Imperial College London

COURSEWORK

IMPERIAL COLLEGE LONDON

DEPARTMENT OF NATURAL SCIENCES

Market Microstructure - Lead Lag Estimation and Alternative Estimators

Author:
Thomas Pink (CID: 06003339)

Date: June 8, 2025

1 Introduction

In this report we attempt to investigate the original 'Toy' lead lag estimator as proposed in a 2013 paper "Estimation of the lead-lag parameter from non-synchronous data" by Hoffman, Rosenbaum & Yoshida (2013) [3] by quantifying the probability of it misidentifying the lag under the 'Toy Bachelier Model',and compare it to an alternative rolling estimator based on a simple extension of the original estimator. We then show that the rolling estimator can be recovered by applying the Hayashi Yoshida (2005)[2] style covariation estimation to general shifts of the original contrast function, rather than just integer values, showing that the alternative does not provide any additional predictive power. Finally, using the lessons of the rolling estimator, we show a small improvement over a Hayashi Yoshida style estimator where the intervals at which we obtain the covariation are discrete.

2 Basics of the Toy Bachelier Model

2.1 Model Definition and Lag Estimator

The Hoffman, Rosenbaum, & Yoshida (2013) Toy Bachelier model [3] assumes that stocks can be modeled as Bachelier processes over short time scales, with some underlying correlation ρ subject to a fixed time delay. Adjusting the notation somewhat from the original paper, we write the two processes as:

$$X_t = x_0 + \sigma_1 W_t, \qquad Y_t = y_0 + \sigma_2 \rho W_{t-\theta} + \sigma_2 \sqrt{(1-\rho^2)} B_t \tag{1}$$

With W_t and B_t Brownian motions, and $\theta > 0$ the lag parameter. With simultaneous observations at $\ldots - \Delta, 0, \Delta, \ldots$ etc

The original Toy lead lag estimator's contrast function, $c(k)$[3] can be computed as follows:

$$c(k) = \sum_{i=1}^{n} (X_{i\Delta} - X_{(i-1)\Delta})(Y_{(i+k)\Delta} - Y_{(i+k-1)\Delta}) \tag{2}$$

With the estimated lag taken to be Δ times the argmax of the absolute value of the contrast function.

3 Alternative Estimator

3.1 Motivation

As seen in Figure 1, the original contrast function can have two peaks around the lag parameter. Therefore it was proposed that an alternative estimator could be obtained by adding two consecutive values together to combine the peaks. The idea

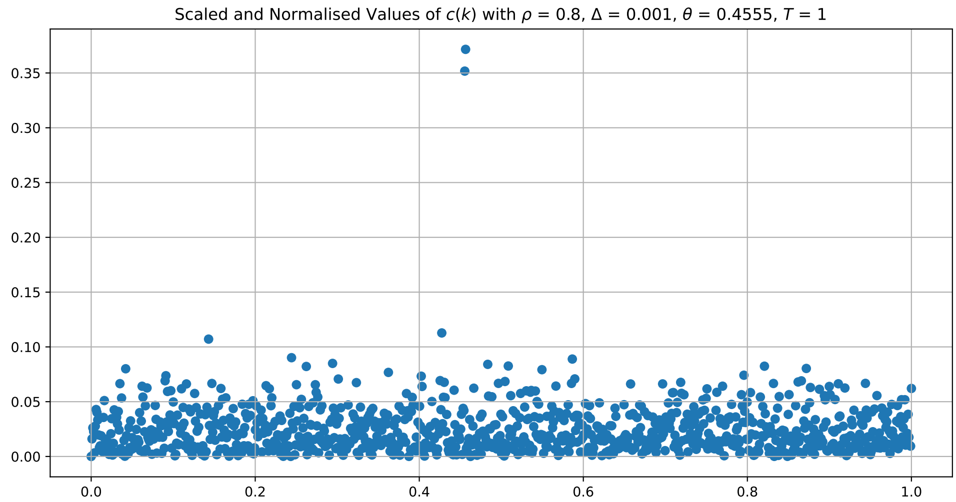


Figure 1: Example of Toy Model Lead Lag Contrast Function on Simulated Data

being that the higher mean outweighs the increased variance of the points that are not around the lag.

More formally we define the alternative 'rolling' contrast function $r(k)$, to be

$$r(k) = c(k) + c(k+1) \tag{3}$$

We obtain an equivalent form by substitution of the original contrast function $c(k)$

$$r(k) = \sum_{i=1}^{n} (X_{i\Delta} - X_{(i-1)\Delta})(Y_{(i+k+1)\Delta} - Y_{(i+k-1)\Delta}) \tag{4}$$

The estimated lag is computed as Δ times the argmax of the absolute value of $r(k)$.

To shorten the notation and for ease of reading throughout this report we write $W_x - W_y = W_y^x$ for the brownian motion increment of W_t.

We desire to characterise the distribution of $c(k)$ for all the values of k. Therefore we substitute in the Equations 1 into equation 2 for $c(k)$, to obtain

$$c(k) = \sum_{i=1}^{n} \rho\sigma_1\sigma_2 W_{(i-1)\Delta}^{i\Delta} W_{(i+k-1)\Delta-\theta}^{(i+k)\Delta-\theta} + \sum_{i=1}^{n} \sqrt{1-\rho^2}\sigma_1\sigma_2 W_{(i-1)\Delta}^{i\Delta} B_{(i+k-1)\Delta}^{(i+k)\Delta} \tag{5}$$

Each brownian increment term within the sums are centred and independent of the term it is multiplied by, except when $[(i+k-1)\Delta-\theta, (i+k)\Delta-\theta]$ has overlap with

$[(i-1)\Delta, i\Delta]$ in the first sum of the product of increments of W_t. This is when

$$\frac{\theta}{\Delta} - 1 < k < \frac{\theta}{\Delta} + 1 \tag{6}$$

There are two values of k (except when θ is an integer multiple of Δ) where overlap occurs. The sum of overlaps over the two possible values of k is Δ. We denote the first overlap between the two increments occurring at $k = k^*$ to be a proportion $\delta \in [0,1]$ of the overall length Δ. The overlap at $k^* + 1$ is therefore $1 - \delta$. In the rolling estimator case, from the definition of $r(k)$ in equation 3 it's clear overlap is experienced at $k^* - 1, k^*$ and $k^* + 1$, of proportions $\delta, 1$ and $1 - \delta$ respectively.

