



# Nearly Unbiased Estimation of Sample Skewness

**DOI:**

[10.1016/j.econlet.2020.109174](https://doi.org/10.1016/j.econlet.2020.109174)

**Document Version**

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

**Citation for published version (APA):**

Li, Y. (2020). Nearly Unbiased Estimation of Sample Skewness. *Economics Letters*, 192(0), Article 109174. <https://doi.org/10.1016/j.econlet.2020.109174>

**Published in:**

Economics Letters

**Citing this paper**

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

**General rights**

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Takedown policy**

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact [uml.scholarlycommunications@manchester.ac.uk](mailto:uml.scholarlycommunications@manchester.ac.uk) providing relevant details, so we can investigate your claim.



# Nearly Unbiased Estimation of Sample Skewness

Yifan Li\*

University of Manchester    Lancaster University

This Version: April 23, 2020

## Abstract

In this paper we examine the finite sample bias of sample skewness estimator for financial returns. We show that the bias of conventional sample skewness comes from two sources: the covariance between past return and future volatility, known as the leverage effect, and the covariance between past volatility and future return, commonly referred to as the volatility feedback effect. We derive explicit expressions for this bias and propose a nearly unbiased skewness estimator under mild assumptions. Our simulation study shows that the proposed estimator leads to almost unbiased skewness estimates with a slightly elevated mean squared error, and can reduce the bias of the skewness coefficient estimates by 40%. In our empirical application, we find that bias-corrected average skewness can better predict future market returns comparing to the case without bias-correction. This leads to an improved performance of skewness-based portfolios in terms of Sharpe ratio, certainty equivalence and transaction cost.

*JEL classification:* C13, C22.

*Keywords:* Skewness, bias, return predictability.

---

\*Corresponding author: Alliance Manchester Business School, Booth Street W, Manchester, M15 6PB, UK. e-mail: yifan.li@manchester.ac.uk.

# 1 Introduction

Skewness, defined as the third central moment<sup>1</sup> of a random variable, is commonly used as a measure of tail risk or extreme events for financial returns. It is typically estimated by the sample skewness estimator and serves as regressors in regression analyses in the literature (see e.g. [Chen et al. \(2001\)](#); [Hutton et al. \(2009\)](#); [Boyer et al. \(2010\)](#); [Kim et al. \(2011\)](#); [Jondeau et al. \(2019\)](#); [Langlois \(2020\)](#) among others.)

In this paper, we show that sample skewness for financial returns is biased in finite sample, and the bias consists of two components: the covariance between past return and future squared return, commonly known as the leverage effect ([Black, 1976](#); [Christie, 1982](#)), and the covariance between past squared return and future return, which can be interpreted as the volatility feedback effect ([French et al., 1987](#); [Campbell and Hentschel, 1992](#)).

To correct for the above bias, we develop a nearly unbiased estimator of skewness inspired by [Vogelsang and Yang \(2016\)](#) and [Yang and Vogelsang \(2018\)](#). We show via simulation that the bias of the sample skewness can be substantial, while our proposed bias-corrected estimator is almost unbiased with a slightly increased mean squared error. We also provide nearly unbiased estimates for the leverage and volatility feedback effects and a bias-reduced sample skewness coefficient estimator.

Using the empirical framework of [Jondeau et al. \(2019\)](#), we show that average skewness can better predict future market return if our bias-correction is applied, which in turn leads to a superior skewness-based portfolio in terms of out-of-sample Sharpe ratio, certainty equivalent return and transaction cost.

The remainder of the paper is structured as follows: Section 2 describes our assumption about the return process. Section 3 states our main result. Simulation and empirical analysis can be found in Sections 4 and 5. Section 6 concludes. All proofs are presented in Appendix A.

## 2 Return Process

We state our assumption about the observed financial returns:

**Assumption 1.** On a filtered probability space in discrete<sup>2</sup> time  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ , the observed return process, denoted as  $R = \{R_t\}_{t \geq 0}$ , is assumed to satisfy the following properties: (1)  $R$  is strictly stationary and ergodic. (2)  $R$  has finite sixth unconditional moments. (3) For arbitrary integers  $h_1 \neq h_2 \neq h_3$ , it holds that:

$$\mathbb{E}[(R_{t+h_1} - \mu)(R_{t+h_2} - \mu)(R_{t+h_3} - \mu)] = 0. \quad (2.1)$$

Conditions 1 and 2 are required for meaningful estimation of the third moments of  $R$ . Condition 3 states that the returns observed at three different times are assumed to be ‘uncorrelated’, which is justified by the stylized facts that financial returns are known to have very weak autocorrelation. This assumption is considerably weaker than the usual martingale assumption imposed on financial returns, which requires  $R_t$  to be independent from  $\mathcal{F}_{t-1}$  for all  $t$ .

We introduce the notation  $\tilde{R}_t = R_t - \mu$  as the centred version of  $R_t$ , and define central moments and cross-moments of  $R_t$  as follows:

$$\mu_k \equiv \mathbb{E}[\tilde{R}_t^k], \quad \gamma_{m,n}(i) \equiv \text{Cov}[\tilde{R}_t^m, \tilde{R}_{t-i}^n], \quad k > 1, m, n, j > 0. \quad (2.2)$$

<sup>1</sup>It is also common to scale the third central moment by the cubic standard deviation of a random variable, which will be referred to as the skewness coefficient in this paper.

<sup>2</sup>By discrete time, we assume that returns are observed equidistantly in terms of business days. Note that throughout this paper, we will assume that the indexing variables  $h, i, j, k, m, n$  and  $t$  only take values in  $\mathbb{N}$  whenever no confusion is caused.

In this paper we focus on estimating the skewness of  $R_t$ , namely  $\mu_3$ . The cross-covariances  $\gamma_{1,2}(j)$  and  $\gamma_{2,1}(j)$  play important roles in the estimation of  $\mu_3$ , which is naturally interpreted as volatility feedback and leverage effects respectively. The skewness coefficient is defined as  $\lambda = \mu_3/\mu_2^{1.5}$ . These moments are frequently estimated by their sample counterparts, which we define as follows:

**Definition 1.** For a realization of  $R$  denoted as  $\{R_t\}_{t=1:T}$ , let  $\hat{\mu}$ ,  $\hat{\mu}_k$  and  $\hat{\gamma}_{m,n}(j)$  denote the naive sample moment estimators for  $\mu$ ,  $\mu_k$  and  $\gamma_{m,n}(j)$  respectively, which are defined as:

$$\begin{aligned}\hat{\mu} &= \frac{1}{T} \sum_{t=1}^T R_t, & \hat{\mu}_k &\equiv \frac{1}{T} \sum_{t=1}^T \hat{R}_t^k, \\ \hat{\gamma}_{m,n}(j) &\equiv \frac{1}{T} \sum_{t=1}^T (\hat{R}_t^m - \hat{\mu}_m)(\hat{R}_{t-j}^n - \hat{\mu}_n),\end{aligned}\tag{2.3}$$

where  $\hat{R}_t \equiv R_t - \hat{\mu}$ . Note that we use the circular definition<sup>3</sup> of negative time index such that  $X_{-j} = X_{T-j}$  for  $j \in \{0, \dots, T-1\}$ . Define the naive skewness coefficient estimator as  $\hat{\lambda} = \hat{\mu}_3/\hat{\mu}_2^{1.5}$ .

By a standard ergodic argument,  $\hat{\mu}$ ,  $\hat{\mu}_k$ ,  $\hat{\gamma}_{m,n}(j)$  and  $\hat{\lambda}$  are consistent estimators for  $\mu$ ,  $\mu_k$ ,  $\gamma_{m,n}(j)$  and  $\lambda$  respectively. When  $R$  is i.i.d., an unbiased estimator of  $\mu_3$  can be constructed by scaling  $\hat{\mu}_3$  (Dodge and Rousson, 1999):

$$\mu_3 = \frac{T^2}{(T-2)(T-1)} \mathbb{E}[\hat{\mu}_3].\tag{2.4}$$

However, under Assumption 1, we show in this paper that the above relationship no longer holds, and more information is required to correct for the bias of  $\hat{\mu}_3$ .

### 3 Main Result

We now present our results regarding the bias of the naive third moment estimators:

**Theorem 1.** The following results hold for all  $1 \leq j \leq T-1$ :

$$\begin{aligned}\mathbb{E}[\hat{\mu}_3] &= \frac{(T-1)(T-2)}{T^2} \mu_3 - \frac{3(T-2)}{T^2} (K_{2,1}^{(T)} + K_{1,2}^{(T)}), \\ \mathbb{E}[\hat{\gamma}_{1,2}(j)] &= \frac{(T-j)(T-2)}{T^2} \gamma_{1,2}(j) - \frac{2j}{T^2} \gamma_{1,2}(T-j) + \frac{j(T-2)}{T^2} \gamma_{2,1}(T-j) - \frac{2(T-j)}{T^2} \gamma_{2,1}(j) \\ &\quad + \frac{2-T}{T^2} \mu_3 + \frac{6-T}{T^2} (K_{2,1}^{(T)} + K_{1,2}^{(T)}), \\ \mathbb{E}[\hat{\gamma}_{2,1}(j)] &= \frac{(T-j)(T-2)}{T^2} \gamma_{2,1}(j) - \frac{2j}{T^2} \gamma_{2,1}(T-j) + \frac{j(T-2)}{T^2} \gamma_{1,2}(T-j) - \frac{2(T-j)}{T^2} \gamma_{1,2}(j) \\ &\quad + \frac{2-T}{T^2} \mu_3 + \frac{6-T}{T^2} (K_{2,1}^{(T)} + K_{1,2}^{(T)}),\end{aligned}\tag{3.1}$$

where  $K_{m,n}^{(T)} = \sum_{j=1}^T \frac{T-j}{T} \gamma_{m,n}(j)$ .

