Journal of Applied Finance & Banking, vol. 3, no. 6, 2013, 143-160
ISSN: 1792-6580 (print version), 1792-6599 (online)
Scienpress Ltd, 2013

Towards Unbiased Portfolio Daily Returns

Yuxing Yan[1]

Abstract

This paper describes a new method to generate unbiased equal-weighted portfolio daily returns by removing the impacts of bid-ask bounce and non-synchronous trading. For example, for the CRSP daily equal-weighted market index over 1964 to 1993 (EWRETD), the annual bias of the time series generated by our method is 0.05%, considerably smaller than 6% as reported by Canina et al. (1998). In addition, we also discuss the research impact by using both biased and unbiased daily EWRETD on beta, alpha, volatility and event study. The paper concludes that the new method should be applied for future estimation of portfolio daily returns which can be either equal-weighted or value-weighted.

JEL classification numbers: G10, G11
Keywords: Equal-weighted market index, Systematic bias, Methodology

1 Introduction

In finance, it has been a common practice for researchers to use portfolios instead of individual stocks. For example, Pastor and Stambaugh (2003) divide all stocks into deciles according to liquidity. In designing their famous 3-factor model, Fama and French (1992, 1993) group stocks into 6 portfolios by size and book-to-market ratio. Easley et al. (2010) analyze the impact of informed trading by classifying stocks into deciles based on Probability of Informed Trading. By suggesting a now-famous momentum effect, Jegadeesh and Titman (1993, 2001) assembly stocks into deciles on their past performance. Compared with individual stocks, portfolios will dampen or remove the influences of extreme cases (stocks). The statistical properties of a portfolio are more stable than those of individual stocks. After portfolios are formed, their returns are estimated and compared based on various strategies or hypotheses.
To estimate portfolio returns, value-weighed and equal-weighted are two commonly used

[1]Department of Economics and Finance, Canisius College, 2001 Main Street, Buffalo, New York, USA. Telephone: (716) 888-2604, e-mail: yuxing.yan@canisius.edu

Article Info: *Received :* September 20, 2013. *Revised :* October 16, 2013.
Published online : November 1, 2013

methods. For various reasons, researchers may prefer one weighting scheme to the other. For instance, Conrad and Kaul (1989) and Bollen and Busse (2004) use a value-weighted scheme, while Jegadeesh and Titman (1993, 2001), Sloan (1996), Amihud (2001), Lesmond et al. (2004) apply an equal-weighted one. However, some researchers apply both weighting schemes, respectively (see Pastor and Stambaugh 2003). Occasionally, a mixture of two is utilized, say, by Fama and French (1992, 1993) who implement a mixed-scheme: value-weighted for their basic 7 portfolios and equal-weighted for their final factors. In comparison, the equal-weighed scheme is relatively easier to implement than the value-weighted one, but the former is more problematic. For example, how to estimate monthly equal-weighted portfolio returns when individual stock's monthly returns available? The answer is obvious: take averages. But if replace "monthly" with "daily" in the above question, what would be the answer then? Most of us would probably offer exactly the same one: take averages. Unfortunately, this answer is incorrect. The reason is that bid-ask bounce (Blume and Stambaugh, 1983) and non-synchronous trading (Roll, 1983) have little impact on monthly returns while exerting a strong influence on daily returns, especially on small stocks. This is the reason why many researchers resort to the monthly returns instead daily ones purely to avoid potential contamination caused by the microstructure.

This is the very purpose of this paper: Generating unbiased portfolio daily returns by using an innovative method that makes it possible for the average annual bias of the CRSP daily equal-weighted index to be only 0.05%, a near-perfect result which is 120 times more accurate than the 6% as reported (Canina et al. 1998). In what follows, the paper will first provide evidence of biased results based on the current method and introduce a new method. It will then present empirical results and statistical properties, followed by a discussion of many practical applications of the new method and some its implications for future research. The paper concludes itself with a strong recommendation that this new method should be applied for future estimation of portfolio daily returns.

2 Evidence of Biased Results Generated

The CRSP equal-weighted market index (*EWRETD*) was chosen to be the focus for analysis for several reasons. To begin, it is well examined and documented in the literature that the CRSP equal-weighted market indices are biased (e.g., Canina et al.1998; Loughran and Ritter 1995). More importantly, a market index itself is a portfolio except with more stocks in it. If it had problems, portfolios formed based on various criterions would be more problematic since the diversifying effect of a market index dampens any adverse effect dramatically. Above and beyond, the CRSP equal-weighted indices are widely used in research; thus their bias-correction will have a significantly positive impact in finance.

The word "bias" is justly used for assessing the CRSP equal-weighted market indices on the basis of two pieces of revealing evidence: a) a daily time series that suffers severely from the negative impact of the microstructure, and b) the inconsistency with its monthly counter-party, which means a totally different result we obtain when we compound biased daily portfolio returns in estimating a long-term return. In their application of the CRSP daily *EWRETD*, for example, Loughran and Ritter (1995) uncovered a hugely upward bias - compounding 60-month CRSP NYSE/AMEX *EWRETD* resulted in a return of 154% between 1974 and 1978. What is worse is that compounding the daily *EWRETD*

gives a value of 243%, which means a 58% overestimation. Canina et al. (1998) claim that compounding the CRSP equal-weighted daily index resulted in an astoundingly huge bias: 6% per year from 1964 to 1993. The contributing factors to such a bias include: a) bid-ask bounce (Blume and Stambaugh, 1983; Venkatesh, 1992), b) non-synchronous trading (Roll, 1983), c) timing of dividends (Canina et al., 1998) and d) conversion itself, which will be explained in section 2.

3 A Buy-and-Hold Multi-Day Method (BHMD)

A new methodology is proposed to remove the impacts of bid-ask bounce and non-synchronous trading. The new method is called Buy-and-Hold Multi-Day (*BHMD*) for its uses multiple days daily returns and form buy-and-hold portfolios. First, define a t-day buy-and-hold equal-weighted index return as:

$$\mathrm{EWRETD}_{\mathrm{BH}}^{\mathrm{t}} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\prod_{\mathrm{s}=1}^{\mathrm{t}}\left(1+\mathrm{R}_{\mathrm{i,s}}^{\mathrm{d}}\right)-1 \tag{1}$$

where $EWRETD^{t}_{BH}$ is a buy-and-hold (*BH*) equal-weighted portfolio return from day 1 to day *t*, *t* is the number of trading days from the beginning of a month, $R^{d}_{i,s}$ is a daily return with dividends for stock *i* on day *s*, and *N* is the number of stocks in the portfolio. For a daily equal-weighted index ($EWRETD^{t}_{daily}$) on day *t* the following formula is applied:

$$\mathrm{EWRETD}_{\mathrm{daily}}^{\mathrm{t}} = \frac{1+\mathrm{EWRETD}_{\mathrm{BH}}^{\mathrm{t}}}{1+\mathrm{EWRETD}_{\mathrm{BH}}^{\mathrm{t-1}}}-1 \tag{2}$$

This *BHMD* method borrows heavily from Blume and Stambaugh (1983): a buy-and-hold strategy reduces or eliminates the impacts of bid-ask bounce and non-synchronous trading.[2] Despite its close association, five aspects distinguish the *BHMD* method from Blume and Stambaugh (1983). First, the *BHMD* method is operationally straightforward: only daily returns are needed, while they use daily prices and assume equality of the initial prices. Second, they use two days' prices, while the *BHMD* method utilizes multiple days' returns, i.e., using more information. Third, the performance of the *BHMD* method is much better: *BHMD* removes more than 99% of the bias, while they achieve 85%.[3] Fourth, they assume no dividends, while the *BHMD* method relaxes this constraint by applying returns with dividends. Fifth, the new generated daily *EWRETD* based on the *BHMD* method is consistent with the CRSP monthly *EWRETD*, while their setting implies a daily rebalancing which differs from that of monthly *EWRETD*. In summary, the *BHMD* method has solved, once and for all, the inconsistency between daily and monthly portfolio returns.