4 Distributions of Estimators

4.1 Central Limit Theorem and Covariances

As seen in equation 2 we have sums of products of identically distributed normal random variables. However they are not independent, so we cannot apply the standard Central Limit Theorem. We therefore state a theorem from Probability and Measure by Billingsley (1995)[1].

4.1.1 CLT Under Strong Mixing (Billingsley 1995)

Suppose that $\{X_1, \ldots, X_n, \ldots\}$ is stationary and α-mixing with $\alpha_n = O\left(n^{-5}\right)$ and that $\mathrm{E}[X_n] = 0$ and $\mathrm{E}[X_n^{12}] < \infty$ Denote $S_n = X_1 + \cdots + X_n$ then the limit

$$\sigma^2 = \lim_{n\to\infty} \frac{\mathrm{E}\left(S_n^2\right)}{n} \tag{7}$$

exists, and if $\sigma \neq 0$ then $\frac{S_n}{\sigma\sqrt{n}}$ converges in distribution to $\mathcal{N}(0,1)$

4.1.2 Application of Theorem

Observing the form of each term in the sum:

$$c(k) = \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \left(\sigma_1 W_{(i-1)\Delta}^{i\Delta} \sigma_2 (\rho W_{(i-1+k)\Delta-\theta}^{(i+k)\Delta-\theta} + \sqrt{(1-\rho^2)} B_{(i-1+k)\Delta-\theta}^{(i+k)\Delta-\theta}) \right) \tag{8}$$

It's clear that the X_i are stationary. For the second condition, if $k\Delta - \theta > 0$ then if $j\Delta > (i+k)\Delta - \theta$ then X_i and X_j are independent. Similarly if $k\Delta - \theta \leq 0$ we require $(j-1+k)\Delta - \theta \geq i\Delta$, to obtain X_i and X_j are independent. Therefore we have $\alpha_i = 0$ for sufficiently large i, and we have the necessary α-mixing.

Therefore the conditions of Theorem 4.1.1 are satisfied so long as we centre the X_i and we are justified in applying the Central Limit Theorem to approximate the individual distributions of $c(k)$ at each k for sufficiently large n.

There may also be some concern that for small i the covariance between terms may be large, causing convergence to be slow. However, we can show (with calculations in the appendix 9.0.1) that the covariance is zero at points without overlap.

4.2 Limiting Distributions of $c(k)$ and $r(k)$

We now characterise distributions of the contrast functions, with supporting calculations of the means and variances in appendix 9.0.2. After applying Theorem 4.1.1 we find for $c(k)$:

$$\frac{c(k)}{\sqrt{n}\Delta\sigma_1\sigma_2} \simeq \begin{cases} \mathcal{N}(\sqrt{n}\delta\rho, 1+\delta^2\rho^2), & \text{if } k = k^*, \\ \mathcal{N}(\sqrt{n}(1-\delta)\rho, 1+(1-\delta^2)\rho^2), & \text{if } k = k^*+1 \\ \mathcal{N}(0,1), & \text{if } k \neq k^*, k^*+1 \end{cases}$$

And for $r(k)$ we obtain:

$$\frac{r(k)}{\sqrt{2n}\Delta\sigma_1\sigma_2} \simeq \begin{cases} \mathcal{N}\left(\sqrt{\frac{n}{2}}\delta\rho, 1+\frac{\delta^2\rho^2}{2}\right), & \text{if } k = k^*-1, \\ \mathcal{N}(\sqrt{\frac{n}{2}}\rho, 1+\frac{\rho^2}{2}), & \text{if } k = k^* \\ \mathcal{N}\left(\sqrt{\frac{n}{2}}(1-\delta)\rho, 1+\frac{(1-\delta)^2\rho^2}{2}\right), & \text{if } k = k^*+1, \\ \mathcal{N}(0,1), & \text{if } k \neq k^*-1, k^*, k^*+1 \end{cases}$$

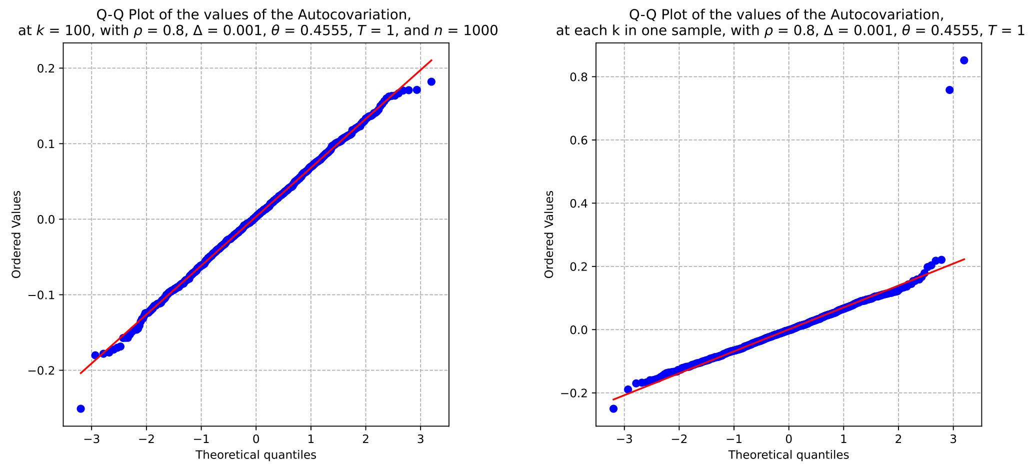
4.2.1 Empirical Evaluation of Convergence in Distribution

Convergence in distribution is supported by Figure 2(a), a QQ plot of the values of $c(100)$ over 1000 samples, closely following the quantiles of a normal distribution.

