}\mathbb{P}(|T_n - C| > \epsilon \mid \mathcal{N}) &\le \epsilon^{-2}\mathbb{E}\left[|T_n - C|^2 \mid \mathcal{N}\right]\\ &\le \epsilon^{-2}\operatorname{var}(T_n \mid \mathcal{N}) + \epsilon^{-2}\left(\mathbb{E}[T_n \mid \mathcal{N}] - C\right)^2 \xrightarrow{\mathbb{P}} 0\,.\end{aligned}$$

This precisely means that T_n converges to C in conditional probability, i.e., for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|T_n - C| > \epsilon \mid \mathcal{N}) = 0\,.$$

Since a probability is bounded by one and $\mathbb{P}(|T_n - C| > \epsilon) = \mathbb{E}[\mathbb{P}(|T_n - C| > \epsilon \mid \mathcal{N})]$, the bounded convergence theorem implies that for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|T_n - C| > \epsilon) = 0\,,$$

i.e., $T_n \xrightarrow{\mathbb{P}} C$.

For simplicity of notation, we also assume that n/m is an integer. For any $a < b$, $N_i(a,b]$ denotes the number of points of N_i in the interval $(a,b]$. Note that $N_i(0,t] = N_i(t)$ for all $t > 0$. Since $B_{1,H}$ and $B_{2,H}$ are independent of $\mathcal{N}$, we have, for any $s < t$,

$$\operatorname{var}\left(B_{i,H}(N_i(t)) - B_{i,H}(N_i(s)) \mid \mathcal{N}\right) = \{N_i(s,t]\}^{2H}\,. \tag{4.13}$$

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

Since, moreover, $B_{1,H}$ and $B_{2,H}$ are mutually independent, this yields

$$\begin{aligned}\mathbb{E}[T_n \mid \mathcal{N}] &= \sigma_1^2 m^{1-2H} n^{-1} \sum_{k=1}^{n/m} \{N_1(((k-1)m, km])\}^{2H} \\ &\quad + \sigma_2^2 \theta^2 m^{1-2H} n^{-1} \sum_{k=1}^{n/m} \{N_2(((k-1)m, km])\}^{2H} \\ &= \sigma_1^2 T_{1,n} + \sigma_2^2 \theta^2 T_{2,n} \,.\end{aligned}$$

We will prove that $T_{i,n}$ converges in probability to λ_i^{2H}, $i = 1, 2$, where λ_i is the intensity of N_i, i.e., $\mathbb{E}\big[N_i(0,1]\big] = \lambda_i$. This will imply (4.11) with $C = \sigma_2^2 \lambda_1^{2H} + \sigma_2^2 \theta^2 \lambda_2^{2H}$.

Since N_i is stationary, we have

$$\mathbb{E}\left[\left|T_{i,n} - \lambda_i^{2H}\right|\right] \le \mathbb{E}\left[\left|\{m^{-1} N_i(m)\}^{2H} - \lambda_i^{2H}\right|\right] \,.$$

For brevity, we now omit the subscript i. By stationarity and ergodicity, $m^{-1}N(m)$ converges almost surely to λ as m goes to infinity. Since $0 < 2H < 1$, it holds that $|a^{2H} - b^{2H}| \le |a-b|^{2H}$ for all real numbers a, b. Thus, for any $\epsilon > 0$,

$$\begin{aligned}&\mathbb{E}\left[\left|\left\{m^{-1}N(m)\right\}^{2H} - \lambda^{2H}\right|\right] \\ &\le \epsilon^{2H} + \mathbb{E}\left[\left|\left\{m^{-1}N(m)\right\}^{2H} - \lambda^{2H}\right| \mathbf{1}_{|m^{-1}N(m)-\lambda|>\epsilon\}}\right] \\ &\le \epsilon^{2H} + \mathbb{E}\left[\left\{m^{-1}N(m)\right\}^{2H} \mathbf{1}_{|m^{-1}N(m)-\lambda|>\epsilon\}}\right] + \lambda^{2H}\mathbb{P}\left(|m^{-1}N(m) - \lambda| > \epsilon\right) \,.\end{aligned}$$

By ergodicity, the last term above tends to zero as m tends to infinity. To deal with the middle term, we apply Hölder's inequality and obtain

$$\begin{aligned}&\mathbb{E}\left[\left\{m^{-1}N(m)\right\}^{2H} \mathbf{1}_{|m^{-1}N(m)-\lambda|>\epsilon\}}\right] \\ &\le \mathbb{E}^{2H}\left[m^{-1}N(m)\right] \mathbb{P}^{1-2H}\left(|m^{-1}N(m) - \lambda| > \epsilon\right) \\ &= \lambda^{2H}\mathbb{P}^{1-2H}\left(|m^{-1}N(m) - \lambda| > \epsilon\right) \to 0 \,,\end{aligned}$$

as $m \to \infty$, again by ergodicity. Thus, we obtain that $\lim_{n\to\infty} \mathbb{E}\big[\big|T_{i,n} - \lambda_i^{2H}\big|\big] = 0$ and this concludes the proof of (4.11).