---

[2]See the proof in Blume and Stambaugh (1983).

[3]By design, the BHMD method is supposed to remove all the biases. The remaining tiny bias is because we could not replicate CRSP monthly EWRETD exactly. If a user estimates her own equal-weighted monthly portfolio returns, she should expect no bias for daily portfolio returns when the BHMD method is applied.

For consistency between daily and monthly data, it should be pointed out that one important contribution of the *BHMD* is that monthly rebalancing is applied in daily portfolio return calculation, which is very likely to replace many existing practices of applying daily rebalancing. While daily rebalancing is not necessarily wrong, it contradicts to the common sense that compounding daily and monthly returns should be equivalent. It is seldom correct when rebalancing frequencies constructing monthly and daily indices are different. Thus, when daily and monthly rebalanced daily *EWRETD*s coexist, the latter will be surely preferable because it can be bias-free. In addition, different rebalancing frequencies add one more uncertainty when results are compared based on monthly and daily indices. To avoid unnecessary complications, researchers should choose the consistent one in the first place.

To summarize the *BHMD*, it can be claimed with confidence that the method has solved, once and for all, the inconsistency between daily and monthly portfolio returns. In other words, compounding the CRSP monthly *EWRETD* over any time period will be equivalent to compounding the daily *EWRETD* generated by the *BHMD*. In addition, with several non-trivial modifications, Equation (2) can be derived from Equation (9) in Blume and Stambaugh (1983) (Details are given in Appendix A).

4 Empirical Results Generated by the BHMD

To avoid confusion, $EWRETD^d_{CRSP}$ is referred to as the current CRSP daily index and $EWRETD^d_{BHMD}$ as the newly constructed one. There are several hypotheses we want to test. The first hypothesis is that the current CRSP daily equal-weighted index suffers statistically significant biases: $\sigma(EWRETD^d_{CRSP}) \neq \sigma(EWRETD^d_{BHMD})$ and $EWRETD^d_{CRSP} - EWRETD^d_{BHMD} > 0$. Several findings emerge from Table 1 which provides sample statistics and testing results on variances and means of those two time series. The first finding is that with no exception, the variances of the CRSP daily *EWRETD* are not statistically different from their true values for all years. This is a piece of good news for researchers, since the potential impact of a different variance is non-existent. The second finding is that annual averages of the mean differences are all positive between $EWRETD^d_{CRSP}$ and $EWRETD^d_{BHMD}$. This confirms that the $EWRETD^d_{CRSP}$ is systemically upward-biased, which is found to be consistent with the predictions made by Blume and Stambaugh (1983), and with the evidence provided by Loughran and Ritter (1995), and Roll (1983). Unlike other researchers, we pin point which years have statistically significant biases and their magnitudes. For 4 years, $EWRETD^d_{CRSP}$ are statistically different from their true values. For example, in 1984, $EWRETD^d_{CRSP}$ was not statistically different from $EWRETD^d_{BHMD}$. The annualized difference was only 1.4%. In contrast, in 1991, $EWRETD^d_{CRSP}$ was statistically biased, with a stunning annualized discrepancy of 29%**.** Such a huge bias could have a dramatic influence on one's conclusion if a researcher uses those 4 years. For unreported result of using 3-year window, we have almost identical results.

Table 1: Statistics of $EWRETD^{d}_{CRSP}$ and $EWRETD^{d}_{BHMD}$ on an annual basis

SAMPLE STATISTICS (*100)						TESTS OF DIFFERENCES					
						Equal variances			Equal means		
year	n	μ_{CRSP}	σ_{CRSP}	μ_{MDBH}	σ_{MDBH}	D_{σ}	F	p	D_{μ}	T-value	P
1964	253	0.0848	*0.2963*	0.0666	0.2976	-0.001	1.01	0.9446	0.018	0.69	0.492
1965	252	0.1358	0.4756	0.1220	0.4749	0.001	1.00	0.9821	0.014	0.33	0.744
1966	252	-0.0095	0.7791	-0.0234	0.7702	0.009	1.02	0.8555	0.014	0.20	0.840
1967	251	0.2389	0.6043	0.2305	0.6098	-0.006	1.02	0.8853	0.008	0.16	0.876
1968	226	0.1646	0.7077	0.1576	0.7074	0.000	1.00	0.9944	0.007	0.10	0.917
1969	250	-0.1046	0.8425	-0.1138	0.8333	0.009	1.02	0.8632	0.009	0.12	0.902
1970	254	-0.0195	1.2098	-0.0386	1.1844	0.025	1.04	0.7359	0.019	0.18	0.857
1971	253	0.0894	0.7934	0.0763	0.7929	0.000	1.00	0.9921	0.013	0.19	0.852
1972	251	0.0351	0.5201	0.0225	0.5256	-0.005	1.02	0.8697	0.013	0.27	0.788
1973	252	-0.1605	0.8479	-0.1722	0.8361	0.012	1.03	0.8242	0.012	0.16	0.876
1974	253	-0.0910	0.8894	-0.1159	0.8818	0.008	1.02	0.8917	0.025	0.32	0.753
1975	253	0.2163	0.7239	0.1989	0.7280	-0.004	1.01	0.9283	0.017	0.27	0.788
1976	253	0.1723	0.5185	0.1598	0.5228	-0.004	1.02	0.8970	0.013	0.27	0.787
1977	252	0.0940	0.3602	0.0856	0.3620	-0.002	1.01	0.9365	0.008	0.26	0.794
1978	252	0.1016	0.7283	0.0964	0.7284	-0.000	1.00	0.9977	0.005	0.08	0.936
1979	253	0.1466	0.6320	0.1396	0.6331	-0.001	1.00	0.9777	0.007	0.12	0.901
1980	253	0.1504	0.8733	0.1452	0.8589	0.014	1.03	0.7922	0.005	0.07	0.947
1981	253	0.0050	0.6963	0.0022	0.6884	0.008	1.02	0.8559	0.003	0.04	0.964
1982	253	0.0947	0.6961	0.0893	0.6979	-0.002	1.01	0.9673	0.005	0.09	0.931
1983	253	0.1294	0.6161	0.1276	0.6233	-0.007	1.02	0.8532	0.002	0.03	0.973
1984	253	-0.0391	0.4954	-0.0466	0.4951	0.000	1.00	0.9920	0.008	0.17	0.864
1985	252	0.1002	0.3821	0.0907	0.3839	-0.002	1.01	0.9417	0.010	0.28	0.779
1986	253	0.0462	0.4861	0.0316	0.4879	-0.002	1.01	0.9537	0.015	0.34	0.737
1987	253	0.0085	1.3313	-0.0266	1.2742	0.057	1.09	0.4869	0.035	0.30	0.762
1988	253	0.1041	0.4990	0.0694	0.5016	-0.003	1.01	0.9329	0.035	0.78	0.435
1989	252	0.0747	0.4047	0.0451	0.4065	-0.002	1.01	0.9460	0.030	0.82	0.412
1990	253	-0.0287	0.6131	-0.0945	0.5980	0.015	1.05	0.6927	0.066	1.22	0.222
1991	253	0.2360	0.5725	0.1666	0.5719	0.001	1.00	0.9877	0.069	1.37	0.173
1992	254	0.1974	0.4809	0.0946	0.4818	-0.001	1.00	0.9752	0.103	2.41**	0.016
1993	253	0.1806	0.4054	0.0948	0.4035	0.002	1.01	0.9395	0.086	2.39**	0.017
1994	252	0.0579	0.4822	-0.0194	0.4775	0.005	1.02	0.8756	0.077	1.81*	0.071
1995	252	0.1744	0.3941	0.1059	0.3939	0.000	1.00	0.9936	0.068	1.95*	0.052
1996	254	0.1188	0.5611	0.0663	0.5491	0.012	1.04	0.7297	0.053	1.07	0.287
1997	253	0.1252	0.6267	0.0746	0.6234	0.003	1.01	0.9342	0.051	0.91	0.363
1998	252	0.0537	1.0075	-0.0070	0.9883	0.019	1.04	0.7611	0.061	0.68	0.496
1999	252	0.1589	0.5854	0.1175	0.5947	0.009	1.03	0.8029	0.041	0.79	0.432
2000	252	0.0039	1.2284	-0.0402	1.1634	-0.065	1.11	0.3895	0.044	0.41	0.679
2001	248	0.1276	1.1213	0.0864	1.1027	-0.019	1.03	0.7933	0.041	0.41	0.680
2002	252	0.0036	1.0399	-0.0408	1.0213	-0.019	1.04	0.7752	0.044	0.48	0.629
2003	252	0.2379	0.7323	0.2195	0.7445	0.012	1.03	0.7938	0.018	0.28	0.780
2004	252	0.0884	0.7228	0.0806	0.7209	-0.002	1.01	0.9681	0.008	0.12	0.903
2005	252	0.0296	0.5905	0.0236	0.5891	-0.001	1.00	0.9694	0.006	0.11	0.909