Figure 2(b) shows that the entries for each k (excluding the cases with overlap at k^* and k^*+1) within a single simulation also appear to follow the quantiles of an independent normal distribution.

5 Accuracy of Estimators

To compare the accuracy of the predictions of the two estimators we compute the probability of each estimator having an error. For the alternative rolling estimator, we consider three points. k^*, k^*+1 and k^*-1 where overlap will be observed in $r(k)$. For the original estimator, we therefore consider the same values for fair comparison. If the argmax of the contrast function evaluates to these points, we consider the



((a)) QQ Plot of a value of $c(k)$ for fixed k over 1000 samples

((b)) QQ Plot of $c(k)$ for k over a single sample

Figure 2: QQ Plots

estimator to have been 'corrrect'. Therefore we define the probability of error as:

$$\mathbb{P}\left(\max_{k \notin \{k^*-1, k^*, k^*+1\}} |c(k)| > \max_{k \in \{k^*, k^*+1, k^*+2\}} |c(k)|\right) \tag{9}$$

Where $c(k)$ can be replaced with $r(k)$ in the case of the rolling estimator. In either case it is clear we can normalise the estimator to recover the approximate normal distributions as done in section 4. For example:

$$\mathbb{P}\left(\max_{k \notin \{k^*, k^*+1, k^*+2\}} \left(\frac{|c(k)|}{\sqrt{n}\Delta\sigma_1\sigma_2}\right) > \max_{k \in \{k^*-1, k^*, k^*+1\}} \left(\frac{|c(k)|}{\sqrt{n}\Delta\sigma_1\sigma_2}\right)\right) \tag{10}$$

We will compute an integral expression of the approximate probability of each estimator's error. For simplicity we assume $\rho > 0$. We are therefore interested in the distributions with the higher means having a greater maximum absolute value, compared to the maximum absolute value of a large sample of centrered distributions. For large enough choices of n, the means of the non-centered distributions are significantly larger than those of the centred distributions. We therefore drop the absolute value on the right hand side of the inequality in equation 10, since the probability of the variable being negative and larger in absolute value than the m centred variables is negligible, given they have comparable variances.

Furthermore, the centred normal distributions will be negative approximately half the time. Therefore we double the number of centred samples and drop the absolute value in the left hand side of equation 10, which if the number of samples is large enough, will have approximately the same distribution.

Under our assumptions we obtain the following approximate expression of Equation

10:

$$\mathbb{P}\left(\max_{i\in\{1,\cdots,2m\}}(Z_i) > \max_{i\in\{2m+1,2m+2,2m+3\}}(\sigma_i Z_i + \mu_i)\right) \tag{11}$$

Where the Z_i are $i.i.d.$ $\mathcal{N}(0,1)$, and μ_i, σ_i correspond to the values of $c(k)$ and $r(k)$ at $\{k^*, k^*+1, k^*+2\}$ respectively. Relabelling $\mu_{2m+i} \to \mu_i$ etc, we derive the following integral expression of the error probability, with details included in the appendix 9.0.4

$$\int_{-\infty}^{\infty} 2m\phi(x)\Phi(x)^{2m-1}\Phi(\frac{x-\mu_1}{\sigma_1})\Phi(\frac{x-\mu_2}{\sigma_2})\Phi(\frac{x-\mu_3}{\sigma_3})dx \tag{12}$$

5.1 Results

We now compare the two estimators using our integral expression and empirical tests on simulated data. Firstly we detail the case $\rho = 0.2, \Delta = 0.001, \theta = 0.4555, T = 1$ over 1000 samples. θ has been chosen specifically so that $\delta = 0.5$.

	Original Estimator	Rolling Estimator
$\hat{\theta} = 0.454$	0	10
$\hat{\theta} = 0.455$	335	847
$\hat{\theta} = 0.456$	316	5
Total	651	862

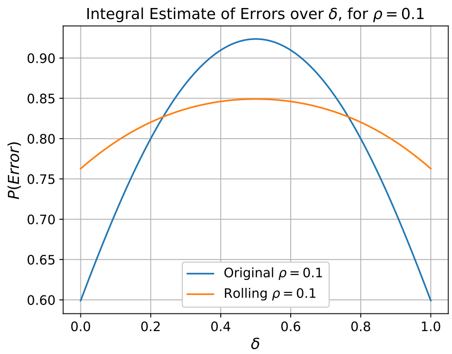
Table 1: Detailed Results and Integral Expressions

Table 1 shows in this case the results are improved under the rolling estimator.

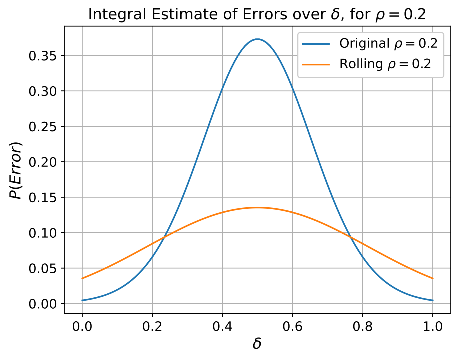
We now simulate results for various different correlations ρ, lag θ and resulting overlaps δ, in each case $\Delta = 0.001$, $T = 1$ over 1000 samples.

