

ORIGINAL ARTICLE

Open Access



# Unbiased weighted variance and skewness estimators for overlapping returns

Stephen Taylor\*  and Ming Fang

## Abstract

This article develops unbiased weighted variance and skewness estimators for overlapping return distributions. These estimators extend the variance estimation methods constructed in Bod et. al. (*Applied Financial Economics* 12:155-158, 2002) and Lo and MacKinlay (*Review of Financial Studies* 1:41-66, 1988). In addition, they may be used in overlapping return variance or skewness ratio tests as in Charles and Darné (*Journal of Economic Surveys* 3:503-527, 2009) and Wong (*Cardiff Economics Working Papers*, 2016). An example using synthetic overlapping returns from a model fit to data from the SPY S&P 500 exchange traded fund is given in order to demonstrate under which circumstances the unbiased correction becomes significant in skewness estimation. Finally, we compare the effect of the HAC weighting schemes of Andrews (*Econometrica* 53:817-858, 1991) as a function of sample size and overlapping return window length.

**Keywords:** Overlapping returns, Variance and skewness estimation, Asset returns, Weighted estimators

## Introduction

Overlapping returns are used in many contexts in the finance and econometrics literature. Applications include variance ratio tests, regression parameter error estimation, and alternative resampling methods. Standard statistical inference and estimation techniques applied to overlapping return financial time series are typically biased. In addition, for such series, recent data is regularly viewed as more relevant than past information, which has resulted in the creation of weighted generalizations of estimation methodologies. This motivates the development of unbiased analogues of such estimators which we explore in the cases of the variance and skewness statistics. Our central aim is to construct unbiased weighted variance and skewness estimators for overlapping return distributions.

Several estimation procedures and hypothesis testing frameworks have been improved through the utilization of overlapping returns. In financial overlapping return applications, Lo and MacKinlay (1988) and Hansen and Hodrick (1980) demonstrate how overlapping returns may be used to increase the efficiency of statistics used in

variance ratio tests. Dunis and Keller (1995) developed a panel regression method based on overlapping returns, and Müller (1993) concludes utilizing overlapping returns in most applications will result in an overall increase in estimation precision of statistics that are a function of the overlapping returns when compared with their analogues for simple returns. In Jackwerth (2000), the author discovered that the overlapping return distribution for the S&P 500 is left-skewed and examined differences between risk neutral and realized distributions between overlapping and non-overlapping returns of the S&P 500 index and observed how associated risk aversion functions changed dramatically around the 1987 stock market crash. Wong (2016) develops skewness and kurtosis ratio tests for overlapping returns. The new weighted unbiased skewness estimator constructed below may be used as an input into any of these applications.

The idea of assigning greater weight to recent data and less weight to past data has been discussed in a number of econometric and financial studies. Past economic data may have little impact or be entirely irrelevant for present projections. In addition, by placing additional weight on recent data, associated estimation procedures tend to react more strongly to structural changes in the underlying assumption about the distribution the sample is drawn from than their uniformly

\*Correspondence: [smt@njit.edu](mailto:smt@njit.edu)  
Martin Tuchman School of Management at the New Jersey Institute of Technology, Newark, NJ, USA

weighted counterparts. For example, Tsokos (2010) shows that under nonstationary economic realization, weighted moving average models perform significantly better than the classical ARIMA model in forecasting stock prices. Andrews (1991) develops weighting schemes used in the estimation of covariance matrices assuming the underlying time series exhibits nontrivial autocorrelation and heteroskedasticity which we will utilize below. Weighted estimators are routinely used in practice as well. In particular, in Longerstaey and Spencer (1996), it is demonstrated that exponentially weighted moving average estimators incorporate external shocks more readily than equally weighted moving averages, thus providing a more realistic measure of current volatility.

Volatility and skewness estimation of financial return distributions has been the subject of a number of articles. Early examples include using maximum likelihood estimation to fit a model distribution to observed data and computing the associated model statistics in Fama (1965) and Mandelbrot (1963). More recent work has focused on the estimation of stochastic volatility models in Broto (2004). Time series techniques have also been widely applied to this task, c.f. (Tsay 2010). Measuring the asymmetry of financial return distributions has also been the central theme of many references. Grigoletto and Lisi (2006) and Wen and Yang (2009) find persistent non-trivial skewness is present in the simple daily return distributions of nearly every major international equity index. Xu (2007) shows that equity return distribution skewness is positively correlated with simultaneous returns and negatively correlated with lagged returns.

When working with overlapping returns, especially when encountering small sample sizes, bias effects from standard estimators, such as the sample variance, become important. In Lo and MacKinlay (1988), the authors provide a consistent but biased overlapping return variance estimator that has been used in several subsequent references, including Liu and He (1991), Fong et al. (1997), and Amélie and Olivier (2009). This estimator was improved in Bod et al. (2002) where the authors constructed an unbiased variance estimator for unweighted overlapping returns. Kluitman and Franses (2002) extended this work to develop an estimator that includes the case where the returns have nontrivial autocorrelation. Our main contribution is to extend these results by developing weighted unbiased variance and skewness estimators for overlapping return time series.

This article is organized as follows. We first fix notation and then derive an unbiased weighted estimator for the variance of a time series of overlapping returns. We give reduced expressions for this estimator in the cases of uniform and exponential weights. Next, we construct a similar weighted unbiased estimator for the skewness of an overlapping return distribution. We then demonstrate the

difference between a normalized version of the skewness estimator and the standard normalized sample skewness in a simulation which models the overlapping return distribution of the S&P 500 index, and then summarize our results. We finally compare the estimation of the weighted volatility and skewness of the overlapping return distribution of the S&P 500 index for various weighting schemes, sample sizes, and overlapping lengths and conclude with potential additional questions to explore.

