

# Final Report

Lo-Bin Chang

November 6, 2013

During the two years of the proposal, we published two finance papers, one probability paper and one machine learning paper, and we finished a draft about vision research. The following are the details of these papers:

**(1) An Invariance for the Large-sample Empirical Distribution of Waiting Time between Successive Extremes**, Lo-Bin Chang, Alok Goswami, Fushing Hsieh, Chii-Ruey Hwang, *Bulletin of the Institute of Mathematics Academia Sinica* (2013) Vol. 8, No. 1, 31-48.

The investigation was motivated by an empirical invariance found in our study of the raw data consisting of the actual trade price of the intraday transactions data (trades and quotes) of companies in S&P500 list from 1998 to 2007 and part of 2008 (see [3]).

The temporal evolution of the price of a stock is given by a stochastic process  $\{S_t\}$ . For example the Black-Scholes model assumes that  $\{S_t\}$  follows a geometric Brownian motion with constant volatility and satisfies the stochastic differential equation

$$dS_t = rS_t dt + \sigma S_t dW_t,$$

where  $r$  and  $\sigma$  are positive constants and  $\{W_t\}$  is standard Brownian motion. An immediate consequence of this model is that the 'returns' over disjoint time intervals of equal duration are independent and identically distributed normal random variables. More specifically, if  $S_1, S_2, \dots, S_{n+1}$  are the prices of a stock at a set of  $n + 1$  equally spaced time points, then the random variables  $(S_2 - S_1)/S_1, (S_3 - S_2)/S_2, \dots, (S_{n+1} - S_n)/S_n$  are i.i.d. normal random variables.

To allow for possible jumps for stock prices, a more general model is considered where the rate of return is assumed to follow a general Lévy process, resulting in the 'geometric Lévy process' model for evolution of stock prices (see [8] for example). However, the

returns over disjoint time intervals of equal duration still remain i.i.d. random variables (though not necessarily normal).

Suppose now that we have available a large set of data on prices of a stock at equal intervals of time and calculate the returns. From this set, say of size  $n$ , of return values, suppose we generate a sequence of 0s and 1s that identifies those among the set of return values which constitute the upper (or lower) ten percent among this set. More specifically, a return value is dubbed as 0 if it falls in the upper (or lower) ten percentile of the set of return values and is dubbed as 1 otherwise. This will lead to a  $n$ -long sequence of 0s and 1s with (roughly)  $0.1n$  many 0s and remaining (roughly)  $0.9n$  many 1s. If we now look at the lengths of the successive runs of 1s in this sequence, including the (possibly empty) runs before the first and after the last 0s, then how are they expected to behave? We prove that if for each  $n$ , the successive returns have an exchangeable joint distribution, as under the Black-Scholes model or geometric Lévy process model, then, in the limit, these run lengths follow a common geometric distribution. Moreover, the run-lengths are asymptotically independent. More specifically for any fixed  $k$ , the lengths of first  $k$  runs of 1s are asymptotically distributed as  $k$  i.i.d. geometric random variables (Theorem 2.1).

Next, we consider the probability histogram generated by the lengths of all the runs of 1s. This will be a histogram of a probability on non-negative integers. We prove a strong law of large numbers (Theorem 2.2) which implies that, under the same hypothesis as in the earlier theorem, this (sample) histogram will converge, with probability one, to that of a geometric distribution. The convergence holds, with probability one, also in total variation norm, in Kolmogorov norm and uniformly (Corollary 2.1). Using results from Chen ([4]), we also obtain an associated central limit theorem (Theorem 2.3) for the (sample) probability histogram. It may be worthwhile to point out in this connection, that by applying a result of Chen ([4], Theorem 2), one may obtain a limit theorem for our probability histogram. However, such a limit theorem is substantially weaker than our Theorem 2.2, in the sense that Chen's result gives convergence only in probability (which is easy in our set-up) and assumes additional conditions.

Another point needs mention here. Under our hypothesis that the returns (over successive equal lengths of times) are exchangeable, it is obvious that the lengths successive runs of 1s, as described above, will also have an exchangeable joint distribution. However, they are not independent, even if the successive returns are assumed to be so. Our Theorem 2.2 shows that the finite sample empirical distribution converges almost surely to an

appropriate geometric distribution, as if it is the empirical distribution based on a finite i.i.d. sample from the geometric distribution. While our Theorem 2.1 shows that the successive run lengths are indeed asymptotically i.i.d geometric, the convergence of the empirical distribution to geometric distribution is not automatic by any means. Further, the de-Finetti decomposition of exchangeable probabilities on infinite product spaces as mixtures of i.i.d. probabilities does not apply here, because, for each  $n$ , we only have a finite-dimensional exchangeable distribution. A natural option would be to try and apply ideas of Diaconis-Freedman for finite exchangeable probabilities (see [7]), but our attempts along this line did not prove fruitful for the strong law.

Although we are not going to present simulation results in this paper, remarks might be made according to simulations.

We prove a strong law of large numbers with the limit being an invariance, a geometric distribution. The associated central limit theorem (Theorem 2.3) easily leads to a Donsker-type theorem (Theorem 2.4) for the empirical distribution function on  $D[0, \infty)$ . This, of course, will characterize the limiting distribution of the corresponding Kolmogorov-Smirnov statistic. As expected, this limiting distribution will not be the usual Kolmogorov distribution, as was already evidenced by our simulations.

