Discovering stock dynamics through multidimensional volatility phases

HSIEH FUSHING*†; SHU-CHUN CHEN‡ and CHII-RUEY HWANG‡

†Department of Statistics, Mathematical Sciences Building, University of California, Davis, CA 95616, USA
‡Institute of Mathematics, Academia Sinica, Taiwan

(Received 30 January 2009; in final form 25 September 2009)

To investigate stock dynamics, we consider volatility as a temporal aggregation of semi-extreme events defined on three dimensions: return, volume and trading number. Onset and offset phases of volatility are computed by means of the hierarchical factor segmentation (HFS) algorithm based on high-frequency data. Through these computed volatility phases we search for dynamic patterns by resolving two questions: Is a return’s volatility closely associated with significant price changes?, and can we derive an early prediction of the sign of the price change at the offset? Can volatility phases reveal which dimension—return, volume or trading number—is the driving force behind the others? Some computed new features of stock dynamics are counter-intuitive. Almost all significant price changes are marked by volatility within the three dimensions. We develop a data-driven potential-based model to make early predictions of the sign of significant price differences at the end of a volatile period. This model recognizes that when a stock’s dynamic enters a volatility state, it typically settles into a subtle imbalance of oscillations between positive and negative returns, and leads to a significant price difference at the offset of volatility. We develop a new statistical analysis to show that a return’s volatility onset is more likely to fall behind the onsets of volume and trading number, while the latter two dimensions are very well-correlated with each other. By incorporating this result with behavioral evidence extracted from scatterplots of the logarithm of volume versus the trading number, we postulate that stock dynamics are chiefly driven by a large group of participants, whose collective large-volume trading action is potentially responsible for stimulating volatility in both return and trading number.

Keywords: Computational finance; Dynamic models; Financial times series; Investor behaviour; Market dynamics; Pattern recognition; Stochastic models

1. Introduction

Paul Samuelson, the Nobel laureate and one of the pioneers of modern financial theory, recently gave the following statement in an article in Business Week magazine (Farrell 2008):

Timing is a tough business. It is easy to sell, but then you have to know when to get back in, ... and we know that hardly anyone is good at it.

This opinion honestly reflects that, given the tremendous advances in financial engineering and mathematics, stock dynamics are still not well-understood.

*Corresponding author. Email: fushing@wald.ucdavis.edu
stable distributions with heavy tails are also popularly considered within Lévy processes and a flights setting (Shlesinger et al. 1993, 1995). However, these modeling techniques are hardly capable of pinpointing the ‘timing’ of the arrival and departure of a volatile period, nor of predicting its up or down price trends.

From a modeling perspective, in order for a stock-dynamics model to explicitly carry ‘timing’ information, it needs to be incorporated with a certain regime-switching-type of structure. Finding such a structure underlying the stock dynamics is at least as challenging as modeling the dynamics itself. Furthermore, this joint modeling of a stock’s dynamics and its volatility regime changes encounters tremendous difficulty in both theoretical and computational development. All of these difficulties will be greatly amplified when the data trajectory is lengthy.

As information technology advances, financial researchers nowadays employ high-frequency data of stock returns, with a time resolution finer than one minute, perhaps down to a tick-by-tick temporal scale. However, such high-frequency data manifest no visual clues regarding how the global trends observable through the price trajectory can possibly be visualized. Figure 1 provides a glimpse of the disconnection between these two trajectories, even though we can reconstruct one based on the other. Since modeling of stock returns is a topic of great importance in finance, an interesting and essential pattern recognition task is: Can we compute and mark all segments on the return time series that precisely correspond to global trends in price? This task is not at all obvious, nor is it immediately possible. Even if we go through the stock return’s definition, our computations will require knowledge of when to start and when to end. Furthermore, it is not clear whether most significant stock price differences occur within volatile periods. Another aspect of the difficulty of this task is due to the magnitude of high-frequency return data. Such a data set is simply too huge to be fit realistically by any statistical model with a global structure.

Thus, without guidance from statistical models and computational methodologies, the study of stock dynamics based on the stock return time series becomes very challenging. Furthermore, any stock dynamic includes two additional time-series dimensions: its trading volume and number. The volume is conventionally viewed as the ‘momentum’ of stock dynamics (Lamoureux and Lasprapes 1994), while the trading number, which is measured as the amount of transactions in a chosen time resolution, is typically taken as a behavioral indicator (Welch 1992). It is intuitive that a better understanding of stock dynamics can be achieved by coupling these three time-series dimensions together. Many methodologies with emphasis on the effect of the volume on return have been proposed and studied by Tauchen and Pitts (1983), Lamoureux and Lasprapes (1990), Gallant et al. (1992) and a survey by Karpoff (1987). Other methodologies prescribing that the trading number and market order, not volume, have a real impact on the return have also been suggested and investigated by Jones et al. (1994), Ane and Geman (2000), Zumbach (2004) and Wyart et al. (2007). All aforementioned methodologies are model based with a focus on directional effects: from volume or trading number to return. Although these models can provide platforms for evaluating specific effects of concern, their model effects are hardly measured and could be misleading if their underlying assumptions differ from reality.

The reality we perceive here is that return, volume and trading number indeed mutually interact with each other in a timely fashion. However, no methodology is available to assist participants and researchers in the financial market to explore and compute pertinent interacting dynamic information. Such exploration and computation are especially critical nowadays since we have accumulated sufficient knowledge about the gaps between the real-world financial market and the hypothetical one embraced by the mathematical finance community a half-century ago.

In this paper we develop computational methodologies that allow us to investigate stock dynamics from the perspective of volatility phases. Here volatility is defined as a temporal aggregation of events which have either large absolute return values, large volume, or large trading number. Thus it is not exactly equivalent to the ‘volatility’ in the Black–Scholes model, even with respect

Figure 1. Time series of (a) the IBM stock price with a 30-second sampling rate in April 2005, and (b) its return.
Discovering stock dynamics

of returns. Specifically, volatility will be distinctively identified as a sequence of segments along the temporal axis of the stock dynamic, rather than as just a purely conceptual variance-like parameter.

These computed phases that mark the volatility not only reflect the missing components in the mathematical and probabilistic modeling of stock prices, but, more importantly, identify realistic driving forces underlying the stock dynamics. Specifically, we explore stock dynamics by computationally resolving the following two questions based on data of a 30-second temporal resolution.

Q1: Is a return’s volatility closely associated with significant price changes, and can we derive an early prediction of the sign of the price change at its offset?

Q2: Can volatility phases reveal which time-series dimension—return, volume or trading number—is the force that drives the others?

To resolve Q1 and Q2, we employ a non-parametric decoding algorithm called hierarchical factor segmentation (HFS) for computation of the volatility phases (Fushing et al. 2006, 2010a, b). On the stock-return dimension, for example, the HFS algorithm segments the whole time series into a composition of volatile periods of return against non-volatile periods. Here, each segment corresponding to a volatile period is identified by a pair of computed onset and offset signature-phase markers. Similar computations are also performed separately on the time series of volume and trading number based on our choice of the events of interest. The three sequences of onset markers along the three time series are our focus for the purpose of exploring stock dynamics. In particular, they are used as the foundation for resolving Q1 and Q2.

The rationale behind the use of a one-per-30-second sampling rate for resolving Q1 and Q2 is as follows. On average, there will be only a few volatility phases computed during a daily trading period. Consequently, the average duration is long enough and suitable for constructing computerized day trading. This sampling rate typically contains a small number of transactions with high probability, such as 40 or fewer for the IBM data used in this paper. That is why it is called high-frequency, but not ultra-high-frequency, data. The collection of transactions within this choice of temporal resolution makes the computed volatility phases nearly invariant with respect to tick time or transaction time. Furthermore, since they are not model based, the sensitivity to microstructure is not a concern here. The model-based realized-variance calculations are sensitive to microstructure pertaining to different choices of the temporal scale (Oomen 2006, Griffin and Oomen 2008). Moreover, a different time resolution will give rise to different volatility phases which carry information concerning different aspects of the stock dynamics. Hence they can be employed for different purposes.

We found from computations for Q1 that the marked return’s volatile periods coincided with most of the significant price changes. For those cases missed by the return’s volatility, they were most likely to be picked up by the volatilities of volume and trading number. Counter-intuitive results found here were: (1) the two distributions of stock returns pertaining to positive and negative price differences within volatile periods are almost identical and centered around zero; and (2) the difference between the volatile and non-volatile periods is not characterized by the mean, but the variation. Furthermore, a data-driven potential model is developed to predict the rise or fall of price at the end of each volatile period by employing only the beginning part of the data. The capability of this model implies that, at least from a stock-return aspect, the stock dynamics as a whole seem to always trend towards equilibrium, although it hardly remains there for long. It repetitively enters a state of volatility in a way that causes a subtle imbalance of oscillations between positive and negative returns. Such an imbalance in turn leads to a significant difference in price at the offset of volatility.

For Q2, a new pattern-recognition technique is devised for easy visualization of how well the three high-frequency time series are coupled together. By means of a Beta-mixture analysis, the onsets of stock return, trading number and volume are found to be well-correlated in the sense that one onset is likely to stimulate the occurrence of the onsets of the other two. The onset phases of volatility for volume and trading number are nearly synchronous. Also, from scatterplots of the trading number and volume, it is observed that stock dynamics chiefly reflect the collective trading behavior of major participants, while a majority of minor participants are passively stimulated. We finally postulate a speculative scenario of stock dynamics in which we depict how the speculative price of a stock is created based on the patterns of computed volatility phases and observed trading behavior.