## Methodology

We begin by establishing notation. Given integers  $n, q > 0$  with  $q < n$ , let  $p_t > 0$  for  $t = 0, \dots, n + q - 1$  denote an asset price time series by  $p_t$  and let  $r_t = p_t/p_{t-1} - 1$  be the associated simple returns where  $t = 1, \dots, n + q - 1$ . Following Bod et al. (2002) and Lo and MacKinlay (1988), we assume that  $r_t$  have zero mean,  $\mathbb{E}[r_t] = 0$ , covariance  $E[r_t r_s] = 0$  for any  $t > s$ , and equal finite variance  $\text{Var}(r_t) = \sigma^2 < \infty$ . We note that in Lo and MacKinlay (1988) and subsequent references, the authors show that these assumptions, referred to as the random walk version of the martingale hypothesis, do not hold for a variety of financial time series. This is achieved by assuming a null hypothesis that they hold and then demonstrating how variance ratios of overlapping returns may be used to reject this assertion. In this spirit, we proceed by defining the  $q$ -period overlapping returns  $y_t$  associated with  $r_s$  by

$$y_t = \sum_{s=t}^{t+q-1} r_s, \quad \text{for } t = 1, \dots, n. \quad (1)$$

We construct weighted unbiased estimators of the variance and skewness of  $y_t$  and pair a weight  $w_t$  with each  $y_t$  such that  $w_t > 0$  and  $\sum_{t=1}^n w_t = 1$ . Let  $W^{ts} = \sum_{k=t}^s w_k$  be the sum of the  $t$ -th through  $s$ -th weight, and note  $W^{1n} = 1$ .

We first derive an unbiased weighted overlapping return variance estimator  $\hat{\sigma}_y^2$ . We seek an estimator of the form

$$\hat{\sigma}_y^2 = C_1(n, q, w)^{-1} \sum_{t=1}^n w_t (y_t - \bar{y}^w)^2, \quad \bar{y}^w \equiv \sum_{t=1}^n w_t y_t, \quad (2)$$

where we find  $C_1$  such that  $\hat{\sigma}_y^2$  is an unbiased estimator of  $q\sigma^2$ . Note that since  $r_i$  are independent, the true variance of each  $y_t$  is  $\text{Var}(y_t) = q\sigma^2$  so that  $\hat{\sigma}^2$  is an unbiased estimator of the variance of the random variable from which  $y_t$  are sampled from if

$$\mathbb{E}[\hat{\sigma}_y^2] = q\sigma^2. \quad (3)$$

This may be viewed as a constraint that defines the constant  $C_1(n, q, w)$  which we determine by computing

$$\mathbb{E} \left[ \sum_t w_t (y_t - \bar{y}^w)^2 \right] = \sum_t w_t \left( \mathbb{E} y_t^2 - 2\mathbb{E} (y_t \bar{y}^w) + \mathbb{E} [(\bar{y}^w)^2] \right), \tag{4}$$

and noting that

$$\sum_t w_t \mathbb{E} (y_t \bar{y}^w) = \mathbb{E} \left( \bar{y}^w \sum_t w_t y_t \right) = \mathbb{E} [(\bar{y}^w)^2], \tag{5}$$

yields

$$\sum_t w_t \left( -2\mathbb{E} (y_t \bar{y}^w) + \mathbb{E} [(\bar{y}^w)^2] \right) = -\mathbb{E} [(\bar{y}^w)^2]. \tag{6}$$

Combining the above, we find

$$\mathbb{E} \left[ \sum_t w_t (y_t - \bar{y}^w)^2 \right] = \sum_t w_t \mathbb{E} y_t^2 - \mathbb{E} [(\bar{y}^w)^2] = q\sigma^2 - \text{Var}(\bar{y}^w), \tag{7}$$

where we note that  $\mathbb{E}(y_t) = \mathbb{E}(\bar{y}^w) = 0$ , and the last equality follows from  $\mathbb{E}(y_t^2) = \text{Var}(y_t) = q\sigma^2$  as well as  $\text{Var}(\bar{y}^w) = \mathbb{E} [(\bar{y}^w)^2] - \mathbb{E} [\bar{y}^w]^2 = \mathbb{E} [(\bar{y}^w)^2]$ .

In order to compute the variance term in the previous equation, we decompose the weighted average of  $y_t$  as

$$\bar{y}^w = \sum_{t=1}^{q-1} W^{1t} r_t + \sum_{t=q}^n W^{(t-q+1)t} r_t + \sum_{t=n+1}^{n+q-1} W^{(t-q+1)n} r_t. \tag{8}$$

One may arrive at this decomposition by viewing the individual return terms of the sum  $\bar{y}^w = \sum_t w_t y_t$  as a table with values  $w_t r_s$  whose first row has elements,  $w_1 r_1, w_1 r_2, \dots, w_1 r_q$  and final row is given by  $w_n r_n, w_n r_{n+1}, \dots, w_n r_{n+q-1}$ . Note that  $\bar{y}^w$  is equivalent to the sum of all values in this table. The first term in this decomposition corresponds to grouping all elements of this table above the diagonal whose edge is formed by the  $w_1 r_q$  and  $w_q r_q$  entries and factoring our common returns multiplied into varying weights. The final term can be arrived at by aggregating all terms below the diagonal formed by the  $w_{n-q+2} r_{n+1}$  and  $w_n r_{n+1}$  entries. The middle sum is computed by combining the remaining terms in the table.

