Investigating Fund Return Distribution when Value of Fund under Management is Irregularly Observed

KiHoon Jimmy Hong[a], John Knight [b], and Steve Satchell[c,*]

Current Version: August 2014

[a] *University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Australia*

[b] *University of Western Ontario, Department of Economics, SSC, London, Ontario N6A 5C2, Canada*

[c] *Trinity College, University of Cambridge, Trinity College, Cambridge, CB2 1TQ, U.K.*

The University of Sydney, City Rd, Darlington NSW 2008, Australia

Abstract

We propose a consistent estimation technique that directly utilizes irregularly spaced observations to investigate the statistical properties of irregularly observed monetary value of fund under management. The contribution of our paper is that we provide an econometrically enhanced and more detailed method that improves the existing likelihood based techniques developed in other fields in estimating the parameters of irregularly spaced observations.

JEL classification: C51, C58, G23

Keywords: Irregularly Spaced Time Series, Fund Value under Management, Ornstein Uhlenbeck

[*]Corresponding author. *E-mail address*: ses999gb@yahoo.co.uk

1. Introduction

In this paper, we propose a consistent estimation technique to investigate the statistical properties of irregularly observed monetary value of fund under management. This estimation method takes advantage of return variance conditioned on the value of a fund and directly utilizes irregularly spaced observations without losing any information. Irregularly spaced data is a popular research topic as it can be found in many industrial and scientific areas: Natural disasters such as earthquakes, floods, or volcanic eruptions, astronomical measurements such as spectra of celestial objects, patients' state of health and many more. However we do not find many studies that look at the particular case of irregular fund value observation, which would be highly relevant to investors and regulators.

We give an example to demonstrate how irregularity in fund value observation could be problematic for potential investors and regulators. Assume that a manager receives seed money to start a fund. The manager would first run the fund with the given seed money to establish historical performance record. Once the record is established, the manager would like to attract new investors for various reasons such as achieving economies of scale or attracting more management fees. The potential investors would like to investigate the historical return distribution of the fund in order to decide whether they should invest or not. As the fund is introduced to a larger pool of investors, regulators may feel that the fund needs to be investigated.

However, the historical value of the fund under management does not necessarily be publicly available on regular basis. There could be many reasons for this including that the assets held by the fund are not appraised regularly, the assets are infrequently traded or simply the manager did not publish the value regularly. Then the potential investors and regulators need to evaluate the fund's return distribution properly based on the irregularly observed data. However, Eckner (2012) show that the estimated moments using irregularly observed data could be heavily biased. The proposed estimation method of this paper is useful in this situation.

The fact that some financial data is observed irregularly is often overlooked in two ways; by having investment horizons of more than a few days or by assuming evenly spaced trade intervals. In studies with a relatively longer investment horizon, it is implicitly assumed that any effect that the intervals could have quickly disappear and such an assumption leads to information loss. Eckner (2012) points out that assuming evenly spaced intervals has four potential problems. (1) The estimates of the second moments, variance and auto covariance may be subject to a significant and hard-to-quantify bias. (2) It could change the causal relationship in a multivariate time series. (3) It leads data loss and dilution. (4) Loss of time information.

The contribution of our paper is that we provide an econometrically enhanced and more detailed method that improves the existing likelihood based techniques developed in other fields in estimating the parameters of models when observations are irregularly spaced. Our result is theoretically equivalent to the existing methods, when the sample is large and the autocorrelation is low. However, our estimation method should outperform the existing methods when the sample size is small and there exists significant autocorrelation in the data.

The rest of the paper is organized as follows. Section 2 presents literature review, Section 3 presents the model, Section 4 presents the estimation method, Section 5 investigates simulated results and Section 6 concludes the paper.

2. Research Background and Literature Review

Investigation of irregularly spaced data has for a long time been a research topic in many areas as it has many applications throughout various fields such as marine science (Gulland and Holt, 1959), forest science (Moser, 1972), biology (Chao, 1987), hydrology (Duan and Sorooshian, 1988), astronomy (Scargle, 1989) and many others. More recent literature includes Bos et al. (2002), Thiebaut and Roques (2005), Broersen (2008) and Davidsen and Griffin (2010).

Among these works, Duan and Sorooshian (1988) seems closest to ours and is also done in discrete time and their results are similar to the results we derive in section 3. They developed a new maximum likelihood criterion suitable for hydrological model calibration using data which are recorded at unequal time intervals and contain auto correlated errors. Our results, however, are more detailed in terms of the structure of the time series process. When the sample size is large and the autocorrelation in the data is low, their result will be equivalent to ours. However, when the sample size is small and there is significant autocorrelation, our approach potentially yields more precise estimates.

Analysis of irregularly spaced data could have many applications in economics and finance. Irregularly reported fund returns, for example, where many mutual and hedge funds compute profit and losses whereby investors only observe them on an irregular basis. Economic and financial shocks also happen on an irregular basis. More generally, relevant news happens and also is observed on an unevenly distributed time space. Many illiquid financial assets, such as real estate, art sold by auction etc., are traded on an irregular basis. The announcement date of earnings is another example where the SEC only requires companies to announce within 40 days from the quarter end. Analysts forecast of many small cap companies are reported on an irregular basis. High frequency trading data also is a good example. Muller (1991), Gilles, Zumbach (2001) and Dacorogna et al. (2001) examine irregular time series data in the context of high frequency financial data.

Much existing literature in financial application of irregularly spaced data focus on irregular financial trades. Engle and Russell (1998) propose a statistical model for the analysis of data which arrive at irregular intervals and develop asymptotic properties of the quasi maximum likelihood estimator as a corollary to ARCH model results. Ghysels and Jasiak (1998) develop a class of ARCH models for irregularly spaced trades and quote arrivals. The econometric contribution of this body of work is to combine the temporal aggregation for the GARCH model and the autoregressive conditional duration model. Meddahi, Renault and Werker (2006) use an exact discretization of continuous time stochastic volatility processes observed at irregularly spaced times to investigate how a coherent GARCH model can be specified for such data.

However, unlike the aforementioned works, we are interested in irregular observations in funds that invest trusted wealth. These funds could include mutual funds, hedge funds, private equity funds, venture capitals etc. Capital injection and investment opportunities in such funds are distributed unevenly across time. An especially important case is where a fund is given seed capital by its backer. Whenever actual funds arrive into the fund, which can occur on an occasional basis, an equal amount of funds, together with the profits accrued to them, is returned to the sponsor. Returns are only measured on these occasions. The challenge then is to calculate the distributional properties of the rate of return of the sponsor. We shall use this situation to motivate our analysis.

As previously discussed, irregularly spaced data is often overlooked in two ways: (1) by having investment horizons of more than few days or (2) by assuming evenly spaced trade intervals. The first cannot be applied to many studies other than ones with a relatively longer investment horizon. The second is the most widely used approach, which transforms irregularly spaced data into equally spaced data using some form of interpolation. (see Adorf, 1995 and Beygelzimer et al., 2005) This paper, on the other hand, provides an estimation method that can directly utilize irregularly spaced financial data without losing any information.