The paper is organized as follows. In section 2 we introduce the HFS algorithm. In sections 3 and 4, computations for questions Q1 and Q2 are proposed and illustrated through real data analysis of the high-frequency price return, volume and trading number of the monthly IBM stock in year 2005. In section 5, the computed patterns are synthesized into speculative stock dynamics. In section 6, several related issues are collected and discussed. All technical details, including how to find the optimal event definition and threshold parameters for the HFS algorithm, and the geometric and Beta mixture analyses, are collected and discussed in the appendix.

2. HFS algorithm for volatility phases

Denote a discrete time series of length \( n \), such as that of stock returns, by \( X_n \). For each time point \( t, t=1, \ldots, n \), the variable \( X(t) \) is assumed to be regulated by the state variable \( S(t) \), which takes the value 0 when in a non-volatile period, and the value 1 in a volatile period. The state-space vector \( S_n \) denotes the unknown 0–1 string of length \( n \). Thus \( S_n \) represents the true partition on \( X_n \).
On each constant-value segment of $S_n$, the corresponding segment of $X_n$ isstationarily distributed according to probability measure $P_0$ or $P_1$, corresponding to states $S(t)=0$ and $S(t)=1$, respectively. That is, each continuum corresponding to a volatile segment of $X_n$ is conceptually stationary, as is each continuum corresponding to a non-volatile segment. We impose no parametric assumptions upon $P_{S(t)}$ or the generating dynamics of $S_n$. Here, the non-parametric decoding task refers to the estimation of $S_n$ based on the observed $X_n$.

Since both digital strings $S_n$ and $X_n$ have the same dimension $n$, the decoding task is not at all easy. To make this decoding task more manageable, we now construct a platform for applying the maximum entropy principle on a sequence of recurrence times of a suitably chosen event, which is usually a characteristic event of $P_0$ or $P_1$. Let $A$ be an observable event, and denote a sequence of recurrence times along $X_n$ by $R_{m(A)} = \{R_1, \ldots, R_m(A)\}$.

To simplify the above decoding problem somewhat, we parsimoniously assume that the segment of $X_n$ falling between two successive occurrences of event $A$ is generated according to a single state. Then we define this state as the state of its corresponding recurrence time. Consequently, we have a significantly lower dimension $n$ of the recurrence-time state-space vector $S_{m(A)} = \{S^R_1, \ldots, S^R_m\}$ for $R_{m(A)}$ with $m(A) \ll n$. We then estimate $S_{m(A)}^R$ as a way of decoding the original state-space vector $S_n$ using a decoding algorithm, which will be introduced in the next subsection.

### 2.1. Hierarchical factor segmentation (HFS) algorithm

The hierarchical factor segmentation (HFS) algorithm was originally developed for an animal behavior study by Fushing et al. (2006). Recently, it has been shown to be more favorable for decoding the 0–1 state-space vector under the hidden Markov Model (HMM) than the Viterbi algorithm, and provides an exhaustive search for CpG islands in genetics and a new algorithm for dissecting rhythmic cycles (Fushing et al. 2010a, b). Given a chosen event and using a three-level coding scheme, the HFS algorithm is functionally designed to locate aggregations of the event onset and offset along $X_n$. Consequently, the time series is partitioned into alternating sequences of high and low event-intensity segments.

**Hierarchical factor segmentation (HFS) algorithm:** Given a chosen observable event $A$ on the time series $X_n$, the construction of the HFS algorithm proceeds according to the following steps.

- **HFS-1.** $X_n$ is transformed into a 0–1 digital string of base 2 via the following coding scheme: code 1 for $X(t)$ when event $A$ is observed at time $t$; otherwise, code 0 for $X(t)$. This first-level 0–1 digital string is denoted by the code sequence $C^0$.

- **HFS-2.** Construct a histogram, say $H^*$, of the event recurrence-time distribution from $C^0$, and denote the sequence of recurrence times (inter-event spacing) by $R^*$.

- **HFS-3.** Choose the first threshold value as an upper $p_{\alpha}$th percentile on $H^*$, say $h$, to transform the $R^*$ sequence into a 0–1” digital string via the second-level coding scheme: (1) a recurrence time of less than $h$ is coded by 0; (2) a recurrence time greater than $h$ is coded by 1. The resulting digital string of base 2 is denoted by the code sequence $C^2$.

- **HFS-4.** Upon code sequence $C^2$, we take code word 1 as another ‘new’ event and construct its corresponding recurrence-time histogram $H^3$ and the sequence of inter-1”-event spacing as $R^3$.

- **HFS-5.** Choose the second threshold value as an upper $p_{\beta}$th percentile from $H^3$, say $h^*$, to transform $R^3$ into another 0–1” digital string via the top-level coding scheme: (1) an inter-1”-event spacing less than $h^*$ is coded by 0; (2) an inter-1”-event spacing greater than $h^*$ is coded by 1. The third digital string of base 2 is denoted by $C^3$.

- **HFS-6.** The resultant code sequence $C^3$ is mapped back onto the time series $C^0$ or $X_n$ as a partition of $|R^3|$ $(m)$ segments on the time span $[1, n]$. Denote this partition as $N(X_n) = ([N_{L1}, N_{R1}])_m^{m-1}$.

This partition or segmentation $N(X_n)$ achieves an aggregating pattern by separating the high event-intensity segments from the low event-intensity segments. A segment, say $[N_{L1}, N_{R1})$, corresponding to the 1” code, is a period of time points falling in between two widely separated 1” codes. The wide separation of two successive 1” codes implies that there are many 1 codes on the particular segment of code sequence $C^3$; equivalently, many events are observed on the segment $X_n$. Thus this is a segment of high event-intensity. In contrast, a segment, say $[N_{R1}, N_{L1})$, corresponding to the 0” code (including segments corresponding to the two 1” codes on both ends) would present an extended period of many 0 codes, but very sparse 1 codes. Thus it is a segment of low event-intensity. That is to say, if $A$ is defined as the event in which the absolute value of the stock return $|X(t)|$ exceeds a chosen upper percentile of the corresponding histogram, then segment $[N_{L1}, N_{R1})$ of the 1” code is a volatile period, while $[N_{R1}, N_{L1}+1)$ is a non-volatile period.

The computational feasibility of the HFS algorithm essentially relies on the fact that, under no parametric assumptions, the number of partitions $N(X_n)$ pertaining to all the possible integer threshold values $(h, h^*)$ is much less than $2^n$. It is this small set of candidate partitions on $X_n$ that facilitates the computation via the HFS algorithm to resolve the difficult change-points problem without prior knowledge about the number of changes and related model-selection issues, as discussed in the appendix.

### 2.2. A simulation study

In this subsection, the performance of the HFS algorithm is evaluated numerically using a computer experiment.
We generate a discrete time series $\mathcal{X}_n$ with $n = 10^4$, according to the following simulation plan. The time span $[1, 10,000]$ is partitioned into five segments by four chosen time points $(t_1, t_2, t_3, t_4)$. Within each segment, each $X(t)$ is a normal random variable distributed according to: (1) $X(t) \sim N(0, \sigma_0^2)$ when $t$ is in $[0, t_1)$, $[t_2, t_3)$ and $[t_4, 10,000]$; (2) $X(t) \sim N(0, 1)$ when $t$ is in $[t_1, t_2)$ or $t \in [t_3, t_4)$. Here $\sigma_0^2$ is chosen such that the Receiver–Operator Characteristic (ROC) curve of $N(0, \sigma_0^2)$ versus $N(0, 1)$ is very close to the empirical ROC curve based on empirical distributions shown in figure 2. Based on calculations, a reasonable choice of $\sigma_0$ is in the range $[1.9, 2.1]$. Therefore, this simulation study not only confirms the validity of the HFS algorithm, but also provides some information concerning ‘good’ choices of events for later exploration on real data.

By taking the partition $(t_1, t_2, t_3, t_4)$ of the time points as the unknown parameters, the HFS algorithm is applied to carry out a segmentation of $\mathcal{X}_n$ based on a series of events $\{A_j\}_{j=1}^{10}$ which exceed certain upper percentiles. Precisely, $A_j = \{w \geq \tilde{F}^{-1}(1 - 0.05 \times j)\}$, where $\tilde{F}(x)$ is the empirical distribution function of the absolute values of $X(t)$, that is $|X|_n = \{|X(t)| \mid t = 1, \ldots, 10,000\}$ and $\tilde{F}^{-1}(h)$ is its empirical quantile function.

Here the HFS algorithm is applied to the time series $|X|_n$ instead of the original time series $\mathcal{X}_n$. The reason for doing this is that the volatility of a stock return is defined through the absolute value of the return. After the HFS algorithm carries out the segmentation of $|X|_n$, each segment $\{N_{Li}, N_{Ri}\}$ corresponding to codeword $1^i$ is assigned the color red when the number of positive $X(t)$ values exceeds the number of negative $X(t)$ values within such a segment. The results of the HFS segmentation are presented in figure 3.

From figure 3 we observe that the HFS algorithm works well. Event $A_2$, based on using the upper 10th percentile, seems to give the best segmentation result. The underlying reason supporting this choice of event can be seen from the fact that the ROC curves in figures 2(e) and (f) are nearly horizontal (with slope less than $1/4$), that is the density ratio of the volatile versus non-volatile periods beyond the cutoff point is huge (at least four-fold). In other words, a return that qualifies as an $A_2$ event is an extreme event for the non-volatile period,
but relatively common in the volatile period. It is based on this difference that we can differentiate the volatile from the non-volatile periods. This idea is embraced in the optimal event selection proposed in appendix B. It is worth noting that, from the change-point analysis perspective, the HFS algorithm based on event recurrence time is an effective methodology that relies neither on a normal-distribution assumption, nor on prior knowledge of the number of changes.