We now prove (4.12). We denote the conditional variance and covariance given $\mathcal{N}$ by var_N and cov_N, respectively. Since $A_{k,m}$ is conditionally Gaussian, we have, for all k, k',

$$\mathrm{cov}_N\left(A_{k,m}^2, A_{k',m}^2\right)\Big) = 2\mathrm{cov}_N^2\left(A_{k,m}, A_{k',m}\right) \,.$$

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

Thus, denoting $S_n = \sum_{k=1}^{[n/m]} A_{k,m}^2$, we have

$$\begin{aligned}\mathrm{var}_N(S_n) &= 2\sum_{k=1}^{n/m} \mathrm{var}_N^2\left(A_{k,m}\right) + 4\sum_{k=1}^{n/m-1}\sum_{k'=k+1}^{n/m} \mathrm{cov}_N^2\left(A_{k,m}, A_{k',n}\right)\\ &= 2\times I + 4\times II\,.\end{aligned}$$

Applying (4.13) to compute $\mathrm{var}_N(A_{k,m})$ and taking expectation, we have, by stationarity,

$$m^{2-4H}n^{-2}\mathbb{E}[I] = \sigma_1^2 mn^{-1}\mathbb{E}\left[(N_1(m)/m)^{4H}\right] + \sigma_2^2\theta^2 mn^{-1}\mathbb{E}\left[(N_2(m)/m)^{4H}\right]\,.$$

Since $4H < 2$, by Jensen's inequality, $\mathbb{E}[N_i^{4H}(m)] \le \left\{\mathbb{E}\left[N_i^2(m)\right]\right\}^{2H}$ and by stationarity $\mathbb{E}\left[N_i^2(m)\right] \le m^2\mathbb{E}\left[N_i^2(1)\right]$. Thus $m^{2-4H}n^{-2}\mathbb{E}[I] = O(m/n)$.

Consider now the last term II. For any positive real numbers $s < t < u < v$ and a standard fractional Brownian motion H, we have

$$\begin{aligned}&\mathrm{cov}\left(B_H(t) - B_H(s), B_H(v) - B_H(u)\right)\\ &\qquad = |v-s|^{2H} - |v-t|^{2H} + |u-t|^{2H} - |u-s|^{2H}\,.\end{aligned}$$

Thus,

$$\begin{aligned}\mathrm{cov}_N\left(A_{k,m}, A_{k',n}\right) &= \sigma_1^2\left\{N_1^{2H}((k-1)m, k'm] - N_1^{2H}(km, k'm]\right.\\ &\quad \left. + N_1^{2H}(km, (k'-1)m] - N_1^{2H}((k-1)m, (k'-1)m]\right\}\\ &\quad + \sigma_2^2\theta^2\left\{N_2^{2H}((k-1)m, k'm] - N_2^{2H}(km, k'm]\right.\\ &\quad \left. + N_2^{2H}(km, (k'-1)m] - N_2^{2H}((k-1)m, (k'-1)m]\right\}\\ &= \sigma_1^2 C_{k,k'} + \sigma_2^2\theta^2 C'_{k,k'}\,.\end{aligned}$$

Hereafter, we only deal with the terms related to N_1, the other terms being similarly dealt with, and we omit the subscript 1. For any $a, c \ge 0$ and $b > 0$, since $0 < 2H < 1$, we have

$$\begin{aligned}&0 \le (a+b)^{2H} - b^{2H} - (a+b+c)^{2H} + (b+c)^{2H}\\ &\quad = 2H(1-2H)\int_b^{a+b}\int_z^{z+c} u^{2H-2}\mathrm{d}u\,\mathrm{d}z \le acb^{2H-2}\,.\end{aligned}$$

Applying this bound with $a_k = N((k-1)m, km]$, $b_{k,k'} = N(km, (k'-1)m]$ yields

$$C_{k,k'}^2 \le b_{k,k'}^{4H-4}a_k^2 a_{k'}^2\mathbf{1}_{\{b_{k,k'}>0\}}\,.$$

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

Taking expectation, we have, by stationarity,

$$m^{2-4H}n^{-2}\sum_{1\le k<k''\le n/m}\mathbb{E}\left[C_{k,k'}^2\mathbf{1}_{\{b_{k,k'}>0\}}\right]$$
$$\le m^{1-4H}n^{-1}\sum_{k=1}^{n/m}\mathbb{E}\left[N^{4H-4}(km)\mathbf{1}_{\{N(km)>0\}}N^2(m)N^2((k-1)m,km]\right].$$

Applying Hölder's inequality yields

$$m^{2-4H}n^{-2}\sum_{1\le k<k''\le n/m}\mathbb{E}\left[C_{k,k'}^2\mathbf{1}_{\{b_{k,k'}>0\}}\right]$$
$$\le m^{-2}\mathbb{E}^{1/2}\left[N^4(m)\right]\frac{m}{n}\sum_{k=1}^{n/m}\mathbb{E}^{1/2}\left[\{N(km)/km\}^{8H-8}\mathbf{1}_{\{N(km)>0\}}\right]$$
$$\times k^{4H-4}.$$

By stationarity, $\mathbb{E}[N^4(m)] = O(m^4)$, and applying Assumption (3.4), we obtain, for some constant c,

$$m^{2-4H}n^{-2}\sum_{1\le k<k''\le n/m}\mathbb{E}\left[C_{k,k'}^2\mathbf{1}_{\{b_{k,k'}>0\}}\right]\le c\,\frac{m}{n}\sum_{k=1}^{n/m}k^{4H-4}=o(1).$$