As mentioned in the first paragraph, our raw data consisted of the actual trade price of the intraday transactions data (trades and quotes) of companies in S&P500 list from 1998 to 2007 and part of 2008. The return process was analyzed at five-minute, one-minute, and 30-second intervals for a whole year (see [3]). We repeated the procedure mentioned in previous paragraphs and considered the empirical waiting time distributions of hitting a certain percentile. Closeness of distributions was measured by ROC area and Kolmogorov distance. The invariance of these empirical distributions were found across stocks, time units, and years. This empirical invariance turned out to be different from the geometric distribution (see [3],[10]) .

On the other hand, simulated data points generated from i.i.d. normal and i.i.d. uniform, regarded as returns, were converted into a sequence of 0s and 1s as above. The histograms generated from the lengths of 1-runs were matched against the histogram of a geometric distribution with the parameter corresponding to the width of the percentile. In all cases, there seemed to be perfect match. Of course, this is only to be expected now, in view of our Theorem 2.2. We would like to point out, however, that originally it was this observed invariance in our simulation results that inspired us to formulate and prove

the main theoretical results described in this work.

Our theoretical results when contrasted with our above-mentioned analysis based on real data on prices of various stocks would raise questions on the validity of a large class of widely accepted models on stock-prices. In fact, any model that merely implies exchangeability of successive returns would be put to question. In recent times, models that incorporate long range dependence, have been proposed; these models use fractional Brownian motion instead of Brownian motion (see [11]). Since exchangeability of successive returns no longer hold for these models (except in two special cases), our invariance results do not apply. However, simulations indicate that a law of large numbers holds but with a different limit.

One other point that we like to note here is that, in the above, we have only described the first layer of coding in what is known as hierarchical segmentation of time-series data (see [10]). The data were coded into a binary 0 – 1 sequence identifying times of occurrences of “rare events” (namely, occurrence of extreme values). One can now think of the lengths of successive 1-runs (waiting times between occurrence of rare events) as the data, which is only exchangeable but not i.i.d., and code these further into a “second layer” of 0 – 1 sequence in exactly the same way as before. An occurrence of a 0 at this level corresponds to times of very long waiting between rare events at the earlier level. A long 1-run at this level corresponds to a long wait between two long waits for extreme events at the earlier level. By assuming only exchangeability (instead of i.i.d.) of the underlying random variables, our results would imply that at every level of coding, the same limiting result for the histograms generated by the lengths of 1-runs is valid.

For possible applications to finance, one may refer to ([3], [9], [10]). For example, volatile period may be defined hierarchically up to the third level using the real data: if the length of runs of 1s falls in say upper ten percentile, denote that period  $0^*$ , otherwise  $1^*$ ; repeat the same procedure for the length of runs of  $1^*$ s and denote the period in the upper ten percentile  $0^\circ$ . One may regard  $0^\circ$ s as the volatile periods. Using this segmentation, the dynamics and trading strategy are studied in ([10]).

## References

- [1] ANDREWS, D. W. K. (1988) Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory* **4**, 458–467.
- [2] BILLINGSLEY, P. (1968) *Convergence of Probability Measures*, John Wiley & Sons.

- [3] CHANG, LO-BIN, SHU-CHUN CHEN, FUSHING HSIEH, CHII-RUEY HWANG, MAX PALMER (2009) Empirical invariance in stock market and related problems, draft.
- [4] CHEN, WEN-CHEN (1981) Limit theorems for general size distributions. *Journal of Applied Probability* **18**, 139–147.
- [5] DAVIDSON, J. (1994) An  $L_1$ -convergence theorem for heterogeneous mixingale arrays with trending moments. *Statistics and Probability Letters* **16**, 301–306.
- [6] DE JONG, R. M. (1996) A strong law of large numbers for triangular mixingale arrays. *Statistics and Probability Letter* **27**, 1–9.
- [7] DIACONIS, P. AND D. FREEDMAN (1980) Finite exchangeable sequences. *The Annals of Probability* **8**, 745–764.
- [8] EBERLEIN, E. (2001) *Application of generalized hyperbolic Lévy motions to finance. Lévy Processes: Theory and Applications* [Ed. Barndorff-Nielsen O. E., Mikosch T. and Resnick S. I.], Birkhäuser, 319–336.
- [9] GEMAN, STUART AND LO-BIN CHANG (2009) Stock prices and the peculiar statistics of rare events, preprint.
- [10] HSIEH, FUSHING, SHU-CHUN CHEN, CHII-RUEY HWANG (2009) Discovering stock dynamics through multidimensional volatility-phase, accepted by *Quantitative Finance*.
- [11] HU, Y., B. OKSENDAL, A. SULEM (2000) Optimal portfolio in a fractional Black & Scholes Market. *Mathematical Physics and Stochastic Analysis: Essays in Honor of Ludwig Streit*, World Scientific, 267–279.
- [12] POLLARD, DAVID (1984) *Convergence of Stochastic Processes*, Springer Series in Statistics, xiv, Springer-Verlag
- [13] TEICHER, H. (1985) Almost certain convergence in double arrays. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **69**, 331–345.