3. The resolution of Q1

In this section we try to resolve question Q1 by analysis of real monthly high-frequency return data. The data set consists of 12 months of IBM stock from the year 2005. The sampling time resolution is 30 seconds, and the time series length $n$ is about 20,000 for each month. The HFS algorithm is applied to the stock return which is calculated as the change of price relative to its price 30 seconds previously.

3.1. A counter-intuitive result

Intuitively, the distribution of stock returns within a volatile period, say $[N_L, N_R]$, with a positive price difference (between the time points $N_L$ and $N_R$) should be located to the right of that derived from a volatile period with a negative price difference. Denote the two return distributions corresponding to a volatile period with positive and negative price changes by $F_0(\cdot)$ and $F_1(\cdot)$, as shown in figures 2(a) and (b), respectively. Based on all 12 HFS segmentation analyses of the monthly IBM stock price for the whole year 2005, we find a counter-intuitive result: the two distributions $F_0(\cdot)$ and $F_1(\cdot)$ are almost identical and both centered around zero. This result implies that the formation of price changes in volatile periods is not simply random sampling from $F_0(\cdot)$ and $F_1(\cdot)$, but requires higher-order mechanisms. One higher-order mechanism is described in the next subsection. Next, we denote the return distribution of non-volatile periods by $F_0(\cdot)$. We further find that $F_0(\cdot)$ and $F_1(\cdot)$ have significantly larger variations than $F_0(\cdot)$, but they all appear nearly symmetric around zero, as shown in figures 2(a) and (c).

To further confirm the above findings, we perform diagnostic pairwise comparisons among the three distributions via the Receiver-Operator Characteristic (ROC) curve. The ROC curve of $F_0(\cdot)$ versus $F_0(\cdot)$ is computed as a curve expressed by $1 - F_0(F_0^{-1}(1 - t))$, as shown in figure 2(d). The nearly diagonal ROC curve indicates that $F_0(\cdot)$ and $F_0(\cdot)$ are nearly identical. This implies that there is very little extractable information for predicting the sign of the price difference from the return distributions, even when derived from volatile periods. Both the ROC curve of $F_0(\cdot)$ versus $F_1(\cdot)$, $1 - F_0(F_1^{-1}(1 - t))$, and the ROC curve of $F_0(\cdot)$ versus $F_1(\cdot)$, $1 - F_0(F_0^{-1}(1 - t))$, as shown in figures 2(e) and (f), reveal a common S-shape, and cross the diagonal line segment at point $(1/2, 1/2)$. The S-shape pattern confirms that both $F_0(\cdot)$ and $F_0(\cdot)$ have much larger variation than $F_0(\cdot)$. The crossing pattern indicates that they share the same median. The same phenomena were also observed for all other stocks analysed, including AAPL, GE, GOOG, GS, JNJ, JPM, MSFT, QQQQ, SPY, WMT, XOM and MOT. We anticipate similar results for the S&P 500 and other indices. Also, our computational approach could have similar applicability to the option price, and other financial assets and foreign exchange rates.

3.2. Modeling the transition of the sign of the return

Next we turn to transition patterns of the sign of the return for possibly useful information. Assume a return $X(t)$ has three signs: $+$ ($=1$), 0 ($=0$) and $-$ ($=-1$). Specifically, we separately examine the conditional probability of each of the three signs given the nearest two previous signs in both the non-volatile and the volatile periods. We also seek information about the difference in this conditional probability between the two characteristically different periods. The heuristic idea here is the following: (1) given a combination of any two signs, say $(0, +)$, the conditional probability of either $+$ or $-$ in the non-volatile period should signal a sense of ‘equilibrium’ in its stock dynamics, such as when $Pr[+|0, +] < Pr[-|0, +]$; (2) the same conditional probability is likely to reveal some sort of ‘off-equilibrium’ in a volatile period, and by tilting one way or the other will signal positive price differences, such as when $Pr[+|0, +] > Pr[-|0, +]$, or similarly for negative price differences, such as when $Pr[+|0, -] < Pr[-|0, -]$.

We propose the following synthetic model to extract all possible ‘off-equilibrium’ indications as deviations from ‘equilibrium’. Let the sign of return $X(t)$ be denoted by $s(t)$, taking values 1, 0 or $-1$, for all $t = 1, \ldots, n$. We also classify the nine possible combinations of the nearest two previous signs $(s(t-2), s(t-1))$ into three classes: (1) $B_{-1} = \{(1, 1), (1, -1), (-1, 1), (-1, 0), (0, -1), (0, 0)\}$; (2) $B_0 = \{(0, 0)\}$, and (3) $B_1 = \{(1, 0), (0, 1), (0, 0), (1, 1)\}$. Conceptually, $s(t)$ is more likely to equal 1 (positive return) than to equal $-1$ given that $(s(t-2), s(t-1))$ is in $B_{-1}$ under an ‘equilibrium’ state (non-volatile period). In contrast, $s(t)$ is more likely to equal $-1$ (negative return) than to equal 1 given that $(s(t-2), s(t-1))$ is in $B_1$ under the same ‘equilibrium’ state.

We quantify the above heuristic comparison by proposing the potential function $q(s(t-2), s(t-1))$ to signal the strength or tendency of observing either $s(t) = 1$ or $s(t) = -1$ at time $t$ under the ‘equilibrium’ state. Specifically, $q(s(t-2), s(t-1))$ satisfies the following ordering:

- On $B_1$: $q(1, 1) < q(-1, 1) < q(0, 1) < q(1, 0) < 0,$
- On $B_{-1}$: $q(-1, -1) > q(1, -1) > q(0, -1) > q(-1, 0) > 0,$
- On $B_0$: $q(0, 0) = 0.$

A schematic diagram of this potential function is given in figure 4. The corresponding frequency of sign-transition triples observed in non-volatile periods of the IBM stock, June 2005, evidently demonstrates coherence with the
potential function. Based on this potential function, we then construct the conditional probabilistic model of the sign transition of the stock return in the 'off-equilibrium' state (volatile period).

1) When \((s(t-2), s(t-1)) \in B_1\) and \(j = 1, 0, -1\):

\[
Pr[s(t) = j \mid (s(t-2), s(t-1)), \alpha^+] = \frac{W_j(s(t-2), s(t-1))e^{\alpha^+/(s(t-2), s(t-1))}}{\sum_{h=1}^{\beta} W_h(s(t-2), s(t-1))e^{\alpha^+/(s(t-2), s(t-1))}},
\]

where \(W_j(s(t-2), s(t-1))\) is the frequency of observations of the ordered triple \((s(t-2), s(t-1), j)\) in the non-volatile period.

2) When \((s(t-2), s(t-1)) \in B_{-1}\):

\[
Pr[s(t) = j \mid (s(t-2), s(t-1)), \alpha^-] = \frac{W_j(s(t-2), s(t-1))e^{\alpha^-/(s(t-2), s(t-1))}}{\sum_{h=1}^{\beta} W_h(s(t-2), s(t-1))e^{\alpha^-/(s(t-2), s(t-1))}},
\]

3) When \((s(t-2), s(t-1)) \in B_0\):

\[
Pr[s(t) = j \mid (0, 0), \alpha^0] = \frac{W_j(0, 0)e^{\alpha^0/(0, 0)}}{\sum_{h=1}^{\beta} W_h(0, 0)e^{\alpha^0/(0, 0)}}.
\]

From here on we use the shorthand OFF-EQ-IDX for the above 'off-equilibrium indicator' model of return dynamics. We summarize the key features of the OFF-EQ-IDX model below to better detect the direction of possible 'off-equilibrium' mainly through the parameters \(\alpha^+\) and \(\alpha^-\).

A1) Positive values of \(\alpha^+\) (and \(\alpha^-\)) indicate that the system might remain in the 'equilibrium' state, or might possibly be 'off-equilibrium', but cannot be deciphered clearly from the perspective of \(\alpha^+\) (and \(\alpha^-\)).

A2) If \(\alpha^+\) is negative, then there is a strong indication that the 'off-equilibrium' phase is happening and tilting towards a positive return.

A3) A negative value of \(\alpha^-\) strongly indicates that the dynamic has its phase in 'off-equilibrium' and is tilting towards a negative return.

A4) When the system is in an 'equilibrium' state, we expect \(\alpha^0\) to be near zero; otherwise, significant non-zero values reveal information about the 'off-equilibrium' state and whether it is tilting towards a positive or negative return, depending on whether it is positive or negative.

We illustrate all these key features in the next subsection on OFF-EQ-IDX analysis and the potential for early prediction of the price trend.