**Remark 1.** By setting  $\gamma_{2,1}(j) = \gamma_{1,2}(j) = 0$  for all  $j$ , we retrieve Equation (2.4) as expected. We see that the bias of  $\hat{\mu}_3$  comes from the two quantities  $K_{2,1}^{(T)}$  and  $K_{1,2}^{(T)}$ , which collect the aggregated leverage and volatility feedback effects for the whole sampling period.

Inspired by Vogelsang and Yang (2016), we propose to model  $\gamma_{1,2}(j)$  and  $\gamma_{2,1}(j)$  using the autocovariance structure of an MA( $h$ )-type model. Formally, we introduce the following concept:

<sup>3</sup>Although this is not the conventional definition of covariance, the circular covariance is asymptotically equivalent to its conventional counterpart. Importantly, the circular design greatly simplifies the derivations in this paper. One can potentially use the conventional covariance and proceed with the approach in Vogelsang and Yang (2016), but the derivation becomes much more complicated which is not pursued in this paper.

**Definition 2.** A process  $\{R_t\}_{t \geq 0}$  is  $h$ -correlated of order  $k$  ( $h_k$ -correlated for short) if  $\gamma_{m,n}(j) = 0$  for all  $j > h \geq 0$  and  $m + n = k$ .

We propose a simple linear unbiased estimator of third moments under the  $h_3$ -correlated assumption:

**Theorem 2.** Assume that  $R$  is  $h_3$ -correlated for some  $h < \lfloor \frac{T}{2} \rfloor$ . Let us denote the  $(T-1)$ -by-1 vector of naive estimators as:  $\hat{\Gamma} = \{\hat{\gamma}_{2,1}(1), \hat{\gamma}_{2,1}(2), \dots, \hat{\gamma}_{2,1}(T-1)\}'$ , and the  $(2h+1)$ -by-1 vector of true moments as  $\Gamma_h = \{\mu_3, \gamma_{1,2}(1), \dots, \gamma_{1,2}(h), \gamma_{2,1}(1), \dots, \gamma_{2,1}(h)\}'$ . Define the  $(2h+1)$ -by- $(T-1)$  deterministic matrix  $\mathbf{H}_h = \{H_{i,j}\}_{i=1:2h+1, j=1:T-1}$  as follows:

$$H_{i,j} = \begin{cases} -\frac{T(T-6)}{(T-2)(T-4)} - \frac{3T\mathbb{1}_{\{j \in [h+1, T-h-1]\}}}{(T-4)(T-2h-1)}, & i = 1, \forall j \\ \frac{T^2}{(T-i-1)(T-4)} \left( \frac{2\mathbb{1}_{\{j=i-1\}}}{T} + \frac{(T-2)\mathbb{1}_{\{i+j=T+1\}}}{T} - \frac{\mathbb{1}_{\{j \in [h+1, T-h-1]\}}}{T-2h-1} \right), & i \in [2, h+1], \forall j \\ H_{i-h, T-j}, & i \in [h+2, 2h+1], \forall j \end{cases}, \quad (3.2)$$

where  $\mathbb{1}_{\{\cdot\}}$  is the indicator function.<sup>4</sup> The estimator defined as  $\hat{\Gamma}_h^* \equiv \mathbf{H}_h \hat{\Gamma}$  is an unbiased and consistent estimator of  $\Gamma_h$  such that:

$$\mathbb{E}[\hat{\Gamma}_h^*] = \Gamma_h, \quad \hat{\Gamma}_h^* \xrightarrow{p} \Gamma_h. \quad (3.3)$$

The upper bound of  $h$  is due to the fact that there are only  $T-1$  identifiable parameters in the system of equations (3.1).  $\hat{\Gamma}_h^*$  has the following important properties:

**Proposition 1.** The following results hold:

1. Set  $h = 0$ , then  $\hat{\Gamma}_0^* = \frac{T^2}{(T-2)(T-1)} \hat{\mu}_3$ .
2. For fixed  $h$  as  $T \rightarrow \infty$ , we have:

$$\sqrt{T}(\hat{\Gamma}_h^* - \hat{\Gamma}_h) \sim o_p(1), \quad (3.4)$$

where  $\hat{\Gamma}_h = \{\hat{\mu}_3, \hat{\gamma}_{1,2}(1), \dots, \hat{\gamma}_{1,2}(h), \hat{\gamma}_{2,1}(1), \dots, \hat{\gamma}_{2,1}(h)\}'$  is the naive sample estimator of  $\Gamma_h$ .

**Remark 2.** The first claim above reconciles  $\hat{\Gamma}_0^*$  with the unbiased skewness estimator in the i.i.d. case, which is just a scaled naive skewness estimator. The second claim shows that our estimator is asymptotically equivalent to the naive estimators, which implies that  $\hat{\Gamma}_h^*$  has the same asymptotic variance and distribution as  $\hat{\Gamma}_h$ .

By design, our estimator also provides unbiased estimates for the leverage and volatility feedback effects. Note that  $\hat{\Gamma}_h^*$  is nearly unbiased whenever  $h$  covers all lags with large cross-covariances<sup>5</sup>. Finally, it can be shown that  $\hat{\Gamma}_h^*$  is equivalent to a linear GMM estimator with an identity weighting matrix. For conciseness we leave this result to Appendix B.

## 4 Simulation Study

We simulate financial returns from an asymmetric stochastic volatility-in-mean model (ASV-M), following the work of Harvey and Shephard (1996), Koopman and Hol Uspensky (2002) and Jungbacker and Koopman (2009). The model provides a simple framework to generate persistent leverage and volatility feedback effects, and is specified as follows:

$$\begin{aligned} R_t &= c + \gamma\omega \exp(h_t/2) + \omega \exp(h_t/2)\varepsilon_t, \\ h_{t+1} &= \phi h_t + \sigma\eta_t, \end{aligned} \quad (4.1)$$

<sup>4</sup>Note that by convention, the set  $[a, b] = \emptyset$  if  $a > b$ .

<sup>5</sup>The bias is driven by all cross-covariance terms  $\gamma_{1,2}(j)$  and  $\gamma_{2,1}(j)$  with  $j > h$ , so the estimator is exactly unbiased if the  $h_3$ -correlatedness assumption holds, and almost unbiased if  $\gamma_{1,2}(j) \approx \gamma_{2,1}(j) \approx 0$  for all  $j > h$ . An explicit form of the bias can be found in Appendix B.

where  $|\phi| < 1$  ensures the stationarity of the model. The innovation terms  $\varepsilon_t$  and  $\eta_t$  are i.i.d. standard normal<sup>6</sup> random variables with  $E[\varepsilon_t \eta_{t+1}] = \rho$  and  $E[\varepsilon_t \eta_{t+h}] = 0$  for all  $h \neq 1$ . The parameter  $\rho$  and  $\gamma$  capture the leverage and volatility feedback effects respectively.

Our choice of parameters and the implied moments of returns are presented in Table 1, which follow largely from Jungbacker and Koopman (2009). We consider three different model specifications with  $\gamma \in \{-0.05, 0, 0.05\}$  to explore the impact of negative, neutral and positive volatility feedback effects on our estimator. The Monte Carlo size is chosen to be  $10^5$ .

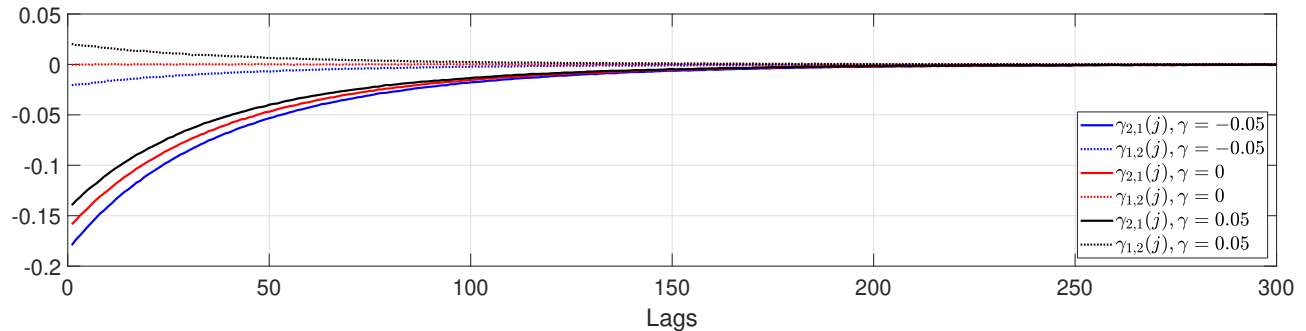
Table 1: Choice of parameters and return moments of the ASV-M model

Choice of Parameters						
$c$	$\gamma$	$\omega^2$	$\phi$	$\sigma^2$	$\rho$	
0.043	$\{-0.05, 0, 0.05\}$	0.681	0.98	0.03	-0.7	
Theoretical Return Moments			Simulated Return Moments			
$\gamma =$	-0.05	0	0.05	-0.05	0	0.05
$\mu$	-0.0024	0.0430	0.0884	-0.0024	0.0429	0.0883
$\mu_2$	0.9950	0.9946	0.9950	0.9952	0.9947	0.9956
$\mu_3$	-0.0623	0	0.0623	-0.0632	-0.0009	0.0620
$\lambda$	-0.0628	0	0.0628	-0.0636	-0.0009	0.0624
$\mu_4$	6.3411	6.3305	6.3411	6.3269	6.3294	6.3689

Note: the simulated return moments are computed based on  $10^5$  simulated sample paths of the ASV-M model with  $T = 2000$ .