Since each individual sum is composed of returns that are independent of all the returns in the other two sums, we find that

$$\text{Var}(\bar{y}^w) = \sigma^2 \left[ \sum_{t=1}^{q-1} (W^{1t})^2 + \sum_{t=q}^n (W^{(t-q+1)t})^2 + \sum_{t=n+1}^{n+q-1} (W^{(t-q+1)n})^2 \right]. \tag{9}$$

Solving for the unbiased constant from Eqs. (7) and (9), we arrive at

$$C_1(n, q, w) = \frac{1}{q\sigma^2} \mathbb{E} \left[ \sum_t w_t (y_t - \bar{y}^w)^2 \right] = 1 - \frac{1}{q} \left[ \sum_{t=1}^{q-1} (W^{1t})^2 + \sum_{t=q}^n (W^{(t-q+1)t})^2 + \sum_{t=n+1}^{n+q-1} (W^{(t-q+1)n})^2 \right]. \tag{10}$$

We note that in the case of uniform weights  $w_t = 1/n$ , this result reduces to that of Bod et. al. (2002), where  $C_B^{-1} = nC_1(n, q, w)$ . Secondly, in the case of exponential weights  $w_s = \alpha^{n-s}/C$  with  $C = (\alpha^n - 1)/(\alpha - 1)$ , one can show

$$C_1(n, q, w) = \frac{2\alpha \alpha^q - \alpha^{2n-q} - 1 + \alpha^{2n} - q\alpha^{n-1}(\alpha^2 - 1)}{q(\alpha^2 - 1)(\alpha^n - 1)^2}. \tag{11}$$

We now derive an unbiased skewness estimator in a similar manner. Assume that  $\mathbb{E}(r_t^3) = \gamma$ , so that  $\mathbb{E}(y_t^3) = q\gamma$ , and consider an estimator of the skewness of the overlapping return distribution of the form

$$\hat{\gamma}_y = C_2(n, q, w)^{-1} \sum_{t=1}^n w_t (y_t - \bar{y}^w)^3. \tag{12}$$

We would like to construct an unbiased estimator of  $q\gamma$  which requires

$$\mathbb{E}[\hat{\gamma}_y] = q\gamma. \tag{13}$$

To determine  $C_2$ , we compute

$$\mathbb{E} \left[ \sum_{t=1}^n w_t (y_t - \bar{y}^w)^3 \right] = \sum_{t=1}^n w_t \mathbb{E} \left[ y_t^3 - 3y_t^2 \bar{y}^w + 3y_t (\bar{y}^w)^2 - (\bar{y}^w)^3 \right]. \tag{14}$$

Decomposing  $y_t$  into independent  $r_t$  sums allows one to calculate the first term with

$$\mathbb{E}(y_t^3) = \mathbb{E} \left[ \sum_{s=t}^{t+q-1} r_s^3 + 3 \sum_{s \neq k} r_s^2 r_k + \sum_{s \neq k \neq l} r_s r_k r_l \right] = \sum_{s=t}^{t+q-1} \mathbb{E}(r_s^3) = q\gamma. \tag{15}$$

For the second term, we note that

$$\mathbb{E}[y_t^2 \bar{y}^w] = \sum_s w_s \mathbb{E}(y_t^2 y_s). \tag{16}$$

Computing the expectation, we find,

$$\begin{aligned} \mathbb{E} [y_t^2 y_s] &= \mathbb{E} \left[ \sum_{k=s}^{s+q-1} r_k \left( \sum_{l=t}^{t+q-1} r_l \right)^2 \right] \\ &= \mathbb{E} \left[ \sum_{k=s}^{s+q-1} r_k \left( \sum_{l=t}^{t+q-1} r_l^2 + \sum_{p \neq m}^{t+q-1} r_m r_p \right) \right] \end{aligned} \tag{17}$$

$$= \mathbb{E} \left[ \sum_{k=t}^{t+q-1} (r_s + \dots + r_{s+q-1}) r_k^2 \right]. \tag{18}$$

Note that only terms of the form  $r_t^3$  are non-trivial in expectation. Splitting into two cases, we have

$$\mathbb{E} [y_t^2 y_s] = \begin{cases} (s-t+q)\gamma & s \leq t \leq s+q-1 \\ (t-s+q)\gamma & t \leq s \leq t+q-1 \end{cases} \tag{19}$$

$$= (q - |t - s|)\gamma, \quad 0 \leq |t - s| \leq q - 1. \tag{20}$$

Combining Eqs. (16) and (20), we find that

$$\sum_{t=1}^n w_t \mathbb{E} [y_t^2 \bar{y}^w] = \gamma \sum_{t,s=1}^n w_t w_s (q - |t - s|) 1_{\{|t-s| \in [0, q-1]\}}. \tag{21}$$

For the final two terms, note

$$\begin{aligned} \mathbb{E} (y_t (\bar{y}^w)^2) &= \sum_{s,k} w_s w_k \mathbb{E}(y_t y_s y_k), \quad \mathbb{E} (\bar{y}^w)^3 \\ &= \sum_{t,s,k} w_t w_s w_k \mathbb{E}(y_t y_s y_k), \end{aligned} \tag{22}$$

so we may simplify

$$\sum_{t=1}^n w_t [3y_t (\bar{y}^w)^2 - (\bar{y}^w)^3] = 2\mathbb{E} [(\bar{y}^w)^3]. \tag{23}$$