Parameters	Original Estimator	Original Integral Estimation	Rolling Estimator	Rolling Integral Estimation
$\theta = 0.4555, \rho = 0.3, \delta = 0.5$	981	0.982	999	0.999
$\theta = 0.4555, \rho = 0.2, \delta = 0.5$	651	0.627	862	0.865
$\theta = 0.4555, \rho = 0.1, \delta = 0.5$	76	0.076	147	0.151
$\theta = 0.4553, \rho = 0.3, \delta = 0.3$	999	0.998	1000	1.000
$\theta = 0.4553, \rho = 0.2, \delta = 0.3$	845	0.833	851	0.890
$\theta = 0.4553, \rho = 0.1, \delta = 0.3$	107	0.132	141	0.163
$\theta = 0.4550, \rho = 0.3, \delta = 0$	1000	1.000	1000	1.000
$\theta = 0.4550, \rho = 0.2, \delta = 0$	998	0.996	944	0.965
$\theta = 0.4550, \rho = 0.1, \delta = 0$	391	0.401	215	0.237

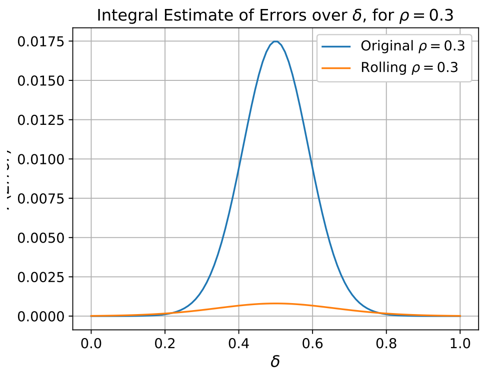
Table 2: Number of times estimator was within k^*-1, k^*, k^*+1 and computed probabilities (3 d.p.)



((a)) Integral Estimate of Error for $\rho = 0.1$



((b)) Integral Estimate of Error for $\rho = 0.2$



((c)) Integral Estimate of Error for $\rho = 0.3$

Figure 3: Error Probability

Table 2 shows our estimated integral captures the probabilities fairly accurately, supporting the assumptions and approximations we made in deriving the expression. The results show that as overlap decreases from 0.5 to 0, the outperformance of the rolling estimator disappears. In the $\delta = 0.5$ case, at $k^*, k^* + 1$ the original estimator has overlap of $\delta = 0.5$ between the two processes, whereas the alternative rolling estimator has overlaps of $\delta = 0.5$ at $k^* - 1, k^* + 1$ and of $\delta = 1$ at k^*. The increased mean at the overlaps compensates for the increased variance, resulting in improved performance. However when $\delta = 0$, the overlap at $k^* + 1$ is 1 in the original estimator, and for the rolling estimator the overlap is 1 at $k^*, k^* + 1$, and 0 otherwise.

Figure 3 shows this more clearly. The starkest difference is in the $\rho = 0.3$ case, where the estimated probability of error for the rolling estimator is much closer to zero on average than the original estimator. Even in the case where $\rho = 0.1$, the maximum error probability across the different overlaps is lower under the rolling estimator.

The key benefit of the rolling estimator, is that it can capture the full overlap between the two correlated parts of X_t and Y_t, W_t and $W_{t-\theta}$, giving a larger signal at the expense of greater variance. This tradeoff is positive when the overlap is close to 0.5, but not when the overlap is close to 0 or 1, with the rolling first outperforming from below $\delta = 0.25$ for each choice of ρ.

6 Insights of the Toy Model

We now attempt to use these insights into possible ways to improve upon the lead-lag estimator for asynchronous data by Hoffman, Rosenbaum and Yoshida (2013) [3]. We therefore begin by applying the concept to the Toy Model.

6.1 Recovery of Rolling Estimator

In the Toy Bachelier Model, if we allow shifts by general $k \in \mathbb{R}$ for $c(k)$ rather than simply integer values, observations of the two processes may no longer be synchronous. Therefore in an asynchronous case, we modify the original definition of $c(k)$ by applying a Hayashi Yoshida style correction of the contrast function [2].

$$c(k) = \sum_{i,j}^{n} X_{t_i}^{t_{i+1}} Y_{s_{j+k\Delta}}^{s_{j+1+k\Delta}} 1_{([t_i,t_{i+1}]\cap[s_{j+k\Delta},s_{j+1+k\Delta}]\neq\emptyset)} \tag{13}$$

Where t_i and s_j are the observations of X_t and Y_t respectively.

We can interpret the Hayashi Yoshida estimator as follows. For each interval $[t_i, t_{i+1}]$ of consecutive observations of X_t, we find the first observation of $Y_{t+k\Delta}$ that precedes t_i, s_l say and the observation that follows t_{i+1}, s_u. We then take the sum of the product of each of the difference of these observations: $X_{t_i}^{t_{i+1}} Y_{s_l}^{s_u}$

In this case where the t_i and s_j are respectively spaced by Δ. With consecutive observations at $n\Delta, (n+1)\Delta$ of X_t, and k some non-integer value, we make consecutive observations of Y_t at $(m-1+k)\Delta, (m+k)\Delta, (m+1+k)\Delta$ with $m-1+k < n < m+k < n+1 < m+k+1$. We add the term $X_{n\Delta}^{(n+1)\Delta} Y_{(m-1+k)\Delta}^{(m+1+k)\Delta}$ to the sum. We obtain an interval of length 2Δ between observations of Y_t, multiplied against an interval of Δ between observations of X_t in each term of the sum. Therefore in some sense the Hayashi Yoshida correction recovers the rolling estimator.

7 Discrete Alternative Estimator

Returning to the continuous case of data arriving at random times, if we choose $k = \frac{\theta}{\Delta}$, and then apply the Hayashi Yoshida style estimator, we already ensure complete overlap without needing to extend the second interval. Therefore the rolling estimator does not appear to provide insight for improvements in this case. However in practice, the k at which the estimator are calculated are not continuous, but discrete, which motivates an adjustment in the discrete case to the Hayashi Yoshida style estimator proposed by Hoffman, Rosenbaum and Hayashi (2013)[3].

We now denote the average separation between observations in the first process X_t as Δ, and compute $c(k)$ at discrete k with separations of Δ then the rolling estimator's insights suggest extending the upper range between observations in Y_t to be larger than those suggested by the Hayashi Yoshida style Estimator [3]. Specifically, as observation times of X_t are random, the overlap at a random displacement for discrete choices of k may not be complete. Therefore we can ensure that we will capture a larger overlap by extending the upper bound of the interval of observations of Y_t.

7.1 Discrete Estimator Definitions

For observations t_i of X_t, s_j of Y_t we define the Hayashi Yoshida Estimator for observations at $k \in \Delta\mathbb{Z}$ to be the Δ times the argmax over k of the absolute value of the contrast function $c(k)$, with $c(k)$ as defined in Equation 13.