The time period is from 1964 to 2005, where $EWRETD^{d}_{CRSP}$ is the daily index directly drawn from the CRSP daily data and $EWRETD^{d}_{BHMD}$ is based on the BHMD method. $\mu_{CRSP}(\sigma_{CRSP})$ and μ_{BHMD} (σ_{BHMD}) are means (standard deviations) for $EWRETD_{CRSP}$ and $EWRETD_{BHMD}$, respectively, $D_{\mu} = \mu_{CRSP} - \mu_{BHMD}$, and $D_{\sigma} = \sigma_{CRSP} - \sigma_{BHMD}$. F and P are for the test of equal variances. T-value and P-value are for the test of equal means. ** for significant at 5% level; * significant at 10% level.

It is reasonable to hypothesize that such a deviation is driven partially by the whole market. The intuition underlying this hypothesis is that when the whole market moves up, more trades are expected to be buyer-initiated. As a result, those trades have a higher probability to be traded at, or close to ask prices; and the opposite is true when the market moves down. To test this hypothesis, the deviation $(EWRETD_{CRSP} - EWRETD_{BHMD})$ is regressed against a market portfolio, plus several other independent variables. Following the

literature, the value-weighted index is used to represent the whole market. The results in Table 2 support the conjecture that the whole market has a strong impact on the bias, since the related coefficients are positive and statistically significant most of the time.

Table 2: Test of the hypothesis: bias of CRSP daily *EWRETD* driven by the market

PANEL A

	Model 1	Model 2	Model 3
N	10574	10574	10574
Intercept	0.000285	0.00028	0.00028
	(57.90)	(56.01)	(57.84)
VWRETD	0.00844	0.00890	0.00847
	(15.18)	(15.92)	(15.24)
Lag(EWRETD$_{BHMD}$)	-0.02128		-0.02274
	(-15.48)		(-17.91)
Lag(VWRETD)	0.01379		0.01626
	(10.19)		(15.81)
Lag(VWRETD)-lag2(VWRETD)	0.00170		
	(2.82)		
R^2	0.0527	0.0233	0.0521

PANEL B

	N	INTERCEPT	X_1	X_2	X_3	X_4	R^2
1964-1968	1232	0.000130	0.0030	-0.0144	0.0165	0.0004	0.040
		(20.39)	(2.46)	(-5.69)	(5.12)	(0.35)	
1969-1973	1258	0.000130	0.0134	-0.0046	0.0004	0.0036	0.121
		(15.24)	(11.66)	(-1.68)	(0.13)	(2.87)	
1974-1978	1261	0.000146	0.0001	-0.0135	0.0104	-0.0016	0.035
		(21.52)	(0.07)	(-6.60)	(5.47)	(-1.87)	
1979-1983	1263	0.000057	-0.0008	-0.0269	0.0209	-0.0023	0.092
		(6.98)	(-0.86)	(-10.69)	(8.37)	(-2.13)	
1984-1988	1262	0.000192	0.0136	-0.0639	0.0334	0.0021	0.318
		(14.46)	(10.33)	(-15.65)	(9.39)	(1.36)	
1989-1993	1263	0.000713	0.0032	-0.0208	0.0121	-0.0000	0.017
		(47.65)	(1.54)	(-4.11)	(2.68)	(-0.01)	
1994-1998	1261	0.000611	0.0043	-0.0131	0.0123	-0.0016	0.021
		(50.04)	(3.01)	(-3.71)	(3.74)	(-1.08)	
1999-2003	1254	0.000389	0.0131	-0.0200	0.0121	0.0033	0.069
		(16.99)	(7.69)	(-4.06)	(2.80)	(1.82)	

Hypothesis: the bias of the CRSP daily *EWRETD* is associated with the whole market, i.e., we expect a high deviation when the market return is high. The following linear regression and its variations are run. $y = \alpha + \sum_i^k \beta_i X_i$ where $y = EWRETD_{CRSP} - EWRETD_{BHMD}$. For example, if *k=4*, $X_1 = VWRETD$, $X_2 = lag(EWRETD_{BHDM})$,$X_3 = lag(VWRETD)$, $X_4 = lag(VWRETD) - lag2(VWRETD)$, *VWRETD* is the CRSP daily value-weighted market index. The time period is from 1964 to 2005. *N* is the number of observations and *T*-values are in the parentheses. Panel B is related to Model 1 for various time periods.

For impacts on beta and Jensen's α, we have the following pair of hypotheses: when the $EWRETD^d_{CRSP}$is used, β_{CRSP} is upward-biased, and Jensen's α is downward-biased. To test these hypotheses, a 3-year moving window is randomly chosen to estimate those parameters from 1964-1969. Beta's and α's estimations are based on the previous 36 months. This scheme results in 36 pairs of alphas and betas for each stock. For a market index, both $EWRETD^d_{CRSP}$ and $EWRETD^d_{BHMD}$ are used. Table 3 lists the results for the first 20 stocks and shows that all differences $(Intercept_{CRSP} - Intercept_{BHMD})$ are negative and statistically significant. This is interpreted as that the CRSP daily *EWRETD*'s upward bias leads to a systematically undervalued intercept (Jensen's α). For beta, the results are mixed, and their significant levels depend on individual stocks.

