Temporal structure and gain-loss asymmetry for real and artificial stock indices

Johannes Vitalis Siven* and Jeffrey Todd Lins†
Saxo Bank A/S, Philip Heymans Allé 15, DK-2900 Hellerup, Denmark

(Received 6 July 2009; published 16 November 2009)

Previous research has shown that for stock indices, the most likely time until a return of a particular size has been observed is longer for gains than for losses. We demonstrate that this so-called gain-loss asymmetry vanishes if the temporal dependence structure is destroyed by scrambling the time series. We also show that an artificial index constructed by a simple average of a number of individual stocks display gain-loss asymmetry—this allows us to explicitly analyze the dependence between the index constituents. We consider mutual information and correlation-based measures and show that the stock returns indeed have a higher degree of dependence in times of market downturns than upturns.

DOI: 10.1103/PhysRevE.80.057102 PACS number(s): 89.65.Gh

I. INTRODUCTION

Inspired by research in the field of turbulence, Simonsen *et al.* [1] considered "inverse statistics" of financial time series: what is the smallest time interval needed for an asset to cross a fixed return level ρ? Figure 1 shows the distribution of this random variable, the *first passage time*, for the Dow Jones Industrial Average index, for $\rho = \pm 5\%$. As noted by Jensen *et al.* [2], the most likely first passage time is shorter for $\rho = -5\%$ than for $\rho = 5\%$, which they refer to as the *gain-loss asymmetry*.

In this Brief Report, we show that the gain-loss asymmetry in the Dow Jones index vanishes if the time series is "scrambled"—that is, if one considers a new time series constructed by randomly permuting the returns. This basic fact, which seems to have gone unnoticed in the literature so far, has important implications: the gain-loss asymmetry is *not* due to properties of the unconditional index returns, such as skewness, but is rather an expression of potentially complex temporal structure. This finding resonates with the results from Siven *et al.* [3], where wavelet analysis is used to demonstrate that the gain-loss asymmetry is a long time scale phenomenon—it vanishes if enough low-frequency content is removed from the index, that is, if the index is sufficiently "detrended." Siven *et al.* [3] also presented a generalization of the asymmetric synchronous market model from Donangelo *et al.* [4], where prolonged periods of high correlation between the individual stocks during index downturns gives rise to a gain-loss asymmetry.

Whether the constituents of, e.g., the Dow Jones index, indeed tend to move with a greater degree of dependence during market downturns could, in principle, be tested empirically by an analysis of the time series of the individual stocks. That is an awkward task, however, since the relative weights for different stocks in these indices have changed over time in complicated ways. To address this issue, we demonstrate that if one defines a new artificial index by simply taking the average of a number of stocks, this index also displays a gain-loss asymmetry. With the constituents readily available, we consider two measures based on correlation and mutual information and show that there indeed is a higher degree of dependence between the stock returns during index downturns than upturns.

II. GAIN-LOSS ASYMMETRY AND TEMPORAL STRUCTURE

For a given process $\{I_t\}_{t\geq 0}$, for instance, daily closing prices of a stock index, the first passage time τ_ρ of the level ρ is defined as

$$\tau_\rho = \begin{cases} \inf\{s>0;\quad \log(I_{t+s}/I_t) \geq \rho\} & \text{if } \rho > 0 \\ \inf\{s>0;\quad \log(I_{t+s}/I_t) \leq \rho\} & \text{if } \rho < 0, \end{cases}$$

and is assumed to be independent of t. The distribution of τ_ρ is estimated in a straightforward manner from a time series $I_0, \ldots, I_T$. Consider $\rho > 0$ and let $t+s$ be the smallest time point such that $\log(I_{t+s}/I_t) \geq \rho$, if such a time point exists. In that case, s is viewed as an observation of τ_ρ. [If $\rho < 0$, take instead $t+s$ such that $\log(I_{t+s}/I_t) \leq \rho$.] Running t from 0 to $T-1$ gives a set of observations from which the distribution of τ_ρ is estimated as the empirical distribution. Given the empirical distribution, we follow Jensen *et al.* [2] and compute a fit of the density function for the generalized gamma distribution. This density is plotted as a solid line together

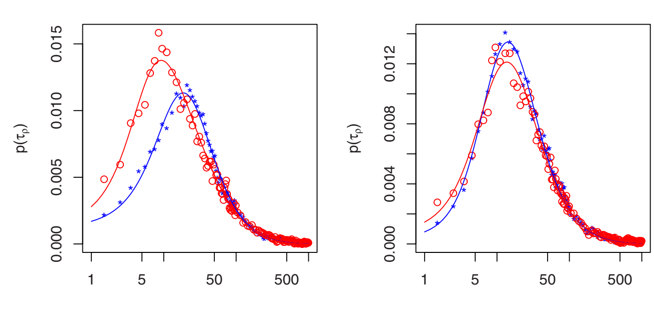

FIG. 1. (Color online) Estimated distribution of the first passage time τ_ρ for the log price of the Dow Jones Industrial Average index (left) and its scrambled version (right). The graphs correspond to $\rho = +5\%$ (stars) and $\rho = -5\%$ (rings). The solid lines are fitted generalized gamma density functions.