Using the decomposition in Eq. (14), by independence of the terms

$$\begin{aligned} \mathbb{E} [(\bar{y}^w)^3] &= \gamma \left[ \sum_{t=1}^{q-1} (W^{1t})^3 + \sum_{t=q}^n (W^{(t-q+1)t})^3 \right. \\ &\quad \left. + \sum_{t=n+1}^{n+q-1} (W^{(t-q+1)n})^3 \right]. \end{aligned} \tag{24}$$

Combining results from Eqs. (15), (21), and (24), we arrive at an expression for  $C_2$  given by

$$\begin{aligned} C_2(n, q, w) &= 1 - \frac{3}{q} \sum_{\substack{t,s=1 \\ |t-s| < q}}^n w_t w_s (q - |t - s|) \\ &\quad + \frac{2}{q} \left[ \sum_{t=1}^{q-1} (W^{1t})^3 + \sum_{t=q}^n (W^{(t-q+1)t})^3 + \sum_{t=n+1}^{n+q-1} (W^{(t-q+1)n})^3 \right]. \end{aligned} \tag{25}$$

In the case of uniform weights  $w_t = 1/n$ , the terms in this expression simplify to

$$\begin{aligned} \sum_{\substack{t,s=1 \\ |t-s| < q}}^n w_t w_s (q - |t - s|) &= \frac{q}{n} + \frac{2}{n^2} \sum_{t=1}^{q-1} (n-t)(q-t) \\ &= \frac{q}{3n^2} (1 + 3nq - q^2) \end{aligned} \tag{26}$$

$$\begin{aligned} \sum_{t=1}^{q-1} (W^{1t})^3 + \sum_{t=q}^n (W^{(t-q+1)t})^3 + \sum_{t=n+1}^{n+q-1} (W^{(t-q+1)n})^3 \\ = \frac{1}{n^3} \left[ \sum_{t=1}^{q-1} t^3 + \sum_{t=q}^n q^3 + \sum_{t=n+1}^{n+q-1} (n-t+q)^3 \right] \end{aligned} \tag{27}$$

$$= \frac{q^2}{2n^3} (1 + 2nq - q^2), \tag{28}$$

which yields a simple form for  $C_2$  given by

$$C_2(n, q, w) = \frac{(n-q+1)(n-q)(n-q-1)}{n^3}. \tag{29}$$

We finally note that it is possible to derive a closed form expression for  $C_2$  in the case of exponential weights that were previously considered for the variance estimator; however, the expression is quite lengthy and we omit it here but it is available upon request. Finally, we turn to a simulation in order to understand for what parameter pairs  $(n, q)$  the effects of the unbiased skewness estimator are most significant.

### Empirical studies and results

We now develop a simulation to compare the relative error of the uniformly weighted unbiased skewness estimator  $\hat{\gamma}_y$  and the standard unbiased sample skewness estimator which may be found in Zwillinger and Kokoska (2000). We first construct a dataset of end of day simple returns calculated from closing prices for the SPY exchange traded fund from January 1, 2012, to December 31, 2016. This was achieved using Bloomberg's Python API and an associated wrapper package named *tia*. We downloaded historical end of day closing prices identified with Bloomberg's `PX_LAST` field that are both split and dividend adjusted. This time series was fully populated with data, and hence, there was no need to fill in missing values.

Next, we fit a model distribution to this data in order to establish a framework to test the weighted unbiased skewness and kurtosis estimators in a setting where the true values of these statistics are known which closely approximates actual market data. To this end, we let  $X$  denote a

skew normal distribution whose probability density function is given by

$$p(x; a, b, c) = \frac{2}{b} \phi\left(\frac{x-a}{b}\right) \Phi\left(c \frac{x-a}{b}\right), \tag{30}$$

where here  $\phi$  and  $\Phi$  are the probability and cumulative distribution functions of a standard normal random variable and  $b > 0$ . The skew normal distribution has mean  $\mu$  and variance  $\sigma^2$  defined by

$$\begin{aligned} \mu &= a + bd\sqrt{\frac{2}{\pi}}, \\ \sigma^2 &= b^2 \left(1 - \frac{2d^2}{\pi}\right), \quad \text{where } d = \frac{c}{\sqrt{1+c^2}}. \end{aligned} \tag{31}$$

The normalized skewness  $\gamma$  of this distribution is given by

$$\gamma = \frac{\text{Skew}(X)}{\text{Var}(X)^{3/2}} = \frac{4-\pi}{2} \frac{(d\sqrt{2/\pi})^3}{(1-2d^2/\pi)^{3/2}}. \tag{32}$$