Consider now the event $\{b_{k,k'}=0\}$. Then, using the above notations, we have

$$\mathbb{E}\left[C_{k,k'}^2\mathbf{1}_{\{b_{k,k'}=0\}}\right]=\mathbb{E}\left[\{a_k^{2H}+a_{k'}^{2H}-(a_k+a_{k'})^{2H}\}^2\mathbf{1}_{\{b_{k,k'}=0\}}\right]$$
$$\le 4\mathbb{E}\left[\left(a_k^{4H}+a_{k'}^{4H}\right)\mathbf{1}_{\{b_{k,k'}=0\}}\right].$$

Thus, by stationarity of N, we have

$$m^{2-4H}n^{-2}\sum_{1\le k<k''\le n/m}\mathbb{E}\left[C_{k,k'}^2\mathbf{1}_{\{b_{k,k'}=0\}}\right]$$
$$\le m^{1-4H}n^{-1}\sum_{k=1}^{n/m}\mathbb{E}[N^{4H}(m)\mathbf{1}_{\{N(m,km]=0\}}]$$
$$+m^{1-4H}n^{-1}\sum_{k=1}^{n/m}\mathbb{E}[N^{4H}((k-1)m,km]\mathbf{1}_{\{N((k-1)m)=0\}}].$$

Applying Hölder's inequality, $\mathbb{E}\left[N^2(m)\right]=O(m^2)$, and stationarity yields

$$m^{1-4H}n^{-1}\sum_{k=1}^{n/m}\mathbb{E}[N^{4H}(m)\mathbf{1}_{\{N(m,km]=0\}}]=m^{1-4H}n^{-1}\mathbb{E}[N^{4H}(m)\mathbf{1}_{\{N(m,n]=0\}}]$$
$$\le m^{1-4H}n^{-1}\mathbb{E}^{2H}[N^2(m)]\mathbb{P}^{1-2H}(N(m,n]=0)$$
$$\le Cmn^{-1}\mathbb{P}^{1-2H}(N(n-m)=0).$$

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

Since $n/m \to \infty$, it holds that $\lim_{n\to\infty} \mathbb{P}(N(n-m)=0)=0$. Similarly,

$$\begin{aligned} m^{1-4H}n^{-1}\sum_{k=1}^{n/m} \mathbb{E}\left[N^{4H}((k-1)m,km\right]\mathbf{1}_{\{N((k-1)m)=0\}}] \\ \le m^{1-4H}n^{-1}\mathbb{E}^{2H}\left[N^2(m)\right]\sum_{k=1}^{n/m}\mathbb{P}^{1-2H}\left(N((k-1)m)=0\right) \\ \le \mathbb{P}^{1-2H}\left(N(n)=0\right), \end{aligned}$$

and this last term tends to 0 as n tends to infinity. This concludes the proof of (4.12) and of (4.7). ■

Proof of Theorem 3.5. The proof is a consequence of the convergence (4.2), the decomposition (4.3), and Lemmas 4.2 and 4.3, whose assumptions are those of the theorem. ■

LEMMA 4.2. *Under the assumptions of Theorem 3.5,*

$$n^{-H-1/2}\sum_{j=1}^{n}\left\{y_1^{\eta}(j)-\theta y_2^{\eta}(j)\right\}y_2(j) \to \Sigma_0\int_0^1 B(s)\,\mathrm{d}B_H(s)\,, \tag{4.14}$$

where B_H is a standard fractional Brownian motion independent of B and

$$\Sigma_0 = \left(\theta^{-2}\lambda_1\sigma_{1,e}^2+\lambda_2\sigma_{2,e}^2\right)\left(c_1^2+\theta^2c_2^2\right)\,.$$

Proof of Lemma 4.2. Denote $S_n=\sum_{j=1}^n\left\{y_1^{\eta}(j)-\theta y_2^{\eta}(j)\right\}y_2(j)$ and write $y_2=y_2^e+y_2^{\eta}$ with obvious notation. Denote $\zeta_j=y_1^{\eta}(j)-\theta y_2^{\eta}(j)=\xi_{1,N_1(j)}-\theta\xi_{2,N_2(j)}$. Then

$$S_n=\sum_{j=1}^n\zeta_j y_2^e(j)+\sum_{j=1}^n\zeta_j\xi_{2,N_2(j)}\,. \tag{4.15}$$

By the last part of Assumption 2.7, the last term in the right-hand side of (4.15) is $O_P(n)$. Consider the first term in the right-hand side of (4.15), say $S_{1,n}$. Write

$$\begin{aligned} S_{1,n} &= \sum_{j=1}^n\zeta_j\sum_{k=1}^{N_2(j)}e_{2,k}+\theta^{-1}\sum_{j=1}^n\zeta_j\sum_{k=1}^{N_1(t_{2,N_2(j)})}e_{1,k} \\ &= \sum_{k=1}^{N_2(n)}e_{2,k}\sum_{\{j\le n:\,N_2(j)\ge k\}}\zeta_j+\theta^{-1}\sum_{k=1}^{N_1(t_{2,N_2(n)})}e_{1,k}\sum_{\{j\le n:\,N_1(t_{2,N_2(j)})\ge k\}}\zeta_j \\ &= T_{1,n}+\theta^{-1}T_{2,n}\,. \end{aligned}$$