3.3. OFF-EQ-IDX analysis and its use in prediction

We apply this OFF-EQ-IDX model to the time series of IBM stock from June 2005 (see figure 5). The frequency \(W_j(s(t-2), s(t-1))\) of sign-transition triples observed in non-volatile periods is shown in figure 4. Two volatile periods are analysed with the results reported in figures 6 and 7. The application of the OFF-EQ-IDX model is based on the following likelihood function of \((\alpha^+, \alpha^-, \alpha^0)\).
derived from a segment of the stock return time series \( \{X(t) \mid N_{t1} \leq t \leq N_{R_0}\} \):

\[
L(\alpha^{+}, \alpha^{-}, \alpha^0) \mid \{X(t) \mid N_{t1} \leq t \leq N_{R_0}\} = \prod_{(s(t-2), s(t-1)) \in B_t} Pr[s(t) = j \mid (s(t-2), s(t-1)), \alpha^+] \\
\times \prod_{(s(t-2), s(t-1)) \in B_{-1}} Pr[s(t) = j \mid (s(t-2), s(t-1)), \alpha^-] \\
\times \prod_{(s(t-2), s(t-1)) \in B_0} Pr[s(t) = j \mid (0, 0), \alpha^0]
\]

Maximum likelihood estimation of \((\alpha^{+}, \alpha^{-}, \alpha^0)\) in the OFF-EQ-IDX model is carried out by maximizing \(L(\alpha^{+}, \alpha^{-}, \alpha^0) \mid \{X(t) \mid N_{t1} \leq t \leq N_{R_0}\}\). The OFF-EQ-IDX analysis of the volatile period \([1281, 1932]\), with a negative price difference, is shown in figure 5(a), and the frequency of sign transition triples is summarized in figure 5(d). Figure 5(b) shows the MLE of \((\hat{\alpha}^+, \hat{\alpha}^-, \hat{\alpha}^0)\) is \((-0.08, -0.19, 0.53)\) and displays the corresponding shape of the likelihood functions \(L_1(\alpha^+)\) (red), \(L_2(\alpha^-)\) (blue), and \(L_0(\alpha^0)\) (green). The parameter \(\alpha^-\) is estimated to be negative in this period, and is much larger than \(\alpha^+\) in absolute value. From the log-likelihood function shapes, we see that the precision of the MLE \(\hat{\alpha}^-\) is much better than that of \((\hat{\alpha}^+\) and \(\hat{\alpha}^0)\). Thus this information strongly indicates a negative price difference, in agreement with observation.

In figure 5(c) we report the trajectories of successive estimates of \((\alpha^+, \alpha^-, \alpha^0)\) using data from the very beginning of the volatile period to its end. The convergence pattern reveals that the separation between \(\hat{\alpha}^+\) and \(\hat{\alpha}^-\), being negative in value, can be perceived by the 100th time point of the sequence of 300 time points. This rather rapid convergence of the estimates of \(\alpha^+\) and \(\alpha^-\) will be useful for early prediction of the price trend.

In figure 6 we report the OFF-EQ-IDX analysis of the volatile period \([10,500, 11,518]\), which has a positive price difference. The MLE \((\hat{\alpha}^+, \hat{\alpha}^-, \hat{\alpha}^0)\) is computed to be \((-0.16, 0.25, 0.39)\). The negative value of \(\hat{\alpha}^+\) together with the positive value of \(\hat{\alpha}^-\) implies a positive price difference, in agreement with observation. It is also seen that \(\hat{\alpha}^+\) converges to \(-0.16\) as early as the 100th time point in the sequence of more than 300 time points.

We emphasize that the potential for early prediction of price trend via successive MLEs under the OFF-EQ-IDX model merits further investigation. Since the computation is relatively simple, the faster computing facility will make such information valuable for traders of financial assets.
4. The resolution of Q2

In this section we explore the potential interaction relationship between the return, trading number and volume time series of IBM stock in the year 2005 with a one-per-30-second sampling rate. We particularly focus on relationships of the relative positions among onsets of the three kinds of volatile periods. Their relative position patterns may suggest how these three time series are possibly coupled together as a major manifestation of the stock dynamics. Here, all onset signature phases are computed using the HFS algorithm applied to the monthly time series, with an event defined as being above the upper 10th percentile. The rationale behind such an event choice is that this event can give rise to a very efficient segmentation, as seen in figure 3. Then we make several behavioral observations from a scatterplot of the volume and trading number, and address their implications with respect to stock dynamics. In the next section, all patterns found in this section are synthesized into a partly speculative and partly realistic scenario of stock dynamics.

We re-emphasize that different sampling rates will lead to events defined with different meanings, and consequently lead to aspects of stock dynamic patterns that can be computed from the HFS algorithm. For instance, if the sampling rate is changed into one-per-5-minutes, then the HFS algorithm would compute a volatile period as an aggregation of correspondingly defined extreme events in a temporal resolution of hours. Then their recurrence times are likely to be measured in days. This kind of pattern hardly belongs to the high-frequency category. Thus such a sequence of volatile periods reveals very different stock dynamics from that resulting from the data of a 30-second sampling rate.

4.1. Coherence via Beta-mixture analysis

In this section a new statistical analysis is proposed to study the relationship among onset phases of volatile periods of return, volume and trading number. The resulting segmentation of the three time series of IBM monthly stock are reported in figure 7 for June 2005. Figure 7(a), (b) and (c) show the segmentations of the return (marked on the price trajectory), trading number and volume. Again, a volatile segment with positive price change is marked in red, and one with negative price change is marked in blue. Figure 7(d) shows three segmentations embedded on the same temporal axis.

Figure 7. Volatile periods computed via the HFS algorithm on the return, volume and trading number of IBM stock for June 2005, with positive price changes marked in red and negative price changes in blue. (a) Marked return’s volatile period marked on stock price trajectory. (b) Volume’s volatile periods marked on volume trajectory. (c) Trading number’s volatile periods marked on trading number trajectory. (d) Volatile periods of the three dimensions marked on the common temporal axis.
perceive the potential coherence among the three time series and then measure the coherence by means of a Beta-mixture analysis.

Denote the three sequences of onset markers computed from the return, volume and trading number time series by \( \{ N_{Li}^{(R)} \}_{i=1}^{M_R} \), \( \{ N_{Li}^{(V)} \}_{i=1}^{M_V} \) and \( \{ N_{Li}^{(TN)} \}_{i=1}^{M_{TN}} \), where \( M_R, M_V \) and \( M_{TN} \) are the respective numbers of onset markers. For instance, we take \( \{ N_{Li}^{(V)} \}_{i=1}^{M_V} \) to partition the temporal axis into a process of segments \( \{ N_{Li}^{(V)} : N_{Li+1}^{(V)} \} \), \( j=1, \ldots, M_V \). Within each segment \( \{ N_{Li}^{(V)} : N_{Li+1}^{(V)} \} \), we mark all contained onset markers of return (in red) and trading number (in blue). In figure 8 we display the process of such segments with its length. Similar constructions for the two processes of segments \( \{ N_{Li}^{(R)} : N_{Li+1}^{(R)} \} \) and \( \{ N_{Li}^{(TN)} : N_{Li+1}^{(TN)} \} \) are likewise performed.

Next, for each of the processes, we then rescale the segment’s length to equal one and record the relative position of all onset markers contained within the segment with values ranging between 0 and 1. Further, by stacking all re-scaled unit intervals together, we create two empirical distributions of the relative position of each of the two marked onset markers versus the position of the onset marker used to segment the time axis. For instance, the relative positions of \( \{ N_{Li}^{(TN)} \}_{i=1}^{M_{TN}} \) and \( \{ N_{Li}^{(R)} \}_{i=1}^{M_R} \) relative to \( \{ N_{Li}^{(V)} \}_{i=1}^{M_V} \) are shown in figure 9(c) and (d). Similar constructions based on return and trading number are presented in figure 9(a, b) and (e, f), respectively.

As shown in figure 9, many histograms reveal a ‘smile’ shape. The sharper the smile, the more coherent the two time series. This smile-shaped histogram importantly indicates that the onset volume’s volatile period is either coherently followed by or preceded by an onset of the trading number’s volatile period, as shown in figure 9(f) and (i). In summary, we see that \( N_{Li}^{(R)} \) is likely to follow \( N_{Li}^{(V)} \) and \( N_{Li}^{(TN)} \), as shown in figure 8(b), (c), (e) and (h). Also, \( N_{Li}^{(V)} \) and \( N_{Li}^{(TN)} \) are very strongly coupled, as seen in figure 8(f) and (i).

To further quantify the degree of the ‘smile’, we propose a Beta-mixture analysis. See appendix D for analytical derivations and related statistical inference. The strength of coherence can be manifested through the two sets of Beta parameter values. The mixture proportion indicates whether one onset is a promoter or follower of the other. Very importantly, these Beta-mixture analyses provide useful extra information when we make predictions regarding the arrival of the return’s next volatile period through their identification criteria. That is, \( N_{Li}^{(R)} \) can be better predicted with information concerning the presence and timing of \( N_{Li}^{(V)} \) and \( N_{Li}^{(TN)} \), based on the Beta-mixture analysis pertaining to the second and third columns of the panels in figure 8, respectively. Therefore, we can be certain that the ‘timing’ issue mentioned at the beginning of section 1 is indeed resolved to a certain extent.

### 4.2. Observed behavior

In this subsection we point out several potential trading behaviors that are related to stock dynamics. The scatterplot of IBM stock trading number versus logarithm of volume (base 10), as shown in figure 10, reveals a common pattern across different months of 2005. In fact, such a pattern is nearly universal across stocks other than IBM. It is observed that \( \log_{10} V \) is well-bounded from below by a very smooth curve:

\[
\log_{10} V = d + c \log_{10} TN
\]

We found that, typically, \( c \approx 1.5 \) and \( d \approx 1.5 \) in the 12 months of 2005 for IBM. This power form of the lower bound, \( V > 10^d TN^c \) with \( c > 1 \), strongly indicates that a significant part of the volume traded in every 30-second interval is primarily contributed by major participants in the financial market. The reasons supporting this speculation are as follows. Let a major participant at a 30-second time resolution be tentatively defined as one who trades more than 1000 shares in one transaction. In 2005, 1000 shares of IBM stock were worth more than one-tenth of a million (\$10^5) dollars. Thus a major participant is likely to be a trader in hedge funds or security companies, in contrast to a minor participant such as an individual investor. The majority of the trading (above 90%) involves volume size ranging from \( 10^3 \) to \( 10^6 \). Hence we can draw a straight line tangent to the lower bound curve with intercept equal to 3 such that the majority of volume remains above this straight line. This means that, on average, volume increases much faster than the slope of this line. By putting the above facts together we can tentatively conclude that the driving forces of stock dynamics should primarily come from the group of major participants, rather than the mass of minor participants.