We define the alternative estimator's contrast function $r(k)$ as an adjustment to the Hayashi Yoshida Estimator as follows: We first try to get as close from below to $t_{i+1} + \Delta$ with s_u, otherwise applying the original style estimator choice of s_u. A more detailed pseudocode for the algorithm for the alternative estimator's contrast function $r(k)$ is included in the appendix.

7.2 Generation of Random Observations

Using the underlying price driven model of Robert and Rosenbaum (2011) [4] of Ultra High-Frequency Style Sensitivity zones, we can generate random observations of prices, based on an underlying driving 'true' stock price. We model our underlying driving stock prices as coupled Bachelier processes as in Equations 1. We take these observations as observations of the underlying Brownian driving motions directly, rather than attempt to apply an estimator to recover them.

We begin by generating processes X_t and Y_t as in the Toy Bachelier Model 1, with $\rho = 0.3$ The observations of X_t are generated using a Bachelier stock process driven by the Brownian Motion X_t with spot at time 0 $S_0 = 100$, volatility $\sigma = 0.5$, stock tick size of 10 and sensitivity zone size $\eta = 0.3$. The observations of Y_t are generated in the same way with Y_t driving the motion, except $\eta = 0.2$ and the stock tick size of 1, resulting in more frequent observations of Y_t. An sample of resultant observations is shown in Figure 5 in the appendix 9.1.

7.3 Testing On Data

An example of the results of the two algorithms are shown for $k \in \Delta\mathbb{Z}^+$ in Figure 4. In this case the Hayashi Yoshida style estimator fails to identify the lag $\theta = 0.5$ correctly, whereas the alternative estimator succeeds.

Writing the predicted lag $\hat{\theta}$, we run the same test over 1000 simulations, we obtain Table 3 of results where $|\hat{\theta} - \theta| < 0.05$

The alternative estimator is correct around 5% more often in these simulations, suggesting it is slightly more robust in this discrete case.

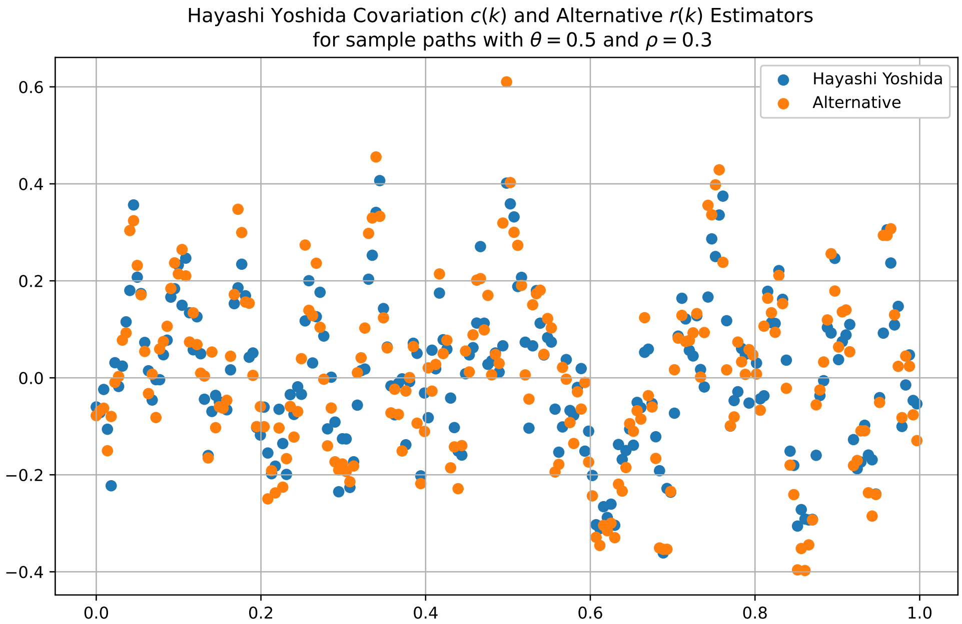


Figure 4: Example of results where the Alternative Estimator Correctly Estimates the Lag and the Original Estimator Fails

Hayashi Yoshida Style Estimator	Alternative Estimator	Total Number of Samples
310	362	1000

Table 3: Number of times the predicted lag $\hat{\theta}$ was within 0.05 of the true value $\theta = 0.5$

8 Conclusion

Under the Toy Bachelier Model [3] we proposed a rolling estimator with an extended interval for observations of the shifted process. The alternative outperforms the original estimator where the 'overlap' δ between the correlated part of the processes is sufficiently close to 0.5. Our analysis showed that this resulted from the rolling estimator capturing a full overlap, increasing the mean of contrast function at the expense of greater variances. If non-integer shifts are permitted in the Toy Model, we show using a Hayashi Yoshida style correction [2] we recover a rolling estimator.

We argue that this does not provide any direct insights on improving the original estimator proposed in the asynchronous random observation times case in the paper by Hoffman, Rosenbaum, Hayashi (2013) [3], but does if we instead restrict ourselves to discrete estimates for the lag, where we proposed an alternative estimator which on simulated data suggested a small improvement.

Extensions of this work could be to investigate the discrete asynchronous case, developing the algorithm with more rigorous arguments, and to identify if the improvement is robust across more varied simulations.

References

[1] P. Billingsley. **Probability and Measure (3^{rd} *ed.*)**. John Wiley & Sons., 1995. ISBN: 0-471-00710-02.

[2] T. Hayashi and N. Yoshida. "On covariance estimation of non-synchronously observed diffusion processes". In: **Bernoulli** 11.2 (2005), pp. 359–379. DOI: `10.3150/bj/1116340299`. URL: `https://doi.org/10.3150/bj/1116340299`.

[3] M. Hoffmann, M. Rosenbaum, and N. Yoshida. "Estimation of the lead-lag parameter from non-synchronous data". In: **Bernoulli** 19.2 (May 2013). ISSN: 1350-7265. DOI: `10.3150/11-bej407`. URL: `http://dx.doi.org/10.3150/11-BEJ407`.