Table 3: Impact on Beta by using the upward biased CRSP *EWRETD*

PERMNO	N	Δα(%)	T-value	Δ β(%)	T-value
10006	36	-0.0113	-3.40***	-0.6689	-1.39*
10014	36	-0.0191	-4.93***	1.1322	1.84**
10030	36	-0.0069	-6.47***	0.2480	0.65
10057	36	-0.0089	-3.19***	0.3268	1.78**
10065	36	-0.0039	-2.91***	0.1200	0.88
10102	36	-0.0116	-4.60***	0.1557	0.75
10137	36	-0.0041	-3.40***	0.1251	1.02
10145	36	-0.0089	-5.14***	0.1043	1.11
10153	36	-0.0174	-3.38***	0.3296	1.05
10161	36	-0.0117	-3.28***	0.2805	1.62*
10188	36	-0.0209	-4.78***	-0.6017	-2.66***
10217	36	-0.0021	-2.65***	-0.3049	-1.14
10225	36	-0.0068	-3.78***	0.0677	0.49
10233	36	-0.0134	-4.80***	0.3703	2.51***
10241	36	-0.0088	-4.75***	0.3530	3.14***
10268	36	-0.0064	-12.52***	-0.2092	-0.97
10276	36	-0.0068	-11.79***	-0.5062	-4.08***
10313	36	-0.0020	-2.51***	-0.2880	-1.26
10321	36	-0.0141	-13.25***	-0.7800	-4.19***
10364	36	-0.0088	-3.53***	-0.0153	-0.05

I randomly choose a 3-year moving window to estimate intercepts and betas and the time period is from 1964-1969. The first intercept and beta based on January 1964 to December 1966 and the second pair based on February 1964 to January 1967. $EWRETD^d_{CRSP}$(drawn from the CRSP daily date) and $EWRETD^d_{BHMD}$ are used as the market indices. N is the number of the observations. α and β from the following regression: $RET_t = \alpha + \beta * RET_t^m + \epsilon$, where, RET_t is a stock's return on day *t*, $RET^m{}_t$ is a market return. Table shows the first 20 stocks. $\Delta\alpha(\%) = 100 * (\alpha_{CRSP} - \alpha_{BHMD})$ and $\Delta\beta(\%) = 100 * (\beta_{CRSP} - \beta_{BHMD})$. Symbols *,**,*** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively, using a 1-tail test.

In this paper, we also test the performance of the method developed by Blume and Stambaugh (1983) to estimate *EWRETD* with daily rebalancing for the whole market (NYSE/AMEX/ NASDAQ).[4] Our empirical results show that their method reduces the

[4] For Blume and Stambaugh (1983), the following formula is applied: $EWRETD_t^{daily} = \sum_{i=1}^{N} \frac{(1+R_{i,t-1})}{\sum_{j=1}^{N}(1+R_{j,t-1})} R_{i,t}$, where $R_{i,t}$ is the CRSP daily return on day t, and N is the number of stocks

bias dramatically. For instance, from 1964 to 1993, the average annual return of the daily *EWRETD*, based on daily rebalancing (Blume and Stambaugh 1983), is 1.032% higher when benchmarked on the monthly CRSP *EWRETD*. Compared with 6.7% annual bias, we conclude that Blume and Stambaugh's methodology removes about 85% biases.

5 Comparison and Discussion

Because the CRSP monthly data is more widely used than its daily data is terms of avoiding the impacts of the microstructure, the monthly CRSP *EWRETD* is utilized as a benchmark in the following tests. The performances of two daily time series are compared: *EWRETD* from CRSP vs. *EWRETD* based on the *BHMD* method. For convenience, Compounding-Method is referred to using the CRSP daily *EWRETD* to get monthly *EWRETD*.

Compounding-Method: compounding the CRSP daily *EWRETD* to obtain a monthly one:

$$\mathrm{EWRETD}_{\mathrm{monthly}} = \prod_{s=1}^{\mathrm{T}}\left(\mathrm{EWRETD}_{\mathrm{daily},s}^{\mathrm{CRSP}} + 1\right) - 1 \tag{3}$$

where $EWRETD_{monthly}$ is a monthly equal-weighted market index, $EWRETD_{daily}^{CRSP}$ is the daily equal-weighted market index retrieved directly from the CRSP on day *s*, and *T* is the number of trading days in each month.

BHMD method: From daily index returns to a monthly one, the following formula is applied:$\left(EWRETD_{montly} + 1\right) = \prod_{s=1}^{T}(EWRETD_{daily} + 1)$, then using Equation (2) for each day's *EWRETD:*

$$\left(EWRETD_{monthly} + 1\right) = (1 + EWRETD_{BH}^{1}) * \frac{\left(1+EWRETD_{BH}^{2}\right)}{\left(1+EWRETD_{BH}^{1}\right)} *** \frac{\left(1+EWRETD_{BH}^{T-1}\right)}{\left(1+EWRETD_{BH}^{T-2}\right)} * \frac{\left(1+EWRETD_{BH}^{T}\right)}{\left(1+EWRETD_{BH}^{T-1}\right)}$$

To simplify it as the final equation:

$$EWRETD_{monthly} = EWRETD_{BHMD}^{T} = \frac{1}{N}\sum_{i=1}^{N}\prod_{s=1}^{T}(1 + R_{i,s}^{d}) - 1 \tag{4}$$

Thanks to the benefits offered by a buy-and-hold strategy (Blume and Stambaugh, 1983), the *BHMD* method is superior to Compounding-Method. Among several undesired microstructure effects, the bid-ask bounce and non-synchronous trading dominate other effects. For Compounding-Method, the impact of bid-ask bounce is severe since the biases are positively correlated among the stocks on a daily basis (Blume and Stambaugh, 1983, Conrad and Kaul, 1993). Since the building block in the *BHMD* method is an *n*-day buy-and-hold return, gains and losses on various days will cancel out each other; and averaging across stocks subsequently makes this "canceling-out" more effective. For a market index, the law of large numbers reduces potential measurement errors even further.

---

in the portfolio. We thank Robert Stambaugh for a detailed discussion.

As the *BHMD* method does not take the average of returns on the same day, non-synchronous trading plays no role. To put it another way, by averaging all of the stocks on a multi-day basis, the impact of non-synchronous trading on each trading day is close to zero.

Table 4: Monthly mean returns on equal-weighed portfolio, by years (NYSE/AMEX)