**(2) Asymptotic Error Bounds for Kernel-based Nystrom Low-Rank Approximation Matrices**, Lo-Bin Chang, Zhidong Bai, Su-Yun Huang, Chii-Ruey Hwang, *Journal of Multivariate Analysis* (2013), Vol. 120, 102-119.

Due to the fast advancement of information technology, kernel-based learning algorithms have become popular nowadays and play an important role in machine learning with ample applications in statistics, biostatistics, medical science, image analysis, pattern recognition, engineering, etc. (See, e.g., Cortes and Vapnik, 1995; Cristianini and Shawe-Taylor, 2000; Schölkopf and Smola, 2002; Alpaydm, 2004; Hastie, Tibshirani and Friedman, 2009). Kernel functions are flexible building blocks for modeling complex and nonlinear data structures. The value of a kernel function  $K(x, y)$  represents a dot product in a kernel-induced Hilbert space, often high-dimensional or even infinite dimensional, and can be interpreted as similarity measure between the two points,  $x$  and  $y$ .

Many kernel-based learning algorithms have computational load scaled with the sample size  $n$  of data collection  $\{X_1, \dots, X_n\}$ . This article considers the Nyström low-rank approximation to kernel-based Gram matrix and provides it a theoretical justification. Asymptotic error bounds on eigenvalues and eigenvectors are derived for the Nyström low-rank approximation matrix. The lack of intuition for the approximation of eigenvectors creates a great surprise on numerical results of asymptotic error. However, to our best knowledge, there is no other existing article mathematically exploring the convergence or error bound of eigenvectors for Nyström approximation matrix. The Nyström method is an easy and yet efficient approach for low-rank approximation, which dramatically cuts down the computational load and memory usage. See, for instance, Lee and Mangasarian (2001), Williams and Seeger (2001), Drineas and Mahoney (2005), Lee and Huang (2007) for studies of Nyström low-rank approximation for kernel matrices.

The underlying kernel function  $K(x, y)$  in this article is assumed continuous, symmetric, non-negative definite, and defined on  $\mathcal{X} \times \mathcal{X}$ . Let  $X$  be a random variable having continuous distribution  $F$  on  $\mathcal{X} \subset \mathbb{R}^p$ . Let  $\mathbf{X}_n$  be the data matrix (random) consisting of i.i.d. copies of  $X$ , i.e.,  $\mathbf{X}_n := (X_1, \dots, X_n)^T$ , which is of size  $n \times p$ ; and let  $\mathbf{K}_n := K(\mathbf{X}_n, \mathbf{X}_n) = [K(X_i, X_j)]_{i,j=1}^n$  be the full kernel matrix. The key idea of Nyström approximation is to employ a reduced kernel. It randomly selects a portion of data set to generate a thin rectangular kernel matrix, called reduced kernel and denoted by  $\widehat{\mathbf{K}}_n := K(\mathbf{X}_n, \widehat{\mathbf{X}}_n)$ , where  $\widehat{\mathbf{X}}_n$  is a data subset matrix formed by a subset of  $\{X_1, \dots, X_n\}$ . Then, it uses this much smaller rectangular kernel matrix to replace or to generate an approximation to the full kernel matrix. See Figure 1 for graphical illustration.

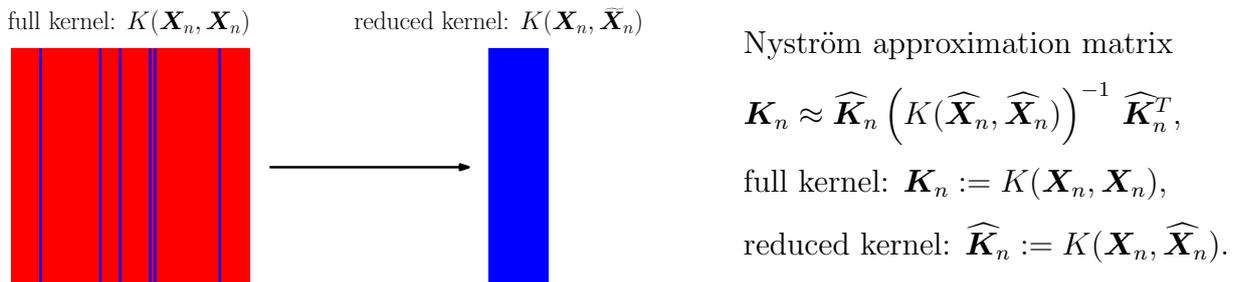


Figure 1: Reduced kernel and Nyström approximation

The technique of using a reduced kernel matrix has been successfully applied to other kernel-based learning algorithms, such as least squares support vector machine (Suykens and Vandewalle, 1999a, 1999b), proximal support vector machine (Fung and Mangasarian,

2001), Lagrangian support vector machine (Mangasarian and Musicant, 2001), active set support vector regression (Musicant and Feinberg, 2004), smooth  $\epsilon$ -support vector regression (Lee, Hsieh and Huang, 2005), kernel sliced-inverse regression (Yeh, Huang and Lee, 2009), robust kernel PCA (Huang, Yeh and Eguchi, 2009), among others.