Furthermore, the scatterplot of \( \log_{10} V \) and \( TN \) in figure 10 is partitioned into six areas according to one threshold on \( \log_{10} V : 5 \), and two threshold values on \( TN : 50 \) and 100. We randomly selected one 30-second data point each from area 1, area 2, area 3 and area 5. They are respectively representative of data in the area from which...
they are selected. By zooming in on these four chosen data points, we construct the corresponding typical histograms of $\log_{10} V$ of all transactions within the 30-second interval. These are shown in figure 11. We see that there exists at least one trading tick involving a huge volume ($10^5$) in area 1 through area 3. The contrast between the higher trading number panel in figure 11(a) and the lower trading number panel in figure 11(c) seems to be in line with speculation concerning herding behavior in the financial market. Furthermore, we also confirm that almost all 30-second intervals with volume larger than $10^4$ and trading numbers larger than 30 are located within a return’s volatile period computed using the HFS algorithm.

Remark 1: In a turbulent era, such as during October and November 2008, the lower-bound function found in the scatterplot of IBM stock’s volume and trading number is $V = 10^{1.83(TN)^{1.08}}$, which becomes much closer to being linear. The maximum trading numbers for this stock in these two months go up as high as 800 per 30 seconds, which is as large as seven times the average maximum trading number during 2005.

5. Speculative stock dynamics

After resolving Q1 and Q2 to some extent in the two previous sections, we synthesize all our findings into partly real and partly speculative stock dynamics as an informative way of summarizing the dynamics results.

The persistent presence of major participants in 30-second high-frequency data implies that a group of major participants or traders constantly and collectively use large leverages (large volume) for carrying out actions and counter-actions of ‘long’ or ‘short’ trading. This collective large-volume trading will be accompanied by intensive large trading numbers, since there exists strong

Figure 9. Histograms of pairwise relative positions of onset markers with superimposed graphic results of the Beta-mixture analysis. (a) $N_{Lj}^{(V)}$ versus $N_{Li}^{(R)}$. (b) $N_{Lk}^{(TN)}$ versus $N_{Li}^{(R)}$. (c) $N_{Lj}^{(R)}$ versus $N_{Lj}^{(V)}$. (d) $N_{Lk}^{(TN)}$ versus $N_{Lj}^{(V)}$. (e) $N_{Lj}^{(R)}$ versus $N_{Lk}^{(TN)}$. (f) $N_{Lj}^{(V)}$ versus $N_{Lk}^{(TN)}$.

Figure 10. Scatterplot of trading number versus logarithm of volume with the lower bound curve superimposed with six marked areas defined by one threshold on the logarithm of volume: 5, and two threshold values on trading number: 50 and 100.
coherence between volume and trading number, as shown in figure 8. When entering volatile periods of volume and trading number, also from figure 8, it is likely that the onset of the return’s volatility will follow. The onset of the return’s volatility has a feedback effect to stimulate more intensive trading with large or small volume, and makes the trading number become larger. Consequently, active trading actions with large volume and trading number keep the momentum of volatile returns going until the stock settles into a significant price change at the offset of volatility. Underlying the offset is usually a typical barrier set by an unseen extremely high asking price or extremely lower bidding price on a huge volume. Then a non-volatile period comes in and lasts for a random length of time until a new onset of the volume’s volatility is created to start a new cycle of stock dynamics.

To a great extent the above rhythmic kind of stock dynamics is reflected in figure 7. These dynamics somehow provide a mechanism for explaining why the histogram of stock returns is too pointed to be bell-shaped (signifying normality). Furthermore, it is not unthinkable that due to different major participants’ differences in perception and sources of information concerning the market and portfolios at hand, a large return with opposite sign can be instantaneously stimulated as a counter-action. This might be the underlying mechanism for the nearly symmetric pattern in the return time series as shown in figure 1. Also, the feedback mechanisms in the financial market, either through trading psychology or strategy or other aspects of stock dynamics, might exist in a certain mysterious manner that makes the intrinsic patterns of stock dynamics very difficult to observe and compute.

The above speculative stock dynamics are certainly not sufficient to fully or quantitatively explain many empirical characteristics of real stock dynamics. However, our computational clues might well shed a very different light from that which pure mathematical and statistical modeling might have on certain empirical facts, such as the too-pointed stock return histogram and its heavy tail, and many others (Mandelbrot 1963). As a matter of fact, our results reported here serve the purpose of pointing out that the endeavors of modeling stock dynamics without taking these behavioral features and their implied trading mechanisms into account are likely to be very unrealistic.

6. Discussion

In this paper, we investigate stock dynamics from the volatility-phases perspective, which is one direction seldom embraced in modern financial theory. No popular statistical finance modeling techniques reported in the literature view the stock dynamics from this realistic angle. As clearly depicted in the scenario discussed in the last section, by neglecting to include major participants’ trading behaviors and potential feedback psychological effects, geometric Brownian motion and Lévy processes and flights might be merely modeling the superficial facets of the dynamics. In contrast, our volatility-phase considerations could possibly open up a window for more programmatic and realistic knowledge of the
financial market. Nonetheless, phase patterns, such as those we have computed here, and many others that are waiting to be discovered, will provide us with sufficiently sophisticated mechanisms that will facilitate a better in-depth understanding of stock dynamics.

The resolutions we have proposed and developed for questions Q1 and Q2 might be very valuable for practitioners and researchers in the financial market. In particular, the OFF-EQ-IDX model and its analysis, and the interacting relationships among a stock’s three dimensions, return, volume and trading number, should allow major and minor participants in the financial market to compute pertinent information concerning ongoing stock dynamics before making any trading decision. One distinct feature of our stock dynamic scenario is that all participants are strongly recommended to think about the potential behavioral effects of their trading actions before they actually take them.

In this paper we do not address the difficult issue of how to make predictions about price difference at the end of a volatile period. Even with more research effort, incorporating more detailed information into the OFF-EQ-IDX model, we may still be a long way from resolving this issue satisfactorily, if it is indeed possible. There are at least three reasons why we take this pessimistic stance at this stage. First, since there is no model assumed to govern, even stochastically, the trajectory of stock prices in a manner like that of the Black–Scholes model, there would be no platform for computational approaches to work out the prediction. Second, the magnitude of the volume of any stock actively involved in the market at any temporal resolution is chiefly determined by a time-varying group of major participants, and their potential perceptual differences regarding a stock’s dynamics are not yet well studied. Hence the potential volume could be highly unpredictable. Third, since world-wide financial markets are so tightly connected nowadays, information about potential influences could have a tremendously wide range of impact that is beyond most people’s imagination. It is likely that a seemingly unrelated news item could cause a significant effect on a stock’s dynamics through a hidden network that we have not yet understood, or simply have never thought of, before it is revealed.

It is also necessary to note that computational capability plays an essential role in computing and extracting information about stock dynamics from the financial market. A timely computed pattern is crucial for carrying out successful trading. Therefore, the simplicity of the Beta-mixture analysis of phase-coherence for predicting the timing of the volatility, and the OFF-EQ-IDX model for prediction of the sign of price difference at the end of volatility, could be very valuable in a real-world financial market.

References


Appendix A: Geometric mixture analysis

We employ the maximum entropy principle as the foundation for estimation of $S_{m(A)}^R$. This realistic principle ensures that a recurrence-time sequence of an observable event on a stationary time series, or any exchangeable process, is approximately i.i.d. geometrically distributed (Fushing et al. 2010a, b). Specifically, we have the following working conditions.
D1: (Conditional independence assumption). \{R_1, \ldots, R_m\} are conditionally independent given \(S_m^R\).

D2: (Geometric distribution assumption). \(P(T_j | S_j^R = s) = G(T_j; \lambda_s), s = 0, 1\), where the intensity parameters \(\lambda_s, s = 0, 1\), are unknown.

In this appendix we extract auxiliary information about the dynamics under study via geometric mixture analysis. In appendices B and C, we separately discuss how to choose an optimal event \(A^*\) and the threshold parameter values needed in the HFS algorithm, as will be developed in the next subsection, by adapting to the computed auxiliary information. The above two conditions D1 and D2 also facilitate a likelihood function based on \(R_m^\infty\). This likelihood function can also serve as another basis for threshold parameter selection. For expository simplicity, we combine both procedures for threshold parameter selection into one. However, it is worth noting that the more we rely on conditions D1 and D2, the more caution we need to exercise in our explorations of real stock dynamics. More reliable patterns should result from fewer imposed parametric conditions, knowing that they may not match well with the dynamics under study.

With the state-space model structure underlying \(R_m^\infty\) and conditions D1 and D2, we derive the mixture analysis for this collection of recurrence times. From this mixture analysis we extract the information concerning the mixture proportion \(\pi_s\) and the component geometric distribution \(G(T_j; \lambda_s)\), where \(s = 0, 1\). This information will be used to fine-tune our decoding algorithm on \(S_m^R = \{S_1^R, \ldots, S_m^R\}\) in the latter section.