	Compounding-Method			*BHMD*		
	Canina et al. (1998)					
YEAR	DAILY	MONTHLY	DIFF	DAILY	MONTHLY	DIFF
1964	1.8068	1.4167	0.3901	1.4310	1.4167	0.0143
1965	2.9399	2.6433	0.2966	2.6344	2.6435	-0.0090
1966	-0.1468	-0.4320	0.2852	-0.4128	-0.4316	0.0188
1967	5.2013	5.0146	0.1867	5.0156	5.0148	0.0008
1968	3.2306	3.0855	0.1451	3.1001	3.0871	0.0130
1969	-2.0420	-2.2326	0.1906	-2.2781	-2.2327	-0.0454
1970	-0.2423	-0.6377	0.3953	-0.5995	-0.6335	0.0340
1971	2.0326	1.7448	0.2879	1.7363	1.7450	-0.0087
1972	0.7502	0.5305	0.2197	0.5897	0.5298	0.0599
1973	-2.7044	-3.2039	0.4996	-3.1739	-3.3509	0.1770
1974	-1.4101	-2.2898	0.8798	-2.3154	-2.2682	-0.0473
1975	5.4477	4.7686	0.6792	4.7533	4.5451	0.2083
1976	4.1184	3.6295	0.4889	3.6971	3.5544	0.1428
1977	1.5324	1.2413	0.2911	1.2030	1.8177	-0.6147
1978	2.0530	1.8210	0.2320	1.7351	2.2278	-0.4927
1979	3.2352	2.9841	0.2511	2.9730	3.0739	-0.1008
1980	2.9984	2.7291	0.2693	2.6677	3.2892	-0.6215
1981	0.7255	0.4832	0.2423	0.4686	0.1006	0.3680
1982	2.6245	2.2817	0.3428	2.2873	2.0105	0.2768
1983	2.9378	2.7194	0.2184	2.7360	2.7996	-0.0636
1984	0.0984	-0.1704	0.2688	-0.1342	-0.9365	0.8024
1985	2.4081	2.0676	0.3405	2.1023	1.9936	0.1087
1986	1.4281	1.0473	0.3807	1.0646	0.7170	0.3476
1987	0.4443	0.0051	0.4392	0.0522	-0.2532	0.3054
1988	2.1429	1.6372	0.5057	1.6746	1.5042	0.1704
1989	1.8012	1.2973	0.5039	1.2967	0.9720	0.3247
1990	-0.8390	-1.7774	0.9384	-1.7716	-1.9006	0.1289
1991	4.1527	2.9331	1.2196	2.9204	3.6170	-0.6966
1992	2.5911	1.6665	0.9246	1.6796	2.1197	-0.4401
1993	2.2647	1.7697	0.4950	1.7874	2.0337	-0.2462
1964-1993	1.7194	1.2924	0.4269	1.297356	1.293521	0.003835

Compounding-Method: compounding the CRSP daily *EWRETD* to get a monthly one. For the *BHMD* method, see Equation (4). The time period covered is from 1964 to 1993 as in Canina et al. (1998). For the *BHMD* method, only stocks traded on the NYSE/AMEX are chosen, which have valid data on the last trading day for each month. Columns 2-4 are borrowed from Canina et al. (1998). All values, except YEAR, are amplified by 100 for a better presentation.

For an easy comparison with Canina et al. (1998), the same time period (1964-1993) and the same markets (NYSE/AMEX) are chosen, and their relevant part is reprinted in Table 4. The last row is virtually the essence of the whole table. According to Canina et al. (1998), the monthly difference between the $EWRETD_{monthly}^{from-daily}$ (estimated from the CRSP daily *EWRETD*) and the $EWRETD_{monthly}^{directly-from-CRSP}$ (retrieved directly from the

CRSP) is 0.4269. This amount is translated into a 6% annual difference while the monthly average difference, based on the *BHMD* method, is a mere 0.003835 or 0.05% for the corresponding annual value.[5] Their value (based on Compounding-Method) is more than 100-fold higher than the *BHMD* (the new method). Canina et al (1998) argue that it is problematic to use a converted monthly *EWRETD*, since 6% is one-third of a typical annual return. But now, with the *BHMD* method, the divergence or bias is only about 0.003% of a typical annual return. Such a small magnitude will be inconsequential for most research applications.

Table 5: Monthly mean returns on equal-weighted index by years
(NYSE/AMEX/NASDAQ)

		Compounding-Method			*BHMD*	
Year	Daily	Monthly	diff	daily	monthly	diff
1964	1.80614	1.41674	0.38940	1.4198	1.4167	0.0030
1965	2.94031	2.64348	0.29682	2.6385	2.6435	-0.0050
1966	-0.14765	-0.43162	0.28397	-0.4411	-0.4316	-0.0095
1967	5.19857	5.01481	0.18376	5.0135	5.0148	-0.0013
1968	3.23117	3.08709	0.14409	3.0897	3.0871	0.0026
1969	-2.04287	-2.23265	0.18978	-2.2368	-2.2327	-0.0042
1970	-0.23833	-0.63352	0.39520	-0.6349	-0.6335	-0.0014
1971	2.02701	1.74502	0.28200	1.7467	1.7450	0.0017
1972	0.78330	0.52982	0.25348	0.5192	0.5298	-0.0106
1973	-3.14316	-3.35093	0.20777	-3.3578	-3.3509	-0.0069
1974	-1.77026	-2.26815	0.49790	-2.2805	-2.2682	-0.0123
1975	4.93050	4.54507	0.38543	4.5471	4.5451	0.0020
1976	3.84782	3.55436	0.29345	3.5630	3.5544	0.0086
1977	2.00662	1.81770	0.18893	1.8167	1.8177	-0.0010
1978	2.34994	2.22778	0.12216	2.2268	2.2278	-0.0009
1979	3.22877	3.07389	0.15487	3.0785	3.0739	0.0046
1980	3.38470	3.28922	0.09548	3.2961	3.2892	0.0069
1981	0.16216	0.10061	0.06155	0.1131	0.1006	0.0125
1982	2.09138	2.01050	0.08088	1.9799	2.0105	-0.0306
1983	2.85247	2.79961	0.05286	2.8102	2.7996	0.0106
1984	-0.77795	-0.93654	0.15859	-0.9307	-0.9365	0.0058
1985	2.18943	1.99358	0.19585	1.9913	1.9936	-0.0023
1986	1.03247	0.71701	0.31546	0.7273	0.7170	0.0103
1987	0.41839	-0.25325	0.67164	-0.2486	-0.2532	0.0047
1988	2.24820	1.50425	0.74396	1.5096	1.5042	0.0053
1989	1.59768	0.97198	0.62570	0.9716	0.9720	-0.0004
1990	-0.53532	-1.90055	1.36523	-1.8937	-1.9006	0.0069
1991	5.15063	3.61697	1.53366	3.6204	3.6170	0.0034
1992	4.35318	2.11970	2.23349	2.1179	2.1197	-0.0018
1993	3.89115	2.03365	1.85750	2.0330	2.0337	-0.0006
1964-1993	1.7688826	1.2935209	0.4753616	1.2935270	1.2935209	.0000073

Compounding-Method: compounding the CRSP daily index to get a monthly *EWRETD*. (for the *BHMD* method, see Equation [4]).The time period covered is from 1964 to 1993 as in Canina et al. (1998). For the *BHMD* method, only stocks traded on NYSE/AMEX/NASDAQ with valid data on the last trading day for each month are chosen. All values are amplified by 100 for a better presentation.

[5]From the monthly averages to an annual one: 1.0129735612-1.0129352112=0.05%.

Table 5 is related to the whole market (NYSE/AMEX/NASDAQ). The same conclusion holds: the *BHMD* method is superior. The monthly difference between the DAILY and MONTHLY average is 0.48% if based the Compounding-Method, while the corresponding value based on the *BHMD* method is a tiny 0.000007%. In terms of annual bias, the Compounding-Method has 6.7%, while the *BHMD* method 0.0008%. Following Canina et al. (1998), the largest 20 monthly differences are presented. Since the *BHMD* method generates both negative and positive differences, those numbers are sorted by absolute values. As the panel shows, the largest variation, based on the Compounding-Method, is 6 times higher than that in the *BHMD* method. In addition, the Compounding-Method has 20 deviations larger than 1%. For the *BHMD* method, the largest difference is less than 0.5% and after just 3 values, the differences fall dramatically to less than 0.086%. For more than 100 months out of 360, the deviations for the Compounding-Method are actually larger than the largest value of the *BHMD* method.