The random subsample  $\{K(\cdot, X_{i_k})\}_{k=1}^m$  is used as a basis subset to replace the full-sample basis set  $\{K(\cdot, X_i)\}_{i=1}^n$ . In a training phase of a kernel algorithm, the thin reduced kernel matrix  $\mathbf{K}_n = K(\mathbf{X}_n, \widehat{\mathbf{X}}_n)$  is used as data inputs, where  $\widehat{\mathbf{X}}_n$  consists of  $\{X_{i_k}\}_{k=1}^m$ . Notice that the number of observations (the column size of  $\mathbf{K}_n$ ) is not reduced, it is the number of basis functions (the row size of  $\mathbf{K}_n$ ) that has been cut down. This uniform random subset for kernel basis selection has a link to the popular uniform design, which is a space filling design. Space filling designs are known to be robust against the worst possible scenario (Fang et al., 2000). Of course, there is always a random luck issue in every random sampling scheme. To improve the quality of the random subsample used as partial kernel basis, a stratified random subset is suggested. For classification problem, the random sampling has to be stratified over classes. For regression problem, the random sampling has to be stratified over the regression responses. Furthermore, the low-rank approximation matrix actually adopts a model with less model complexity, thus a larger penalty is suggested to enforce better data fidelity. See Lee and Huang (2007) for more detailed discussion and suggestions for practical implementation.

The idea of using random subset can also be found in series of works of CUR matrix decompositions (Drineas, Kannan and Mahoney, 2006a, 2006b, 2006c; Mahoney and Drineas, 2009, and references therein). See Figure 2 for illustration of a CUR decomposition.

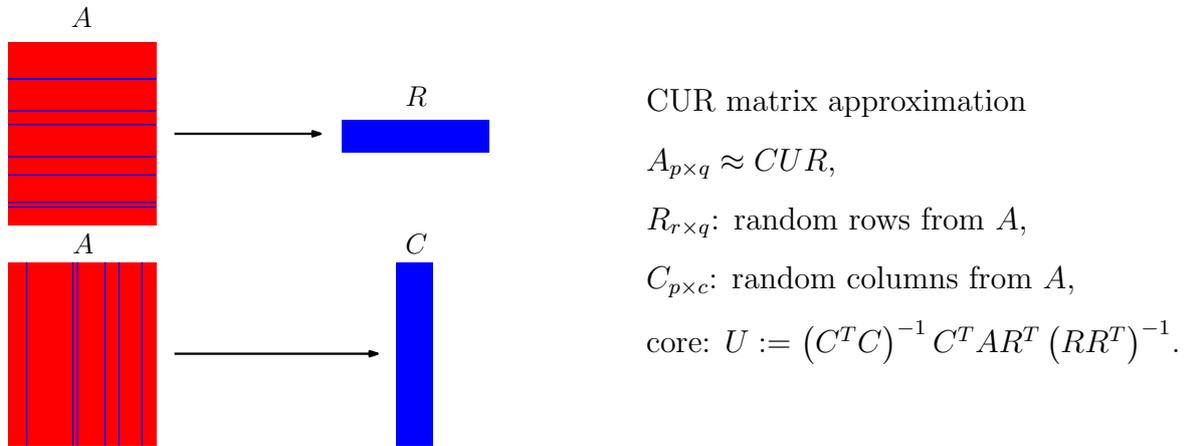


Figure 2: CUR decomposition

In addition to being continuous, symmetric and non-negative definite, the kernel function  $K$  is assumed square-integrable

$$\int_{\mathcal{X} \times \mathcal{X}} K^2(x, y) dF(x) dF(y) = c < \infty, \quad (1)$$

has the following spectrum decomposition

$$K(x, y) = \sum_{k=1}^{\infty} \lambda_k \eta_k(x) \eta_k(y), \quad \text{where} \quad \int_{\mathcal{X}} \eta_k(x) \eta_j(x) dF(x) = \delta_{kj}, \quad (2)$$

and is of trace type, i.e.,

$$\sum_{k=1}^{\infty} \lambda_k < \infty. \quad (3)$$

For simplicity, we assume eigenvalues of  $K$  are strictly positive, distinct, and arranged in descending order

$$\lambda_1 > \lambda_2 > \lambda_3 > \cdots > \lambda_k > \cdots > 0. \quad (4)$$

(Notice that the method is designed to work for symmetric non-negative definite kernels, not for others (see the second example in Section 4). The number of positive eigenvalues of the kernel needs to be greater than  $n$  for  $n \times n$  data kernel matrices defined below. If the eigenvalues are not distinct, the corresponding eigenspaces of non-distinct eigenvalues are of more than one dimension. Thus, the convergence result for eigenvalues still holds, but the convergence for eigenvectors need to be modified in terms of eigenspaces.) Consider the  $n \times n$  kernel data matrix (scaled by  $n^{-1}$ )

$$\mathbf{M}_n := n^{-1} K(\mathbf{X}_n, \mathbf{X}_n) = n^{-1} [K(X_i, X_j)]_{i,j=1}^n = n^{-1} \mathbf{K}_n \quad (5)$$