Table 1 shows that the simulated return moments closely match the theoretical moments. We plot simulated values of  $\gamma_{1,2}(j)$  and  $\gamma_{2,1}(j)$  in Figure 1. The figure clearly shows a negative leverage effect, while the volatility feedback effect changes sign according to the sign of  $\gamma$ . Both effects decay exponentially and become approximately zero at around 200 lags, which indicates that  $h \geq 200$  should yield a nearly unbiased  $\hat{\Gamma}_h^*$ .

Figure 1: Simulated leverage and volatility feedback effect of the ASV-M model



Note: the simulated return moments are computed based on  $10^5$  simulated sample paths of the ASV-M model with  $T = 2000$ . Parameter choices can be found in Table 1.

In the simulation we consider  $T \in \{500, 1000, 1500\}$ , and construct  $\hat{\Gamma}_h^*$  for  $h \in \{0, 100, 200, 400\}$ . For each choice of  $h$  and  $T$ , we extract  $\hat{\mu}_{3,h}^*$  from  $\hat{\Gamma}_h^*$  and also compute  $\hat{\lambda}_h^* = \hat{\mu}_{3,h}^* / \hat{\mu}_2^{1.5}$ . Our simulation results are presented in Table 2.

We firstly focus on the results of  $\hat{\mu}_{3,0}^*$  and  $\hat{\lambda}_0^*$  in Table 2. Both estimators have significantly positive biases for all sample sizes due to a dominating leverage effect. The size of the biases and the mean squared errors (MSE) shrink as  $T$  increases, which is a result of the consistency of the naive estimators. We would like to stress that the biases can be sizeable compared to the skewness itself. For example, in the case  $\rho = -0.7$  and

<sup>6</sup>For simplicity we assume normality of  $\varepsilon_t$  in order to derive the theoretical model moments, however it is also simple to replace this assumption by a non-normal distribution.

Table 2: Simulation results for skewness and skewness coefficient estimators

Model Spec.			T = 500			T = 1000			T = 1500		
$\rho$	$\gamma$	$h$	$Bias(\hat{\mu}_{3,h}^*)$	$MSE(\hat{\mu}_{3,h}^*)$	$\frac{MSE(\hat{\mu}_{3,h}^*)}{MSE(\hat{\mu}_{3,0}^*)}$	$Bias(\hat{\mu}_{3,h}^*)$	$MSE(\hat{\mu}_{3,h}^*)$	$\frac{MSE(\hat{\mu}_{3,h}^*)}{MSE(\hat{\mu}_{3,0}^*)}$	$Bias(\hat{\mu}_{3,h}^*)$	$MSE(\hat{\mu}_{3,h}^*)$	$\frac{MSE(\hat{\mu}_{3,h}^*)}{MSE(\hat{\mu}_{3,0}^*)}$
-0.7	-0.05	0	<b>0.0447</b>	0.2172	1.0000	<b>0.0236</b>	0.1193	1.0000	<b>0.0154</b>	0.0816	1.0000
-0.7	-0.05	100	<b>0.0052</b>	0.2633	1.2122	<u>0.0024</u>	0.1326	1.1123	0.0009	0.0879	1.0779
-0.7	-0.05	200	0.0005	0.2788	1.2834	-0.0001	0.1351	1.1333	-0.0008	0.0890	1.0910
-0.7	-0.05	400				-0.0006	0.1389	1.1645	-0.0011	0.0896	1.0985
-0.7	0	0	<b>0.0368</b>	0.1975	1.0000	<b>0.0183</b>	0.1106	1.0000	<b>0.0121</b>	0.0738	1.0000
-0.7	0	100	<b>0.0057</b>	0.2262	1.1450	0.0017	0.1192	1.0774	0.0008	0.0778	1.0542
-0.7	0	200	0.0017	0.2371	1.2004	-0.0002	0.1206	1.0902	-0.0005	0.0785	1.0633
-0.7	0	400				-0.0005	0.1227	1.1086	-0.0008	0.0792	1.0728
-0.7	0.05	0	<b>0.0260</b>	0.1852	1.0000	<b>0.0124</b>	0.1004	1.0000	<b>0.0085</b>	0.0695	1.0000
-0.7	0.05	100	<u>0.0032</u>	0.2061	1.1127	0.0002	0.1069	1.0650	0.0001	0.0725	1.0437
-0.7	0.05	200	0.0004	0.2147	1.1590	-0.0012	0.1079	1.0749	-0.0008	0.0729	1.0496
-0.7	0.05	400				-0.0014	0.1093	1.0892	-0.0010	0.0732	1.0532
Model Spec.			T = 500			T = 1000			T = 1500		
$\rho$	$\gamma$	$h$	$Bias(\hat{\lambda}_h^*)$	$MSE(\hat{\lambda}_h^*)$	$\frac{MSE(\hat{\lambda}_h^*)}{MSE(\hat{\lambda}_0^*)}$	$Bias(\hat{\lambda}_h^*)$	$MSE(\hat{\lambda}_h^*)$	$\frac{MSE(\hat{\lambda}_h^*)}{MSE(\hat{\lambda}_0^*)}$	$Bias(\hat{\lambda}_h^*)$	$MSE(\hat{\lambda}_h^*)$	$\frac{MSE(\hat{\lambda}_h^*)}{MSE(\hat{\lambda}_0^*)}$
-0.7	-0.05	0	<b>0.0802</b>	0.0864	1.0000	<b>0.0503</b>	0.0604	1.0000	<b>0.0366</b>	0.0471	1.0000
-0.7	-0.05	100	<b>0.0507</b>	0.0923	1.0691	<b>0.0325</b>	0.0630	1.0439	<b>0.0239</b>	0.0487	1.0339
-0.7	-0.05	200	<b>0.0473</b>	0.0965	1.1181	<b>0.0306</b>	0.0637	1.0553	<b>0.0224</b>	0.0491	1.0406
-0.7	-0.05	400				<b>0.0302</b>	0.0656	1.0870	<b>0.0222</b>	0.0494	1.0479
-0.7	0	0	<b>0.0628</b>	0.0800	1.0000	<b>0.0388</b>	0.0561	1.0000	<b>0.0285</b>	0.0440	1.0000
-0.7	0	100	<b>0.0402</b>	0.0854	1.0680	<b>0.0250</b>	0.0585	1.0417	<b>0.0186</b>	0.0454	1.0304
-0.7	0	200	<b>0.0376</b>	0.0888	1.1097	<b>0.0236</b>	0.0589	1.0499	<b>0.0175</b>	0.0456	1.0357
-0.7	0	400				<b>0.0234</b>	0.0604	1.0755	<b>0.0173</b>	0.0459	1.0424
-0.7	0.05	0	<b>0.0446</b>	0.0779	1.0000	<b>0.0278</b>	0.0547	1.0000	<b>0.0206</b>	0.0429	1.0000
-0.7	0.05	100	<b>0.0288</b>	0.0833	1.0698	<b>0.0180</b>	0.0570	1.0408	<b>0.0135</b>	0.0442	1.0295
-0.7	0.05	200	<b>0.0272</b>	0.0863	1.1079	<b>0.0170</b>	0.0574	1.0481	<b>0.0127</b>	0.0444	1.0340
-0.7	0.05	400				<b>0.0168</b>	0.0582	1.0640	<b>0.0125</b>	0.0445	1.0375

Note: MSE stands for mean squared error. The biases that are significantly different from zero at 1% and 5% are in **bold** and underlined fonts, respectively. Each statistic is computed based on  $10^5$  replications of the ASV-M model, with full parameter specifications given in Table 1.

$\gamma = -0.05$ , the true skewness is  $\mu_3 \approx \lambda \approx -0.063$ . However, the biases of  $\hat{\mu}_{3,0}^*$  and  $\hat{\lambda}_0^*$  for the case  $T = 500$  are 0.0447 and 0.0802 respectively.