[4] C. Robert and M. Rosenbaum. "A new approach for the dynamics of ultra-high-frequency data: the model with uncertainty zones". In: **Journal of Financial Econometrics** 9.2 (2011), pp. 344–366. DOI: `10.1093/jjfinec/nbq023`. URL: `https://doi.org/10.1093/jjfinec/nbq023`.

9 Appendix

9.0.1 Computing Covariances

There may be some concern that although we have eventual independence, the codependence for X_i and X_j is high when i and j are close. Therefore the convergence to a normal distribution may be quite slow. We therefore compute the covariance between terms in order to assess this.

For $c(k)$ at $k \notin \{k^*, k^*+1\}$, dropping constants σ_1, σ_2, ρ the covariance between any individual terms in the sum of $c(k)$ consists of evaluating sums of terms of the following forms:

$$\mathbb{E}[W^{i\Delta}_{(i-1)\Delta} W^{(i+k)\Delta-\theta}_{(i+k-1)\Delta-\theta} W^{j\Delta}_{(j-1)\Delta} W^{(j+k)\Delta-\theta}_{(j+k-1)\Delta-\theta})] \tag{14}$$

$$\mathbb{E}[W^{i\Delta}_{(i-1)\Delta} B^{(i+k)\Delta}_{(i+k-1)\Delta} W^{j\Delta}_{(j-1)\Delta} W^{(j+k)\Delta-\theta}_{(j+k-1)\Delta-\theta}] \tag{15}$$

$$\mathbb{E}[W^{i\Delta}_{(i-1)\Delta} B^{(i+k)\Delta}_{(i+k-1)\Delta} W^{j\Delta}_{(j-1)\Delta} B^{(j+k)\Delta}_{(j+k-1)\Delta}] \tag{16}$$

If $i \neq j$ then by the original definition in Equations 1 the increment $B^{(i+k)\Delta}_{(i+k-1)\Delta}$ is always independent of the other terms in the product.

Therefore equations (16) and (15) evaluate to 0.

From now on we only need consider equation (14).

Each term is a product of zero mean brownian motion increments. Therefore as long as one of the intervals doesn't overlap with the others, or there is another independent term, the brownian motion increment will be independent, and so the overall mean will evaluate to zero.

For $c(k)$ evaluated at values excluding k^*, k^*+1, we know that the first two terms and the last two terms do not overlap with each other. We therefore only need to consider overlap with the last two terms

If the increment $W^{i\Delta}_{(i-1)\Delta}$, overlaps with $W^{j\Delta}_{(j-1)\Delta}$, then $i = j$.

If it overlaps with $W^{(j+k)\Delta-\theta}_{(j+k-1)\Delta-\theta}$, then if the unshifted term $W^{i\Delta}_{(i-1)\Delta}$ was earlier than the shifted term with index i, in which case there can be no overlap, from the shifted term with index i, as it is later than all of the j indexed terms. Similarly if it was later, then the shifted term with index i must be earlier than the index j terms. And so equation 14 also evaluates to 0.

So the dependence indicated by the covariance is low for $k \notin \{k^*, k^*+1\}$, and therefore we hope that convergence to a normal distribution is relatively robust even for smaller values of n. Similar arguments hold in the case of $r(k)$ which is clear from it's expression in equation 4

9.0.2 Calculation of Distribution of $c(k)$

Throughout the below, Z_i are i.i.d. $\mathcal{N}(0,1)$.
In $c(k)$ we have sums of the following form:

$$X_i = \sigma_1 W^{i\Delta}_{(i-1)\Delta} \sigma_2 (\rho W^{(i+k)\Delta-\theta}_{(i-1+k)\Delta-\theta} + \sqrt{(1-\rho^2)} B^{(i+k)\Delta-\theta}_{(i-1+k)\Delta-\theta}) \tag{17}$$

We consider the mean and variance of each X_i. If there is no overlap from the choice of k then the left hand side of the product is fully independent of the right hand side. The right hand side is then the sum of two normal random variables in itself. Therefore the sum simplifies to:

$$\sigma_1 \sqrt{\Delta} Z_1 \sigma_2 \sqrt{\Delta} Z_2 = \sigma_1 \sigma_2 \Delta Z_1 Z_2 \tag{18}$$

Where the Z_i s are independent standard normal distributions. From this we can see the overall term has mean 0 and variance $\Delta^2 \sigma_1^2 \sigma_2^2$.

In the case where we have overlap δ, (w.l.o.g. from below) we can split the terms as follows:

$$\sigma_1(W^{i\Delta}_{(i-1+\delta)\Delta} + W^{(i-1+\delta)\Delta}_{(i-1)\Delta})\sigma_2\left(\rho(W^{(i-1+\delta)\Delta}_{(i-1)\Delta} + W^{(i-1)\Delta}_{(i-2+\delta)\Delta}) + \sqrt{(1-\rho^2)}B^{(i-1+\delta)\Delta}_{(i-2+\delta)\Delta}\right) \tag{19}$$

As before we can simplify the term as:

$$\sigma_1\sigma_2(\sqrt{(1-\delta)\Delta}Z_1 + \sqrt{\delta\Delta}Z_2)\left(\rho\sqrt{\delta\Delta}Z_2 + \rho\sqrt{(1-\delta)\Delta}Z_3 + \sqrt{(1-\rho^2)\Delta}Z_4\right) \tag{20}$$

It's clear than when computing the mean, by expanding out the product the only term that will contribute is $\sigma_1\sigma_2\Delta\rho\sqrt{\delta}Z_2^2$, which has mean $\sigma_1\sigma_2\Delta\rho\sqrt{\delta}$.