Now, let us turn to the properties of the newly generated daily equal-weighted index based on the *BHMD* method. For the NYSE/AMEX market from January 1973 to December 1978, Loughran and Ritter (1995) report that compounding the daily CRSP *EWRETD* will generate a huge upward bias, compared with compounding the monthly *EWRETD*.[6] By using the current vintage of the CRSP database (last trading date is 12//30/2005), those two values, based on the Compounding-Method, are 236% and 179%. The corresponding overestimation is 33%. By using the newly generated daily *EWRETD*, the total return is 183%. Over this 6-year period, the percentage difference is only 1.8%. This translates into 0.36% per year. Such a small deviation demonstrates the quality of the newly generated daily equal-weighted market index. For a robust check, the daily index for the whole market (NYSE/AMEX/NASDAQ) is generated. Over the same time period (1973-1978), compounding the CRSP daily and monthly *EWRETD* leads to 236%, and 179%. The percentage difference is 32%, about 6% per year. The total return of compounding the new daily series is 176%. The percentage variation is quite small, only about -0.28% per year. Thus, such a magnitude should have a minimum impact on most research topics. These two results confirm that the newly generated daily index is consistent with a monthly one than the current CRSP daily index.

[6]Dividend timing refers to for a daily return estimation, a dividend is included on the dividend paying day, while for monthly returns, it is treated as paid at the end of the month.

Table 6: Panel A: 20 highest differences between daily and monthly *EWRETD*

Panel A	Compounding-Method				*BHMD*			
RANK	YYYYMM	DAILY	MONTHLY	DIFF	YYYYMM	DAILY	MONTHLY	DIFF
1	199207	6.4740	2.7615	3.71243	198211	8.7926	9.2037	-0.4112
2	199012	7.0515	3.5677	3.48378	197602	10.8017	10.6291	0.1726
3	199101	10.0163	6.6686	3.34768	198301	11.8125	11.7062	0.1063
4	199011	4.8832	1.6225	3.26074	197502	5.3059	5.3918	-0.0860
5	197412	3.8167	1.1049	2.71179	197311	-17.6075	-17.5252	-0.0823
6	199010	2.2404	-0.4526	2.69299	197411	-5.2161	-5.1395	-0.0766
7	199112	0.2899	-2.3898	2.67973	197601	19.0803	19.0037	0.0766
8	198710	6.1262	3.5468	2.57945	197810	-17.8024	-17.7280	-0.0744
9	199111	-1.3447	-3.9147	2.56997	198610	2.4490	2.3777	0.0713
10	199103	-3.3098	-5.7155	2.40572	197609	1.6282	1.6990	-0.0708
11	197409	5.2613	3.1411	2.12013	197606	2.5851	2.6510	-0.0659
12	199110	8.8610	6.7612	2.09984	197702	0.3795	0.4400	-0.0605
13	199205	-25.2037	-27.2248	2.02105	197809	0.5635	0.6200	-0.0565
14	197311	3.6301	1.6190	2.01105	197608	-1.7217	-1.6673	-0.0544
15	197410	10.3816	8.4175	1.96408	197603	1.8652	1.8109	0.0543
16	199108	5.4661	3.5353	1.93078	197412	-8.1260	-8.0744	-0.0516
17	199102	3.2357	1.3719	1.86385	197808	9.4053	9.3545	0.0508
18	199203	0.7155	-1.1112	1.82672	197407	-5.2421	-5.1915	-0.0507
19	197005	5.7898	3.9698	1.81992	196401	3.2589	3.2092	0.0497
20	199107	5.6838	3.8740	1.80976	198212	2.3935	2.3447	0.0488

NYSE/AMEX/NASDAQ (1994-2003). Compounding-Method: compounding the daily index to get a monthly *EWRETD*. (for the *BHMD*, see Equation [4]). The last column is sorted according to the absolute value of the difference (*DIFF*) which is defined as *DAILY* minus *MONTHLY*. *DAILY* refers to monthly returns generated from the CRSP daily data, and *MONTHLY* refers to monthly returns drawn directly from CRSP.

Panel B: percentages of positive, negative values for monthly *EWRETDs* (retrieved director from CRSP and its estimate from daily data), and their differences. Time period is 1964 to 1993 for stocks listed on NYSE and AMEX.

Panel B: Equal-Weighted Market indices (*EWRETD*) 1964-1993 (NYSE/AMEX)						
	Compounding-Method			*BHMD*		
	DAILY	MOHTHLY	DIFF	DAILY	MONTHLY	DIFF
# of months return>0	234	223	339	223	223	173
% Positive	65.00%	61.94%	94.17%	61.94%	61.94%	48.06
Panel C: Value-weighted market indices (*VWRETD*)						
	DAILY	MOHTHLY	DIFF			
# of months return>0	217	217	197			
% Positive	60.28	60.28	54.72			

For a replication, the deviation from the original time series should behave like a random variable. This implies that the mean of deviations should be small, and being positive and negative with roughly equal probabilities. Panel B shown in Table 6 represents proportions of positive returns for the equal-weighted indices based on two methods. For 360 months, from 1964 to 1993, the Compounding-Method results in 339 positive differences. The number counts for 94.17% of all months, a far cry from 50%. The *BHMD* method leads to only 173 positive differences, counting for 48.06% of all months and only 1.94% away

from 50%. Thus, its performance is at par with the monthly *value-weighted* index generated from the daily data (54.72%). For the value-weighted NYSE/AMEX index, its percentage of positive differences is 53.6%, about 3.6% away from the ideal value (see Table 4 in Canina et al. 1998).

Table 7: Monthly mean returns on CRSP equal-weighed portfolio, by years (1994-2005)

PANEL A: Comparison between two methods

	Compounding-Method			*BHMD*		
year	DAILY	MONTHLY	DIFF	DAILY	MONTHLY	DIFF
1994	1.2334	-0.3963	1.6297	-0.4058	-0.3963	-0.0095
1995	3.7371	2.2517	1.4854	2.2607	2.2517	0.0091
1996	2.5904	1.4633	1.1271	1.4647	1.4633	0.0014
1997	2.7321	1.6429	1.0892	1.6440	1.6429	0.0011
1998	1.3159	0.0274	1.2885	0.0284	0.0274	0.0010
1999	3.4503	2.5510	0.8992	2.5544	2.5510	0.0033
2000	0.1473	-0.7374	0.8847	-0.7347	-0.7374	0.0027
2001	2.9555	2.0798	0.8757	2.0832	2.0798	0.0035
2002	0.1575	-0.7645	0.9220	-0.7607	-0.7645	0.0038
2003	5.1329	4.7313	0.4016	4.7339	4.7313	0.0026
2004	1.8889	1.7258	0.1631	1.7260	1.7258	0.0002
2005	0.6418	0.5074	0.1344	0.5091	0.5074	0.0017
1994-2005	2.1653	1.2569	0.9084	1.2586130	1.256870	0.0017428

Compounding-Method: compounding the daily index to get monthly *EWRETD*. (for the *BHMD* method, see Equation [4]). Similar to Table 1, Panel A shows the comparison between two Methods from 1994 to 2005. Panel B shows the largest 20 monthly differences.