3. The Model

3.1. The Discrete i.i.d. Case

As mentioned earlier, to motivate our paper, we use a situation where a fund is given seed capital by its backer. Whenever actual funds arrive into the fund, which can occur on an occasional basis, an equal amount of funds, together with the profits accrued to them, is returned to the sponsor. We want to calculate the distributional properties of the rate of return to the sponsor.

Our problem can be cast in the following manner. Let x_i, where i = (1, …,m), be the returns to the fund observed on m irregular dates. The time between observing the i th and i-1th return we denote by $t_i – t_{i-1} = \Delta t_i$. If the underlying data are independently distributed, with mean μ and variance σ^2 then it is immediate that the observed random variable x_i has mean $\mu\Delta t_i$ and variance $\sigma^2\Delta t_i$; we shall write this in a regression format as

$$x_i = \mu\Delta t_i + \sigma\sqrt{\Delta t_i} z_i \quad (1)$$

where i = (1, …, m) and z_i is distributed independently (0,1). We note in passing that we cannot assert that z_i are identically distributed without additional assumptions. We can calculate a new regression

$$y_i = \mu\sqrt{\Delta t_i} + \sigma z_i \quad (2)$$

This regression satisfies Gauss-Markov assumptions and so, denoting total elapsed time by T,

$$\hat{\mu} = \frac{\sum_{i=1}^{m} y_i\sqrt{\Delta t_i}}{\sum_{i=1}^{m} \Delta t_i} = \frac{\sum_{i=1}^{m} x_i}{T} \tag{3}$$

Furthermore, the “optimal”, i.e. the unbiased, OLS estimated variance will be

$$s^2 = \frac{\sum_{i=1}^{m}(y_i - \hat{\mu}\sqrt{\Delta t_i})^2}{m-1} = \frac{\sum_{i=1}^{m}(\frac{x_i}{\sqrt{\Delta t_i}} - \hat{\mu}\sqrt{\Delta t_i})^2}{m-1} \tag{4}$$

We have said nothing about the error distribution; if we assume that the underlying error is Gaussian, then z_i is independent $N(0,1)$. And, as is well-known, s^2 is distributed as $\sigma^2 \frac{\chi^2(m-1)}{m-1}$ which in words says that the sample variance is distributed as the population variance multiplied by a chi-squared variable with m-1 degrees of freedom divided by m-1.

Let x_i, $i = 1 \ldots m$, be now described by a continuous time process with finite second moments; we denote its non-constant mean by $\mu(t)$ and its auto-covariance function $K(r,s)$ which can be defined even if the process is not weakly stationary as long as second moments exist. In particular, $\mathrm{E}(x_i) = \int_{t_{i-1}}^{t_i} \mu(s)\, ds$ and $\mathrm{Cov}(x_i, x_j) = \int_{t_{i-1}}^{t_i} \int_{t_{j-1}}^{t_j} K(r,s)\, dr\, ds$. And in particular, $\mathrm{Var}(x_i) = \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t_i} K(r,s)\, dr\, ds$. We could make a further weak stationarity assumption,$K(r,s) = K(r-s)$, however it is hard to say much else without going to a particular example.

3.2. Ornstein Uhlenbeck Model

In this section, we study the effect and the estimation of irregularly spaced data on an OU Model. The OU process is often used to model mean reverting financial processes. Mean reverting processes are naturally attractive to model financial asset prices because of the economic argument that when prices are ‘too high’, demand will reduce, and supply will increase, hence prices decrease, producing a counter-balancing effect. This process has been extensively studied by Bergstrom (1990) and it has been successfully used by many others including Lo and Wang (1995) in studying financial asset price processes. We present some general results for OU processes before looking at our particular model. We assume that

$$dy(t) = \theta(\kappa - y(t))dt + \sigma dW(t) \tag{5}$$

Where $\theta > 0$ and $\sigma > 0$. It is well-known that

$E(y(t)) = y(0)e^{-\theta t} + \kappa(1 - e^{-\theta t})$

$K(s,r) = Cov(y(s),y(r)) = \frac{\sigma^2}{2\theta} e^{-\theta(s+r)}(e^{2\theta \min(s,r)} - 1)$

We shall motivate our analysis in terms of our fund where $P(t)$ is the value of the fund at time t; it is only observed at a point at which there is an inflow or outflow. At this time, we compute the current pre-transaction value and assess the return relative to the previous observed value. Because new funds are replacing old (seed) funds the net impact of flows is zero. Consequently, the post-transaction value is equal to the pre-transaction value (as long as the amount of new funds does not exceed the residual seed money) but it is at these times that fund value is measured. New entrant funds are effectively locked in and cannot be withdrawn. At some point all the seed money will all be returned and so the flow of funds will jump at this time, that is pre and post values will differ.

In more general cases, we can assume that at the particular time that new funds arrive and/or leave, the continuous time process stops and instantaneously restarts from the new level. A knowledge of these levels is available to fund managers so these computations are straight-forward. We assume that inter-arrival times are given and non-stochastic but this can be easily generalized to stochastic times. Knight and Satchell (2013) (reference to be added) provide an analysis of multivariate log-normal processes wherein arrival times are i.i.d. negative exponential so that the overall number of events (defined as either additions or subtractions) follows a homogeneous Poisson process. Modelling the monetary value of a fund, as well as the return of a fund is advantageous because we can observe both the level and return and it allows us to condition on the monetary value.

We now turn to our specific case. Assume that the logarithm of the asset prices $\log P(t)$ has linear trends. We consider the process

$$q(t) \coloneqq \log P(t) - \mu t \tag{6}$$

Assume that $q(t)$ satisfies the following stochastic differential equation,

$$dq(t) = -\theta q(t) dt + \sigma dW(t) \tag{7}$$

where the parameter μ is for the market and θ and σ are the parameters of the OU process. We assume $\sigma > 0$, and W denotes a Wiener process. Equation (7) does not presume stationarity or trending. OU processes and more general stochastic differential equation systems have been analyzed econometrically by Sargan (1974), Robinson (1977), and Phillips (1974, 1983). As previously discussed, the OU process has several advantages in investigating financial assets. It can be explicitly solved and there are exact solutions for discretized versions of this model.

Furthermore, the OU process allows for stationary behavior, random walks and explosive behavior in q. Mean reverting processes are naturally attractive to model prices of financial assets since they embody the economic argument that when prices are ‘too high’, demand will reduce, supply will increase thereby reducing prices and producing a counter-balancing effect. When prices are 'too low' the opposite will happen, again pushing prices back towards some kind of long term mean. The OU process is the continuous-time analogue of the discrete-time AR(1) process. It can be interpreted as a scaling limit of a discrete process, in the same way that Brownian motion is a scaling limit of random walks.