with eigenvalues  $\lambda_{n1} \geq \lambda_{n2} \geq \lambda_{n3} \geq \cdots \geq \lambda_{nn} \geq 0$  and corresponding unit eigenvectors  $\mathbf{u}_{nk}$ ,  $k = 1, 2, \dots, n$ . Matrices  $\mathbf{K}_n$  and  $\mathbf{M}_n$  are both called a full kernel matrix. The eigenvalue decomposition problem for a full kernel matrix  $\mathbf{M}_n$  is computationally costly. An alternative is to resort to a reduced kernel by random subset. Since data are i.i.d. copies of  $X$ , without loss of generality, we may assume that the random subset, denoted by  $\widehat{\mathbf{X}}_n^{(m)}$ , is formed by  $\{X_1, \dots, X_m\}$  for some  $m < n$ . Consider the partition of  $\mathbf{M}_n$  as

$$\mathbf{M}_n = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix} \begin{matrix} m \\ n - m \end{matrix}. \quad (6)$$

The following rank- $m$  approximation matrix is called Nyström approximation to  $\mathbf{M}_n$

$$\widehat{\mathbf{M}}_n^{(m)} := \begin{bmatrix} \mathbf{M}_{11} \\ \mathbf{M}_{21} \end{bmatrix} \mathbf{M}_{11}^{-1} [\mathbf{M}_{11} \quad \mathbf{M}_{12}] = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} \mathbf{M}_{11}^{-1} & \mathbf{M}_{21} \mathbf{M}_{11}^{-1} \mathbf{M}_{12} \end{bmatrix}. \quad (7)$$

This approximation is based on a reduced kernel  $\frac{1}{n}K(\mathbf{X}_n, \widehat{\mathbf{X}}_n^{(m)}) = [\mathbf{M}_{11}^T, \mathbf{M}_{21}^T]^T$ , which is of size  $n \times m$ . Denote the  $k^{\text{th}}$  eigenvalue and its associated eigenvector of  $\widehat{\mathbf{M}}_n^{(m)}$  by  $\widehat{\lambda}_{nk}^{(m)}$  and  $\widehat{\mathbf{u}}_{nk}^{(m)}$ . The aims of this article are

- (i) to study the asymptotic orders of magnitude for the full kernel  $\{\lambda_{nk}, \mathbf{u}_{nk}\}$  and the reduced kernel  $\{\widehat{\lambda}_{nk}^{(m)}, \widehat{\mathbf{u}}_{nk}^{(m)}\}$ , as compared with the ideal ones  $\{\lambda_k, \frac{1}{\sqrt{n}}\eta_k(\mathbf{X}_n)\}$ , where  $\eta_k(\mathbf{X}_n)$  is an  $n$ -vector given by  $\eta_k(\mathbf{X}_n) := (\eta_k(X_1), \dots, \eta_k(X_n))^T$ ; and
- (ii) to find asymptotic bounds for  $\mathbf{M}_n - \widehat{\mathbf{M}}_n^{(m)}$  in terms of their eigenvalues and eigenvectors, where  $\mathbf{M}_n - \widehat{\mathbf{M}}_n^{(m)}$  is the difference between the full kernel matrix and the reduced-rank approximation matrix.

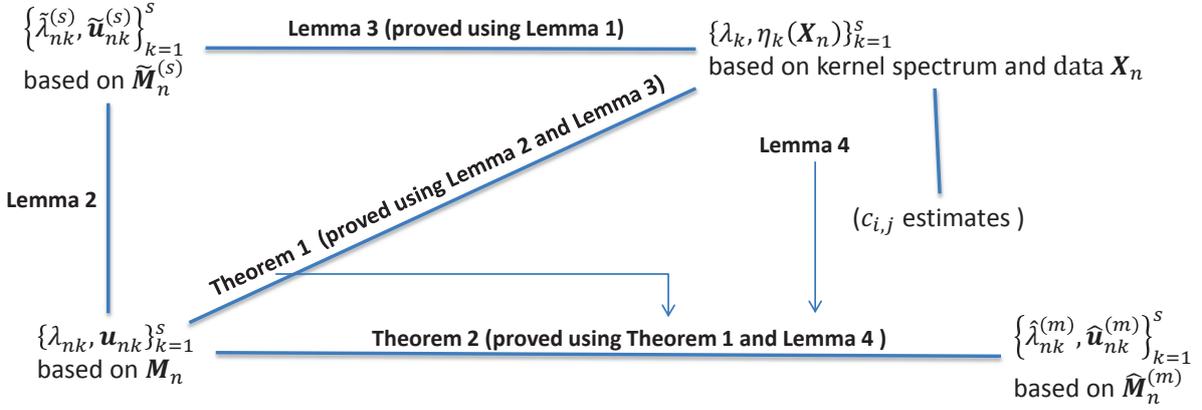


Figure 3: Connection diagram. The correspondences among variables, lemmas and theorems can be viewed in this diagram. The major goal is to connect  $\{\lambda_{nk}, \mathbf{u}_{nk}\}$  (with full kernel matrix) to  $\{\widehat{\lambda}_{nk}^{(m)}, \widehat{\mathbf{u}}_{nk}^{(m)}\}$  (with reduced-rank approximation matrix).