*jvs@saxobank.com
†jtl@saxobank.com

with the empirical distribution in all figures, to guide the eyes—we do not discuss the fitted parameters nor claim that τ_ρ truly follows a generalized gamma distribution.

III. GAIN-LOSS ASYMMETRY FOR THE DOW JONES INDEX

Figure 1 shows the estimated first passage time for the Dow Jones index. As discussed in the introduction, there is a gain-loss asymmetry in that the most likely first passage time is shorter for $\rho=-5\%$ than for $\rho=5\%$. Next, we construct a *scrambled* version of the index by randomly rearranging the log returns. Formally, if $I_0,\ldots,I_T$ denotes the time series of daily closing prices of the Dow Jones index, let $\delta I_t=\log(I_t/I_{t-1})$ for $t=1,\ldots,T$ and draw a random permutation $j_1,\ldots,j_T$ of $\{1,\ldots,T\}$. We define

$$\delta\tilde{I}_t=\delta I_{j_t},\quad \text{for } t=1,\ldots,T,$$

and let the scrambled index be given by $\tilde{I}_0=I_0$ and

$$\tilde{I}_t=\tilde{I}_0\exp\left(\sum_{s=1}^{t}\delta\tilde{I}_s\right),\quad \text{for } t=1,\ldots,T.$$

Figure 1 shows that the scrambled index does *not* display a gain-loss asymmetry. This result is surprisingly strong: since the empirical return distributions are identical for an index and any of its scrambled versions, it shows that the gain-loss asymmetry is an expression of potentially complex temporal structure in the index. This fits nicely with the results from Siven *et al.* [3], where a multiscale decomposition is used to demonstrate that the gain-loss asymmetry is a long rather than short scale phenomenon.

The gain-loss asymmetry in the asymmetric synchronous market model from Donangelo *et al.* [4] does not disappear when the index returns are scrambled. This is to be expected since the daily returns in that model are independent and identically distributed, so all statistical properties remain the same when the time series is scrambled. However, for the generalized model proposed in Siven *et al.* [3], the asymmetry *does* vanish, in perfect agreement with the Dow Jones index (see Fig. 2).

IV. GAIN-LOSS ASYMMETRY FOR AN ARTIFICIAL INDEX

Consider N stocks and let $S_{n,t}$ denote the closing price of the nth stock on day t, for $t=0,1,\ldots,T$. We consider the artificial index constructed by averaging all the stocks,

$$I_t=\frac{1}{N}\sum_{n=1}^{N}\frac{S_{n,t}}{S_{n,0}}.$$

The denominators $S_{n,0}$ give all stocks equal weight in the index at time $t=0$.

We consider historical stock prices from January 1970 until December 2008 for the following 12 Dow Jones constituents: Boeing Co. (BA), Citigroup, Inc. (C), El DuPont de Nemours & Co. (DD), General Electric Co. (GE), General Motors Corporation (GM), International Business Machines Corp. (IBM), Johnson & Johnson (JNJ), JPMorgan Chase & Co. (JPM), The Coca-Cola Co. (KO), McDonald's Corp. (MCD), Proctor & Gamble Co. (PG), and Alcoa, Inc. (AA). These companies are chosen since long-time series of stock returns are available; but our results are stable in the sense that adding or removing companies gives very similar results.

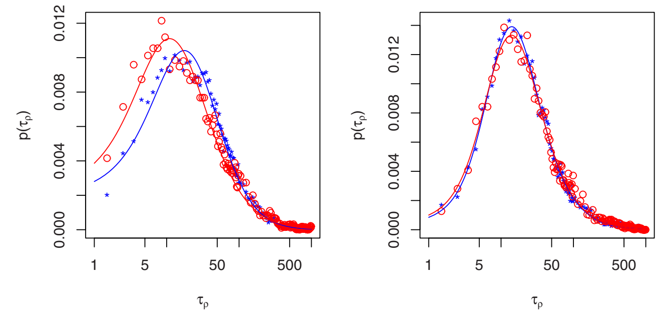

FIG. 2. (Color online) Estimated distribution of the first passage time τ_ρ for the log price in a realization of the generalized asymmetric synchronous market model from Siven *et al.* [3] (left) and its scrambled version (right). The graphs correspond to $\rho=+5\%$ (stars) and $\rho=-5\%$ (rings). The solid lines are fitted generalized gamma density functions.

Figure 3 shows that the index constructed from these stocks display a gain-loss asymmetry, much like the Dow Jones index, and that the asymmetry vanishes if we scramble the time series.