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

Denote $W_n(t) = \sum_{j=1}^{[nt]} \zeta_j$. Since $N_2(j) < k$ iff $j < t_{2,k}$, we obtain

$$T_{1,n} = y_2^{e_2}(n) W_n(1) - \sum_{k=1}^{N_2(n)} e_{2,k} W_n\left(t_{2,k}/n\right) .$$

By Assumption 2.7 and Theorem 3.1, $n^{-1/2-H} y_2^{e_2}(n) W_n(1) \to \sqrt{\lambda_2}\sigma_2 B_2(1) Z(1)$ with $Z = c_1 B_H^{(1)} - \theta c_2 B_H^{(2)} \overset{law}{=} \sqrt{c_1^2 + \theta^2 c_2^2}\, B_H$. Let the last term be denoted by U_n. Since the shocks $e_{i,k}$ are i.i.d. Gaussian, we can compute the characteristic function of U_n,

$$\begin{aligned}\mathbb{E}[\exp\{\mathrm{i}tn^{-1/2-H} U_n\}] &= \mathbb{E}\left[\exp\left\{-\frac{\sigma_{2,e}^2 t^2}{2}\frac{1}{n}\sum_{k=1}^{N_2(n)}\left(n^{-H} W_n(t_{2,k}/n)\right)^2\right\}\right] \\ &\to \mathbb{E}\left[\exp\left\{-\frac{\lambda_2 \sigma_{2,e}^2 t^2}{2}\int_0^1 Z^2(s)\,\mathrm{d}s\right\}\right] .\end{aligned}$$

The convergence is actually joint with that of $n^{-1/2-H} y_2^{e_2}(n) W_n$, thus we have

$$n^{-1/2-H} T_{1,n} \to \sqrt{\lambda_2}\sigma_{2,e} B_2(1) Z(1) - \sqrt{\lambda_2}\sigma_{2,e}\int_0^1 Z(s)\,\mathrm{d}B_2(s) .$$

The limit can also be written as $\sqrt{\lambda_2}\sigma_{2,e}\int_0^1 B_2(s)\,\mathrm{d}Z(s)$. Consider now the term $T_{2,n}$. Note that $N_1(t_{2,N_2(j)}) < k$ iff $j \le t_{2,N_2(t_{1,k})+1}$. Thus,

$$T_{2,n} = \sum_{k=1}^{N_1(t_{2,N_2(n)})} e_{1,k} W_n(1) - \sum_{k=1}^{N_1(t_{2,N_2(n)})} e_{1,k} W_n\left(t_{2,N_2(t_{1,k})+1}/n\right) .$$

By similar arguments as previously, we obtain

$$n^{-H-1/2} T_{2,n} \to \sqrt{\lambda_1}\sigma_1 B_1(1) Z(1) - \sqrt{\lambda_1}\sigma_1\int_0^1 Z(s)\,\mathrm{d}B_1(s) .$$

All convergences hold jointly, thus (4.14) holds. ■

LEMMA 4.3. *Under the assumptions of Theorem 3.5,*

$$\sum_{j=1}^{n}\left\{r_{1,j} - \theta r_{2,j}\right\} y_2(j) = O_P(n) , \tag{4.16}$$

Proof of Lemma 4.3. We first study the term with $r_{1,j}$ and split it into three parts,

$$\sum_{j=1}^{n} r_{1,j} y_2(j) = \sum_{j=1}^{n} r_{1,j} y_2^{e_1}(j) + \sum_{j=1}^{n} r_{1,j} y_2^{e_2}(j) + \sum_{j=1}^{n} r_{1,j} y_2^{\eta}(j)$$

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

We start with the last one. Recall that $N_1(t_{2,N_2(j)}) < k \le N_1(j)$ iff $t_{1,k} \le j \le t_{1,k} + A_2(t_{1,k})$. Thus,

$$\sum_{j=1}^{n} r_{1,j} y_2^{\eta}(j) = \sum_{j=1}^{n} \xi_{2,N_2(j)} \sum_{N_1(t_{2,N_2(j)}) < k \le N_1(j)} e_{1,k} \,. \tag{4.17}$$

If the microstructure shocks are independent of the counting processes, then

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{j=1}^{n} e_{1,k} \sum_{t_{1,k} \le j < t_{1,k} + A_2(t_1,k)} \xi_{2,N_2(j)}\right)^2 \mid \mathcal{N}\right] \\
&\quad = \sigma_{1,e}^2 \sum_{k=1}^{N_1(n)} \mathbb{E}\left[\left(\sum_{t_{1,k} \le j < t_{1,k} + A_2(t_1,k)} \xi_{2,N_2(j)}\right)^2 \mid \mathcal{N}\right] \\
&\quad \le C \sum_{k=1}^{N_1(n)} (A_2(t_{1,k}) + 1)^2 \sup_{\ell} \mathbb{E}\left[\xi_{2,\ell}^2\right] \,.
\end{aligned}$$

Conditioning on N_1 and then taking expectation yields

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sum_{j=1}^{n} e_{1,k} \sum_{t_{1,k} \le j < t_{1,k} + A_2(t_1,k)} \xi_{2,N_2(j)}\Big)^2\Big] \\
&\quad \le C \mathbb{E}\big[N_1(n)\big] \sup_{t} \mathbb{E}\big[\{1 + A_2(t)\}^2\big] \sup_{\ell} \mathbb{E}\big[\xi_{\ell}^2\big] = O(n) \,.
\end{aligned}$$