Results established in this article are summarized in Figure 3. The  $\widehat{\lambda}_{nk}^{(s)}$ ,  $\widehat{\mathbf{u}}_{nk}^{(s)}$  and  $\widehat{\mathbf{M}}_n^{(s)}$  in the diagram are intermediate variables defined later in the article to prove our main results. Notice that two opposite directions can be chosen for an eigenvector. For convenience, the directions of the eigenvectors,  $\mathbf{u}_{nk}$ ,  $\widehat{\mathbf{u}}_{nk}^{(m)}$ ,  $\widehat{\mathbf{u}}_{nk}^{(s)}$  are chosen toward the direction of  $\eta_k(\mathbf{X}_n)$  for each  $k$ . The rest of the article is organized as follows. Main results of eigenvalues and eigenvectors error bounds for Nyström approximation matrix are given in Section 2. Technical lemmas and proofs are placed in Section 3. Two numerical examples are displayed in Section 4. A list of notation usage is appended at the end of the article.

## References

- Alpaydm, E. (2004). *Introduction to Machine Learning*. MIT Press, Cambridge, MA.
- Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley.
- Bai, Z.D. (1999). Methodologies in spectral analysis of large dimensional random matrices, a review. *Statistica Sinica*, 9: 611-677.
- Cortes, C. and Vapnik, V.N. (1995). Support-vector networks, *Machine Learning*, 20: 273-297.
- Cristianini, N. and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines*. Cambridge University Press, Cambridge.
- Drineas, P., Kannan, R. and Mahoney, M.W. (2006a). Fast Monte Carlo algorithms for matrices I: approximating matrix multiplication. *SIAM J. Computing*, 36: 132-157.
- Drineas, P., Kannan, R. and Mahoney, M.W. (2006b). Fast Monte Carlo algorithms for matrices II: computing a low-rank approximation to a matrix. *SIAM J. Computing*, 36: 158-183.
- Drineas, P., Kannan, R. and Mahoney, M.W. (2006c). Fast Monte Carlo algorithms for matrices III: computing a compressed approximate matrix decomposition. *SIAM J. Computing*, 36: 184-206.
- Drineas, P. and Mahoney, M.W. (2005). On the Nyström method for approximating a Gram matrix for improved kernel-based learning. *J. Machine Learning Research*, 6: 2153-2175.
- Fang, K.T., Lin, D.K.J., Winker, P. and Zhang, Y. (2000). Uniform design: theory and application. *Technometrics*, 42: 237-248.
- Fung, G. and Mangasarian, O.L. (2001). Proximal support vector machine classifiers. In *Proceedings KDD-2001: Knowledge Discovery and Data Mining*, F. Provost and R. Srikant, Eds., San Francisco, CA, Aug. 26V29, 2001, pp. 77-86.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second Edition. Springer, New York.

- Huang, S.Y., Yeh, Y.R. and Eguchi, S.(2009). Robust kernel principal component analysis. *Neural Computation*, 21: 3179-3213.
- Lee, Y.J., Hsieh, W.F. and Huang, C.M. (2005).  $\epsilon$ -SSVR: A smooth support vector machine for  $\epsilon$ -insensitive regression. *IEEE Transactions on Knowledge and Data Engineering.*, 17: 678-685.
- Lee, Y.J. and Huang, S.Y. (2007). Reduced support vector machines: a statistical theory. *IEEE Trans. Neural Networks*, 18: 1-13.
- Lee, Y.J. and Mangasarian, O.L. (2001). RSVM: Reduced support vector machines. In *Proceedings of the First SIAM International Conference on Data Mining*. Philadelphia: SIAM.
- Mahoney, M.W. and Drineas, P. (2009). CUR matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences, USA*, 106(3): 697-702.
- Mangasarian, O.L. and Musicant, D.R. (2001). Lagrangian support vector machines, *J. Machine Learning Research*, 1: 161-177.
- Musicant, D.R. and Feinberg, (2004). A. Active set support vector regression, *IEEE Trans. Neural Networks*, 15: 268-275.
- Schölkopf, B. and Smola, A. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA.
- Suykens, J.A.K. and Vandewalle, J. (1999a). Least squares support vector machine classifiers. *Neural Process. Lett.*, 9: 293-300.
- Suykens, J.A.K. and Vandewalle, J. (1999b). Multiclass least squares support vector machines. In *Proc. IJCNN*, Washington, DC, 1999, 900-903.
- Williams, C. and Seeger, M. (2001). Using the Nyström method to speed up kernel machines. In *Advances in Neural Information Processing Systems*, 13: 682-688, T.K. Leen, T.G. Dietterich, and V. Tresp, Eds. MIT Press, Cambridge, MA.
- Yeh, Y.R., Huang, S.Y. and Lee, Y.J. (2009). Nonlinear dimension reduction with kernel sliced inverse regression. *IEEE Transactions on Knowledge and Data Engineering*, 21: 1590-1603.

Zhang, F. (1999). *Matrix Theory: Basic Results and Techniques*. Springer, New York.

**(3) Empirical Scaling Laws and the Aggregation of Non-stationary Data**, Lo-Bin Chang, Stuart Geman, *Physica A* (2013), Vol. 392, Issue 20, 5046-5052.