The bias of  $\hat{\mu}_{3,h}^*$  drops substantially with  $h = 100$ , and becomes insignificant for all model specifications with  $h \geq 200$  as expected. A sizeable bias reduction of  $\hat{\lambda}_h^*$  is also observed (about 40% for all cases) with  $h \geq 200$ . However,  $\hat{\lambda}_h^*$  is still significantly biased, which is mainly driven by a non-zero correlation between  $\hat{\mu}_{3,h}^*$  and  $1/\hat{\mu}_2^{1.5}$ . Importantly, the MSE ratios also shrink monotonically towards 1 as  $T$  increases for each  $h$ . This is clear evidence supporting the asymptotic equivalence result as derived in Proposition 1.

Unsurprisingly, the bias correction comes at the cost of an increase in MSE. A larger  $h$  always increases the MSE and reduces the bias, suggesting a classic bias-variance trade-off. Therefore, a good choice of  $h$  should be just large enough to correct for the most noticeable leverage and volatility feedback effects. We argue that the price to pay for such bias correction is a small increase (less than 15% for most of the cases) in MSE, which is perhaps much less relevant empirically than correcting for a potentially large bias for the skewness estimates.

## 5 Empirical Application

In this section, we follow [Jondeau et al. \(2019\)](#) and show that the predictability of average skewness on future market return can be improved by our bias-correction procedure, which in turn leads to a better skewness-based trading strategy. We obtain daily prices for all U.S. listed stocks<sup>7</sup> from the Center for Research in Security

<sup>7</sup>To ensure that our estimator can be reliably constructed, for each month we only include stocks that: (1) are traded on NYSE, AMEX or Nasdaq; (2) have share code 10 or 11; (3) are traded above \$1 throughout the month; (4) have a positive daily variance; (5) have valid price records for every trading day. Stock prices are adjusted for corporate actions.

Prices (CRSP) for the period 08/1963-12/2016. Value-weighted market returns and risk-free rates are retrieved from the Kenneth French Data Library<sup>8</sup>.

Let  $D_t$  denote the number of trading days in month  $t$ , and denote the daily excess returns for stock  $i$  and the market as  $r_{i,d}$  and  $r_{m,d}$  respectively, where  $d \in [1, D_t]$ . We compute the following monthly measures of variance and skewness for the  $i$ -th stock:

$$V_{i,t} = \sum_{d=1}^{D_t} (r_{i,d} - \bar{r}_{i,t})^2 + 2 \sum_{d=2}^{D_t} (r_{i,d} - \bar{r}_{i,t})(r_{i-1,d} - \bar{r}_{i,t}), \quad (5.1)$$

$$Sk_{i,t,h} = \hat{\lambda}_{i,h}^* / \sqrt{D_t},$$

where  $\bar{r}_{i,t}$  is the mean daily excess return of stock  $i$  in month  $t$  and  $\hat{\lambda}_{i,h}^*$  is the bias-corrected skewness coefficient constructed from daily excess returns of stock  $i$ . For  $r_{m,d}$ ,  $V_{m,t}$  and  $Sk_{m,t,h}$  are computed analogously. We then construct monthly value-weighted cross-sectional average variance and skewness, defined as  $V_t = \sum_{i=1}^{N_t} w_{i,t} V_{i,t}$  and  $Sk_{t,h} = \sum_{i=1}^{N_t} w_{i,t} Sk_{i,t,h}$  respectively, where  $w_{i,t}$  and  $N_t$  denote the monthly relative market capitalization of stock  $i$  and the number of stocks used in month  $t$ , respectively. Finally, denote  $r_{m,t}$  as the market excess return in month  $t$ .

We firstly replicate the predictive regression in Table 3 of Jondeau et al. (2019), which is specified as follows:

$$r_{m,t+1} = \beta_0 + \beta_1 r_{m,t} + \beta_2 V_{m,t} + \beta_3 Sk_{m,t,h} + \beta_4 V_t + \beta_5 Sk_{t,h} + e_{m,t+1}. \quad (5.2)$$

We consider  $h \in [0, 3]$ , with  $h = 0$  being the baseline case with no bias correction<sup>9</sup>. We present our regression outputs in Table 3.

Table 3: Predictive regression of market return

$h =$	0	1	2	3	0	1	2	3
Panel A: 1963-2016:								
Constant	0.012*** (4.791)	0.012*** (4.943)	0.012*** (4.934)	0.012*** (4.951)	0.013*** (4.778)	0.014*** (4.895)	0.014*** (4.825)	0.014*** (4.840)
$r_{m,t}$	0.090** (2.069)	0.093** (2.151)	0.093** (2.134)	0.093** (2.141)	0.102** (2.274)	0.106** (2.366)	0.106** (2.349)	0.106** (2.362)
$V_{m,t}$					0.730 (0.625)	0.708 (0.605)	0.741 (0.640)	0.747 (0.639)
$Sk_{m,t,h}$					0.014 (1.021)	0.015 (1.159)	0.014 (1.072)	0.013 (1.014)
$V_t$	-0.411* (-1.828)	-0.412* (-1.841)	-0.419* (-1.876)	-0.421* (-1.878)	-0.707 (-1.296)	-0.701 (-1.284)	-0.720 (-1.331)	-0.723 (-1.341)
$Sk_{t,h}$	-0.120** (-2.461)	-0.127*** (-2.741)	-0.121*** (-2.711)	-0.118*** (-2.782)	-0.146*** (-2.734)	-0.155*** (-3.068)	-0.147*** (-3.023)	-0.143*** (-3.060)
Adj. $R^2$	1.694%	1.870%	1.830%	1.818%	1.618%	1.825%	1.770%	1.745%
Panel B: 1990-2016:								
Constant	0.013*** (4.320)	0.014*** (4.456)	0.013*** (4.395)	0.013*** (4.342)	0.016*** (4.697)	0.016*** (4.767)	0.015*** (4.664)	0.015*** (4.648)
$r_{m,t}$	0.070 (1.159)	0.076 (1.243)	0.075 (1.227)	0.073 (1.184)	0.092 (1.393)	0.097 (1.458)	0.097 (1.446)	0.096 (1.419)
$V_{m,t}$					1.822* (1.676)	1.771 (1.618)	1.800* (1.656)	1.845* (1.699)
$Sk_{m,t,h}$					0.009 (0.468)	0.010 (0.542)	0.007 (0.401)	0.007 (0.394)
$V_t$	-0.454 (-1.456)	-0.437 (-1.418)	-0.454 (-1.483)	-0.459 (-1.492)	-1.066** (-2.117)	-1.033** (-2.030)	-1.059** (-2.113)	-1.077** (-2.166)
$Sk_{t,h}$	-0.117* (-1.754)	-0.133** (-2.111)	-0.123** (-2.016)	-0.114* (-1.926)	-0.135* (-1.789)	-0.151** (-2.073)	-0.136* (-1.951)	-0.129* (-1.899)
Adj. $R^2$	2.018%	2.342%	2.249%	2.112%	2.270%	2.573%	2.466%	2.368%

Note: \*\*\*, \*\* and \* represent significance at 1%, 5% and 10% respectively. Newey-West  $t$ -statistics are presented in parentheses, with lag choice  $\lfloor 0.75T^{1/3} \rfloor$ , and  $T$  is the sample size. Adj.  $R^2$  stands for adjusted  $R^2$ .

Our results corroborate the findings in Jondeau et al. (2019) that  $Sk_{t,0}$  can indeed predict market return, as

<sup>8</sup>[https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)

<sup>9</sup>Larger  $h$  further deteriorates the results and is not reported in this paper.



the coefficient of  $Sk_{t,0}$  is significantly negative for both the full sample and the half subsample, after controlling for  $V_{m,t}$  and  $Sk_{m,t,0}$ . Importantly, substituting  $Sk_{t,0}$  with  $Sk_{t,h}$  for  $h > 0$  results in more extreme  $t$ -statistics and magnified adjusted  $R^2$  in all scenarios, which indicates that bias-corrected average skewness measures strengthen the predictability of future market returns. The gain in adjusted  $R^2$  is the most prominent for  $h = 1$  (about 15% in relative terms) but decreases as  $h$  increases, possibly due to an increased variance of  $\hat{\lambda}_{i,h}^*$ .