PANEL B: Largest 20 differences based on the *BHMD*

RANK	YYYYMM	$EWRET_{EST}$	*EWRETD*	*DIFF*
1	199412	-1.4936	-1.3424	-0.1512
2	199404	-1.1328	-1.1728	0.0400
3	200108	-3.4625	-3.5003	0.0378
4	199409	0.4502	0.4865	-0.0362
5	200008	5.9659	5.9309	0.0349
6	199507	5.5019	5.4715	0.0304
7	200303	0.9778	0.9495	0.0283
8	200203	7.4835	7.4587	0.0249
9	199509	2.6608	2.6370	0.0238
10	199901	6.4573	6.4802	-0.0228
11	200002	12.0617	12.0396	0.0220
12	200204	-0.2983	-0.3190	0.0207
13	199511	1.6852	1.6666	0.0186
14	200201	1.7994	1.8179	-0.0185
15	199808	-19.6592	-19.6408	-0.0184
16	199502	2.8196	2.8369	-0.0172
17	199512	0.9114	0.8943	0.0172
18	200007	-1.9517	-1.9346	-0.0171
19	200104	7.6254	7.6425	-0.0170
20	200207	-10.7373	-10.7539	0.0166

Then, the robustness of the conclusions can be explored by expanding into other time periods and various stock markets. From 1964 to 2005, based on the *BHMD* method, the annual values for DAILY and MONTHLY are 1.2826 and 1.2831, respectively. The annualized variation is -0.008%, smaller than the corresponding value of 0.021% for the base period of 1964 to 1993. Table 7 shows the results for the years after 1993. The annualized average bias is only 0.024%. The time period before 1964 (1925-1963) also is investigated. The performance of the *BHMD* method is at par with the two more recent periods just discussed, with a tiny 0.06% annual bias.

By now, the first order of returns has been explained. What follows will be the next most important statistic – variance. Table 8 offers the variance comparisons for monthly *EWRETD* based on two methods. Again, the performance of the *BHMD* method is much better. A 5-year window is used from 1925 to 2004. For the Compounding-Method, the largest percentage difference of variance is about 16.0% when compared with its true value, while 1.2% for the *BHMD* method. After 1964, the Compounding-Method has three 5-year periods with the above 5% deviations, while the *BHMD* method has less than 0.6% for all periods. The most troublesome is the uncertainty of a bias direction for the Compounding-Method. For the window of 1975 to 1979, the Compounding-Method under-represents the true volatility by 7.7%, but overstates it by 7.4% over 1995 to 1999. The trend of the Compounding-Method 1 is not encouraging either since the error remains at an astounding level of 6% for 2000-2004. For the *BHMD* method, the largest deviation after 1990 is a meager 0.05%.

Table 8: Comparison of variances of monthly *EWRETD* based on two methods

		$EWRETD_{CRSP}$	Compounding-Method		*BHMD*	
WINDOW	N	σ_0^2	σ_1^2	DIFF (%)	σ_2^2	DIFF (%)
1925-1929	48	0.003449	0.002897	-15.9965	0.003447	-0.06828
1930-1934	60	0.035671	0.029969	-15.9836	0.036107	1.22490
1935-1939	60	0.012535	0.011827	-5.6430	0.012593	0.46918
1940-1944	60	0.004408	0.004256	-3.4517	0.004463	1.23479
1945-1949	60	0.002971	0.002943	-0.9436	0.002978	0.23806
1950-1954	60	0.001321	0.001287	-2.5747	0.001329	0.58990
1955-1959	60	0.000991	0.000973	-1.8963	0.000995	0.31000
1960-1964	60	0.001642	0.001663	1.2890	0.001652	0.57998
1965-1969	60	0.003271	0.003200	-2.1793	0.003271	0.00647
1970-1974	60	0.004647	0.004659	0.2626	0.004656	0.19079
1975-1979	60	0.004015	0.003706	-7.6842	0.004034	0.46945
1980-1984	60	0.003101	0.003021	-2.5765	0.003094	-0.19833
1985-1989	60	0.002823	0.002936	3.9999	0.002821	-0.08199
1990-1994	60	0.001992	0.001967	-1.2586	0.001993	0.03399
1995-1999	60	0.002343	0.002516	7.3947	0.002344	0.04606
2000-2004	60	0.004538	0.004261	-6.0834	0.004536	-0.02704
Means				-3.3328		0.31362

Compounding-Method: compounding the CRSP daily *EWRETD* drawn from the CRSP daily data to get a monthly *EWRETD*. (for the *BHMD* method, see Equation [4]). A 5-year window is used (1925- 2004). $\sigma^2(EWRETD_{CRSP})$ is the variance based on the CRSP monthly *EWRETD* drawn directly from CRSP. $DIFF_k$ is defined as 100* [σ^2 (Method *i*) - σ^2 ($EWRETD_{CRSP}$)] / σ^2 ($EWRETD_{CRSP}$), where *i*=1 and 2 and *N* is the number of observations (months).

In terms of dividend timing,[6] our empirical results show that even after this adjustment, the improvement is marginal. This result is consistent with Canina et al. (1998). Below are three reasons why the timing of dividends is inconsequential. First, the number of stocks paying dividends is quite small. From 1964 to 2005, only 5.58% of returns satisfying the condition of $|RET\text{-}RETX| > 1\%$, where *RET* and *RETX* are monthly returns with and without dividends. It is worth noticing that the cut-off point of 1% is quite small. Second, when dividend timing is ignored, some stocks enjoy overestimated returns, while others have underestimated ones. From 1964 to 2005, the positive differences count for 49.99%, while negative ones 49.96%, almost half and half. This reduces the impact of dividends on an index further since the positive and negative biases will be cancelled out. Third, the numerator (related to returns) has a scale of 0.1 to 10, while the denominator (number of stocks) has a scale of several thousands. Thus, any minor deviations in returns because of timing of dividends will be completely diversified away.

The consensus of the literature is that value-weighted daily indices are less problematic with issues associated with the microstructure. In discussing strategies available to avoid the problems associated with the CRSP equal-weighted index, Canina et al. (p414, 1998) recommend "The first is to use the value-weighted index as the benchmark portfolio. This portfolio does not suffer from any compounding related issues". The empirical work on market indices confirms this. However, researchers should be aware of the limitation of this statement since most work done so far has focused only on market indices where big stocks dominate. For example, the mean size of the largest CRSP cap-decile in 2004 is 950 times higher than that of the smallest cap-decile. Thus, in constructing daily value–weighted portfolios with small stocks, the above statement might not hold. For the smallest size portfolio (cap-decile 1), the total return difference (compounding monthly vs. daily value-weighted) over 1964-1993 is 29.4%, about 1% per year. The second counterexample is related to the CRSP smallest beta portfolio. The total returns over 1964 – 1993, are 651.3% and 406.6% based on monthly and daily value-weighted portfolios: a quite big bias of -3.4% per year. Thus, when estimating portfolio daily returns, researchers should apply the methodology developed in this paper whether it is an equal-weighed or value-weighed portfolio.[7]

6 Conclusion

Because of bid-ask bounce (Blume and Stambaugh, 1983) and non-synchronous trading (Roll, 1983), almost all equal-weighted portfolio daily returns and a few value-weighted portfolio daily returns are not estimated correctly. In addition, they are not consistent with their monthly counterparties. The CRSP daily *EWRETD* is a typical example, see Canina et al. (1998). In this paper, a better method called Buy-and-Hold Multi-Day (*BHMD*) method is proposed in order to obtain unbiased portfolio daily returns, free of errors associated with the microstructure and consistent with the monthly ones. When estimating daily portfolio returns, especially equal-weighted, researchers should apply the *BHMD* method to eliminate potential biases.

---

[7]Again, compounding a value-weighted daily index based on daily rebalancing is conceptually inconsistent with compounding its corresponding monthly value-weighted index based on monthly rebalancing.