We now compute the second moment:

$$\mathbb{E}\left[\sigma_1^2\sigma_2^2\Delta^2\left(\sqrt{(1-\delta)}Z_1 + \sqrt{\delta}Z_2\right)^2\left(\rho\sqrt{\delta}Z_2 + \rho\sqrt{(1-\delta)}Z_3 + \sqrt{(1-\rho^2)}Z_4\right)^2\right] \tag{21}$$

Retaining only the square terms since any non-square contribution will have a Z_i which is centred and independent from the other terms in the product we obtain:

$$= \sigma_1^2\sigma_2^2\Delta^2\mathbb{E}\left[\left((1-\delta)Z_1^2 + \delta Z_2^2\right)\left(\rho^2\delta Z_2^2 + \rho^2(1-\delta)Z_3^2 + (1-\rho^2)Z_4^2\right)\right] \tag{22}$$

Apart from the Z_2^4 term in the expansion, which has mean 3, each of the $Z_i^2Z_j^2, i \neq j$ has mean 1. We therefore obtain:

$$= \sigma_1^2\sigma_2^2\Delta^2\left((1-\delta)\delta\rho^2 + (1-\delta)^2\rho^2 + (1-\delta)(1-\rho^2) + 3\delta^2\rho^2 + \delta(1-\delta)\rho^2 + \delta(1-\rho^2)\right) \tag{23}$$

Subtracting off the mean squared $\sigma_1^2\sigma_2^2\Delta^2\rho^2\delta$ and collecting like terms, we obtain variance of:

$$\sigma_1^2\sigma_2^2\Delta^2(1+\delta^2) \tag{24}$$

For overlap of, at $k = k^* + 1$ $(1-\delta)$ simply substitute $(1-\delta)$ for δ.

Therefore after applying the central limit theorem (and abusing notation), we find that for $c(k)$

$$\frac{c(k^*)}{\sqrt{n}\Delta\sigma_1\sigma_2} \simeq \mathcal{N}(\sqrt{n}\delta\rho, 1 + \delta^2\rho^2) \tag{25}$$

$$\frac{c(k^*+1)}{\sqrt{n}\Delta\sigma_1\sigma_2} \simeq \mathcal{N}(\sqrt{n}(1-\delta)\rho, 1 + (1-\delta)^2\rho^2) \tag{26}$$

$$\frac{c(k)}{\sqrt{n}\Delta\sigma_1\sigma_2} \simeq \mathcal{N}(0,1), k \neq k^*, k^*+1 \tag{27}$$

9.0.3 Calculation of Distribution of $r(k)$

The calculations are analogous in the $r(k)$ case, except due to the interval in the second term being $[(i+k-1)\Delta, (i+k+1)\Delta]$ rather than $[(i+k-1)\Delta, (i+k)\Delta]$ the variances are doubled. Therefore we omit all but the computations in the full overlap case for brevity.

We again have each term in the sum being the following form:

$$\sigma_1 W^{i\Delta}_{(i-1)\Delta}\sigma_2\left(\rho W^{(i+1+k)\Delta-\theta}_{(i-1+k)\Delta-\theta} + \sqrt{(1-\rho^2)}B^{(i+1+k)\Delta-\theta}_{(i-1+k)\Delta-\theta}\right) \tag{28}$$

The mean in the case of full overlap, at k^*, of each term is computed in the same way, and therefore $\sigma_1\sigma_2\Delta\rho$.

We now compute the second moment, again by splitting into independent normal distributions:

$$\sigma_1 W^{i\Delta}_{(i-1)\Delta}\sigma_2\left(\rho W^{(i+\delta)\Delta}_{(i-1)\Delta} + \rho W^{i\Delta}_{(i-1)\Delta} + \rho W^{(i-1)\Delta}_{(i-2+\delta)\Delta} + \sqrt{(1-\rho^2)}B^{(i+\delta)\Delta}_{(i-2+\delta)\Delta}\right) \tag{29}$$

Taking expectations of the second moment:

$$\mathbb{E}\left[\sigma_1^2\sigma_2^2\Delta^2\Big(Z_1\Big)^2\left(\rho\sqrt{(1-\delta)}Z_2 + \rho Z_1 + \rho\sqrt{\delta}Z_3 + \sqrt{(1-\rho^2)}Z_4\right)^2\right] \tag{30}$$

We retain only the square terms by the same argument as before, and therefore obtain:

$$\mathbb{E}\left[\sigma_1^2\sigma_2^2\Delta^2\left(\rho^2(1-\delta)Z_1^2Z_2^2 + \rho^2 Z_1^4 + \rho^2\delta Z_1^2 Z_3^2 + (1-\rho^2)Z_1^2Z_4^2\right)\right] \tag{31}$$

Which again we can evaluate as follows:

$$= \sigma_1^2\sigma_2^2\Delta^2\left(\rho^2(1-\delta) + 3\rho^2 + \rho^2\delta + (1-\rho^2)\right) \tag{32}$$

Subtracting the mean squared, and collecting terms we obtain the variance at $k = k^*$

$$\sigma_1^2\sigma_2^2\Delta^2(1+2\rho^2) \tag{33}$$

Summarising in the case of our rolling estimator, and again applying the central limit theorem, for $r(k)$ we obtain

$$\frac{r(k^*-1)}{\sqrt{2n}\Delta\sigma_1\sigma_2} \simeq \mathcal{N}\left(\sqrt{\frac{n}{2}}\delta\rho, 1 + \frac{\delta^2\rho^2}{2}\right) \tag{34}$$

$$\frac{r(k^*)}{\sqrt{2n}\Delta\sigma_1\sigma_2} \simeq \mathcal{N}\left(\sqrt{\frac{n}{2}}\rho, 1 + \frac{\rho^2}{2}\right) \tag{35}$$

$$\frac{r(k^*+1)}{\sqrt{2n}\Delta\sigma_1\sigma_2} \simeq \mathcal{N}\left(\sqrt{\frac{n}{2}}(1-\delta)\rho, 1+\frac{(1-\delta)^2\rho^2}{2}\right) \tag{36}$$

$$\frac{r(k)}{\sqrt{2n}\Delta\sigma_1\sigma_2} \simeq \mathcal{N}(0,1), k \neq k^*-1, k^*, k^*+1 \tag{37}$$

9.0.4 Integral Expression of Probability

We obtained the following expression for the probability of error.

$$\mathbb{P}\left(\max_{i\in\{1,\cdots,2m\}}(Z_i) > \max_{i\in\{2m+1,2m+2,2m+3\}}(\sigma_i Z_i + \mu_i)\right) \tag{38}$$

Where the Z_i are $i.i.d.$ $\mathcal{N}(0,1)$, and μ_i, σ_i correspond to the values of $c(k)$ and $r(k)$ at $\{k^*, k^*+1, k^*+2\}$ respectively.