The solution for $q(t)$ can be written as

$$q(t) = q(0)e^{-\theta t} + \int_0^t \sigma e^{-\theta(t-u)} dW(u) \qquad (8)$$

From our general results above, $E(q(t)) = q(0)e^{-\theta t}$ and

$K(s,r) = Cov(q(s),q(r)) = \frac{\sigma^2}{2\theta} e^{-\theta(s+r)}(e^{2\theta \min(s,r)} - 1),\ s \neq r$

$= Var\big(q(s)\big) = \frac{\sigma^2}{2\theta} e^{-2s\theta}(e^{2\theta s} - 1),\ s = r$

First, assume that returns are computed over evenly spaced time periods. The one period return, (which may not be observable) at time t can be written as

$$r(t+1,t) = \mu + q(0)e^{-\theta t}(e^{-\theta} - 1) + \int_t^{t+1} \sigma e^{-\theta(t+1-u)} dW(u) \qquad (9)$$

Assume that we observe the return at irregular intervals, Δt_i, then the return between two time points; t_i and t_{i+1} can be written as

$$r(t_{i+1}, t_i) = \mu \Delta t_i + q(t_i)(e^{-\theta \Delta t_i} - 1) + \int_{t_i}^{t_{i+1}} \sigma e^{-\theta(t_{i+1}-u)} dW(u) \qquad (10)$$

A more convenient form for estimation is to use the autoregressive representation of (8). We can rewrite (8) as

$$q(t) = q(t-1)e^{-\theta} + \int_{t-1}^{t} \sigma e^{-\theta(t-u)} dW(u) \tag{11}$$

Generalizing,

$$q(t_{i+1}) = q(t_i)e^{-\theta \Delta t_i,} + \int_{t_i}^{t_{i+1}} \sigma e^{-\theta(t_{i+1}-u)} dW(u) \tag{12}$$

This is the analogue of (46) of Lo and Wang (1995),which is the explicit solution of the univariate trending OU process in a recursive representation of q. Lo and Wang (1995) show that the maximum likelihood estimator of the discrete-time parameters of this process is asymptotically equivalent to the ordinary least squares estimator applied to detrended prices. The continuous-time parameters μ, σ, and θ may then be obtained from the discrete-time parameter estimates. Returns from t to $t+s$ can now be simply expressed as

$$r(t+s,t) = \mu s + q(t)(e^{-\theta s} - 1) + \int_{t}^{t+s} \sigma e^{-\theta(t+s-u)} dW(u) \tag{13}$$

The advantage of the expression in (13) is that the error terms are uncorrelated; the disadvantage is that the expression is mixed, involving both de-trended prices and returns. The conditional (conditioning on $q(t)$) mean, μ_r, and the variance of $r(t + s, t)$, σ_r^2, can be written as

$$E(r(t+s,t)|q(t)) = \mu_r = \mu s \tag{14}$$

$$Var(r(t+s,t)|q(t)) = \sigma_r^2 = \frac{\sigma^2}{2\theta}\left(1-e^{-2\theta s}\right) \tag{15}$$

Since in what follows $t \geq j+k > j$, and using the independence property of non-overlapping increments property of Brownian motion,

$$\text{Cov}(r(t+s,t), r(j+k,j)|q(t)) = 0 \qquad \text{for } t > j \tag{16}$$

We call the net change of the fund due to flow at time t_i, $X(t_i)$. So at times t_i , the level of funds under management changes by $X(t_i)$. Then the level equation for the fund under management is $P(t_i^+) = P(t_i^-) + X(t_i)$. Assuming t_i and t_{i+1} are adjacent event times and define $\Delta t_i = t_{i+1} - t_i$, the return between t_i and t_{i+1} can be written as

$$r(t_{i+1}, t_i) = \ln\left(\text{P}((t_{i+1}^-))\right) - \ln\left(P({t_i}^+)\right)$$

$$r(t_{i+1}, t_i) = \mu\Delta t_i + q(t_i^+)(e^{-\theta\Delta t_i} - 1) + \int_{t_i}^{t_{i+1}} \sigma e^{-\theta(t_{i+1}-u)} dW(u) \tag{17}$$

We do not enter into the calculations except to adjust the funds to the level which is earning the returns. Also note, technically that $q(t^+)$=$\ln\left(P(t^+)\right) - \mu t$ whilst $\text{q}(t^-) = \ln\left(P(t^-)\right) - \mu t$. We can also, if we wished, include withdrawals over and above the repayment of seed money.

4. Estimation of the Model

Model parameters may best be estimated but using the $q(t_i^+)$ in (17) as we can use generalized least squares noting that while the error terms are uncorrelated they are heteroscedastic. Using our previous result in (15) and (16), weighted Least Squares can be applied iteratively to a linear regression of $r(t_{i+1}, t_i)$ on Δt_i and $q(t_i^+)$. This would involve an initial estimate of μ to calculate $q(t)$. However, we can rewrite (17) as

$$r(t_{i+1}, t_i) = \mu\Delta t_i + \ln\left(P(t_i^+)\right)\left(e^{-\theta\Delta t_i} - 1\right) - t_i\mu\left(e^{-\theta\Delta t_i} - 1\right) + \int_{t_i}^{t_{i+1}} \sigma e^{-\theta(t_{i+1}-u)} dW(u) \quad (18)$$

So an alternative and possibly better estimation involves a weighted linear regression of $\mathrm{r}(\mathrm{t}_{\mathrm{i}+1}, \mathrm{t}_\mathrm{i})$ on Δt_i, t_i and $\ln\left(P(t_i^+)\right)$, using the square-root of the diagonal of the conditional covariance matrix as weights.

We can generalize Lo and Wang (1995) who show that the maximum likelihood estimator of the discrete-time parameters of the equally-spaced $q(t)$ process is asymptotically equivalent to the ordinary least squares estimator applied to detrended prices to demonstrate that our iterated WLS procedure is equivalent to maximum-likelihood. Our proposed estimation method is as follows.

From Equation (18) we have,

$$\frac{r(t_{i+1},t_i)}{Stdev(r(t_{i+1},t_i)|q(t))} = \frac{\mu\left(\Delta t_i - t_i\left(e^{-\theta\Delta t_i}-1\right)\right)}{Stdev(r(t_{i+1},t_i)|q(t))} + \frac{\ln\left(P(t_i^+)\right)}{Stdev(r(t_{i+1},t_i)|q(t))}\left(e^{-\theta\Delta t_i} - 1\right) + \varepsilon_i \quad (19)$$

where $Stdev(r(t_{i+1}, t_i)|q(t)) = \sqrt{\frac{\sigma^2}{2\theta}(1 - e^{-2\theta\Delta t_i})}$ and $\varepsilon_i \sim N(0,1)$. The purpose of scaling by the conditional return standard deviation is to control potential heteroskedasticity. This results in more efficient estimates, i.e. smaller estimated standard error. This will be further discussed in section 5.2 with simulated results. We notice that for a given value of θ and σ^2, μ can be easily estimated using:

$$\frac{r(t_{i+1},t_i)}{Stdev(r(t_{i+1},t_i)|q(t))} - \frac{\ln\left(P(t_i^+)\right)}{Stdev(r(t_{i+1},t_i)|q(t))}\left(e^{-\theta\Delta t_i} - 1\right) = \frac{\mu\left(\Delta t_i - t_i\left(e^{-\theta\Delta t_i} - 1\right)\right)}{Stdev(r(t_{i+1},t_i)|q(t))} + \varepsilon_i \quad (20)$$