The mixture likelihood of \(\varphi = (\pi_0, \lambda_1, \lambda_0)\) is computed as follows:

\[
L(\varphi | R_m^\infty, S_m^R) = \left\{ \prod \pi_1 f(R_i | S_i^R = 1) \right\} \left\{ \pi_0 f(R_i | S_i^R = 0) \right\}^{1 - S_i^R} \\
= \left\{ \pi_1 (1 - e^{-\lambda_1}) \right\} \sum S_i^R e^{-\lambda_1} \sum (R_i - 1) S_i^R \\
\times \left\{ \pi_0 (1 - e^{-\lambda_0}) \right\} \sum S_i^R e^{-\lambda_0} \sum (R_i - 1) (1 - S_i^R).
\]

Let \(M_1 = \sum S_i^R = m - M_0\). Then the log-likelihood of \(\varphi\) is

\[
\log L(\varphi | R_m^\infty, S_m^R) = M_1 \log \pi_1 (1 - e^{-\lambda_1}) - \lambda_1 \sum (R_i - 1) S_i^R \\
+ M_0 \log \pi_0 (1 - e^{-\lambda_0}) - \lambda_0 \sum (R_i - 1) (1 - S_i^R).
\]

To find the MLE for \(\varphi\), we apply the EM algorithm with \(R_m^\infty\) as the missing data. The expected log-likelihood in the E-step at the \(j\)th iteration is computed as

\[
\gamma^{(j)}(\varphi | R_m^\infty) = E_{\psi^{(j-1)}}[M_1 | R_m^\infty]\log[\pi_1 (1 - e^{-\lambda_1})] \\
- \lambda_1 \sum E_{\psi^{(j-1)}}[S_i^R | R_m^\infty](R_i - 1) \\
+ E_{\psi^{(j-1)}}[M_0 | R_m^\infty]\log[\pi_0 (1 - e^{-\lambda_0})] \\
- \lambda_0 \sum E_{\psi^{(j-1)}}[1 - S_i^R | R_m^\infty](R_i - 1),
\]

where the so-called responsibility of state \(s = 1\) for the \(i\)th data point is

\[
\gamma^{(j)}_i = \frac{\pi_1^{(j-1)} e^{-\lambda_1^{(j-1)}(R_i - 1)}}{\sum_{s=0,1} \pi_s^{(j-1)} e^{-\lambda_s^{(j-1)}(R_i - 1)}}.
\]

Then the M-step at the \(j\)th iteration is computed as follows:

- the estimate of the mixture proportion is
  \[
  \pi_1^{(j)} = \sum \gamma^{(j)}_i / m;
  \]
- and the estimates of the geometric distribution parameters are
  \[
  \lambda_s^{(j)} = \log \frac{\mu_s^{(j)} / \mu_s^{(j-1)}}{1 - \mu_s^{(j-1)}}, \quad \mu_0^{(j)} = \frac{1}{m} \sum \gamma^{(j)}_i R_i, \\
  \mu_1^{(j)} = \frac{1}{m} \sum 1 - \gamma^{(j)}_i, R_i.
  \]

Let \(\hat{\pi}_1^{(M)}, \hat{\lambda}_1^{(M)}\), and \(\hat{\lambda}_0^{(M)}\) denote the MLE estimates of the mixture proportion, geometric intensity and mean recurrence time parameters. We can compute the sample Fisher information matrix of \(\psi = (\pi_1, \lambda_1, \lambda_0)^T\) using the missing information matrix proposed by Louis (1982). Denote this sample Fisher information by \(I_0\). Some computational issues related to \(I_0\) are collected and discussed in appendix D in the setting of Beta-mixture analysis.

Appendix B: Finding the optimal event

Upon the choice of event \(A\), we further denote the mixture parameter vector as \(\psi(A) = (\pi_1(A), \lambda_1(A), \lambda_0(A))\)' and its MLE derived from the geometric mixture analysis as \(\hat{\psi}(A) = (\hat{\pi}_1(A), \hat{\lambda}_1(A), \hat{\lambda}_0(A))\)' Below we give a proposal for choosing an optimal event \(A^*\).

Let the marginal mixture density be

\[
f(R | \psi(A)) = \pi_1(A) f(R | \lambda_1(A)) + \pi_0(A) f(R | \lambda_0(A)),
\]

where \(f(R | \lambda(A)), s = 0, 1\), are the two component geometric densities. Heuristically we would like to find an event that can make the marginal density \(f(R | \psi(A))\) as distant as possible from the two component densities \(f(R | \lambda_1(A)), s = 0, 1\). Under the geometric setting,
this heuristic can be expressed well using the following re-scaled difference between \( \hat{\lambda}_1(A) \) and \( \hat{\lambda}_2(A) \): Take the normal asymptotic result \( \tilde{\psi}(A) \sim N(\psi(A), I_0^{-1}(A)) \). Then

\[
D(A | \tilde{\psi}(A)) = \frac{\hat{\lambda}_1(A) - \hat{\lambda}_0(A)}{\tilde{\sigma}_0(A)},
\]

\[
\tilde{\sigma}_0^2(A) = (0, 1, -1)^T I_0^{-1}(A) (0, 1, -1)^T.
\]

Here the re-scaling factor \( \tilde{\sigma}_0(A) \) is the sample standard deviation of \( \hat{\lambda}_1(A) - \hat{\lambda}_0(A) \).

Therefore, we compare two events \( A_1 \) and \( A_2 \) using the values of \( D(A_1 | \tilde{\psi}(A_1)) \) and \( D(A_2 | \tilde{\psi}(A_2)) \). By this comparison we can realistically find the data-driven optimal event \( A'' \) within a pre-selected collection of potential events, denoted by \( \mathcal{A} \):

\[
A'' = \arg \max_{A \in \mathcal{A}} D(A | \tilde{\psi}(A)).
\]

Here such an event \( A \) has to be identified by researchers in accordance with the real-world dynamics under study. Typically this selection of \( A \) will require well-grounded subject-matter knowledge and understanding of the dynamics of interest.

**Appendix C: Finding the optimal threshold parameters via the adaptive model selection criterion**

Computationally, the HFS algorithm transforms the segmentation on \( X_n \) into an event-intensity change-point analysis problem. Very importantly, we assume here no prior knowledge of the number of changes throughout this time series. One key step in the HFS algorithm is the selection of the optimal threshold values \((h, h^*)\). Model selection techniques will be employed for this purpose. However, we have no prior knowledge of a suitable penalty size when adding a parameter. The penalty in the Schwarz information criterion (BIC) can be too heavy, whereas in Akaike’s information criterion (AIC) may be too light. Thus we propose to choose the optimal threshold values by selecting the resultant segmentation that best adapts to the geometric mixture information derived in appendix A.

We develop the following adaptive model selection technique for choosing the threshold values \((h_0, h_\ast)\). Recall that the output of the HFS algorithm is an estimate of the the state-space sequence \( S_n \). From the point of view of pattern recognition, this estimate is a segmentation \( N(X_n) \) of \( X_n \). It can also be viewed as a solution to the event-\( A \)-intensity change-point analysis problem that makes no assumptions about the number of changes. Under the maximum entropy principle, the recurrence time distribution in the segmentation \( N(X_n) \) is again geometric. The recurrence time within all \( \{N_{L_1}, N_{R_0}\} \) segments of \( 1^\circ \) is \( G(\cdot : \lambda_1) \)-distributed, while the recurrence time in the remaining \( \{N_{R_1}, N_{L_\ast + 1}\} \) is distributed \( G(\cdot : \lambda_0) \). Therefore, any point of \( \{(N_{L_1}) \) that separates \( 0^\circ \) and \( 1^\circ \) segments can be defined as an intensity change-point. Denote the number of change-points by \( \omega(h_0, h_\ast) \). Let \( \{R_{i_1}^{m_1}\}_{i_1=1}^{m_1} \) and \( \{R_0^{m_0}\}_{m_0=1}^{m_0} \) denote the observed recurrence times within identified volatile and non-volatile segments, respectively, where \( m_0 + m_1 = m \) is the length of \( R_m \) in step HFS-2 of the algorithm.

Next, we compute the likelihood function of \((\lambda_0, \lambda_1)\) based on the observed recurrence times within the two classes of segments:

\[
L(\lambda_0, \lambda_1; X_n) = P_{m_0}^{m_0}(\lambda_0) \prod e^{-\lambda_0(R_0-1)} P_{m_1}^{m_1}(\lambda_1) \prod e^{-\lambda_1(R_1-1)},
\]

where \( P_{m_0}(\lambda_0) = (1 - e^{-\lambda_0}) \), \( s = 1, 0 \). Note that this estimate of \((\lambda_0, \lambda_1)\) is independent of the estimate in the mixture geometric analysis in section 3.1. With \( L(\lambda_0, \lambda_1; X_n) \), the maximum likelihood estimates of \((\lambda_0, \lambda_1)\) are derived as follows:

\[
\hat{\lambda}_0 = \log \left\{ \frac{\tilde{c}_0}{\tilde{c}_0 - 1} \right\},
\]

\[
\hat{\lambda}_1 = \log \left\{ \frac{\tilde{c}_1}{\tilde{c}_1 - 1} \right\},
\]

where \( \tilde{c}_1 = \sum R_{j_1}/m_1 \) and \( \tilde{c}_0 = \sum R_0/m_0 \) are the average recurrence times on core and non-core segments, respectively.

In the null case when no segmentation is imposed on \( X_n \), the maximum entropy principle implies that the the recurrence time distribution is again geometric. Denote this geometric distribution by \( G(\cdot : \lambda) \) with maximum likelihood estimator \( G(\cdot : \lambda) = L(\lambda | \sigma) = P_{m}^{(\lambda)} e^{-\lambda(n-m)} \) with \( \lambda = \log [\tilde{c}/(\tilde{c} - 1)] \), \( \tilde{c} = n/m \), and \( m = m_0 + m_1 \). Since the number of events on \( X_n \) is a fixed constant, the estimator \( L(\lambda | \sigma) \) is the maximum entropy distribution.