We fit this distribution to the SPY daily return data using maximum likelihood estimation. Let the likelihood function of this model be denoted by  $\mathcal{L}$ . We find the model parameters by directly maximizing the log-likelihood function

$$\begin{aligned} \hat{a}, \hat{b}, \hat{c} &= \underset{a,b,c}{\operatorname{argmax}} \ln \mathcal{L} = n \ln \left(\frac{2}{b}\right) \\ &+ \sum_{t=1}^n \ln \phi\left(\frac{x_t - a}{b}\right) + \Phi\left(c \frac{x_t - a}{b}\right) \end{aligned} \tag{33}$$

using a BFGS optimizer developed in Fletcher (1987) and determine that the best fit parameters are given by  $\hat{a} = 0.00640$ ,  $\hat{b} = 0.01006$ , and  $\hat{c} = -1.1029$ , which have associated mean, variance, and normalized skewness given by  $4.528 \times 10^{-4}$ ,  $6.5789 \times 10^{-5}$ , and  $-0.1689$ , respectively. Returns will be drawn from this distribution in the Monte Carlo simulation considered below. Specifically, this simulation consists of sampling  $n$  values from the MLE distribution, computing the  $q$ -period overlapping returns of these time series, then calculating the normalized unbiased sample skewness  $\tilde{\gamma}_s$  (see Zwillinger and Kokoska (2000)) and the normalized unbiased skewness  $\tilde{\gamma}_y$  with uniform weights given by  $\tilde{\gamma}_y = \hat{\gamma}_y / \hat{\sigma}_y^3$ . We then compute the relative error  $|\tilde{\gamma}_s - \tilde{\gamma}_y| / \tilde{\gamma}_y$ , between the different estimators for each simulation, repeat the above process 10,000 times, and display the mean percentage errors in Table 1 for distinct  $(n, q)$  pairs.

We omit cases where the number of overlapping returns  $n - q \leq q$ , and first note that as the sample size increases, the error between the two estimators decreases for any fixed  $q$  value. However, when  $q/n$  is relatively large, say greater than 5%, then there are significant differences between the two estimators.

Next, we explore several weighting schemes described in Andrews (1991) which are widely used for covariance matrix estimation in the presence of heteroskedasticity and autocorrelation. Specifically, we consider weights constructed by Bartlett (Oppenheim et al. 1999), Parzen (White 1980), Tukey-Hamming (Blackman and Tukey 1958), and the Quadratic Spectral weights of Priestley (1962) and Epanechnikov (1969). These weights are defined in terms of a kernel function  $k(\cdot)$  and are given by  $w_t = k(bt/T)$  where  $T$  is a bandwidth parameter and  $b$  is a scaling constant in Zeileis (2004) and Zwillinger (2000). There are many references that study the problem of optimal bandwidth selection c.f. (Lazarus et al. 2017; Newey and West 1994; Stock and Watson 2011; Wooldridge 2006); however, we are interested in constructing reasonable weighting schemes to place importance on more recent over prior data. We found that setting the bandwidth to the sample size and  $b = 1.2$  achieves this aim.

In Fig. 1, we plot the kernel functions associated with these weighting schemes given in Andrews (1991) which are defined by

$$k_{BT}(x) = \begin{cases} 1 - |x| & |x| \leq 1 \\ 0 & |x| > 1, \end{cases} \tag{34}$$

$$k_{PR}(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & |x| \leq 1/2 \\ 2(1 - |x|)^3 & 1/2 \leq |x| \leq 1 \\ 0 & |x| > 1, \end{cases} \tag{35}$$

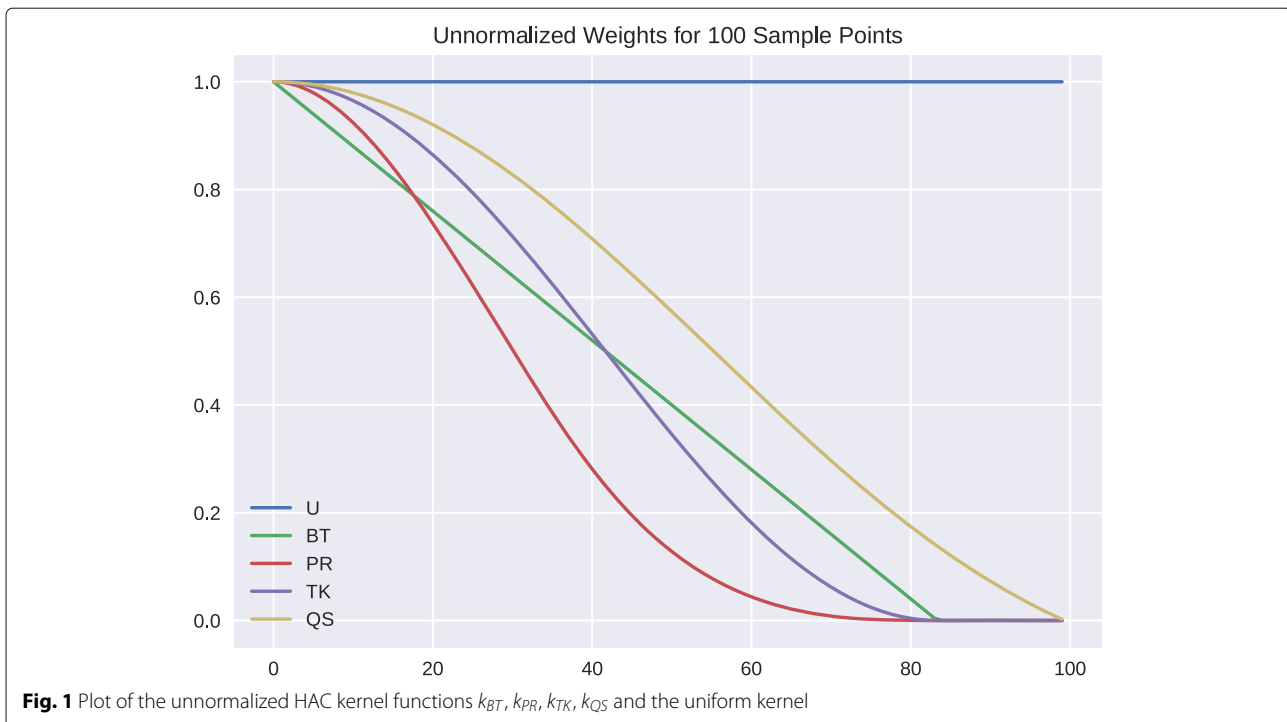
$$k_{TH}(x) = \begin{cases} (1 + \cos(\pi x))/2 & |x| \leq 1 \\ 0 & |x| > 1, \end{cases} \tag{36}$$

$$k_{QS}(x) = \frac{25}{12\pi^2 x^2} \left( \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right). \tag{37}$$