As it consists of entirely normal distributions, we compute an integral expression to evaluate numerically.

Writing $X = \max_{i\in\{1,\cdots,2m\}}(Z_i)$, and $Y = \max_{i\in\{2m+1,2m+2,2m+3\}}(\sigma_i Z_i + \mu_i)$ then

$$\mathbb{P}\left(X \leq x\right) = \mathbb{P}\left(\max_{i\in\{1,\cdots,2m\}}(Z_i) \leq x\right) = \mathbb{P}\left(Z_i \leq x, i \in \{1,\cdots,2m\}\right) = \prod_{i=1}^{2m}\mathbb{P}\left(Z_i \leq x\right) = \Phi(x)^{2m} \tag{39}$$

$$\mathbb{P}\left(Y \leq y\right) = \mathbb{P}\left(\max_{i\in\{2m+1,2m+2,2m+3\}}(\sigma_i Z_i + \mu_i) \leq y\right) = \mathbb{P}\left(Z_i \leq \frac{y-\mu_i}{\sigma_i}, i \in \{2m+1, 2m+2, 2m+3\}\right) = \prod_{i=2m+1}^{2m+3}\mathbb{P}\left(Z_i \leq \frac{y-\mu_i}{\sigma_i}\right) = \prod_{i=2m+1}^{2m+3}\Phi(\frac{y-\mu_i}{\sigma_i}) \tag{40}$$

Thus we have obtained the CDFs of the two distributions, and by differentiating, the PDFs.

Now computing $\mathbb{P}(X > Y)$, and using that X and Y are independent, writing F_X, f_X and F_Y, f_Y for the CDFs and PDFs of X and Y we compute:

$$\mathbb{P}\left(X > Y\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{x} f_Y(y)f_X(x)dydx = \int_{-\infty}^{\infty} F_Y(x)f_X(x)dx \tag{41}$$

Substituting back in the CDF and PDF for Y and X, and relabelling $\mu_{2m+i} \to \mu_i, \sigma_{2m+i} \to \sigma_i$ we obtain:

$$\int_{-\infty}^{\infty} 2m\phi(x)\Phi(x)^{2m-1}\Phi(\frac{x-\mu_1}{\sigma_1})\Phi(\frac{x-\mu_2}{\sigma_2})\Phi(\frac{x-\mu_3}{\sigma_3})dx \quad (42)$$

9.1 Asynchronous Observations of Brownian Motions

An example of the processes generated by the method described in secction 7.1 is included below in Figure 5

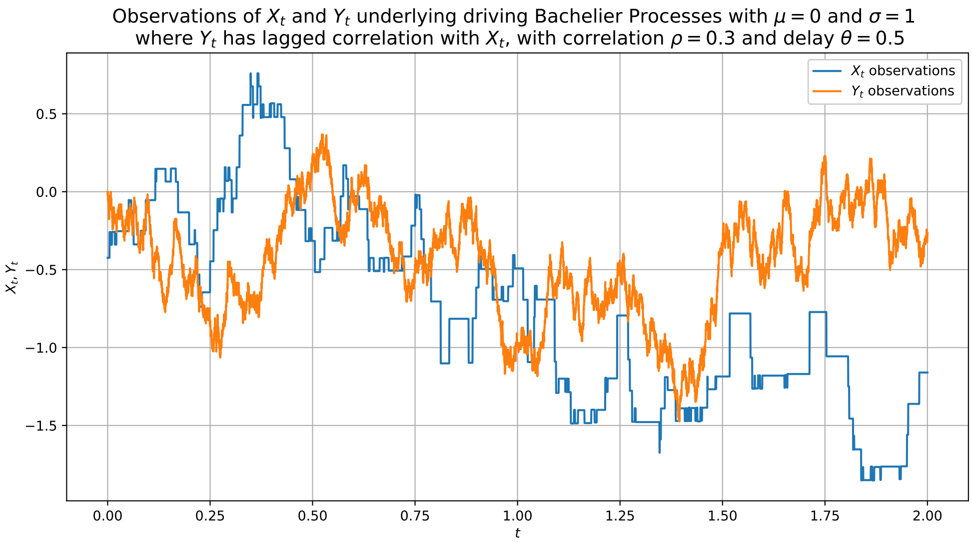


Figure 5: Observations of Lead Lag Bachelier Processes at Random Times based on Observations of Stocks following the Sensitivity Zones Model [4]

9.2 Pseudocode for Alternative Estimator $r(k)$

We detail the pseudocode for the alternative estimator's contrast function $r(k)$ on the random observation time data case below

1. W.l.o.g. let X_t be the process with less frequent observations (so we have more control over our choice of observations of Y_t)
2. Compute Δ the average time between observations of X_t
3. For a given $k \in \{... - 2, -1, 0, 1, 2...\}$
4. Initialise a cumulative sum
5. For every pair of consecutive observations at t_i, t_{i+1} of X_t,
6. Choose observation s_l of $Y_{t+k\Delta}$ such that it is the most recent observation before t_i. (As before)

7. For s_u. First attempt to choose observation s_u of $Y_{t+k\Delta}$ such that $s_u \in [t_{i+1}, t_{i+1} + \Delta]$, where s_u is as close to $t_{i+1} + \Delta$ as possible.
8. If there is no such observation, choose the next closest to t_{i+1} with $s_u > t_{i+1}$ (as in Hayashi Yoshida).
9. Add the product $X_{t_i}^{t_{i+1}} Y_{s_l}^{s_u}$ to the rolling sum.
10. After all consecutive observation pairs have been computed and added to the sum (by repeating steps 5-9), return the value of the cumulative sum, $r(k)$.