The 1900 dissertation of Louis Bachelier, on The Theory of Speculation (Bachelier, 1900), proposed a random-walk model for security prices. The basic model, elaborated to accommodate heavy-tailed distributions and stochastic volatilities, still provides a compelling and nearly universally accepted foundation for a theory of price movements. At the same time, a salient and much-discussed feature of the data is the remarkably precise self-similarity of the returns, relative to the return interval, of many of these securities. This was first observed by Mandelbrot (1963) and has since been found in multiple data sets involving a range of securities and time periods (cf. Evertsz, 1995, Mantegna & Stanley, 1995, Guillaume et al., 1997, Gopikrishnan et al., 1999, Podobnik et al., 2000, Ivanov et al., 2001, Wang and Hui, 2001, Gencay et al., 2001, Xu & Gencay, 2003, Ivanov et al., 2004, Matteo et al., 2005, Matteo, 2007, Glattfelder et al., 2011, and Podobnik et al., 2011, to name a few). With a straightforward calculation we will conclude that state-of-the-art models of price movements do not generate self-similar processes, and are therefore at odds with the empirical scaling of returns. We will then show that scaling of *empirical* distributions is likely to be a statistical consequence of the aggregation of large amounts of *non-stationary* data.

Bachelier’s remarkable thesis included a first construction of Brownian motion, and proposed a suitably scaled version as a model for the price dynamics of securities:  $S(t) = S(0) + \sigma w(t)$ , in which  $w$  is a “standard” Brownian motion and  $\sigma$  is the standard deviation of the change in price after one unit of time. The model has evolved, incrementally, to better accommodate theoretical and empirical constraints. For example, the realization that the scale of an ensuing price increment is typically and logically proportional to the current price, rather than independent of it, leads to the geometric (instead of linear) Brownian motion:

$$R(t) \doteq \ln S(t) - \ln S(0) = \sigma w(t) \tag{8}$$

after correcting for a possible drift associated with risk-free investment.

Additionally, the common observation that returns are too peaked and heavy-tailed to be consistent with the normal distribution led Mandelbrot (1963) to seek a replacement

for the Brownian motion, while preserving the compelling argument that increments of prices arise from large numbers of small influences. As stable processes are the only possible limits of rescaled sums of independent random variables (the “generalized central limit theorem,” Lévy, 1925), and as the resulting theoretical return distributions are a better, and often excellent, fit to empirical returns, Mandelbrot proposed models of the same form as (8) but with  $w(t)$  interpreted more generally as an  $\alpha$ -stable Lévy process,  $\alpha \in (0, 2]$ . The special case  $\alpha = 2$  recovers ordinary Brownian motion.

Further refinements are dictated by the fact that volatilities, modeled by the scaling factor  $\sigma$  (which is a standard deviation only in the case  $\alpha = 2$ ), are almost never constant ( $\sigma = \sigma(t)$ , “stochastic volatility” cf. Shephard, 2005 and Shapira, 2011 ). And in fact the evidence is for very rapid fluctuations in  $\sigma(t)$  (e.g. the left-hand panel in Figure ?? is typical). A parsimonious extension of (8), whether or not  $\alpha = 2$ , is through the stochastic integral

$$R(t) = \ln S(t) - \ln S(0) = \int_0^t \sigma(s)dw(s) \quad (9)$$

which falls out of the same thought experiment that took us from discrete and small price movements to the stable process  $w(t)$ , except that a step at time  $t$  has scale proportional to  $\sigma(t)$  rather than  $\sigma$ .

Many lines of thought lead to more-or-less the same thing. For example, the function  $\sigma(t)$  can be thought of as itself a stochastic process, dependent or independent of  $w$ , or as a given deterministic (perhaps historical) volatility trajectory. Many authors prefer to think of  $\sigma(t)$  as a proxy for, or measure of, market activity or “market time,” and in fact under very general conditions the result of a random time change can also be expressed by (9), cf. Clark (1973), Geman et al. (2001) and Veraart & Winkel (2010). We will assume that either  $\sigma(t)$  is deterministic or, if stochastic, it is independent of  $w$ , in which case we will condition on a sample path of  $\sigma(t)$  so that the two situations amount to the same thing. In any case, it would be a mistake to think of  $\sigma(t)$  as statistically *stationary*, given the prototypical intraday volatility profile (including high values in the opening and closing thirty minutes) and the overall rise in volatility with rise in volume through the years and decades over which return profiles are studied.

The question we wish to examine is an apparent incompatibility between the class of models embodied by equation (9) and the widely cited evidence for scaling of returns on stock prices and other financial processes.