Note that when using weighted estimators, one effectively reduces the original sample size. For example, in the extreme case of binary zero or one valued weights, only the weights with value one contribute to the estimator which reduces the sample size to the percentage of one valued weights. In this example, we can find the percentage sample size reduction by approximating the

**Table 1** Mean relative percentage errors between the normalized unbiased sample skewness  $\tilde{\gamma}_s$  and the normalized unbiased skewness  $\tilde{\gamma}_y$  for varying sample sizes  $n = 32, \dots, 16384$ , and overlapping return periods  $q = 2, \dots, 128$

$n/q$	2	4	8	16	32	64	128
32	14.24	33.14	68.01	*	*	*	*
64	7.08	16.55	35.36	69.79	*	*	*
128	3.53	8.25	17.71	36.47	70.67	*	*
256	1.76	4.11	8.83	18.29	37.03	71.12	*
512	0.88	2.05	4.41	9.12	18.58	37.31	71.34
1024	0.44	1.03	2.20	4.55	9.27	18.72	37.45
16384	0.03	0.06	0.14	0.28	0.58	1.16	2.34



area under each weight curve. In reference to the uniform weights which we take to have normalized area of 1, the HAC weights effectively reduce the sample size by PR: 31%, BT,TK: 41%, and QS: 54% so that uniform weights have approximately two to three times the sample size of these weighting schemes.

Next, we examine how estimation of the unbiased weighted standard deviation and skewness estimators varies as a function of the overlapping return period  $q$ , sample size  $n$ , and weighting scheme using the SPY dataset previously described. We consider overlapping return periods of 5, 21, and 63 sample points which correspond to weekly, monthly, and quarterly aggregation windows for our example daily return data. Next, we truncate the sample size to 256, 512, and 1024 data points which roughly correspond to 1, 2, and 5-year time periods, using a trailing truncation window on the SPY returns.

Estimation results are presented in Table 2 where overlapping return standard deviations are displayed as percentages. We first note that estimates for uniform weights tend to be outliers as they include two to three times the effective sample size as the other weighting schemes. Next, note that overall standard deviation values are greatest in the  $n = 512$  case, slightly lower for  $n = 256$ , and considerably lower for  $n = 1024$ . This is due to the historical volatility of the S&P 500 being relatively high during the mid 2015 to early 2016 time period and lower in prior years over the five year window being considered. One may also note that the overlapping return distribution becomes increasingly negatively skewed as a result.

Finally, we compare how skewness estimation varies over time for different weighting schemes. In particular, we consider a 4-year window from 1/1/2012 to 1/1/2016 and estimate the unbiased skewness using a 252-day lookback period for each of the HAC estimators and a 126-day lookback period for the uniformly weighted estimator which is selected to ensure that the sample sizes are on par with one another. This procedure is carried out on a rolling basis, and results are plotted in Fig. 2.

We note that the general forms of the time series in Fig. 2 tend to be similar for the majority of dates displayed. The HAC weighted estimators are more reactive than the uniform estimator and do not exhibit single day jumps that are as large in magnitude as the HAC uniform estimator.

### Discussion and conclusions

In summary, we have derived closed form expressions for weighted unbiased variance and skewness estimators. We also developed simplified expressions for these estimators in the case of exponential weights for the variance estimator and uniform weights for both estimators. The differences between the standard unbiased sample skewness and new normalized unbiased skewness estimators were demonstrated to be significant in the case of skewness estimation for SPY end of day return data for HAC weighting schemes.

We note that as in Bod et al. (2002) and Lo and MacKinlay (1988), we assume returns satisfy the random walk version of the martingale hypothesis which generally does not hold for financial time series. An