and letting $w_i(\theta,\sigma^2) = \frac{r(t_{i+1},t_i) - \ln\left(P(t_i^+)\right)}{Stdev(r(t_{i+1},t_i)|q(t))}\left(e^{-\theta\Delta t_i} - 1\right)$ and $z_i(\theta,\sigma^2) = \frac{\left(\Delta t_i - t_i\left(e^{-\theta\Delta t_i} - 1\right)\right)}{Stdev(r(t_{i+1},t_i)|q(t))}$, then (20) can be simply written as

$$w_i(\theta,\sigma^2) = \mu z_i(\theta,\sigma^2) + \varepsilon_i \quad (21)$$

Consequently, OLS gives $\hat{\mu}(\theta,\sigma^2) = \frac{\sum w_i(\theta,\sigma^2) z_i(\theta,\sigma^2)}{\sum z_i^2(\theta,\sigma^2)}$. The residual sum of squares (RSS) from (21) is now

$$\sum\left[w_i(\theta,\sigma^2) - \hat{\mu}(\theta,\sigma^2) z_i(\theta,\sigma^2)\right]^2 \quad (22)$$

and minimizing with respect to θ, σ^2 gives $\hat{\theta}$, $\hat{\sigma}^2$ with $\hat{\mu} = \hat{\mu}(\theta,\sigma^2)$. These estimates are now used as the starting values for the full non-linear optimization of (21).

5. Simulated Investigation

5.1. Simulation

Good data in fund value observation are hard to find. We choose to simulate our result as this would deliver the contribution of the proposed method clearly. Rather than trying to choose plausible values of μ, θ and σ^2 it may be easier to choose plausible values for the observed (population) mean, variance and autocorrelations. Following Lo and Wang (1995) we denote by $\bar{r}$, s^2 and $\rho(1)$ the unconditional mean, variance and first-order autocorrelation of returns respectively. These can be defined without any reference to a particular data-generating process. Then using results in Lo and Wang (1995) and setting their τ =1, we then have the results:

$$\mu = \bar{r} \;, \quad \theta = -\log(1+2\rho(1)) \quad \text{and} \quad \sigma^2 = s^2 \frac{\log(1+2\rho(1))}{2\rho(1)} \tag{23}$$

Using these relationships we can then examine the effect of small and large autocorrelation along with varying the sample size. We doubt that μ will have any effect or at least little so we can choose a suitable value, 0.001. This assumes the expected one period (daily) fund return of 0.1%. We also set s = 0.005 hence the fund return one period standard deviation of 0.5%. These figures could be considered reasonable. Table 1 gives the OU parameters values associated with various first order return autocorrelations ($\rho(1)$).

(Insert Table 1 here)

We refer to these values as the ‘underlying parameter values’. We simulate 200 period observations where during the first 100 periods, there is no fund inflow. In the second 100 periods, a fund experiences random capital injection.

We set the initial value of fund under management, $P(0)$, as $100,000,000 which is the seed money. We generate i.i.d. observations of a negative exponential with a mean of 3 (corresponding to days).We sample from a Bernouilli distribution which equals 1 with a probability of 0.9 corresponding to a measurement without flow and equals 0, with probability of 0.1 corresponding to a measurement with flow. We call the net change of the fund due to flow at time t_i, $X(t_i)$. If there is flow, we sample $X(t_i)$ from a scaled chi-squared 3 with a mean of 10,000,000. Note that the degree of freedom, 3, is an arbitrary choice of a positive distribution. We keep the $P(t)$ at 100,000,000 until all the seed capital is returned (about 18 months) then we increase the capital as the inflows occur.

Then at times t_i, the level of funds under management changes by $X(t_i)$. Also as noted, $q(t^+)=\ln\left(P(t^+)\right)-\mu t$ whilst $q(t^-)=\ln\left(P(t^-)\right)-\mu t$. We simulated 200-period regularly observed data first, and then randomly choose from the regularly observed data, as described in the previous paragraph, to construct the irregularly observed data. As the average length of irregular period is 3 (the mean of a negative exponential distribution), we have 67 observations for irregular data on average. Therefore the irregular data could be seen as regularly observed data with missing observations.

We simulate 3000 sets of data. We can also, if we wished, include withdrawals over and above the repayment of seed money. The technical details of the simulation are provided in Appendix A. Finally inflows and outflows could be related to past performance, in which case we would use a Cox process i.e. a Poisson process whose intensity function is dependent on past returns.

5.2. Simulated Result

Let us define three cases of interest.

(1) Case 1: The value of a fund is regularly observed for each period. This is considered as the true case in the sense that estimates based on this framework should be consistent and asymptotically efficient.

(2) Case 2: The value of a fund is observed irregularly. We use our estimation method.

(3) Case 3: The value of a fund is observed irregularly. We take the irregularly observed data and compute return moments ignoring the irregularities.

Case 1 is based on regularly observed data in every period while case 2 and 3 are based on irregularly observed data. In order to help readers distinguish the parameters for three cases. Although the choice of initial values of the estimation do not affect the result as we estimate iteratively and the estimated parameters converge very quickly, we make sure that the initial values are computed based only on the available information set for each case. In case 1, information from regularly observed data is used and we assume that the underlying case information is not available. For case 2, information from irregularly observed data is used to set the initial estimation values and we assume that the underlying and the regularly observed data are not available. Note that we do not need to assume initial values in case 3.

In case 1, we denote $\bar{\hat{\mu}}$, $\bar{\hat{\sigma}}$, $\bar{\hat{\theta}}$, $\overline{\widehat{\mu_r}}$ and $\overline{\widehat{\sigma_r}}$ as mean of estimated μ, σ, θ, return mean and return standard deviation respectively. Parameters μ, σ and θ are the OU parameters and return mean and standard deviation are estimated from regularly observed data. In case 2, we use notation $\bar{\tilde{\mu}}$, $\bar{\tilde{\sigma}}$, $\bar{\tilde{\theta}}$, $\overline{\tilde{\mu}_r}$ and $\overline{\tilde{\sigma}_r}$. In case 3, we use notations $\bar{\dot{\mu}}$, $\bar{\dot{\sigma}}$, $\bar{\dot{\theta}}$, $\overline{\dot{\mu}_r}$ and $\overline{\dot{\sigma}_r}$.