Such improvements in the in-sample fit can generate non-trivial out-of-sample financial gain. Following [Jondeau et al. \(2019\)](#), we form  $Sk_{t,h}$ -based portfolios and evaluate their performances across different choices of  $h$ . Firstly, let  $s_0 \in [1, T]$  split  $[1, T]$  into an in-sample and an out-of-sample period. The portfolio only consists of market return and a risk-free asset. At each  $t \in [s_0, T - 1]$ , we decide the portfolio weight<sup>10</sup> for market return at time  $t + 1$  by:

$$w_{p,t}^{(h)} = \frac{\hat{\mu}_t^{(h)}}{\alpha \hat{\sigma}_{m,t}^2}, \quad (5.3)$$

where  $\alpha$  is the risk-aversion coefficient,  $\hat{\sigma}_{m,t}^2$  is the rolling window sample variance estimates of  $r_{m,t}$  over the past 5 years, and  $\hat{\mu}_t^{(h)} = \hat{c}_0 + \hat{c}_1 r_{m,t} + \hat{c}_2 Sk_{t,h}$  where  $\hat{c}_0$ ,  $\hat{c}_1$  and  $\hat{c}_2$  are estimated from the following expanding window predictive regression:

$$r_{m+1,s} = c_0 + c_1 r_{m,s} + c_2 Sk_{s,h} + u_{s+1}, \quad s = 1 : t - 1. \quad (5.4)$$

The out-of-sample excess portfolio return at time  $t + 1$  is simply  $r_{p,t+1}^{(h)} = w_{p,t}^{(h)} r_{m,t+1}$ . We evaluate the portfolios by their Sharpe ratios ( $SR^{(h)}$ ), certainty equivalent returns ( $CE^{(h)}$ ) and annual transaction costs ( $TC^{(h)}$ ) defined below:

$$\begin{aligned} SR^{(h)} &= \bar{r}_p^{(h)} / \hat{\sigma}_p^{(h)}, \\ CE^{(h)} &= \bar{r}_p^{(h)} - \alpha (\hat{\sigma}_p^{(h)})^2, \\ TC^{(h)} &= \frac{12f}{T - s_0 + 1} \sum_{t=s_0}^{T-1} |w_{p,t}^{(h)} - w_{p,t-1}^{(h)}|, \end{aligned} \quad (5.5)$$

where  $\bar{r}_p^{(h)}$  and  $\hat{\sigma}_p^{(h)}$  are the out-of-sample sample mean and standard deviation of  $r_{p,t}^{(h)}$ , and  $f$  is the transaction fee per dollar. Our benchmark portfolio is the simple buy-and-hold strategy. We choose  $s_0 \in \{240, 480\}$  to represent 20- and 40-year in-sample periods, and set  $\alpha = 2$ ,  $f = 0.001$  following [Jondeau et al. \(2019\)](#). Our results are presented in Table 4.

Table 4: Out-of-sample portfolio performance

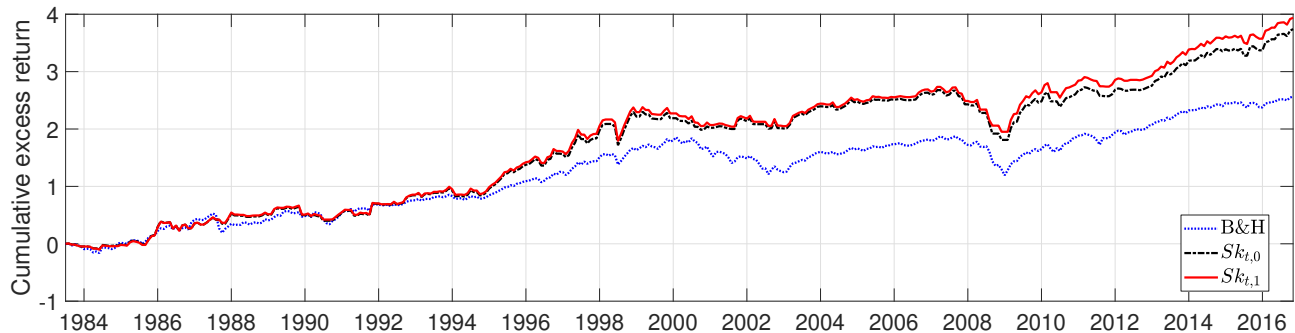
Portfolio	B&H	$Sk_{t,0}$	$Sk_{t,1}$	$Sk_{t,2}$	$Sk_{t,3}$	$\frac{Sk_{t,1}}{Sk_{t,0}}$	$\frac{Sk_{t,2}}{Sk_{t,0}}$	$\frac{Sk_{t,3}}{Sk_{t,0}}$
Panel A: 20-year in-sample period								
Annualized $\bar{r}_p^{(h)}$	7.679%	11.194%	11.785%	11.650%	11.359%	1.0528	1.0407	1.0148
Annualized $\hat{\sigma}_p^{(h)}$	0.1523	0.1947	0.1943	0.1954	0.1970	0.9978	1.0036	1.0118
Annualized $SR^{(h)}$	0.5042	0.5749	0.6066	0.5962	0.5766	1.0551	1.0369	1.0029
Annualized $CE^{(h)}$	5.359%	7.403%	8.011%	7.831%	7.479%	1.0821	1.0578	1.0102
Annual $TC^{(h)}$	-	0.884%	0.874%	0.855%	0.876%	0.9884	0.9674	0.9908
Panel B: 40-year in-sample period								
Annualized $\bar{r}_p^{(h)}$	8.402%	11.115%	12.327%	11.949%	11.787%	1.1091	1.0750	1.0605
Annualized $\hat{\sigma}_p^{(h)}$	0.1414	0.1953	0.1929	0.1946	0.1944	0.9874	0.9960	0.9949
Annualized $SR^{(h)}$	0.5941	0.5690	0.6391	0.6141	0.6065	1.1232	1.0793	1.0659
Annualized $CE^{(h)}$	6.402%	7.299%	8.607%	8.163%	8.009%	1.1792	1.1184	1.0974
Annual $TC^{(h)}$	-	0.908%	0.895%	0.862%	0.884%	0.9859	0.9500	0.9738

Note: B&H stands for the buy-and-hold strategy. We multiply  $\bar{r}_p^{(h)}$  and  $\hat{\sigma}_p^{(h)}$  by 12 to annualize them, and the portfolio performance measures are computed using the annualized version. Columns with title  $\frac{Sk_{t,h}}{Sk_{t,0}}$  compute the ratio of the corresponding portfolio performance statistics for the  $Sk_{t,h}$ -based portfolio to those for the  $Sk_{t,0}$ -based portfolio.

<sup>10</sup>The weight invested in the market return is constrained to be between 0 and 2 to prohibit short-selling and allow a maximum borrowing of 100%.

From Table 4, the  $Sk_{t,1}$ -based portfolio clearly dominates the  $Sk_{t,0}$ -based one in the Markovitz sense with a higher mean return and a lower variance for both choices of  $s_0$ . As can be seen from Figure 2, such advantage in return is persistent throughout the out-of-sample period. The improvements of  $SR^{(h)}$  and  $CE^{(h)}$  can be substantial - larger than 10% in the  $s_0 = 480$  case. Interestingly, the  $Sk_{t,1}$ -based portfolio also has smaller  $TC^{(h)}$  thus should always be preferred in practice. As  $h$  increases, the performances of  $Sk_{t,h}$ -based portfolios deteriorate, but  $SR^{(h)}$ ,  $CE^{(h)}$  and  $TC^{(h)}$  can still outperform those of the  $Sk_{t,0}$ -based portfolio. Overall, the  $Sk_{t,h}$ -based portfolios clearly outperform the buy-and-hold strategy, and  $Sk_{t,0}$ -based portfolio is in general inferior to its bias-corrected version.

Figure 2: Cumulative portfolio excess return: 20-year in-sample period



Note: The figure plots cumulative  $r_{p,t}^{(h)}$  over the out-of-sample period with choice of portfolio specified in the legend, where B&H stands for the buy-and-hold strategy.

## 6 Concluding Remarks

In this paper we demonstrate that the sample skewness estimator for financial returns is biased in the presence of leverage and volatility feedback effects. We derive explicit expressions for the bias, and develop a nearly unbiased skewness estimator. The simulation study confirms our theoretical results that the proposed estimator can greatly reduce the bias for the skewness and skewness coefficient estimators with a moderate MSE increase.

Our empirical application demonstrates that bias-corrected average skewness of individual stocks can better predict future market returns than the naive version, which leads to sizeable financial gains in an out-of-sample portfolio analysis. These promising results have profound implications for investors with a skewness-based trading strategy which deserves further investigation, as we have demonstrated that it is possible to extract more precise signals of the market by bias-correcting sample skewness of individual stocks.

## Acknowledgements

I am grateful for Kevin Aretz, Jiayu Jin, Aleksey Kolokolov, and an anonymous referee for their constructive comments.