Consider now $R_{2,n} = \sum_{j=1}^{n} r_{1,j} y_2^{e_2}(j)$.

$$R_{2,n} = \sum_{j=1}^{n} y_2^{e_2}(j) \sum_{N_1(t_{2,N_2(j)})+1}^{N_1(j)} e_{1,k} = \sum_{k=1}^{N_1(n)} e_{1,k} \sum_{t_{1,k} \le j < t_{1,k} + A_2(t_{1,k})} y_2^{e_2}(j) \,.$$

By independence of the efficient shocks and the counting processes, we have

$$\mathbb{E}[R_{2,n}^2 \mid \mathcal{N}] \le C N_1(n) \sum_{k=1}^{N_1(n)} (A_2(t_{1,k}) + 1)^2 = O_P(n^2) \,.$$

This proves that $R_{2,n} = O_P(n)$. Consider, finally, $R_{1,n} = \sum_{j=1}^{n} r_{1,j} y_2^{e_1}(j)$. By definition, $e_{1,k}$ is independent of $y_2^{e_1}(j)$ for j such that $N_1(t_{2,N_2(j)}) < k$. Thus, we can compute the conditional variance given $\mathcal{N}$,

$$\begin{aligned}
\mathbb{E}[R_{1,n}^2 \mid \mathcal{N}] &= \sigma_{1,e}^2 \sum_{k=1}^{N_1(n)} \mathbb{E}\left[\left(\sum_{t_{1,k} \le j < t_{1,k} + A_2(t_{1,k})} y_2^{e_1}(j)\right)^2 \mid \mathcal{N}\right] \\
&\le C N_2(n) \sum_{k=1}^{N_1(n)} (A_2(t_{1,k}) + 1)^2 = O_P(n)
\end{aligned}$$

by (3.5). This concludes the proof of Lemma 4.3. ■

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

4.3. Proof of Theorems 3.6 and 3.7

Proof of Theorem 3.6. Write

$$\tilde{\theta}_n = \theta + \frac{\sum_{\ell=1}^m d_{\Delta r,\ell}\,\bar{d}_{\Delta y_2,\ell}}{\sum_{\ell=1}^m |d_{\Delta y_2,\ell}|^2} + \frac{\sum_{\ell=1}^m d_{\Delta y^\eta,\ell}\,\bar{d}_{\Delta y_2,\ell}}{\sum_{\ell=1}^m |d_{\Delta y_2,\ell}|^2}$$

with $y^\eta(j) = y_1^\eta(j) - \theta y_2^\eta(j)$, $r(j) = r_1(j) - \theta r_2(j)$, and

$$r_1(j) = \sum_{k=N_1(t_{2,N_2(j)})+1}^{N_1(j)} e_{1,k}\,, \quad r_2(j) = \sum_{k=N_2(t_{1,N_1(j)})+1}^{N_2(j)} e_{2,k}\,.$$

By summation by parts, since $h(0) = h(1) = 0$, for any time series $\{x_j\}$, we can write

$$d_{\Delta x,\ell} = \sum_{j=0}^{n-1} \{h_\ell(j/n) - h_\ell((j+1)/n)\}\, x_j = -\frac{1}{n}\sum_{j=0}^{n-1} w_\ell(j,n) x_j, \tag{4.18}$$

with $w_\ell(j,n) = n\{h_\ell((j+1)/n) - h_\ell(j/n)\}$. Applying (4.18) to y_2 yields

$$d_{\Delta y_2,\ell} = -\frac{1}{n}\sum_{j=0}^{n-1} w_\ell(j,n) y_2(j)\,.$$

Since the assumptions of Theorems 3.6 and 3.7 imply those of Theorem 3.1, the continuous mapping theorem yields

$$\{n^{-1/2} d_{\Delta y_2,\ell}, 1 \le \ell \le m\} \to \left\{ -\Sigma_e \int_0^1 h_\ell'(s) B(s)\,\mathrm{d}s\,, 1 \le \ell \le m \right\}, \tag{4.19}$$

where B is a standard Brownian motion and $\Sigma_e^2 = \theta^{-2}\lambda_1\sigma_{1,e}^2 + \lambda_2\sigma_{2,e}^2$. By integration by parts, the integral can also be expressed as

$$-\int_0^1 h_\ell'(s) B(s)\,\mathrm{d}s = \int_0^1 h_\ell(s)\,\mathrm{d}B(s)\,.$$

This, in turn, implies

$$n^{-1}\sum_{\ell=1}^m |d_{\Delta y_2,\ell}|^2 \to \Sigma_e^2 \sum_{\ell=1}^m \left| \int_0^1 h_\ell(s)\,\mathrm{d}B(s) \right|^2. \tag{4.20}$$

Applying now (4.18) to y^η, we obtain

$$d_{\Delta y^\eta,\ell} = -\frac{1}{n}\sum_{j=0}^{n-1} w_\ell(j,n)\left\{ y_1^\eta(j) - \theta y_2^\eta(j) \right\}\,.$$