Note that all the estimated parameters in case 1, 2 and 3 are denoted as averages. This is because these are the averages of the estimated parameters over 3000 simulations. For example, in case 1, these can be expressed as

$$\bar{\hat{\mu}} = \frac{\sum_{i=1}^{N} \hat{\mu}}{N},\ \bar{\hat{\sigma}} = \frac{\sum_{i=1}^{N} \hat{\sigma}}{N},\ \bar{\hat{\theta}} = \frac{\sum_{i=1}^{N} \hat{\theta}}{N},\ \overline{\widehat{\mu_r}} = \frac{\sum_{i=1}^{N} \widehat{\mu_r}}{N},\ \overline{\widehat{\sigma_r}} = \frac{\sum_{i=1}^{N} \widehat{\sigma_r}}{N},$$

$$\mathrm{SE}(\bar{\hat{\mu}}) = \sqrt{\frac{\sum_{i=1}^{N} (\hat{\mu} - \bar{\hat{\mu}})^2}{N-1}},\ \mathrm{SE}(\bar{\hat{\sigma}}) = \sqrt{\frac{\sum_{i=1}^{N} (\hat{\sigma} - \bar{\hat{\sigma}})^2}{N-1}},\ \mathrm{SE}(\bar{\hat{\theta}}) = \sqrt{\frac{\sum_{i=1}^{N} (\hat{\theta} - \bar{\hat{\theta}})^2}{N-1}}, \qquad (24)$$

$$\text{SE}(\overline{\widehat{\mu_r}}) = \sqrt{\frac{\sum_{i=1}^{N}(\widehat{\mu_r}-\overline{\widehat{\mu_r}})^2}{N-1}},\ \text{SE}(\overline{\widehat{\sigma_r}}) = \sqrt{\frac{\sum_{i=1}^{N}(\widehat{\sigma_r}-\overline{\widehat{\sigma_r}})^2}{N-1}}$$

where N is the number of simulations, which is 3000 in our case. The equivalent formulae can be obtained for case 2 and 3 by replacing $\hat{\mu},\ \hat{\sigma},\ \hat{\theta}\ ,\ \widehat{\mu_r},\ \widehat{\sigma_r}$ with $\tilde{\mu},\ \tilde{\sigma},\ \tilde{\theta},\ \tilde{\mu}_r,\ \tilde{\sigma}_r$ and $\dot{\mu},\ \dot{\sigma},\ \dot{\theta},\ \dot{\mu}_r,\ \dot{\sigma}_r$, respectively. Table 2 reports the accuracy of the OU parameter estimation in Case 1. Panel A, B and C present the estimated OU parameters, their standard errors and mean squared errors of Case 1, respectively. Panel D presents the p-values of statistical difference between the underlying case and case 1 OU parameters. The p-values are to test the hypothesis that the expected values of the estimated OU parameters of case 1 are not statistically significantly different from those of the underlying case. The test is a standard mean difference test assuming the normality in the estimated parameter distribution. All subsequent p-values are computed using the same assumption.

(Insert Table 2 here)

Estimated OU parameters and their standard errors reported in panel A and panel B are computed using (24). Comparing the results in panel A of Table 2 and Table 1, we can see that the estimated OU parameters based on regularly observed data are not much different from those of the underlying OU parameters. The results in panel D indicates that the case 1 estimates of the OU parameters are not statistically different from the underlying OU parameters at the 1% significance level.

Therefore, Table 2 concludes that the case 1 estimation of the OU parameters is statistically accurate. Table 3 compares one period return mean and standard deviation of the underlying case and case 1. Panel A reports the underlying return mean and standard deviation. Panel B presents the average of case 1 return mean and standard deviation estimates. And panel C reports the p-values testing the hypothesis that the return means and standard deviations estimates of the underlying case are not statistically significantly different from those of case 1.

(Insert Table 3 here)

The first and the third rows of panel A are computed using (14) and (15) respectively, where the observation interval, s = 1 and using the OU parameters in Table 1, which are $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\theta}$. These are the analytical return moments of (13) and therefore standard errors do not exist. The return mean and standard deviation in panel B are computed using (24) with case 1 parameters, $\hat{\mu}$, $\hat{\sigma}$, $\hat{\theta}$, $\widehat{\mu_r}$ and $\widehat{\sigma_r}$.

Comparing the underlying return mean and standard deviation in panel A to case 1 estimated return mean and standard deviation in panel B, we can see that the case 1 return mean and standard deviation are very close to those of the underlying case with very small standard errors. Panel C shows that all p-values of difference test between the expected values of the underlying return mean and standard deviation and the expected values of case 1 return mean and standard deviation and show that they are not statistically different at the 1% significance level. From Table 3, we observe that the case 1 estimated one period (regularly observed) return mean and standard deviation are unbiased compare to the underlying return mean and standard deviation.

Table 4 reports the return mean and standard deviation for case 2 ($\overline{\tilde{\mu}_r}$ and $\overline{\tilde{\sigma}_r}$) and for case 3 ($\overline{\dot{\mu}_r}$ and $\overline{\dot{\sigma}_r}$), respectively. In order to help readers in comparing the results from the irregularly observed data to those of regularly observed data, we present in panel A the estimated return mean and standard deviation of case 1, which was included in panel B of

Table 3. Panel B and C of Table 4 show the estimated return mean and standard deviation of case 2 and 3 respectively.

(Insert Table 4 here)

Panel A is the same as panel B of Table 3. Panel B is computed using (24) but the parameters for case 2, $\tilde{\mu}$, $\tilde{\sigma}$, $\tilde{\theta}$, $\tilde{\mu}_r$, $\tilde{\sigma}_r$, are used instead of the parameters of case 1, $\hat{\mu}$, $\hat{\sigma}$, $\hat{\theta}$, $\widehat{\mu_r}$, $\widehat{\sigma_r}$. Panel C is computed using (24) but the parameters for case 3, $\dot{\mu}$, $\dot{\sigma}$, $\dot{\theta}$, $\dot{\mu_r}$, $\dot{\sigma_r}$, are used.

Note that the expected frequency of the irregular observations is 3, which is given in the simulation with the parameter of the negative exponential distribution. Therefore if we compute return mean and standard deviation, ignoring the irregularity as we do in case 3, the average interval is 3 periods. In order to compute comparable means and variances for one period, the estimated return mean and the variance in case 3, reported in panel C, are scaled by the average observed irregular frequency.

By comparing the row 1 and 3 of panel A to row 1 and 3 of panel B, we note that case 2 return means and standard deviations are very close to those of case 1. However from comparing the row 1 and 3 of panel A to row 1 and 3 of panel C, we can see that the return standard deviations of case 3 are significantly overestimated compare to those of case 1 while the return means of case 3 are somewhat overestimated compared to those of case 1.

Table 5 investigates the statistical significance of the estimated parameter differences. Panel A of Table 5 shows the relative difference of one period return mean and standard deviation of case 2 from case 1. Panel B reports the relative difference of one period return mean and standard deviation of case 3 from case 1. Panel C reports the p-values of testing the hypothesis that the expected return means and standard deviations of case 2 are not statistically different from those of case 1. Panel D reports the p-values of testing the hypothesis that the expected return means and standard deviations of case 2 are not statistically different from those of case 1.

(Insert Table 5 here)

As indicated in the first column, panel A is computed as the ratio of the estimated return moments of case 2 to those of case 1. Panel B is computed as the ratio of the estimated return moments of case 3 to those of case 1. Panel C shows that the expected values of the return means and standard deviations in case 2 are not statistically significantly different from those of case 1 while panel D shows that the expected values of the return standard deviations in case 3 are statistically significantly different from those of case 1 at the 1% confidence level. Panel D also shows that the estimated return means of case 3 are not statistically significantly different from those of case 1 at the 1% significance level. This may be a result of the fact that the test is based on normality assumption, after all the bias in the mean returns is quite high.