# Appendices

## A Proofs

*Proof to Theorem 1.* We introduce some notations:  $\tilde{R}_t = R_t - \mu$ ,  $\check{R} = \frac{1}{T} \sum_{t=1}^T \tilde{R}_t = \hat{\mu} - \mu$ . We therefore have  $R_t - \hat{\mu} = \tilde{R}_t - \check{R}$ . We now start from the first claim:

$$\begin{aligned} \mathbb{E}\left[\frac{1}{T} \sum_{t=1}^T (\tilde{R}_t - \check{R})^3\right] &= \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\tilde{R}_t^3 - 3\tilde{R}_t^2\check{R} + 2\check{R}^3]. \\ &= \mu_3 - \frac{3}{T} \sum_{t=1}^T \mathbb{E}[\tilde{R}_t^2\check{R}] + 2\mathbb{E}[\check{R}^3]. \end{aligned} \quad (\text{A.1})$$

We now look at the second term in the above equation:

$$\begin{aligned} \frac{3}{T} \sum_{t=1}^T \mathbb{E}[\tilde{R}_t^2\check{R}] &= \frac{3}{T^2} \sum_{t=1}^T \mathbb{E}[\tilde{R}_t^2 \sum_{t=1}^T \tilde{R}_t] \\ &= \frac{3}{T} \mathbb{E}[\tilde{R}_t^3] + \frac{3}{T^2} \sum_{t=1}^T \sum_{j=1}^{t-1} \mathbb{E}[\tilde{R}_t^2\tilde{R}_j] + \frac{3}{T^2} \sum_{t=1}^T \sum_{j=t+1}^T \mathbb{E}[\tilde{R}_t^2\tilde{R}_j] \\ &= \frac{3}{T} \mu_3 + \frac{3}{T} (K_{2,1}^{(T)} + K_{1,2}^{(T)}), \end{aligned} \quad (\text{A.2})$$

where we have used repeatedly the stationarity assumption of  $R$ . For the third term, we have:

$$\begin{aligned} 2\mathbb{E}[\check{R}^3] &= \frac{2}{T^3} \mathbb{E}\left[\sum_{i,j,k \in \{1, \dots, T\}} \tilde{R}_i \tilde{R}_j \tilde{R}_k\right] \\ &= \frac{2}{T^2} \mu_3 + \frac{2}{T^3} \mathbb{E}\left[\sum_{i \neq j \neq k \in \{1, \dots, T\}} \tilde{R}_i \tilde{R}_j \tilde{R}_k\right] \\ &= \frac{2}{T^2} \mu_3 + \frac{6}{T^3} \mathbb{E}\left[\sum_{i=j > k \in \{1, \dots, T\}} \tilde{R}_i \tilde{R}_j \tilde{R}_k\right] + \frac{6}{T^3} \mathbb{E}\left[\sum_{i=j < k \in \{1, \dots, T\}} \tilde{R}_i \tilde{R}_j \tilde{R}_k\right] \\ &= \frac{2}{T^2} \mu_3 + \frac{6}{T^2} (K_{2,1}^{(T)} + K_{1,2}^{(T)}). \end{aligned} \quad (\text{A.3})$$

where we have used condition 3 of Assumption 1 in the paper. Therefore, claim 1 can be easily proved by substituting Equations (A.3) and (A.2) into Equation (A.1).

For claim 2, we have the following result:

$$\begin{aligned} \mathbb{E}[\hat{\gamma}_{1,2}(j)] &= \frac{1}{T} \sum_{t=1}^T \mathbb{E}[(\tilde{R}_t - \check{R})(\tilde{R}_{t-j} - \check{R})^2] \\ &= \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\tilde{R}_t \tilde{R}_{t-j}^2 - 2\tilde{R}_t \tilde{R}_{t-j} \check{R}] + 2\mathbb{E}[\check{R}^3] - \frac{1}{T} \sum_{t=1}^T \tilde{R}_t^2 \check{R}. \end{aligned} \quad (\text{A.4})$$

We deal with the terms one-by-one:

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\tilde{R}_t \tilde{R}_{t-j}^2] &= \frac{j}{T} \gamma_{2,1}(T-j) + \frac{T-j}{T} \gamma_{1,2}(j) \\ -\frac{2}{T} \sum_{t=1}^T \mathbb{E}[\tilde{R}_t \tilde{R}_{t-j} \check{R}] &= -\frac{2j}{T^2} \gamma_{2,1}(T-j) - \frac{2(T-j)}{T^2} \gamma_{2,1}(j) - \frac{2(T-j)}{T^2} \gamma_{1,2}(j) - \frac{2j}{T^2} \gamma_{1,2}(T-j) \\ -\frac{1}{T} \sum_{t=1}^T \tilde{R}_t^2 \check{R} &= -\frac{1}{T} \mu_3 - \frac{1}{T} (K_{2,1}^{(T)} + K_{1,2}^{(T)}). \end{aligned} \quad (\text{A.5})$$

Again we invoke condition 3 of Assumption 1 in the paper to derive the second relationship. Substituting the above and Equation (A.3) into Equation (A.4) yield the desired result. For claim 3, we note that by symmetry of the circular cross-covariance, we always have  $\hat{\gamma}_{2,1}(j) = \hat{\gamma}_{1,2}(T-j)$ . By interchanging the index  $t$  and  $t-j$  in the expression of claim 2, we obtain the expression for claim 3. This completes the proof.  $\square$

*Proof to Theorem 2.* We firstly show that the matrix  $\mathbf{H}_h$  solves the linear equation  $\mathbf{H}_h \mathbf{E}[\hat{\mathbf{\Gamma}}] = \mathbf{\Gamma}$  under the  $h_3$ -correlatedness assumption, which implies the unbiasedness of  $\hat{\mathbf{\Gamma}}_h^*$ . From Theorem 1 in the paper, for all  $j \in [h+1, T-h-1]$ , it holds under the  $h_3$ -correlatedness assumption that:

$$\mathbf{E}[\hat{\gamma}_{2,1}(j)] = \frac{2-T}{T^2} \mu_3 + \frac{6-T}{T^2} (K_{2,1}^{(T)} + K_{1,2}^{(T)}). \quad (\text{A.6})$$

Therefore by averaging the above equation over all  $j \in [h+1, T-h-1]$ , we have:

$$\frac{1}{T-2h-1} \sum_{j=h+1}^{T-h-1} \mathbf{E}[\hat{\gamma}_{2,1}(j)] = \frac{2-T}{T^2} \mu_3 + \frac{6-T}{T^2} (K_{2,1}^{(T)} + K_{1,2}^{(T)}). \quad (\text{A.7})$$

Now for all  $j \in [1, h]$ , we see that:

$$\begin{aligned} \mathbf{E}[\hat{\gamma}_{1,2}(j)] &= \frac{(T-j)(T-2)}{T^2} \gamma_{1,2}(j) - \frac{2(T-j)}{T^2} \gamma_{2,1}(j) + \frac{2-T}{T^2} \mu_3 + \frac{6-T}{T^2} (K_{2,1}^{(T)} + K_{1,2}^{(T)}), \\ \mathbf{E}[\hat{\gamma}_{2,1}(j)] &= \frac{(T-j)(T-2)}{T^2} \gamma_{2,1}(j) - \frac{2(T-j)}{T^2} \gamma_{1,2}(j) + \frac{2-T}{T^2} \mu_3 + \frac{6-T}{T^2} (K_{2,1}^{(T)} + K_{1,2}^{(T)}). \end{aligned} \quad (\text{A.8})$$

Substituting Equation (A.7) into the above, we arrive at a system of two equations with two unknowns  $\gamma_{1,2}(j)$  and  $\gamma_{2,1}(j)$ . Solving the above system of equations yields:

$$\begin{aligned} \gamma_{1,2}(j) &= \frac{T^2}{(T-j)(T-4)} \left( \frac{T-2}{T} \mathbf{E}[\hat{\gamma}_{2,1}(T-j)] + \frac{2}{T} \mathbf{E}[\hat{\gamma}_{2,1}(j)] - \frac{1}{T-2h-1} \sum_{j=h+1}^{T-h-1} \mathbf{E}[\hat{\gamma}_{2,1}(j)] \right), \\ \gamma_{2,1}(j) &= \frac{T^2}{(T-j)(T-4)} \left( \frac{T-2}{T} \mathbf{E}[\hat{\gamma}_{2,1}(j)] + \frac{2}{T} \mathbf{E}[\hat{\gamma}_{2,1}(T-j)] - \frac{1}{T-2h-1} \sum_{j=h+1}^{T-h-1} \mathbf{E}[\hat{\gamma}_{2,1}(j)] \right), \end{aligned} \quad (\text{A.9})$$

where we have used the relationship  $\hat{\gamma}_{2,1}(j) = \hat{\gamma}_{1,2}(T-j)$ . We now solve for  $\mu_3$  using the following result from Theorem 1:

$$\mathbf{E}[\hat{\mu}_3] = \frac{(T-1)(T-2)}{T^2} \mu_3 - \frac{3(T-2)}{T^2} (K_{2,1}^{(T)} + K_{1,2}^{(T)}). \quad (\text{A.10})$$

Combining the above equation with Equation (A.7), we again form a system of two equations with two knowns  $\mu_3$  and  $(K_{2,1}^{(T)} + K_{1,2}^{(T)})$ . Solving for  $\mu_3$  yields:

$$\mu_3 = \frac{T(T-6)}{(T-2)(T-4)} \mathbf{E}[\hat{\mu}_3] - \frac{3T}{T-4} \frac{1}{T-2h-1} \sum_{j=h+1}^{T-h-1} \mathbf{E}[\hat{\gamma}_{2,1}(j)]. \quad (\text{A.11})$$

Finally, as  $\hat{\mu}_3$  is not included in the vector  $\hat{\mathbf{\Gamma}}$ , we use the following lemma:

**Lemma 1.** It holds that  $\sum_{j=1}^{T-1} \hat{\gamma}_{2,1}(j) = -\hat{\mu}_3$ .

*Proof.* We prove this lemma algebraically:

$$\begin{aligned} \sum_{j=1}^{T-1} \hat{\gamma}_{2,1}(j) &= \frac{1}{T} \sum_{j=1}^{T-1} \sum_{t=1}^T (\hat{R}_t^2 - \hat{\mu}_2) \hat{R}_{t-j} \\ &= \frac{1}{T} \sum_{j=1}^{T-1} \sum_{t=1}^T \hat{R}_t^2 \hat{R}_{t-j} = \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^{T-1} \hat{R}_t^2 \hat{R}_{t-j} \\ &= \frac{1}{T} \sum_{j=1}^{T-1} \sum_{t=1}^T \hat{R}_t^2 (-\hat{R}_t) = -\hat{\mu}_3. \end{aligned} \quad (\text{A.12})$$