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

In the case of weak fractional cointegration, we apply Assumption 2.6, the continuous mapping theorem and integration by parts to obtain

$$n^{-H}\ell(n)d_{\Delta y^\eta,\ell} = -n^{-1-H}\ell(n)\sum_{j=0}^{n-1} w_\ell(j,n)\left\{y_1^\eta(j)-\theta y_2^\eta(j)\right\} \to \int_0^1 h_\ell(t)\,\mathrm{d}Z_H(t)\,, \tag{4.21}$$

where, by independence of $B_H^{(1)}$ and $B_H^{(2)}$,

$$Z_H(t) = c_1 B_H^{(1)}(\lambda_1 t) - \theta c_2 B_H^{(2)}(\lambda_2 t) \overset{law}{=} \sqrt{\lambda_1^{2H}c_1^2+\lambda_2^{2H}\theta^2 c_2^2}\,B_H,$$

and B_H is a standard fractional Brownian motion. The first part of Lemma 4.4 shows that $d_{\Delta r,\ell}$ is negligible under the assumptions of Theorem 3.6. This, and the convergences (4.19), (4.20), and (4.21), concludes the proof of Theorem 3.6. ■

Proof of Theorem 3.7. Since $h_\ell(0) = h_\ell(1) = 0$, we have $\sum_{j=0}^{n-1} w_\ell(j,n) = 0$, hence

$$\sum_{j=0}^{n-1} w_\ell(j,n)y_i^\eta(j) = \sum_{j=0}^{n-1} w_\ell(j,n)\left(\xi_{i,N_i(j)}-\xi_{i,0}\right) = \sum_{j=0}^{n-1} w_\ell(j,n)\left(\xi_{i,N_i(j)}-\mu_i^*\right)\,.$$

Denote $S_{i,0} = 0$ and for $k \geq 1$, $S_{i,k} = \sum_{j=1}^{k}\left(\xi_{i,N_i(j)}-\mu_i^*\right)$. Define $\omega_\ell(j,n) = n\left\{w_\ell(j+1,n)-w_\ell(j,n)\right\}$. Applying again summation by parts, we have

$$\sum_{j=0}^{n-1} w_\ell(j,n)y_i^\eta(j) = -\frac{1}{n}\sum_{j=1}^{n-1}\omega_\ell(j,n)S_{i,j} + w_\ell(n,n)S_{i,n-1} + w_\ell(0,n)\left(\xi_{i,0}-\mu_i^*\right)\,,$$

Under Assumption 2.7, by the continuous mapping theorem, we obtain

$$\begin{aligned} n^{1-\gamma}\ell(n)d_{\Delta y^\eta,\ell} &= -n^{-\gamma}\ell(n)\sum_{j=1}^{n} w_\ell(j,n)y_i^\eta(j) \\ &\to \int_0^1 h_\ell''(t)B_H^{(i)}(t)\,\mathrm{d}t - h'(1)B_H^{(i)}(1) \overset{law}{=} -\int_0^1 h_\ell'(s)\,\mathrm{d}B_H^{(i)}(s)\,. \end{aligned} \tag{4.22}$$

The second part of Lemma 4.4 implies that the term $d_{\Delta r,\ell}$ does not contribute to the limit under the assumptions of Theorem 3.7. This, and the convergences (4.19), (4.20), and (4.22), concludes the proof of Theorem 3.7. ■

LEMMA 4.4. *Under the assumptions of Theorem 3.6, then* $d_{\Delta r,\ell} = O_P(1)$. *Under the assumptions of Theorem 3.7, then* $d_{\Delta r,\ell} = O_P(n^{-1/2})$.

Proof. Applying (4.18) to r, we see that we only need to prove that the independence between the counting processes and the efficient shocks and (3.3)

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

implies that $\sum_{j=1}^{n} w_\ell(j,n) r_{i,j} = O_p(n)$ and (3.5) implies that $\sum_{j=1}^{n} w_\ell(j,n) r_{i,j} = O_p(n^{1/2})$. We start with r_1.

$$\sum_{j=1}^{n} w_\ell(j,n) r_{1,j} = \sum_{k=1}^{N_1(n)} e_{1,k} \sum_{t_{1,k} \le j < t_{1,k} + A_2(t_{1,k})} w_\ell(j,n) \, .$$

Taking conditional expectation yields, for $q = 1, 2$,

$$\mathbb{E}\left[\left| \sum_{j=1}^{n} w_\ell(j,n) r_{1,j} \right|^q \mid \mathcal{N} \right] \le C \sum_{k=1}^{N_1(n)} (A_2(t_{1,k}) + 1)^q \, .$$

Applying (3.3) if $q = 1$ and (3.5) if $q = 2$ shows that the last term is $O_P(n)$. This proves that $\sum_{j=1}^{n} w_\ell(j,n) r_{1,j} = O_P(n)$ under the assumptions of Theorem 3.6 and $O_P(\sqrt{n})$ under the assumptions of Theorem 3.7. The term $\sum_{j=1}^{n} w_\ell(j,n) r_{2,j}$ is dealt with similarly. ■