Consider the null hypothesis that there is no aggregation of the chosen event in \( X_n \). A reasonable and effective test statistic is the log-likelihood ratio,

\[
\Delta^0(h_0, h_\ast) = \log L(\lambda_0, \lambda_1 | \sigma) - \log L(\lambda | \sigma)
\]

\[
= \sum_{k=0}^{l_0 + l_1} \left[ m_k \log \left\{ \frac{1}{c_k - 1} \right\} - l_k \log \left\{ \tilde{c}_k \right\} \right]
\]

\[
- m \log \left\{ \frac{1}{\tilde{c} - 1} \right\} - n \log \left\{ \tilde{c} \right\},
\]

where \( l_0 + l_1 = n \), \( l_0 = \sum k_0 \) and \( l_1 = \sum k_1 \) are the total lengths of the two kinds of segments.

To make use of \( \Delta^0(h_0, h_\ast) \) for selecting an optimal threshold parameter \((h_0, h_\ast)\) in the HFS algorithm, we perform the following optimization:

\[
\Gamma(h_0, h_\ast) = \Delta^0(h_0, h_\ast) - \frac{k}{2} (\omega(h_0, h_\ast) + 1),
\]

\[
(h_0^{(k)}; h_\ast^{(k)}) = \arg \max_{(h_0, h_\ast)} \Gamma(h_0, h_\ast),
\]

where \( k \) is the penalty for adding one extra parameter. For instance, \( k = 1 \) is the AIC model selection criterion, while \( k = \log m \) corresponds to the BIC criterion. Empirically, we know that \( k = 1 \) would be likely to give rise to too many cores (1\(^\circ\)), while \( k = \log m \) tends to produce too few cores. Hence, we propose to choose the \( k \) value for which the HFS segmentation based on the threshold values \((h_0^{(k)}, h_\ast^{(k)})\) provides a pair of intensity
estimates, say \((\hat{z}_1, \hat{z}_0)\), that is closest to \((\hat{z}_1, \hat{z}_0)\) derived in the geometric mixture analysis. Specifically, we choose the integer \(k\) minimizing the following sum of signal-to-noise (S/N) ratios:

\[
\text{arg min}_{1<k<\log n} \left[ \frac{\hat{z}_1^{(\text{M})} - \hat{z}_1^{(k)}}{|\hat{z}_1^{(M)} - \hat{z}_1^{(k)}|} + \frac{\hat{z}_0^{(\text{M})} - \hat{z}_0^{(k)}}{|\hat{z}_0^{(M)} - \hat{z}_0^{(k)}|} \right].
\]

Denote the corresponding optimal threshold parameters by \((\hat{h}^{(k)}_0, \hat{h}^{(k)}_1)\) and let the segmentation \(N(X_n \mid \hat{h}^{(k)}_0, \hat{h}^{(k)}_1)\) refer to the output of the HFS algorithm performed using these thresholds.

### Appendix D: Beta-mixture analysis

Let the Beta density function be denoted as follows:

\[
f(x \mid \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1},
\]

where

\[
B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} \, dx.
\]

Let \(\theta = (\alpha_0, \theta_0, \alpha_0)\) and \(\theta_1 = (\alpha_1, \theta_1)\). A set of observable random variables from a Beta-mixture setting is denoted by \(X_{n,k} = [X_1, \ldots, X_n]\), where each \(X_i\) is coupled with an unobservable mixture status random variable \(\delta_i\) taking values 0 or 1 to indicate from which population this random variable is generated. The complete data vector is denoted as \(\Delta = \{\delta_1, \ldots, \delta_n\}\). Here, \(Pr[\delta_i = 1] = \pi\) is the mixture proportion of population 1.

The complete likelihood contributed from data pair \((\Delta, X_n)\) and the marginal likelihood of \(X_n\) and their corresponding log-likelihood versions, are calculated respectively as

\[
L^*(\theta \mid X_n, \delta_i) = [\pi f(x \mid \alpha_1, \beta_1)]^{\delta_i} [1-\pi] f(x \mid \alpha_0, \beta_0)]^{1-\delta_i},
\]

\[
L(\theta \mid X_n, \delta_i) = \delta_i \log [\pi f(x \mid \alpha_1, \beta_1)] + (1-\delta_i) \log [1-\pi] f(x \mid \alpha_0, \beta_0),
\]

\[
L(\theta \mid X_n) = \pi f(x \mid \alpha_1, \beta_1) + (1-\pi) f(x \mid \alpha_0, \beta_0),
\]

\[
l(\theta \mid X_n) = \pi \log [\pi f(x \mid \alpha_1, \beta_1)] + (1-\pi) \log [1-\pi] f(x \mid \alpha_0, \beta_0)].
\]

We then denote the corresponding likelihood functions of \(\theta \in R^5\) given data \(X_n\) by

\[
L^*(\theta \mid X_n) = \prod_{i=1}^{n} L^*(X_i \mid \theta),
\]

\[
L(\theta \mid X_n) = \prod_{i=1}^{n} L(X_i \mid \theta),
\]

\[
l(\theta \mid X_n) = \prod_{i=1}^{n} l(X_i \mid \theta).
\]

We introduce some notation needed throughout this section. Denote two \(5 \times 1\) vectors \(V_j(x, \theta) = (v_{j1}(x, \theta),\ldots, v_{j5}(x, \theta))\), where \(j = 0, 1\), such that

\[
v_{j1}(x, \theta) = 1/\pi, \quad v_{j2}(x, \theta) = \log x, \quad v_{j3}(x, \theta) = -E_0[\log X],
\]

\[
v_{j4}(x, \theta) = \log(1-x), \quad v_{j5}(x, \theta) = 0.
\]

The partial derivative of a real-valued function \(g(\theta)\) with respect to \(\theta\) is the 5 \(\times\) 1 vector

\[
\frac{\partial g(\theta)}{\partial \theta} = \begin{pmatrix}
\frac{\partial g(\theta)}{\partial \pi} & \frac{\partial g(\theta)}{\partial \alpha_1} & \frac{\partial g(\theta)}{\partial \beta_1} & \frac{\partial g(\theta)}{\partial \alpha_0} & \frac{\partial g(\theta)}{\partial \beta_0}
\end{pmatrix}.
\]

Thus the system of score equations of the complete likelihood function of \(\theta\) is computed to be

\[
\begin{align*}
\frac{\partial \log l^{c}(\theta)}{\partial \theta} & = \frac{\partial}{\partial \theta} \sum_{i=1}^{n} \log[p(X_i \mid \alpha_1, \beta_1)]w_{j1}(\theta) \\
& = \sum_{j=1}^{5} V_j(X_i, \theta) \delta_i + V_0(X_i, \theta)(1-\delta_i) \\
& = 0.
\end{align*}
\]

Let \(\hat{\theta} = (\hat{\pi}, \hat{\alpha}_0, \hat{\beta}_0)\) denote one version of the estimate of \(\theta\). We compute the responsibility of \(X_i\) to populations 1 and 0 as follows:

\[
w_{j1}^{c}(\theta) = Pr[\delta_i = 1 \mid X_i] = \frac{\pi f(x \mid \alpha_1, \beta_1)}{L(X_i \mid \theta)},
\]

\[
w_{j2}^{c}(\theta) = Pr[\delta_i = 0 \mid X_i] = \frac{(1-\pi) f(x \mid \alpha_0, \beta_0)}{L(X_i \mid \theta)}.
\]

The conditional expectation of the complete log-likelihood function of \(\theta\) evaluated at \(\hat{\theta}\) is calculated as

\[
E_{\hat{\theta}}[\log l^{c}(\theta \mid X_n, \Delta)] = \sum_{i=1}^{n} \log[p(X_i \mid \alpha_1, \beta_1)]w_{j1}^{c}(\theta)
\]

\[
+ \log[(1-\pi) f(x \mid \alpha_0, \beta_0)]w_{j2}^{c}(\theta).
\]

The corresponding conditional expectation of \(\Delta(\theta)\) with respect to \(\Delta\), given \(X_n\), as a system of score equations of \(\hat{\theta}\) evaluated at \(\hat{\theta}\), is computed as

\[
\begin{align*}
\hat{s}_{\alpha}(\Delta) &= E_{\hat{\theta}}[\frac{\partial}{\partial X} \log l^{c}(\theta \mid X_n, \Delta)] \\
& = \sum_{i=1}^{n} V_j(X_i, \theta) w_{j1}^{c}(\theta) + V_0(X_i, \theta) w_{j2}^{c}(\theta) \\
& = (s_{\alpha}(\pi \mid \theta), s_{\alpha}(\alpha_1 \mid \theta), s_{\alpha}(\beta_1 \mid \theta), s_{\alpha}(\alpha_0 \mid \theta), s_{\alpha}(\beta_0 \mid \theta))'.
\end{align*}
\]

Specifically, we have

\[
s_{\alpha}(\pi \mid \theta) = \frac{\partial}{\partial \pi} E_{\hat{\theta}}[\log l^{c}(\theta \mid X_n, \Delta)]
\]

\[
= \frac{1}{\pi} \sum_{i=1}^{n} w_{j1}^{c}(\theta) + \frac{1}{1-\pi} \sum_{i=1}^{n} w_{j2}^{c}(\theta),
\]

\[
s_{\alpha}(\alpha_j \mid \theta) = \frac{\partial}{\partial \alpha_j} E_{\hat{\theta}}[\log l^{c}(\theta \mid X_n, \Delta)]
\]

\[
= \sum_{i=1}^{n} \log(X_i - E_0[\log X]) w_{j1}^{c}(\theta),
\]

\[
s_{\alpha}(\beta_j \mid \theta) = \frac{\partial}{\partial \beta_j} E_{\hat{\theta}}[\log l^{c}(\theta \mid X_n, \Delta)]
\]

\[
= \sum_{i=1}^{n} \log(1-X_i) - E_0[\log(1-X)] w_{j2}^{c}(\theta).
\]
With the above system of equations, the EM algorithm can be implemented using the following iterations.