In what follows, we will use this artificial index as a kind of proxy for a real stock index. This has the advantage that the individual index constituents are readily available for analysis. This is unlike the Dow Jones index, for which the relative weights and, indeed, the set of constituents have changed over time.

V. DEPENDENCE BETWEEN CONSTITUENTS DURING PERIODS OF INDEX UPTURNS AND DOWNTURNS

Here, and in what follows, $\{I_t\}_{t=0,\ldots,T}$ denotes the artificial index defined in the previous section.

Inspired by the generalized asymmetric synchronous market model from Siven *et al.* [3], our general intuition is that

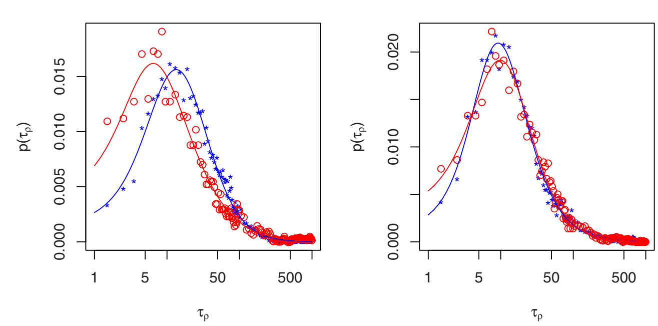

FIG. 3. (Color online) Estimated distribution of the first passage time τ_ρ for the log price of the artificial index I (left) and its scrambled version (right). The graphs correspond to $\rho=+5\%$ (stars) and $\rho=-5\%$ (rings). The solid lines are fitted generalized gamma density functions.

the individual stocks tend to "move together" to a greater degree during index downturns than during upturns, resulting in more violent downturns than upturns. To quantify this, we first divide the price history of our artificial index I into two parts, corresponding to upturns and downturns, respectively.

Fix a window length L and consider the index return over the kth window,

$$\Delta I_k = I_{kL} - I_{(k-1)L},$$

for $k=1,\ldots,\lfloor T/L \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x. We define the set of indices, for which the daily returns belong to a window over which the index went up,

$$U = \bigcup_{\{k;\Delta I_k>0\}} \{(k-1)L+1,\ldots,kL\},$$

respectively, went down,

$$D = \bigcup_{\{k;\Delta I_k<0\}} \{(k-1)L+1,\ldots,kL\}.$$

Note that the sets U and D are disjoint.

We will consider two measures of dependence between all the individual stocks and evaluate it for the returns corresponding to days $t \in U$ and compare that to the same measures evaluated for days $t \in D$. Before describing the first measure, the mean of mutual information, we establish some additional notation. Let the nth *index* be defined by

$$I_{n,t} = \frac{1}{N-1} \sum_{m \neq n} \frac{S_{m,t}}{S_{m,0}}.$$

The nth index is simply the artificial index constructed by averaging all stocks except the nth. Denote the log return at day t in the nth stock and index by $\delta S_{n,t} = \log(S_{n,t}/S_{n,t-1})$ and $\delta I_{n,t} = \log(I_{n,t}/I_{n,t-1})$.

A. Mean mutual information

The mutual information of two discrete stochastic variables X and Y is defined as

$$M(X,Y) = \sum_x \sum_y p_{XY}(x,y) \log\left(\frac{p_{XY}(x,y)}{p_X(x) p_Y(y)} \right),$$

where p_{XY} denotes the joint and p_X and p_Y as the marginal probability functions of X and Y. Mutual information can be written as $M(X,Y) = H(X) + H(Y) - H(X,Y)$, where $H(X)$ and $H(Y)$ are the marginal entropies, and $H(X,Y)$ is the joint entropy of X and Y, and it is a measure of dependence in the sense that X and Y are independent if and only if $M(X,Y) = 0$. Mutual information can be estimated from a finite set $\{(X_t, Y_t)\}_{t=1,\ldots,n}$ of joint samples of (X,Y) in a number of different ways (see Paninski [5]). In the computations below, we apply the most straightforward estimator: the so-called plug-in estimator.