Note that we have applied the relation  $\sum_{t=1}^T \hat{R}_t = 0$  repetitively and used the circular definition of negative index in the above derivation. This completes the proof.  $\square$

Substituting  $\sum_{h=1}^{T-1} \hat{\gamma}_{2,1}(j) = -\hat{\mu}_3$  into Equation (A.11), we find that:

$$\mu_3 = -\frac{T(T-6)}{(T-2)(T-4)} \sum_{h=1}^{T-1} \mathbb{E}[\hat{\gamma}_{2,1}(j)] - \frac{3T}{T-4} \frac{1}{T-2h-1} \sum_{j=h+1}^{T-h-1} \mathbb{E}[\hat{\gamma}_{2,1}(j)]. \quad (\text{A.13})$$

Now write the  $2h+1$  equations in Equations (A.9) and (A.13) in matrix form, we clearly have  $\mathbb{E}[\hat{\Gamma}_h^*] = \mathbf{H}_h \mathbb{E}[\hat{\Gamma}] = \mathbf{\Gamma}$  as desired. This proves the unbiasedness of  $\hat{\Gamma}_h^*$ . The consistency of  $\hat{\Gamma}_h^*$  follows from the consistency of  $\hat{\Gamma}$  and the continuous mapping theorem. This completes the proof.  $\square$

*Proof to Proposition 1.* For claim 1, we firstly note that as  $h = 0$ ,  $\mathbf{H}_0$  becomes a row vector with the following structure:

$$\mathbf{H}_0 = -\left(\frac{T(T-6)}{(T-2)(T-4)} + \frac{3T}{(T-4)(T-1)}\right) \boldsymbol{\nu}'_{(T-1) \times 1},$$

where  $\boldsymbol{\nu}_{(T-1) \times 1}$  is a  $(T-1)$ -by-1 vector of ones. Simplifying the above expression and multiplying both sides by  $\hat{\Gamma}$ , we have:

$$\hat{\Gamma}_0^* = \mathbf{H}_0 \hat{\Gamma} = -\frac{T^2}{(T-2)(T-1)} \sum_{j=1}^{T-1} \hat{\gamma}_{2,1}(j). \quad (\text{A.14})$$

Claim 1 therefore follows by a direct application of Lemma 1. For claim 2, we express each term of  $\hat{\Gamma}_h^*$  explicitly:

$$\begin{aligned} \hat{\mu}_{3,h}^* &= \frac{T(T-6)}{(T-2)(T-4)} \hat{\mu}_3 - \frac{3T}{T-4} \frac{1}{T-2h-1} \sum_{j=h+1}^{T-h-1} \hat{\gamma}_{2,1}(j), \\ \hat{\gamma}_{1,2,h}^*(j) &= \frac{T^2}{(T-j)(T-4)} \left( \frac{T-2}{T} \hat{\gamma}_{2,1}(T-j) + \frac{2}{T} \hat{\gamma}_{2,1}(j) - \frac{1}{T-2h-1} \sum_{j=h+1}^{T-h-1} \hat{\gamma}_{2,1}(j) \right), \\ \hat{\gamma}_{2,1,h}^*(j) &= \frac{T^2}{(T-j)(T-4)} \left( \frac{T-2}{T} \hat{\gamma}_{2,1}(j) + \frac{2}{T} \hat{\gamma}_{2,1}(T-j) - \frac{1}{T-2h-1} \sum_{j=h+1}^{T-h-1} \hat{\gamma}_{2,1}(j) \right), \end{aligned} \quad (\text{A.15})$$

Examining the order of  $T$  for the coefficients, it is obvious that for all fixed  $h$  and  $j \leq h$ :

$$\begin{aligned} \hat{\mu}_{3,h}^* - \hat{\mu}_3 &\sim \mathcal{O}_p(T^{-1}), \\ \hat{\gamma}_{1,2,h}^*(j) - \hat{\gamma}_{1,2}(j) &\sim \mathcal{O}_p(T^{-1}), \\ \hat{\gamma}_{2,1,h}^*(j) - \hat{\gamma}_{2,1}(j) &\sim \mathcal{O}_p(T^{-1}). \end{aligned} \quad (\text{A.16})$$

This implies that  $\hat{\Gamma}_h^* - \hat{\Gamma}_h \sim \mathcal{O}_p(T^{-1})$  which further implies claim 2. This completes the proof.  $\square$

## B A GMM Representation of $\hat{\Gamma}_h^*$

This section shows that  $\hat{\Gamma}_h^*$  has a linear GMM representation with an identity weighting matrix, and derive its bias when the  $h_3$ -correlatedness assumption fails. We start by writing  $\mathbf{\Gamma}_h^+ = \{\gamma_{1,2}(h+1), \dots, \gamma_{1,2}(T-1), \gamma_{2,1}(h+1), \dots, \gamma_{2,1}(T-1)\}$ , and focus on the following system of equations for all  $j \in [1, T-1]$ :

$$\begin{aligned} \mathbb{E}[\hat{\gamma}_{2,1}(j)] &= \frac{(T-j)(T-2)}{T^2} \gamma_{2,1}(j) - \frac{2j}{T^2} \gamma_{2,1}(T-j) + \frac{j(T-2)}{T^2} \gamma_{1,2}(T-j) - \frac{2(T-j)}{T^2} \gamma_{1,2}(j) \\ &\quad + \frac{2-T}{T^2} \mu_3 + \frac{6-T}{T^2} (K_{2,1}^{(T)} + K_{1,2}^{(T)}), \end{aligned} \quad (\text{B.1})$$

As the r.g.s. of each equation can be represented as a linear combination of  $\mathbf{\Gamma}_h$  and  $\mathbf{\Gamma}_h^+$ , we can write the system of equations in a compact matrix form:

$$\mathbb{E}[\hat{\Gamma}] = \mathbf{G}_h \mathbf{\Gamma}_h + \mathbf{G}_h^+ \mathbf{\Gamma}_h^+, \quad (\text{B.2})$$

where  $\mathbf{G}_h$  and  $\mathbf{G}_h^+$  are the corresponding deterministic coefficient matrices, which can be constructed explicitly. Specifically,  $\mathbf{G}_h = \{G_{i,j}\}_{i=1:T-1, j=1:2h+1}$  has the following structure:

$$G_{i,j} = \begin{cases} \frac{2-T}{T^2}, & i \in [1, T-1], j = 1 \\ \frac{(T-i)(T-2)\mathbb{1}_{\{i=j-1\}} - 2i\mathbb{1}_{\{i+j=T+1\}} + \frac{(6-T)(T-j-1)}{T^3}, & i \in [1, T-1], j \in [2, h+1] \\ G_{T-i, j-h}, & i \in [1, T-1], j \in [h+2, 2h+1] \end{cases}, \quad (\text{B.3})$$

and the expression of  $\mathbf{G}_h^+$  can be derived similarly. The representation in Equation (B.2) is a classic linear GMM estimation problem, where we have  $T-1$  available moment conditions on the l.h.s but  $2T-1$  moments to be estimated ( $\mathbf{\Gamma}_h$  and  $\mathbf{\Gamma}_h^+$ ). For the model to be at least exactly identified, we can only estimate at most  $T-1$  number of parameters. Using the  $h_3$ -correlatedness assumption as an identification restriction, it is clear that  $h$  must satisfy  $2h+1 \leq T-1$ . This justifies the upper bound of  $h$  in Theorem 2 of the paper.