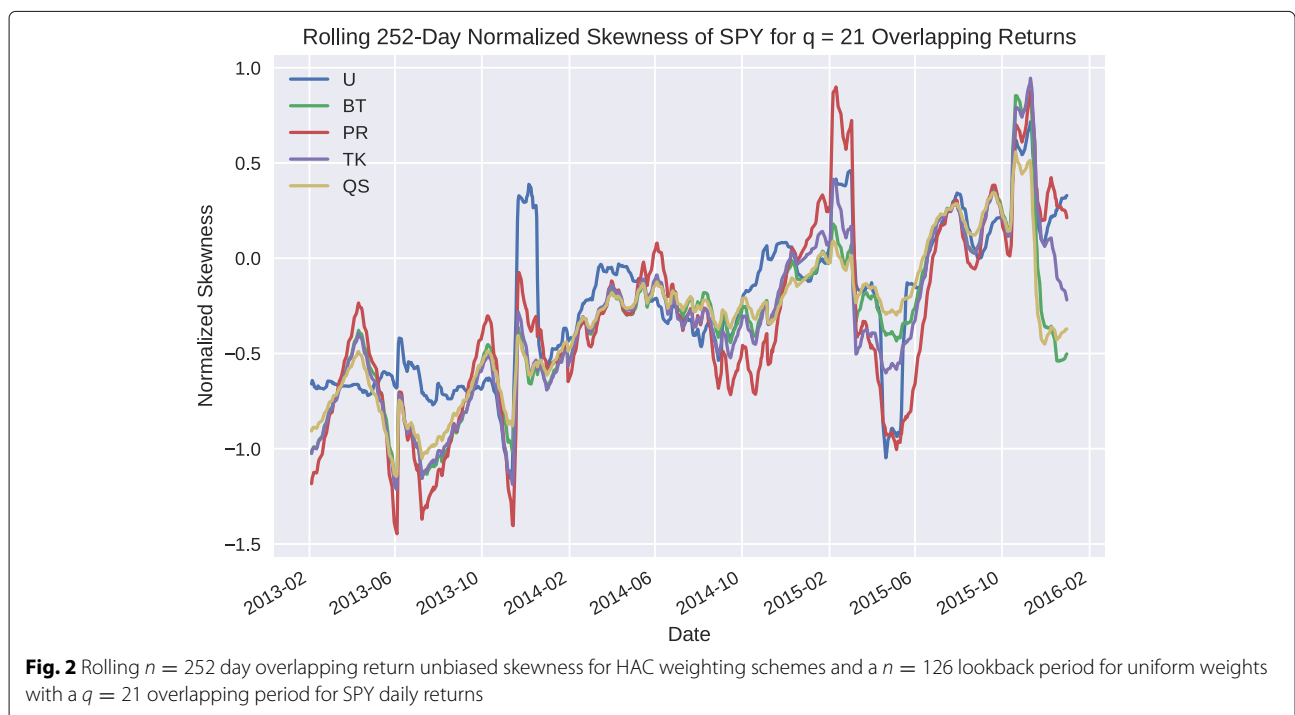
**Table 2** Comparison of unbiased overlapping return standard deviation and skewness estimators as a function of weighting scheme, sample size  $n$ , and overlapping return period  $q$

	$n = 256$			$n = 512$			$n = 1024$		
	$q = 5$	$q = 21$	$q = 63$	$q = 5$	$q = 21$	$q = 63$	$q = 5$	$q = 21$	$q = 63$
Std(%)									
<i>U</i>	1.764	3.146	4.086	1.885	3.459	5.225	1.729	3.130	4.387
<i>BT</i>	2.167	3.810	4.359	1.999	3.519	5.330	1.611	2.847	3.845
<i>PR</i>	2.409	4.321	4.679	1.894	3.196	4.963	1.512	2.634	3.260
<i>TK</i>	2.215	3.892	4.388	1.997	3.480	5.328	1.556	2.730	3.501
<i>QS</i>	2.038	3.561	4.246	2.001	3.555	5.380	1.648	2.930	3.978
Skewness									
<i>U</i>	-0.489	-0.296	0.356	-1.038	-0.356	-0.467	-0.858	-0.400	-0.866
<i>BT</i>	-0.612	-0.525	0.299	-1.200	-0.310	-0.343	-0.748	-0.401	-0.985
<i>PR</i>	-0.582	-0.752	0.294	-1.355	-0.387	-0.668	-0.508	-0.335	-0.488
<i>TK</i>	-0.616	-0.623	0.334	-1.280	-0.308	-0.316	-0.610	-0.328	-0.801
<i>QS</i>	-0.601	-0.446	0.318	-1.184	-0.298	-0.318	-0.791	-0.401	-0.986

Here, we use the following abbreviations for different kernel based weighting schemes: *U* uniform, *BT* Bartlett, *PR* Parzen, *TK* Tukey-Hanning, and *QS* Quadratic Spectral

interesting future application of the skewness estimator would be to develop a hypothesis test for this assumption which may compliment the results in Lo and MacKinlay (1988). For additional future work, it would be of interest to consider as in Kluitman and Franses (2002) analogues of the weighted unbiased variance and skewness estimators under the assumption that the return process satisfies an AR(1), MA(1) or more

general time series model. This will require repeating the above derivations and retaining terms of the form  $\mathbb{E}[r_t r_s]$  and  $\mathbb{E}[r_t r_s r_k]$  for  $t \neq s \neq k$  which are no longer trivial but depend on the underlying model one assumes for the return process. Then, one could fit such models to market data and compare values of the two estimators. We anticipate the estimator will strongly depend on the sign of the AR(1) lag parameter and the



white noise parameter of the MA(1) model as shown in Kluitman and Franses (2002) but leave this for a future study.

#### Acknowledgements

We thank the referees for their comments which greatly improved both the language and content of this article. ST would like to thank Michael Ehrlich for comments that improved the exposition of this article.

#### Authors' contributions

ST developed and derived the main estimators in this article. MF performed the literature review, simulations, and a verification of calculations. Both authors read and approved the final manuscript.

#### Authors' information

ST is an Assistant Prof. of Finance at the New Jersey Institute of Technology. MF is an Assistant Prof. of Accounting at the New Jersey Institute of Technology. Both ST and MF are new faculty members in the first years of their academic career.