Let $M_{U,n}$ and $M_{D,n}$ denote the mutual information of the returns of the nth stock and index estimated from the samples $\{(\delta S_{n,t}, \delta I_{n,t})\}_{t \in U}$, respectively, $\{(\delta S_{n,t}, \delta I_{n,t})\}_{t \in D}$. We average over n to obtain the *mean mutual information*, which can be seen as a measure of the degree of dependence between all the stocks over periods of upturns and, respectively, downturns of the index I,

$$M_U = \frac{1}{N} \sum_{n=1}^{N} M_{U,n},$$

$$M_D = \frac{1}{N} \sum_{n=1}^{N} M_{D,n}.$$

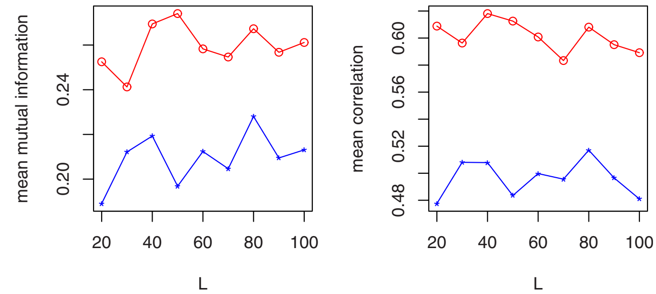

FIG. 4. (Color online) The mean mutual information (left) and the mean correlation (right) for the artificial index, as function of the window length L. The graphs show the mean mutual information and correlation corresponding to index upturns (stars), respectively, downturns (rings).

Figure 4 shows the mean mutual information for varying window length—there is clearly a higher degree of dependence between the stocks returns during index downturns. However, given the hypothesis that stocks tend to move together to a greater degree during index downturns, with the result that downturns are more dramatic than upturns, there is a potential problem with the measure: the mutual information between the nth stock and index is large whenever there is a high degree of dependence, not only when they tend to move in the same direction. If some stocks tend to move up when the index moves down, this would moderate the downturns, contrary to our intuition, and yet result in high values for the mean mutual information. For this reason, we also consider a correlation-based measure.

B. Mean correlation

Let $C_{U,n}$ and $C_{D,n}$ denote the correlation between the returns of the nth stock and index estimated from the samples $\{(\delta S_{n,t}, \delta I_{n,t})\}_{t \in U}$, respectively, $\{(\delta S_{n,t}, \delta I_{n,t})\}_{t \in D}$. We average over n to obtain the *mean correlation*, which can be seen as a measure of the degree of dependence between all the stocks over periods of upturns, respectively, downturns of the index I,

$$C_U = \frac{1}{N} \sum_{n=1}^{N} C_{U,n},$$

$$C_D = \frac{1}{N} \sum_{n=1}^{N} C_{D,n}.$$

Figure 4 shows the mean correlation for varying window length; this measure of dependence between the stock returns also shows markedly higher values during index downturns that during index upturns. Contrary to the mean mutual information, however, the presence of "defensive" stocks that move up during index downturns would give negative contributions.

VI. CONCLUSION

If the gain-loss asymmetry observed for stock indices were a property of the unconditional distribution of returns, then the phenomenon should remain invariant under random permutations of the returns—this is not the case, as we have demonstrated. We may begin to rely more confidently on expectations derived from the generalized asymmetric synchronous market model, which have previously demonstrated that differences in correlated movements in index constituents for down moves and up moves can give rise to the kind of temporal dependence structure that produces such asymmetry. However, there are practical difficulties in exploring the correlations between the time series of the individual constituents of real stock indices since these are not readily available, so we have shown that the gain-loss asymmetry can also be reproduced in an artificial stock index constructed as a simple average of a number of individual stocks.

Considering two different measures of dependence, mean mutual information and mean correlation, we concluded that there indeed is a greater degree of dependence between the constituents of the artificial index during downturns than upturns. This part of our analysis can be seen as an attempt to overcome some of the general difficulties in formulating tractable ways of analyzing nonstationary dependence structure in multivariate stochastic processes. Future work in the direction of analyzing the dynamics of the changes in the level of dependence between asset prices would certainly be interesting—not least from the perspective of investors who seek diversification that does not break down at the worst possible time. For instance, is it possible to design a localized measure of the level of dependence between stock prices and zoom in even more on the points in time where it is changing?

Future research could also investigate the connection between the gain-loss asymmetry and other stylized facts concerning the temporal structure of financial time series—particularly, the asymmetric *leverage effect* (see Bouchaud *et al.* [6]), as attempted by Alghren *et al.* [7].

[1] I. Simonsen, M. H. Jensen, and A. Johansen, Eur. Phys. J. B **27**, 583 (2002).
[2] M. H. Jensen, A. Johansen, and I. Simonsen, Int. J. Mod. Phys. B **17**, 4003 (2003).
[3] J. Siven, J. Lins, and J. Lundbek Hansen, J. Stat. Mech.: Theory Exp. 2009, P02004 (2009).
[4] R. Donangelo, M. H. Jensen, I. Simonsen, and K. Sneppen, J. Stat. Mech.; Theory Exp. 2006, L11001 (2006).
[5] L. Paninski, Neural Comput. **15**, 1191 (2003).
[6] J.-P. Bouchaud, A. Matacz, and M. Potters, Phys. Rev. Lett. **87**, 228701 (2001).
[7] P. T. H. Ahlgren, M. H. Jensen, I. Simonsen, R. Donangelo, and K. Sneppen, Physica A **383**, 1 (2007).