Under the  $h_3$ -correlatedness when the system of equations is at least just-identified, we have the following classic linear GMM estimator of  $\mathbf{\Gamma}_h$ :

**Theorem 3.** Assume  $R$  is  $h_3$ -correlated with  $h < \lfloor \frac{T}{2} \rfloor$ . Denote  $\mathbf{W}$  as a  $(T-1)$ -by- $(T-1)$  non-random positive definite matrix, and define  $\hat{\mathbf{\Gamma}}_h^{(\mathbf{W})}$  as:

$$\hat{\mathbf{\Gamma}}_h^{(\mathbf{W})} = (\mathbf{G}'_h \mathbf{W} \mathbf{G}_h)^{-1} \mathbf{G}'_h \mathbf{W} \hat{\mathbf{\Gamma}}. \quad (\text{B.4})$$

Then  $\hat{\mathbf{\Gamma}}_h^{(\mathbf{W})}$  satisfies the following properties for all valid choices of  $\mathbf{W}$  and  $h$ :

$$\mathbb{E}[\hat{\mathbf{\Gamma}}_h^{(\mathbf{W})}] = \mathbf{\Gamma}_h, \quad \hat{\mathbf{\Gamma}}_h^{(\mathbf{W})} \xrightarrow{p} \mathbf{\Gamma}_h. \quad (\text{B.5})$$

*Proof.* Firstly, by comparing the columns of  $\mathbf{G}_h$  directly, we see that  $\mathbf{G}_h$  has full column rank as long as  $h < \lfloor \frac{T}{2} \rfloor$ , which implies for any positive definite matrix,  $\mathbf{G}'_h \mathbf{W} \mathbf{G}_h$  must also have full rank and is hence invertible. Also, under the  $h_3$ -correlatedness assumption, we have  $\mathbf{\Gamma}_h^+$  by construction. The following result is therefore obvious:

$$\mathbb{E}[\hat{\mathbf{\Gamma}}_h^{(\mathbf{W})}] = (\mathbf{G}'_h \mathbf{W} \mathbf{G}_h)^{-1} \mathbf{G}'_h \mathbf{W} \mathbb{E}[\hat{\mathbf{\Gamma}}] = (\mathbf{G}'_h \mathbf{W} \mathbf{G}_h)^{-1} \mathbf{G}'_h \mathbf{W} \mathbf{G}_h \mathbf{\Gamma}_h = \mathbf{\Gamma}_h, \quad (\text{B.6})$$

which proves the unbiasedness of  $\hat{\mathbf{\Gamma}}_h^{(\mathbf{W})}$ . Note that the results hold for arbitrary valid  $\mathbf{W}$  and  $h$ . The consistency follows from the consistency of  $\hat{\mathbf{\Gamma}}$  and the continuous mapping theorem, and the proof is complete.  $\square$

We now show that the estimator  $\hat{\mathbf{\Gamma}}_h^*$  proposed in our paper is a special case of  $\hat{\mathbf{\Gamma}}_h^{(\mathbf{W})}$ :

**Proposition 2.** It holds that  $\hat{\mathbf{\Gamma}}_h^* = \hat{\mathbf{\Gamma}}_h^{(\mathbf{I})}$ , where  $\mathbf{I}$  is a  $(T-1)$ -by- $(T-1)$  identity matrix.

*Proof.* We need to prove that  $\mathbf{H}_h = (\mathbf{G}'_h \mathbf{G}_h)^{-1} \mathbf{G}'_h$ , where  $\mathbf{H}_h$  is defined in Theorem 2 of the paper. This is equivalent to proving  $\mathbf{G}'_h \mathbf{G}_h \mathbf{H}_h = \mathbf{G}'_h$ . We start with a result obtained from manual calculation:

$$\mathbf{G}_h \mathbf{H}_h = \begin{bmatrix} \mathbf{I}_h, & \mathbf{0}_{h \times (T-2h-1)}, & \mathbf{0}_{h \times h} \\ \mathbf{0}'_{h \times (T-2h-1)} & \frac{\mathbf{\iota}_{(T-2h-1) \times (T-2h-1)}}{T-2h-1}, & \mathbf{0}'_{h \times (T-2h-1)} \\ \mathbf{0}_{h \times h} & \mathbf{0}_{h \times (T-2h-1)}, & \mathbf{I}_h, \end{bmatrix}, \quad (\text{B.7})$$

where  $\mathbf{I}_h$  is the  $h$ -by- $h$  identity matrix,  $\mathbf{0}_{m \times n}$  and  $\mathbf{\iota}_{m \times n}$  are  $m$ -by- $n$  matrices of zeros and ones, respectively. We now write as a block structure:  $\mathbf{G}'_h = (\mathbf{G}_h^{(1)}, \mathbf{G}_h^{(2)}, \mathbf{G}_h^{(3)})$ , where  $\mathbf{G}_h^{(1)}$  and  $\mathbf{G}_h^{(3)}$  are the first and last  $h$  columns of  $\mathbf{G}'_h$ , and  $\mathbf{G}_h^{(2)}$  is the  $h+1$  to the  $T-h-1$  columns, respectively. From the block-diagonal structure of  $\mathbf{G}_h \mathbf{H}_h$ , it becomes clear that  $\mathbf{G}'_h \mathbf{G}_h \mathbf{H}_h = \mathbf{G}'_h$  is equivalent to the following relationship:

$$\mathbf{G}_h^{(2)} = \mathbf{G}_h^{(2)} \frac{\mathbf{\iota}_{(T-2h-1) \times (T-2h-1)}}{T-2h-1}. \quad (\text{B.8})$$

From the structure of  $\mathbf{G}_h$  in Equation (B.3), we see that each row of  $\mathbf{G}_h^{(2)}$  has the structure  $a \cdot \mathbf{1}_{1 \times (T-2h-1)}$  for some constant  $a$ . Also, the following identity holds:

$$\mathbf{1}_{1 \times (T-2h-1)} \frac{\mathbf{1}_{(T-2h-1) \times (T-2h-1)}}{T-2h-1} = \mathbf{1}_{1 \times (T-2h-1)}. \quad (\text{B.9})$$

Therefore, Equation (B.8) holds by row-wise application of the above equation. This implies  $\mathbf{G}'_h \mathbf{G}_h \mathbf{H}_h = \mathbf{G}'_h$  and hence the proposition, and the proof is complete.  $\square$

Finally, using the GMM representation, it is simple to derive the bias of  $\hat{\mathbf{\Gamma}}_h^*$ :

**Corollary 1.** Suppose  $R$  is  $T_3$ -correlated, then it holds that:

$$\text{E}[\hat{\mathbf{\Gamma}}_h^*] - \mathbf{\Gamma}_h = \mathbf{H}_h \mathbf{G}_h^+ \mathbf{\Gamma}_h^+. \quad (\text{B.10})$$

*Proof.* Multiplying both sides of Equation (B.2) by  $\mathbf{H}_h$ , we have:

$$\mathbf{H}_h \text{E}[\hat{\mathbf{\Gamma}}_h] = \mathbf{H}_h \mathbf{G}_h \mathbf{\Gamma}_h + \mathbf{H}_h \mathbf{G}_h^+ \mathbf{\Gamma}_h^+. \quad (\text{B.11})$$

The l.h.s. is equal to  $\text{E}[\hat{\mathbf{\Gamma}}_h^*]$  by construction, and we have  $\mathbf{H}_h \mathbf{G}_h = (\mathbf{G}'_h \mathbf{G}_h)^{-1} \mathbf{G}'_h \mathbf{G}_h = \mathbf{I}_h$ . This completes the proof.  $\square$

The corollary clearly shows that the bias of  $\hat{\mathbf{\Gamma}}_h^*$  is a linear combination of  $\mathbf{\Gamma}_h^+$ , which are the cross-moments with lags larger than  $h$ . Therefore, if  $h$  is set to be large enough such that  $\mathbf{\Gamma}_h^+$  is close to zero elementwise,  $\hat{\mathbf{\Gamma}}_h^*$  will be a nearly unbiased estimator as claimed in the paper.

## References

- Black, F. (1976). Studies of Stock Price Volatility Changes. In *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics*, pages 177–181.
- Boyer, B., Mitton, T., and Vorkink, K. (2010). Expected idiosyncratic skewness. *Review of Financial Studies*, 23(1):170–202.
- Campbell, J. Y. and Hentschel, L. (1992). No news is good news : An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31:281–318.
- Chen, J., Hong, H., and Stein, J. C. (2001). Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices. *Journal of Financial Economics*, 61(3):345–381.
- Christie, A. (1982). The stochastic behavior of common stock variances Value, leverage and interest rate effects. *Journal of Financial Economics*, 10:407–432.
- Dodge, Y. and Rousson, V. (1999). The Complications of the Fourth Central Moment. *The American Statistician*, 53(3):267–269.
- French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19:3–29.
- Harvey, A. C. and Shephard, N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics*, 14(4):429–434.
- Hutton, A. P., Marcus, A. J., and Tehranian, H. (2009). Opaque financial reports, R2, and crash risk. *Journal of Financial Economics*, 94(1):67–86.
- Jondeau, E., Zhang, Q., and Zhu, X. (2019). Average skewness matters. *Journal of Financial Economics*, 134(1):29–47.
- Jungbacker, B. and Koopman, S. J. (2009). Parameter Estimation and Practical Aspects of Modeling Stochastic Volatility. In Mikosch, T., Kreiß, J. P., Davis, R., and Andersen, T. G., editors, *Handbook of Financial Time Series*, pages 810–838. Springer Berlin Heidelberg.
- Kim, J.-B., Li, Y., and Zhang, L. (2011). Corporate tax avoidance and stock price crash risk: Firm-level analysis. *Journal of Financial Economics*, 100(3):639–662.
- Koopman, S. J. and Hol Uspensky, E. (2002). The stochastic volatility in mean model: empirical evidence from international stock markets. *Journal of Applied Econometrics*, 17(6):667–689.
- Langlois, H. (2020). Measuring skewness premia. *Journal of Financial Economics*, 135(2):399–424.
- Vogelsang, T. J. and Yang, J. (2016). Exactly/Nearly Unbiased Estimation of Autocovariances of a Univariate Time Series With Unknown Mean. *Journal of Time Series Analysis*, 37(6):723–740.
- Yang, J. and Vogelsang, T. J. (2018). Finite sample performance of a long run variance estimator based on exactly (almost) unbiased autocovariance estimators. *Economics Letters*, 165:21–27.