This result is consistent with the previous literature. For example, Davidsen and Griffin (2010), find that standard estimates of the volatility of uneven sampling intervals can be strongly biased by applying fractional Brownian motion to ice core records. Eckner (2012) also points out that the estimates of the second moments, variance and autocovariance may be subject to a significant and hard-to-quantify bias.

Table 6 demonstrates the relative gain in estimation accuracy from using the proposed method. Panel A of Table 6 shows the relative difference of one period return mean and standard deviation of case 3 from case 2. Panel B reports the p-values of testing the hypothesis that the expected return means and standard deviations of case 2 are not statistically different from those of case 3. . Panel C reports the relative gain in return mean and standard deviation when case 2 estimation is used against case 3 in percentage term. These values are computed by subtracting 1 from the relative difference of return means and standard deviations of case 3 against 2 subtracted by 1, therefore can be expressed as

$$\frac{\overline{\dot{\mu}_r}}{\overline{\tilde{\mu}_r}} - 1 \ \text{ and } \ \frac{\overline{\dot{\sigma}_r}}{\overline{\tilde{\sigma}_r}} - 1 \tag{25}$$

(Insert Table 6 here)

Panel A provides the relative gain in estimation of return means and standard deviations of case 2 to case 3 as defined in (24). From Table 6, we see 15% accuracy gain in the expected mean estimates on average when the proposed estimation method of this paper is employed compared to when we overlook the irregularity in irregularly observed fund under management data. The relative gain in return standard deviations ranges from 19.86% to 41.39%. Panel B compares the return means and standard deviations estimates of case 2 and 3 to show that both mean and standard deviations could be statistically significantly different.

When the estimation method proposed in this paper is employed, the estimated means and standard deviations of returns are more accurate compared to case 3 when the irregularity in the data is overlooked. Moreover, the estimated return means and standard deviations of the newly proposed method are not statistically significantly different from the regularly observed return means and standard deviations. Therefore, we conclude that the proposed method provides unbiased and efficient estimates of both return means and standard deviations by directly utilizing the irregularly spaced fund value observations.

Lastly, we make a brief further note on the heteroskedasticity in the irregularly observed return distribution. As discussed in section 4, if we do not scale the both sides of the (19), we still get the unbiased return means and standard deviations estimates. However, the standard errors of the estimated return standard deviations increase significantly. This result shows that conditioning on past level data (fund value) increases the estimation efficiency. Hence it provides evidence that the use of return variance conditioned on the value of a fund adds value. Although the result without the scaling factor is not shown in this section for the brevity of the paper, it is available upon request from the authors.

6. Conclusion

In this paper, we demonstrate a consistent estimation technique to investigate the statistical properties of irregularly observed fund return, when the monetary value of a fund under management follows a stochastic process. Modelling the monetary value of a fund, as well as the return of a fund is advantageous because we can observe both the level and return and it allows us to condition on the monetary value. Good data in fund value observation are hard to find. Consequently, we choose to simulate our result as this would deliver the contribution of the proposed method clearly.

The simulated result shows that the irregularly observed return has unbiased return mean but significantly biased variance. This is consistent with the existing literature. We show that the estimated return moments using the proposed estimation technique that takes advantage of return variance conditioned on the value of a fund yields unbiased and more efficient return variance estimate.

References

Adorf, H. M. (1995). Interpolation of irregularly sampled data series - a survey. In Astronomical Data Analysis Software and Systems IV, pp. 460–463.

Bergstrom, A. R. 1990. Continuous Time Econometrics Modelling. Oxford University Press: Oxford.

Beygelzimer, A., E. Erdogan, S. Ma, and I. Rish (2005). Statictical models for unequally spaced time series. In SDM.

Bos, R., S. de Waele, and P. M. T. Broersen (2002). Autoregressive spectral estimation by application of the burg algorithm to irregularly sampled data. IEEE Transactions on Instrumentation and Measurement 51(6), 1289–1294.

Broersen, P. M. T. (2008). Time series models for spectral analysis of irregular data far beyond the mean data rate. Measurement Science and Technology 19, 1–14.

Chao, A. (1987) "Estimating the Population Size for Capture-Recapture Data with Unqual Catchability", Biometrics, 43, 783-791.

Dacorogna, M. M., R. Gen¸cay, U. A. M¨uller, R. B. Olsen, and O. V. Pictet (2001). An Introduction to High-Frequency Finance. Academic Press.

Davidsen, J., Griffin, J., 2010. "Volatility of unevenly sampled fractional Brownian motion: an application to ice core records", Phys Rev E Stat Nonlin Soft Matter Phys, 81(1), 016107

Duan, Q. and Sorooshian, S.R., 1988. A maximum likelihood criterion for use with data collected at unequal time intervals Water Resources Research, Volume 24, Issue 7, pages 1163–1173,

Eckner, A. (2012). A Framework for the Analysis of Unevenly Spaced Time Series Data. working paper.

Engle, R. F. (2000). The econometrics of ultra-high-frequency data.Econometrica, 68(1), 1–22.

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

Gilles Zumbach, U. A. M. (2001). Operators on inhomogeneous time series. International Journal of Theoretical and Applied Finance 4, 147–178.

Gulland, J. A. and Holt, S. J. (1959), "Estimation of Growth Parameters for Data at Unequal Time Intervals", ICES Journal of Marine Science, 25(1), 47-49.

Ghysels, E. and Jasiak, J. (1998) GARCH for Irregularly Spaced Financial Data: The ACD-GARCH Model, Studies in Nonlinear Dynamics and Econometrics, 2(4), 133-149

Hong, K. & Satchell, S. (2014). The Sensitivity of Beta to the Time Horizon when Log Prices follow an Ornstein-Uhlenbeck Process. European Journal of Finance, 20(3), 264-290.

Lo, Andrew W and Wang, J. (1995). “Implementing Option Pricing Models When Asset Returns Are Predictable,” Journal of Finance, American Finance Association, vol. 50(1), pages 87-129, March.

Meddahi, N., E. Renault and Werker, B. (2006) GARCH and Irregularly Spaced Data, Economics Letters, 90, 200-204.

Moser, J. W. (1972) "Dynamics of an Uneven-Aged Forest Stand", Forest Science, 18(3), 184-191.

Muller, U. A. (1991). Spacially weighted moving averages with repeated application of the ema operator. Working Paper, Olsen and Associates, Zurich, Switzerland.

Phillips, P.C.B., 1974, “The estimation of some continuous time models”, Econometrica 42, 803-823.

Phillips, P.C.B., 1983, "Exact Small Sample Theory in the Simultaneous Equations Model", Chapter 8 in Z. Griliches and M. D. Intriligator (eds.) Handbook of Econometrics, Volume 1,

North-Holland, Amsterdam, 449-516.
Robinson, Peter M., 1977. "The construction and estimation of continuous time models and discrete approximations in econometrics," Journal of Econometrics, Elsevier, vol. 6(2), pages 173-197.
Sargan, J.D., 1974, “Some discrete approximations to continuous time stochastic models”, Journal of the Royal Statistical Society B 36, pages 74-90.
Scargle, J., 1981. Studies in Astronomical Time Series Analysis: I. Modeling Random Processes in the Time Domain The Astrophysical Journal Supplement Series, 45, 1-7.
Scargle, J., 1989. Studies in Astronomical Time Series Analysis: III. Fourier Transforms, Autocorrelation Functions, and Cross-Correlation Functions of Unevenly Spaced Data Astrophysical Journal, 343, 874-887.
Thiebaut, C. and S. Roques. 2005. Time-Scale and Time-Frequency Analyses of Irregularly Sampled Astronomical Time Series. EURASIP Journal on Applied Signal Processing 15, 2486–2499.