4.4. Proof of Theorem 3.8

Write

$$\tilde{\theta}_{n,\delta} = \theta + \frac{\sum_{\ell=1}^{m} d_{\Delta\tilde{r},\ell} \bar{d}_{\Delta\tilde{y}_2,\ell}}{\sum_{\ell=1}^{m} |d_{\Delta\tilde{y}_2,\ell}|^2} + \frac{\sum_{\ell=1}^{m} d_{\Delta\tilde{y}^\eta,\ell} \bar{d}_{\Delta\tilde{y}_2,\ell}}{\sum_{\ell=1}^{m} |d_{\Delta\tilde{y}_2,\ell}|^2}$$

with $\tilde{y}(j) = \tilde{y}_1^\eta(j) - \theta \tilde{y}_2^\eta(j)$, $\tilde{r}(j) = \tilde{r}_1(j) - \theta \tilde{r}_2(j)$ and

$$r_1(s) = \sum_{k=N_1(t_{2,N_2(s)})+1}^{N_1(s)} e_{1,k} \, , \quad r_2(s) = \sum_{k=N_2(t_{1,N_1(s)})+1}^{N_2(s)} e_{2,k} \, .$$

and the DFT is defined as in (3.7). Applying summation by parts as in (4.18), we obtain

$$\begin{aligned} d_{\Delta\tilde{y}_2,\ell} &= -\frac{1}{n} \sum_{j=0}^{n-1} w_\ell(j,n) \tilde{y}_2(j) = -\frac{1}{n} \int_0^{n\delta} w_\ell(\lceil s/\delta \rceil, n) y_2(s) \, \mathrm{d}s \\ &= -\int_0^{\delta} w_\ell(\lceil nt/\delta \rceil, n) y_2(ns) \, \mathrm{d}t \, , \end{aligned}$$

with $w_\ell(j,n) = n\{h_\ell((j+1)/n) - h_\ell(j/n)\}$ as before, and $\lceil t \rceil$ is the smallest integer larger than or equal to t. This yields

$$n^{-1/2} d_{\tilde{y}_2,\ell} \to -\Sigma_e \int_0^1 h_\ell'(s) B(\delta s) \, \mathrm{d}s \overset{law}{=} \Sigma_e \int_0^1 h_\ell(s) \, \mathrm{d}B(s) \, .$$

Since $\eta_j = \xi_j - \xi_{j-1}$, we have

$$\tilde{y}_i^\eta(j) = \int_{(j-1)\delta}^{j\delta} \xi_{i,N_i(s)} \, \mathrm{d}s - \delta \xi_{i,0} \, .$$

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

Differencing cancels the term $\delta\xi_0$. Applying (4.18) and summation by parts and the property that $\sum_{j=0}^{n-1} w_\ell(j,n) = 0$, we obtain

$$\begin{aligned} d_{\Delta\tilde{y}_i^\eta,\ell} &= -\frac{1}{n}\sum_{j=0}^{n-1} w_\ell(j,n)\int_{(j-1)\delta}^{j\delta} \xi_{i,N_i(s)}\,\mathrm{d}s \\ &= -\frac{1}{n}\sum_{j=0}^{n-1} w_\ell(j,n)\int_{(j-1)\delta}^{j\delta} \left\{\xi_{i,N_i(s)} - \mu_i^*\right\}\mathrm{d}s \\ &= \frac{1}{n^2}\sum_{j=1}^{n-1} \omega_\ell(j,n)\int_0^{j\delta} \left\{\xi_{i,N_i(s)} - \mu_i^*\right\}\mathrm{d}s \\ &\quad -\frac{1}{n} w_\ell(n,n)\int_0^{(n-1)\delta} \left\{\xi_{i,N_i(s)} - \mu_i^*\right\}\mathrm{d}s\,. \end{aligned}$$

Under Assumption 2.8, we thus have, with $Z = B_H^{(1)} - \theta B_H^{(2)}$,

$$n^{1-H}\left\{d_{\Delta\tilde{y}_1^\eta,\ell} - \theta d_{\Delta\tilde{y}_2^\eta,\ell}\right\} \to \int_0^1 h_\ell''(s) Z(\delta s)\,\mathrm{d}s - h'(1)Z(\delta)\,.$$

We must now deal with the remaining terms of the cointegrating error. If $H > 1/2$, Lemma 4.5 implies that the term $d_{\Delta\tilde{r},\ell}$ does not contribute to the limit. If $H = 1/2$, both terms are of the same order. This concludes the proof of Theorem 2.8. ■

LEMMA 4.5. *Under the assumptions of Theorem 3.8,*

$$d_{\Delta\tilde{r}_i,\ell} = O_P\left(n^{-1/2}\right)\,.$$

Proof. Applying as usual summation by parts, we obtain

$$\begin{aligned} d_{\Delta\tilde{r}_1,\ell} &= -\frac{1}{n}\sum_{k=1}^{N_1(n\delta)} e_{1,k}\sum_{j=1}^{n} w_\ell(j,n)\int_{(j-1)\delta}^{j\delta} \mathbf{1}_{\{t_{1,k}\le s < t_{1,k}+A_2(t_{1,k})\}}\,\mathrm{d}s \\ &= -\frac{1}{n}\sum_{k=1}^{N_1(n\delta)} e_{1,k}\int_0^{n\delta} w_\ell(\lceil s/\delta\rceil, n)\mathbf{1}_{\{t_{1,k}\le s < t_{1,k}+A_2(t_{1,k})\}}\,\mathrm{d}s \\ &= -\frac{1}{n}\sum_{k=1}^{N_1(n\delta)} e_{1,k}\int_{t_{1,k}\wedge(n\delta)}^{\{t_{1,k}+A_2(t_{1,k})\}\wedge(n\delta)} w_\ell(\lceil s/\delta\rceil, n)\,\mathrm{d}s \\ &= -\sum_{k=1}^{N_1(n\delta)} e_{1,k}\int_{(t_{1,k}/n)\wedge\delta}^{\{(t_{1,k}+A_2(t_{1,k}))/n\}\wedge\delta} w_\ell(\lceil nt/\delta\rceil, n)\,\mathrm{d}t\,. \end{aligned} \tag{4.23}$$