- **(E-step):** Evaluate the $ith$ step responsibilities $(w_i^1(\hat{\theta}))(n)_{i=1}^n$.
- **(M-step):** Solve the system of equations $\bar{\gamma}_{m,n}(\theta) = 0$ for $\theta^{(k+1)}$.

In particular, the M-step is equivalent to the following procedure.

1. The parameter $\hat{\theta}_0^{(k+1)}$ is computed using
   \[
   \hat{\theta}_0^{(k+1)} = \frac{\sum_{i} w_i^1(\hat{\theta}^{(k)})}{\sum_{i} w_i^1(\hat{\theta}^{(k)})}.
   \]

2. The parameter $\hat{\theta}_1^{(k+1)} = (\hat{\alpha}_1^{(k+1)}, \hat{\beta}_1^{(k+1)})$ is computed by finding $\theta_1$ such that
   \[
   E_{0n}[\log X] = \frac{\sum_{i} \log X_i w_i^1(\hat{\theta}^{(k)})}{\sum_{i} w_i^1(\hat{\theta}^{(k)})},
   \]
   \[
   E_{0n}[\log(1 - X)] = \frac{\sum_{i} \log(1 - X_i) w_i^1(\hat{\theta}^{(k)})}{\sum_{i} w_i^1(\hat{\theta}^{(k)})}.
   \]

3. Likewise the parameter $\hat{\theta}_0^{(k+1)} = (\hat{\alpha}_0^{(k+1)}, \hat{\beta}_0^{(k+1)})$ is computed by finding $\theta_0$ such that
   \[
   E_{0n}[\log X] = \frac{\sum_{i} \log X_i w_i^0(\hat{\theta}^{(k)})}{\sum_{i} w_i^0(\hat{\theta}^{(k)})},
   \]
   \[
   E_{0n}[\log(1 - X)] = \frac{\sum_{i} \log(1 - X_i) w_i^0(\hat{\theta}^{(k)})}{\sum_{i} w_i^0(\hat{\theta}^{(k)})}.
   \]

Furthermore, we take the second partial derivatives to derive the sample Fisher information matrix
\[
i_{0,0} = (-1) \frac{\partial^2}{\partial \theta^2} \sum_{i} I(\theta | X_i)
\]
\[
= (-1) \frac{\partial}{\partial \theta} \sum_{i} V_i(X_i, \theta)w_i^1(\theta) + V_0(X_i, \theta)w_i^0(\theta)
\]
\[
= E[\theta, \theta | X_n] - \sum_{i} \left[ \left( V_i(X_i, \theta) \cdot \left\{ \frac{\partial}{\partial \theta} w_i^1(\theta) \right\} \right) + \left[ V_0(X_i, \theta) \cdot \left\{ \frac{\partial}{\partial \theta} w_i^0(\theta) \right\} \right] \right]
\]
\[
= E[\theta, \theta | X_n] - \sum_{i} \left[ w_i^1(\theta)V_i(X_i, \theta) \right] - \left[ V_0(X_i, \theta) \cdot \left\{ V_i(X_i, \theta) - V_0(X_i, \theta) \cdot w_i^0(\theta) \right\} \right]
\]
where the 5 × 5 block diagonal symmetric matrix $E[\theta, \theta | X_n]$ is computed as
\[
E[\theta, \theta | X_n]_{11} = \sum_{i} \left[ \frac{1}{\pi(1 - \pi)} \left( \frac{w_i^1(\theta)}{1 - \pi} + \frac{1 - w_i^0(\theta)}{\pi} \right) \right],
\]
\[
E[\theta, \theta | X_n]_{22} = \sum_{i} \left[ \text{Var}_i[\log X] w_i^1(\theta) \right],
\]
\[
E[\theta, \theta | X_n]_{33} = \sum_{i} \left[ \text{Cov}_i[\log X, \log(1 - X)] w_i^1(\theta) \right],
\]
\[
E[\theta, \theta | X_n]_{44} = \sum_{i} \left[ \text{Var}_i[\log X] w_i^0(\theta) \right],
\]
\[
E[\theta, \theta | X_n]_{55} = \sum_{i} \left[ \text{Cov}_i[\log X, \log(1 - X)] w_i^0(\theta) \right].
\]

For instance, the second row of the matrix $i_{0,0}$ is calculated as follows:
\[
i_{0,0} = -\frac{\partial}{\partial \theta} s_0(\theta | \theta)
\]
\[
= -\frac{1}{\pi(1 - \pi)} \sum_{i} \left[ \log X_i - E_{00}[\log X] \right] w_i^1(\theta) w_i^0(\theta),
\]
\[
i_{0,1} = -\frac{\partial}{\partial \theta} s_1(\theta | \theta)
\]
\[
= \sum_{i} \left[ \text{Var}_i[\log X] w_i^1(\theta) \right] - \sum_{i} \left[ \log X_i - E_{00}[\log X] \right] w_i^1(\theta) w_i^0(\theta),
\]
\[
i_{0,2} = -\frac{\partial}{\partial \theta} s_2(\theta | \theta)
\]
\[
= \sum_{i} \left[ \text{Cov}_i[\log X, \log(1 - X)] w_i^1(\theta) \right] - \sum_{i} \left[ \log X_i - E_{00}[\log X] \right] w_i^1(\theta) w_i^0(\theta),
\]
\[
i_{0,3} = -\frac{\partial}{\partial \theta} s_3(\theta | \theta)
\]
\[
= \sum_{i} \left[ \text{Var}_i[\log X] w_i^0(\theta) \right] - \sum_{i} \left[ \log X_i - E_{00}[\log X] \right] w_i^1(\theta) w_i^0(\theta),
\]
\[
i_{0,4} = -\frac{\partial}{\partial \theta} s_4(\theta | \theta)
\]
\[
= \sum_{i} \left[ \text{Cov}_i[\log X, \log(1 - X)] w_i^0(\theta) \right] - \sum_{i} \left[ \log X_i - E_{00}[\log X] \right] w_i^1(\theta) w_i^0(\theta).
\]
It should be noted from the above derivation that the sample Fisher information matrix $i_{\theta, \phi}$ is different from the conditional expectation of the complete sample Fisher information matrix $E[\delta_{\theta, \phi}|X_n]$ by the matrix

$$H(\theta | X_n) = \sum_{i=1}^{n} (V_{1}(X_i, \theta) - V_{0}(X_i, \theta)) \cdot (V_{1}(X_i, \theta) - V_{0}(X_i, \theta))^{\prime} \cdot Var[\delta_i | X_i]$$

Here $H(\theta | X_n)$ is equal to $Var[\tilde{z}^{\theta}(\theta) | X_n]$, the conditional variance of the complete score $\tilde{z}^{\theta}(\theta)$. That is,

$$Var[\tilde{z}^{\theta}(\theta) | X_n] = Var\left[ \sum_{i=1}^{n} V_{1}(X_i, \theta) \delta_i + V_{0}(X_i, \theta) (1 - \delta_i) | X_n \right]$$

$$= \sum_{i=1}^{n} (V_{1}(X_i, \theta) - V_{0}(X_i, \theta)) \cdot (V_{1}(X_i, \theta) - V_{0}(X_i, \theta))^{\prime} \cdot Var[\delta_i | X_i]$$

$$= \sum_{i=1}^{n} (V_{1}(X_i, \theta) - V_{0}(X_i, \theta)) \cdot (V_{1}(X_i, \theta) - V_{0}(X_i, \theta))^{\prime} \cdot w_{1}^{\prime}(\theta)w_{1}(\theta)$$

$$= H(\theta | X_n).$$

Thus the above equality provides an alternative proof of Louis’ formula (Louis 1982) from the marginal likelihood perspective.

There are two computational issues that arise when applying the EM algorithm for statistical inference. (1) Is there a natural way to evaluate $H(\theta | X_n)$? (2) In the M-step, should we use $E[\tilde{r}_{\theta, \phi} | X_n]$ or $E[\tilde{r}_{\theta, \phi} | X_n] - H(\theta | X_n) = i_{\theta, \phi}$?

In the EM algorithm, if the Newton–Raphson method is used in the M-step, the Hessian matrix is $E[\tilde{r}_{\theta, \phi} | X_n]$, not $i_{\theta, \phi}$. Ideally, the computations for maximizing the marginal likelihood function $\sum_{i=1}^{n} L(\theta | X_i)$ will be numerically more stable and computationally more economical when using $i_{\theta, \phi}$ rather than $E[\tilde{r}_{\theta, \phi} | X_n]$.

An intuitive idea is as follows. Since $E[\tilde{r}_{\theta, \phi} | X_n]$ can be significantly ‘larger’ than $i_{\theta, \phi}$, the adjustment at each iteration step using $E[\tilde{r}_{\theta, \phi} | X_n]$ in the M-step of the EM algorithm can be unrealistically too small, especially when the likelihood function cannot be evaluated precisely. This could be one the chief causes for non-convergence of the EM algorithm. In contrast, for maximizing the marginal likelihood function, each iteration step using $i_{\theta, \phi}$ is much larger, so that convergence will be slower but steadier.