Appendix

Appendix A. Simulation Process.

Initial price: $P(0)$ = 100,000,000

Probability distribution function of random observation:

$pdf(\Delta t_i; \lambda) = \lambda e^{-\lambda \Delta t_i}$ where $\lambda = \frac{1}{3}$

Simulation: $\Delta t_i = \frac{-\ln(u)}{\lambda}$ where u ~ uniform(0,1)

$$pdf(k;p) = \begin{cases} 0.9 \; if \; k = 1 \\ 0.1 \; if \; k = 0 \end{cases}$$

If k = 0 at time t_i, $P(t_i) \sim \frac{10000000}{3} \chi^2(3)$, $E[P(t_i)] = 10000000$

Simulation: $P(t_i) = \frac{10000000}{3} \sum_{j=1}^{3} z_j^2$ where $z \sim N(0,1)$

No fund flow for the first 100 observations.

Allow only fund inflow for next 100 observations

Tables

Table 1

Underlying Parameters

This table reports the OU parameters, θ, σ and μ for return first order autocorrelation, ρ(1) of -0.1, -0.2, -0.3, -0.4 and -0.45.

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
θ	0.2231	0.5108	0.9163	1.6094	2.3026
σ	0.0053	0.0057	0.0062	0.0071	0.0080
μ	0.0010	0.0010	0.0010	0.0010	0.0010

Table 2

Underlying vs. Case 1: The OU Parameters Accuracy

This table reports the accuracy of the OU parameters using regularly observed simulated series. Parameters $\bar{\hat{\theta}}$, $\bar{\hat{\sigma}}$ and $\bar{\hat{\mu}}$ are the average of the estimated OU parameters in case 1 computed using (24). Underlying value is the value assumed in simulation, estimated value is the parameter value estimated from the simulated data for return first order autocorrelation, ρ(1) of -0.1, -0.2, -0.3, -0.4 and -0.45. Panel A reports the estimated OU parameters in Case 1, where the value of a fund is observed on regular basis. Panel B reports the standard errors of Case 1 estimated parameters. Panel C reports Mean Squared Errors of the estimated parameters in Case 1. Panel D reports the p-values whether the estimated values of the Case 1 parameters are not statistically significantly different from the underlying parameter values. The notation '< 0.0001' indicates that the value is less than 0.0001. 200 observation period and 3000 simulation are used.

Panel A: Case 1 Parameters Estimation

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
$\bar{\hat{\theta}}$	0.2165	0.5068	0.9109	1.6289	2.3727
$\bar{\hat{\sigma}}$	0.0053	0.0056	0.0062	0.0071	0.0081
$\bar{\hat{\mu}}$	0.0010	0.0010	0.0010	0.0010	0.0010

Panel B: Case 1 Estimation Standard Errors

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
$\bar{\hat{\theta}}$	0.0675	0.1076	0.1790	0.4249	0.8329
$\bar{\hat{\sigma}}$	0.0003	0.0004	0.0005	0.0008	0.0013
$\bar{\hat{\mu}}$	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Panel C: Case 1 Estimation MSE

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
$\bar{\hat{\theta}}$	0.0046	0.0116	0.0321	0.1808	0.6984
$\bar{\hat{\sigma}}$	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
$\bar{\hat{\mu}}$	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Panel D: P-values of statistical difference between the underlying and Case 1 OU parameters

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
$\bar{\hat{\theta}}$	0.9215	0.9700	0.9758	0.9635	0.9329
$\bar{\hat{\sigma}}$	0.9311	0.9599	0.9866	0.9514	0.9564
$\bar{\hat{\mu}}$	0.9824	0.9946	0.9813	0.9502	0.9755

Table 3

Underlying vs. Case 1: One Period Return Mean and Standard Deviation Comparison

This table reports the estimated mean, its estimated standard error, estimated standard deviation and estimated its standard error of returns in underlying values and in Case 1 for return first order autocorrelation, ρ(1) of -0.1, -0.2, -0.3, -0.4 and -0.45. Parameters μ_r and σ_r are the return mean and standard deviation in the underlying case and parameters $\overline{\widehat{\mu_r}}$ and $\overline{\widehat{\sigma_r}}$ are the averages of the return mean and standard deviation in case 1 computed using (24). SE indicates standard error. Panel A reports the return mean and return standard deviation and their standard errors for underlying values. Panel B reports the return mean and standard deviation and their standard errors in Case 1. Panel C reports p-values whether the estimated values of the Case 1 parameters are not statistically significantly different from the underlying parameter values. The notation '< 0.0001' indicates that the value is less than 0.0001. 200 observation period and 3000 simulation are used.

Panel A: Underlying Return Mean and Return Standard Deviation

ρ(1)	0	-0.1	-0.2	-0.3	-0.4	-0.45
μ_r	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
SE(μ_r)	NA	NA	NA	NA	NA	NA
σ_r	0.0050	0.0047	0.0045	0.0042	0.0039	0.0037
SE(σ_r)	NA	NA	NA	NA	NA	NA

Panel B: Case 1 Return Mean and Return Standard Deviation

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
$\overline{\widehat{\mu_r}}$	0.0010	0.0010	0.0010	0.0010	0.0010
SE($\overline{\widehat{\mu_r}}$)	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
$\overline{\widehat{\sigma_r}}$	0.0047	0.0045	0.0042	0.0039	0.0037
SE($\overline{\widehat{\sigma_r}}$)	0.0002	0.0002	0.0002	0.0002	0.0002

Panel C: P-values of statistical difference between the underlying and Case 1 return mean and standard deviation

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
$\overline{\widehat{\mu_r}}$	0.9824	0.9946	0.9813	0.9502	0.9755
$\overline{\widehat{\sigma_r}}$	0.9727	0.9774	0.9835	0.9173	0.9423

Table 4

One Period Return Mean and Standard Deviation: Case 1, Case 2 and Case 3

This table reports the estimated mean, estimated mean standard error, estimated standard deviation (Std Deviation) and estimated standard deviation standard error of returns for Case 2 and Case 3 for return first order autocorrelation, ρ(1) of -0.1, -0.2, -0.3, -0.4 and -0.45. Parameters $\overline{\widehat{\mu_r}}$ and $\overline{\widehat{\sigma_r}}$ are the averages of the return mean and standard deviation in case 1 computed using (24). Parameters $\overline{\tilde{\mu}_r}$ and $\overline{\tilde{\sigma}_r}$ are the averages of the return mean and standard deviation in case 2 computed using (24). Parameters $\overline{\dot{\mu}_r}$ and $\overline{\dot{\sigma}_r}$ are the averages of the return mean and standard deviation in case 3 computed using (24). SE indicates standard error. Panel A reports the return mean and return standard deviation and their standard errors for Case 1. Panel B reports the return mean and return standard deviation and their standard errors for Case 2. Panel C reports the return mean and standard deviation and their standard errors in Case 3. The notation '< 0.0001' indicates that the value is less than 0.0001. 200 observation period and 3000 simulation are used.