Taking conditional expectation and applying (3.5), we obtain

$$\mathbb{E}\left[|d_{\Delta\tilde{r}_1,\ell}|^2 \mid \mathcal{N}\right] \le \frac{C}{n^2}\sum_{k=1}^{N(n)} A_2^2\left(t_{1,k}\right) = O_P\left(n^{-1}\right)\,.$$

■

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

REFERENCES

Arcones, M.A. (1994) Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *The Annals of Probability* 22(4), 2242–2274.

Baccelli, F. & P. Brémaud (2003) Elements of queueing theory, vol. 26 of *Applications of Mathematics (New York)*, 2nd ed. Springer-Verlag.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, & N. Shephard (2008) Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* 76(6), 1481–1536.

Bauwens, L. & D. Veredas (2004) The stochastic conditional duration model: A latent variable model for the analysis of financial durations. *Journal of Econometrics* 119(2), 381–412.

Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley.

Carrasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18(1), 17–39.

Chen, W.W. & C.M. Hurvich (2003a) Estimating fractional cointegration in the presence of polynomial trends. *Journal of Econometrics* 117(1), 95–121.

Chen, W.W. & C.M. Hurvich (2003b) Semiparametric estimation of multivariate fractional cointegration. *Journal of the American Statistical Association* 98(463), 629–642.

Daley, D.J. & D. Vere-Jones (2003) *An Introduction to the Theory of Point Processes, vol. I: Elementary Theory and Methods*, 2nd ed. Probability and Its Applications. Springer.

Deo, R., M. Hsieh, & C.M. Hurvich (2010) Long memory in intertrade durations, counts and realized volatility of NYSE stocks. *Journal of Statistical Planning and Inference* Special Issue in Honor of Emanuel Parzen on the Occasion of his 80th Birthday and Retirement from the Department of Statistics, Texas A&M University - Emmanuel Parzen. 140(12), 3715–3733.

Deo, R., C. Hurvich, & Y. Lu (2006) Forecasting realized volatility using a long-memory stochastic volatility model: Estimation, prediction and seasonal adjustment. *Journal of Econometrics* 131(1-2), 29–58.

Deo, R., C.M. Hurvich, P. Soulier, & Y. Wang (2009) Conditions for the propagation of memory parameter from durations to counts and realized volatility. *Econometric Theory* 25(3), 764–792.

Engle, R.F. & J.R. Russell (1998) Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* 66(5), 1127–1162.

Ghysels, E., P. Santa-Clara, & R. Valkanov (2006) Predicting volatility: Getting the most out of return data sampled at different frequencies. *Journal of Econometrics* 131(1-2), 59–95.

Giraitis, L., P.M. Robinson, & D. Surgailis (1999) Variance-type estimation of long memory. *Stochastic Processes and their Applications* 80(1), 1–24.

Hamilton, J.D. (1994) *Time Series Analysis*. Princeton University Press.

Hurvich, C.M. & Y. Wang (2009) A pure-jump transaction-level price model yielding cointegration, leverage, and nonsynchronous trading effects. Working paper, New York University.

Hurvich, C.M. & Y. Wang (2010) A pure-jump transaction-level price model yielding cointegration. *Journal of Business & Economic Statistics* 28(4), 539–558.

Ibragimov, I.A. & Y.A. Rozanov (1978) *Gaussian Random Processes*, vol. 9 of *Applications of Mathematics*. Springer-Verlag. Translated from the Russian by A.B. Aries.

Marinucci, D. & P.M. Robinson (2000) Weak convergence of multivariate fractional processes. *Stochastic Processes and Their Applications* 86(1), 103–120.

Prigent, J.-L. (2001) Option pricing with a general marked point process. *Mathematics of Operations Research* 26(1), 50–66.

Prigent, J.-L., O. Renault, & O. Scaillet (2002) An autoregressive conditional binomial option pricing model. In *Mathematical finance—Bachelier Congress, 2000 (Paris)*, pp. 353–373. Springer.

Prigent, J.-L., O. Renault, & O. Scaillet (2004) Option pricing with discrete rebalancing. *Journal of Empirical Finance* 11(1), 133–161.

Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes*, vol. 4 of *Applied Probability. A Series of the Applied Probability Trust*. Springer-Verlag.

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press

Resnick, S.I. (2007) *Heavy-Tail Phenomena*. Springer Series in Operations Research and Financial Engineering. Springer.

Robinson, P.M. & D. Marinucci (2001) Narrow-band analysis of nonstationary processes. *The Annals of Statistics* 29(4), 947–986.

Vervaat, W. (1972) Functional central limit theorems for processes with positive drift and their inverses. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 23, 245–253.

Whitt, W. (2002) *Stochastic-Process Limits*. Springer Series in Operations Research. Springer-Verlag.

https://doi.org/10.1017/S0266466613000406 Published online by Cambridge University Press