There are many potential applications of the *BHMD* method. This method makes many research topics feasible in the first place. Doing research by using an unbiased daily index will definitely shed new insights on many topics that have not been studied before because of the impact of the microstructure. Along the same line, the impacts of the microstructure on cap-, beta- or standard deviation-based portfolios can be analyzed. Another implication of the *BHMD* method is to make research results comparable. The methodology discussed throughout this paper will help bridge the gaps between research projects analyzing monthly data and those using high-frequency data. Moreover, the *BHMD* method can assist researchers to convert the daily data into weekly or semi-monthly ones.[8]

ACKNOWLEDGEMENTS: We thank Mark Keintz, Michael Boldin, Marshall Blume, Robert Stambaugh, Sinan Tan, and Shuguan Zhang for their comments and helpful suggestions. The paper was presented at the 2007 Eastern Financial Association annual conference and Wharton Research Data Services internal seminar and we thank the participants for their helpful comments. In addition, we thank Fengying Xu and Jing Yan for their editorial help. All remaining errors are our sole responsibility. The former title of the paper was "Research impacts and correction of the upward-biased CRSP daily equal-weighted index".

References

[1] Amihud, Yakov, Illiquidity and Stock returns, *Journal of Financial Markets*, **5**, (2002), 31-56.

[2] Blume, Marshall, and Robert Stambaugh, Biases in computed returns: An application to the size effect, *Journal of Financial Economics*, **12**, (1983) 387-404

[3] Bollen, Nicolas P. and Jeffrey A. Busse, Short-Term Persistence in Mutual Fund Performance, *Review of Financial Studies*, **18**(2), (2004), 569-597.

[4] Canina, Linda, Roni Michaely, Richard Thaler, and Kent Womack, Caveat Compounder: A Warning about Using the Daily CRSP Equal-Weighted Index to Compute Long –Run Excess Returns, *Journal of Finance*, **53**(1), (1998), 403-416.

[5] Conrad, Jennifer, Gautam Kaul, Mean Reversion in Short-Horizon Expected Returns, *Review of Financial Studies*, **2**(2), (1989), 225-240.

[6] Conrad, Jennifer, Gautam Kaul, Long-term Market Overreaction or Biases in Computed Returns, *Journal of Finance*, **48**, (1993), 39-63.

[7] Easley, D., Hvidkjaer, S., O'Hara, M., Factoring information into returns. *Journal of Financial and Quantitative Analysis*, **45**, (2010), 293-309.

[8] Fama, Eugene and Kenneth R. French, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics*, **33**, (1993), 3-56.

[9] Fama, Eugene and Kenneth R. French, The cross-section of expected stock returns, *Journal of Finance*, **47**, (1992), 427-465.

[10] Jegadeesh, Narasimhan, and Sheridan Titman, Returns to buying winners and selling losers: Implications for market efficiency, *Journal of Finance*, **48**, (1993), 65-92.

[8]We thank an anonymous referee for pointing out this.

[11] Jegadeesh, Narasimhan, and Sheridan Titman, Profitability of Momentum Strategies: An Evaluation of Alternative Explanations, *Journal of Finance*, **46**(2), (2001), 699-720.
[12] Lesmond, David,Michael J. Schill and Chunsheng Zhou, The Illusory Nature of Momentum Profits, *Journal of Financial Economics*, **71**(2), (2004), 349-380.
[13] Loughran, Tim and Jay R. Ritter, The New Issues Puzzle, *Journal of Finance*, **50**(1), (1995), 23-51.
[14] Pastor, Lubos and Robert Stambaugh, Liquidity Risk and Expected Stock Returns, *Journal of Political Economy*, **111**(3), (2003), 642-685.
[15] Roll, Richard, On computing mean returns and the small firm premium, *Journal of Financial Economics*, **12**, (1983), 371-386.
[16] Sloan, R. G., Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings? *Accounting Review*, **71**(3), (1996), 289-315.
[17] Venkatesh, P. C., Empirical Evidence on the Impact of the Bid-Ask Spread on the Characteristics of CRSP Daily Returns, *Journal of Financial Research*, **15**(2), (1992), 113-125.

Appendix

Derivation of the Buy-and-Hold Multi-Day (*BHMD*) method from Blume and Stambaugh (1983**)**
Below is Equation (9) from Blume and Stambaugh (1983, p. 392):

$$r_{BH,t} = \frac{\sum_{i=1}^{N} P_{i,t}}{\sum_{i=1}^{N} P_{i,t-1}} - 1 \tag{9}$$

where, $P_{i,t}$ is the price of stock *i* on day *t*, *N* is the number of stocks, and *BH* stands for buy-and-hold. Blume and Stambaugh (1983) assume the initial stock prices are the same.
Modification 1: starting from the first day of each month, we have,

$$r_{BH,t} = \frac{\sum_{i=1}^{N} P_{i,0} \prod_{s=1}^{t}(1+R_{i,s})}{\sum_{i=1}^{N} P_{i,0} \prod_{s=1}^{t-1}(1+R_{i,s})} - 1 \tag{10}$$

where, $P_{i,0}$ is the closing price of stock *i* on the last trading day in previous month; $R_{i,s}$ is the return for stock *i* on day *s*.

Modification 2: relaxing the assumption of no dividends by applying returns with dividends.

Modification 3: relaxing the assumption of equal initial prices by adding a weight (w_i), we have:

$$r_{BH,t} = \frac{\sum_{i=1}^{N} w_{i,0} P_{i,0} \prod_{s=1}^{t}(1+R_{i,s})}{\sum_{i=1}^{N} w_{i,0} P_{i,0} \prod_{s=1}^{t-1}(1+R_{i,s})} - 1 \tag{11}$$

where w_i is the number of shares for stock *i* and defined as: $w_{i,0} = C/(N * P_{i,0})$ and *C* is the dollar amount of the total investment. Assume the total investment is \$1,000 and 10 stocks. If the initial price of stock A is \$100, we will buy one share ($w_{A,0}$*= 1000/(10*100*). If the initial price of stock B is \$50, we will purchase 2 shares ($w_{B,0}$*=1000/(10*50*). Plugging (12) into (11), we have:

$$r_{BH,t} = \frac{\sum_{i=1}^{N} \frac{C}{N*P_{i,0}} P_{i,0} \prod_{s=1}^{t}(1+R_{i,s})}{\sum_{i=1}^{N} \frac{C}{N*P_{i,0}} P_{i,0} \prod_{s=1}^{t-1}(1+R_{i,s})} - 1 = \frac{\frac{1}{N}\sum_{i=1}^{N} \prod_{s=1}^{t}(1+R_{i,s})}{\frac{1}{N}\sum_{i=1}^{N} \prod_{s=1}^{t-1}(1+R_{i,s})} - 1 = \frac{EWRETD_{BH}^{t}+1}{EWRETD_{BH}^{t-1}+1} - 1 \tag{12}$$

The last step in (12) uses Equation (4) of $EWRETD_{BH}^{t} + 1 = \frac{1}{N}\sum_{i=1}^{N} \prod_{s=1}^{t}(1 + R_{i,s})$ in Section 5. Finally, we have:

$$r_{BH,t} = EWRTED_{daily}^{t} \tag{13}$$

since $EWRETD_{daily}^{t}$ is defined as $\frac{\mathrm{EWRTED_{BH}^{t}}+1}{\mathrm{EWRETD_{BH}^{t-1}}+1} - 1$**,** see Equation (2) in Section 3.