Panel A: Case 1 Return Mean and Return Standard Deviation

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
$\overline{\widehat{\mu_r}}$	0.0010	0.0010	0.0010	0.0010	0.0010
SE($\overline{\widehat{\mu_r}}$)	0.0000	0.0000	0.0000	0.0000	0.0000
$\overline{\widehat{\sigma_r}}$	0.0047	0.0045	0.0042	0.0039	0.0037
SE($\overline{\widehat{\sigma_r}}$)	0.0002	0.0002	0.0002	0.0002	0.0002

Panel B: Case 2 Return Mean and Return Standard Deviation

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
$\overline{\tilde{\mu}_r}$	0.0010	0.0010	0.0010	0.0010	0.0010
SE($\overline{\tilde{\mu}_r}$)	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
$\overline{\tilde{\sigma}_r}$	0.0048	0.0045	0.0043	0.0040	0.0038
SE($\overline{\tilde{\sigma}_r}$)	0.0003	0.0003	0.0003	0.0003	0.0003

Panel C: Case 3 Return Mean and Return Standard Deviation

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
$\overline{\dot{\mu}_r}$	0.0012	0.0012	0.0012	0.0012	0.0012
SE($\overline{\dot{\mu}_r}$)	0.0001	0.0001	0.0001	0.0001	0.0001
$\overline{\dot{\sigma}_r}$	0.0062	0.0055	0.0050	0.0047	0.0045
SE($\overline{\dot{\sigma}_r}$)	0.0003	0.0003	0.0003	0.0003	0.0003

Table 5

Relative Difference and Difference Test

This table reports the relative difference and statistical difference of return mean and standard deviation for Case 2 and Case 3 from Case 1 for return first order autocorrelation, ρ(1) of -0.1, -0.2, -0.3, -0.4 and -0.45. Parameters $\overline{\widehat{\mu_r}}$ and $\overline{\widehat{\sigma_r}}$ are the averages of the return mean and standard deviation in case 1 computed using (24). Parameters $\overline{\tilde{\mu}_r}$ and $\overline{\tilde{\sigma}_r}$ are the averages of the return mean and standard deviation in case 2 computed using (24). Parameters $\overline{\dot{\mu}_r}$ and $\overline{\dot{\sigma}_r}$ are the averages of the return mean and standard deviation in case 3 computed using (24). SE indicates standard error. Panel A reports the relative difference of return mean and standard deviation between Case 1 and Case 2. Panel B reports the relative difference of return mean and standard deviation between Case 1 and Case 3. Panel C reports p-values whether the estimated values of Case 2 are statistically significantly different from Case 1. Panel D reports p-values whether the estimated values of Case 3 are not statistically significantly different from Case 1. The notation '< 0.0001' indicates that the value is less than 0.0001. 200 observation period and 3000 simulation are used.

Panel A: Relative Difference of Return Mean and Standard Deviation of Case 2 from Case 1

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
$\overline{\tilde{\mu}_r}/\overline{\widehat{\mu_r}}$	1.0017	1.0006	1.0002	0.9996	0.9996
$\overline{\tilde{\sigma}_r}/\overline{\widehat{\sigma_r}}$	1.0062	1.0138	1.0190	1.0220	1.0113

Panel B: Relative Difference of Return Mean and Standard Deviation of Case 3 from Case 1

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
$\overline{\dot{\mu}_r}/\overline{\widehat{\mu_r}}$	1.1562	1.1572	1.1533	1.1503	1.1551
$\overline{\dot{\sigma}_r}/\overline{\widehat{\sigma_r}}$	1.3062	1.2281	1.1983	1.2005	1.2099

Panel C: P-values of statistical difference between the Case 2 and Case 1 return mean and standard deviation

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
$\overline{\tilde{\mu}_r}$ and $\overline{\widehat{\mu_r}}$	0.9924	0.9906	0.9939	0.9695	0.9635
$\overline{\tilde{\sigma}_r}$ and $\overline{\widehat{\sigma_r}}$	0.9100	0.8264	0.7949	0.7738	0.8784

Panel D: P-values of statistical difference between the Case 3 and Case 1 return mean and standard deviation

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
$\overline{\dot{\mu}_r}$ and $\overline{\widehat{\mu_r}}$	0.2708	0.2524	0.2542	0.2557	0.2368
$\overline{\dot{\sigma}_r}$ and $\overline{\widehat{\sigma_r}}$	< 0.0001	0.0003	0.0068	0.0088	0.0047

Table 6

Relative Difference and Difference Test

This table reports the relative difference and statistical difference of return mean and standard deviation for Case 2 and Case 3for return first order autocorrelation, ρ(1) of -0.1, -0.2, -0.3, -0.4 and -0.45. Parameters $\overline{\tilde{\mu}_r}$ and $\overline{\tilde{\sigma}_r}$ are the averages of the return mean and standard deviation in case 2 computed using (24). Parameters $\overline{\dot{\mu}_r}$ and $\overline{\dot{\sigma}_r}$ are the averages of the return mean and standard deviation in case 3 computed using (24). SE indicates standard error. Panel A reports the relative difference of return mean and standard deviation between Case 2 and Case 3. Panel B reports p-values whether the estimated values of Case 2 are statistically significantly different from Case 3. Panel C reports the relative gain in return mean and standard deviation when Case 2 estimation method is used against Case 3 in percentage term as defined in (25). The notation '< 0.0001' indicates that the value is less than 0.0001. 200 observation period and 3000 simulation are used.

Panel A: Relative Difference of Return Mean and Standard Deviation of Case 3 from Case 2

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
$\overline{\dot{\mu}_r}/\overline{\tilde{\mu}_r}$	1.1711	1.1542	1.1565	1.1531	1.1507
$\overline{\dot{\sigma}_r}/\overline{\tilde{\sigma}_r}$	1.4139	1.2982	1.2114	1.1759	1.1746

Panel B: P-values of statistical difference between the Case 2 and Case 3 return mean and standard deviation

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
$\overline{\dot{\mu}_r}$ and $\overline{\tilde{\mu}_r}$	0.0025	0.0022	< 0.0001	< 0.0001	< 0.0001
$\overline{\dot{\sigma}_r}$ and $\overline{\tilde{\sigma}_r}$	< 0.0001	0.0007	0.0144	0.0197	0.0075

Panel C: Relative Gain in Return Mean and Standard Deviation of Case 2 from Case 3

ρ(1)	-0.1	-0.2	-0.3	-0.4	-0.45
$\overline{\dot{\mu}_r}/\overline{\tilde{\mu}_r}$	17.11%	15.42%	15.65%	15.31%	15.07%
$\overline{\dot{\sigma}_r}/\overline{\tilde{\sigma}_r}$	41.39%	29.82%	21.14%	17.59%	17.46%