#### Competing interests

The authors declare that they have no competing interests.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 16 November 2017 Accepted: 1 June 2018

Published online: 17 November 2018

#### References

- Andrews, DWK (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, 53(3), 817–858.
- Bod, P, Blitz, D, Franses, PH, Kluitman, R (2002). An Unbiased Variance Estimator for Overlapping Returns. *Applied Financial Economics*, 12(3), 155–158.
- Broto, C (2004). Estimation methods for stochastic volatility models: a survey. *Journal of Economic Surveys*, 18(5), 613–649.
- Amélie, C, & Olivier, D (2009). Variance ratio tests of random walk: An overview. *Journal of Economic Surveys*, 3, 503–527.
- Dunis, C, & Keller, A (1995). Efficiency Tests with Overlapping Data: An Application to the Currency Option Market. *European Journal of Finance*, 1, 345–66.
- Epanechnikov, VA (1969). Non-parametric Estimation of a Multivariate Probability Density. *Theory of Probability and Its Applications*, 14, 153–158.
- Fama, E (1965). The behavior of stock prices. *Journal of Business*, 38, 34–105.
- Fletcher, R (1987). *Practical Methods of Optimization*. Wiley. <https://www.amazon.com/Practical-Methods-Optimization-R-Fletcher/dp/0471494631>.
- Fong, WM, Koh, SK, Ouliaris, S (1997). Joint variance-ratio tests of the martingale hypothesis for exchange rates. *Journal of Business and Economic Statistics*, 15, 51–59.
- Grigoletto, M, & Lisi, F (2006). Looking for skewness in financial time series. *Working Paper Series*, 7. [http://paduaresearch.cab.unipd.it/7094/1/2006\\_7\\_20070123084924.pdf](http://paduaresearch.cab.unipd.it/7094/1/2006_7_20070123084924.pdf).
- Hansen, L, & Hodrick, RJ (1980). Forward exchange rates as optimal predictors of future spot rates: an econometric analysis. *Journal of Political Economy*, 88, 829–853.
- Jackwerth, JC (2000). Recovering Risk Aversion from Options Prices and Realized Returns. *Review of Financial Studies*, 13(2), 433–451.
- Xu, J (2007). Price Convexity and Skewness. *The Journal of Finance*, 62(5), 2521–2552.
- Kluitman, R, & Franses, PH (2002). Estimating volatility on overlapping returns when returns are autocorrelated. *Applied Mathematical Finance*, 9(3), 179–188.
- Lazarus, E, Lewis, DJ, Stock, JH, Watson, MW (2017). HAR Inference: Recommendations for Practice. *JBES Invited Paper*. <https://scholar.harvard.edu/elazarus/publications/har-inference-recommendationspractice-jbes-invited-paper>.
- Liu, CY, & He, J (1991). A variance ratio test of random walks in foreign exchange rates. *Journal of Finance*, 46, 773–785.
- Lo, AW, & MacKinlay, AC (1988). Stock Market Prices do not Follow Random Walks; Evidence from a Simple Specification Test. *Review of Financial Studies*, 1, 41–66.
- Mandelbrot, B (1963). The variation of certain speculative prices. *Journal of Business*, 36(4), 394–419. [https://web.williams.edu/Mathematics/sjmiller/public\\_html/341Fa09/econ/Mandelbroit\\_VariationCertainSpeculativePrices.pdf](https://web.williams.edu/Mathematics/sjmiller/public_html/341Fa09/econ/Mandelbroit_VariationCertainSpeculativePrices.pdf).
- Müller, UA (1993). Statistics of variables observed over overlapping intervals. [https://ideas.repec.org/p/wop/olaswp/\\_010.html](https://ideas.repec.org/p/wop/olaswp/_010.html).
- Newey, WK, & West, KD (1994). Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies*, 61, 631–653.
- Oppenheim, AV, Schaffer, RV, Buck, JR (1999). *Discrete-Time Signal Processing*. Prentice-Hall Signal Processing Series. <https://www.amazon.com/Discrete-Time-Signal-Processing-3rd-Prentice-Hall/dp/0131988425>.
- Priestley, MB (1962). Basic Considerations in the Estimation of Spectra. *Technometrics*, 4, 551–564.
- Longerstae, J, & Spencer, M (1996). RiskMetrics Technical Document. Fourth Edition. <https://www.msci.com/documents/10199/5915b101-4206-4ba0-ae2-3449d5c7e95a>.
- Stock, JH, & Watson, MW (2011). *Introduction to Econometrics Third Edition*. Addison-Wesley. Pearson Education Limited, Edinburgh Gate, Harlow, Essex CM20 2JE, England.
- Tsay, RS (2010). *Analysis of Financial Time Series*. 111 River St. Hoboken: Wiley.
- Tsokos, CP (2010). K-th Moving, Weighted and Exponential Moving Average for Time Series Forecasting Models. *European Journal of Pure and Applied Mathematics*, 3(3), 406–416.
- Blackman, RB, & Tukey, JW (1958). *The measurement of power spectra*. Dover: Dover Publications.
- White, H (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test of Heteroskedasticity. *Econometrica*, 48, 817–838.
- Wong, W (2016). Skewness and Kurtosis Ratio Tests: With Applications to Multiperiod Tail Risk Analysis. *Cardiff Economics Working Papers*. <https://www.econstor.eu/handle/10419/174121>.
- Wooldridge, JM (2006). *Introductory Econometrics: A Modern Approach Third Edition*. Mason: Thomson.
- Wen, F, & Yang, X (2009). Skewness of Return Distribution and Coefficient of Risk Premium. *Jr Syst Sci & Complexity*, 22, 360–371.
- Zeileis, A (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software*, 11(10), 1–17.
- Zwillinger, D, & Kokoska, S (2000). *CRC Standard Probability and Statistics Tables and Formulae*. Boca Raton: Chapman & Hall.

Submit your manuscript to a SpringerOpen journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)