## References

- [1] A. Amarasingham, M. Harrison, N. Hatsopoulos, and S. Geman. Conditional modeling and the jitter method of spike resampling. *Journal of Neurophysiology*, 107:517–531, 2012.
- [2] L. Bachelier. Théory de la spéculation. *Annales Scientifiques de l’Ecole Normale Supérieure*, 3(17):21–86, 1900.
- [3] L.-B. Chang, S. Geman, F. Hsieh, and C.-R. Hwang. Invariance in the recurrence of large returns and the validation of models of price dynamics. submitted for publication, 2013.
- [4] P.K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1):135–155, 1973.
- [5] C. Evertsz. Fractal geometry of financial time series. *Fractals*, 3(3):609–616, 1995.
- [6] H. Geman, D.B. Madan, and M. Yor. Time changes for Lévy processes. *Mathematical Finance*, 11(1):79–96, 2001.
- [7] R. Gencay, F. Selcuk, and B. Whitcher. Scaling properties of foreign exchange volatility. *Physica A*, 289:249–266, 2001.
- [8] J.B. Glattfelder, A. Dupuis, and R.B. Olsen. Patterns in high-frequency fx data: discovery of 12 empirical scaling laws. *Quantitative Finance*, 11:599–614, 2011.
- [9] P. Gopikrishnan, V. Plerou, L. Amaral, M. Meyer, and H. Stanley. Scaling of the distribution of fluctuations of financial market indices. *Physical Review E*, 60:5305–5316, 1999.
- [10] D. Guillaume, M. Dacorogna, R. Davé, Ulrich A. Müller, R. Olsen, and O. Pictet. From the bird’s eye to the microscope: A survey of new stylized facts of the intraday foreign exchange markets. *Finance and Stochastics*, 1:95–129, 1997.
- [11] F. Hsieh, S.-C. Chen, and C.-R. Hwang. Discovering stock dynamics through multi-dimensional volatility phases. *Quantitative Finance*, 12:213–230, 2012.

- [12] P.C. Ivanov, B. Podobnik, Y. Lee, and H.E. Stanley. Truncated levy process with scale-invariant behavior. *Physica A*, 299:154–160, 2001.
- [13] P.C. Ivanov, A. Yuen, B. Podobnik, and Y. Lee. Common scaling patterns in inter-trade times of u. s. stocks. *Physical Review E*, 69:056107(7 pages), 2004.
- [14] P. Lévy. *Calcul des probabilités*. Gauthier-Villars, Paris, 1925.
- [15] B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36:394–419, 1963.
- [16] R.N. Mantegna and H.E. Stanley. Scaling behaviour in the dynamics of an economic index. *Nature*, 376:46–49, 1995.
- [17] T.D. Matteo. Multi-scaling in finance. *Quantitative Finance*, 7:21–36, 2007.
- [18] T.D. Matteo, T. Aste, and M.M. Dacorogna. Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *Journal of Banking and Finance*, 29:827–851, 2005.
- [19] B. Podobnik, P.C. Ivanov, and H.E. Stanley. Scale-invariant truncated levy process. *Europhysics Letters*, 52(5):491–497, 2000.
- [20] B. Podobnik, A. Valentincic, D. Hovatic, and H.E. Stanley. Asymmetric levy flight in financial ratios. *Proceedings of the National Academy of Sciences*, 108:17883, 2011.
- [21] T. Preis, J.J. Schneider, and H.E. Stanley. Switching processes in financial markets. *Proceedings of the National Academy of Sciences*, 108(19):7674, 2011.
- [22] Y. Shapira, D. Y. Kenett, Ohad Raviv, and E. Ben-Jacob. Hidden temporal order unveiled in stock market volatility variance. *AIP Advances*, 1(2):022127, 2011.
- [23] N. Shephard. *Stochastic Volatility*. Oxford University Press, Oxford, 2005.
- [24] A.E. Veraart and M. Winkel. Time change. In Rama Cont, editor, *Encyclopedia of Quantitative Finance*. Wiley Online Library, 2010.
- [25] B.H. Wang and P.M. Hui. The distribution and scaling of fluctuations for hang seng index in hong kong stock market. *The European Physical Journal B*, 20:573–579, 2001.

[26] Z. Xu and R. Gencay. Scaling, self-similarity and multifractality in fx markets. *Physica A*, 323:578–590, 2003.

**(4) Invariance in the Recurrence of Large Returns and the Validation of Models of Price Dynamics**, Lo-Bin Chang, Stuart Geman, Fushing Hsieh, Chii-Ruey Hwang, *Physical Review E* (2013), Vol. 88, Issue 2, 022116.

Given a sequence of stock prices  $s_0, s_1, \dots$  recorded at fixed intervals, say every five minutes, let  $r_n \doteq \log \frac{s_n}{s_{n-1}}$ ,  $n = 1, 2, \dots$ , be the corresponding sequence of returns. Fix  $N$  and define an excursion to be a return that is large, in absolute value, relative to the set  $\{r_1, r_2, \dots, r_N\}$ . Specifically, following Hsieh et al. (2012), define the *excursion process*,  $z_1, z_2, \dots, z_N$ :

$$z_n = \begin{cases} 1 & \text{if } r_n \leq l \text{ or } r_n \geq u \\ 0 & \text{if } r_n \in (u, l) \end{cases}$$

where  $l$  and  $u$  are, respectively, the 10'th and 90'th percentiles of  $\{r_1, \dots, r_N\}$ . We call the event  $z_n = 1$  an excursion, since it represents a large movement of the stock relative to the chosen set of returns. We will study the distribution of waiting times between large stock returns by studying the distribution of the number of zeros between successive ones of the excursion process. Our motivation includes:

1. The empirical observation (cf. Chang et al., 2013) that this waiting-time distribution is nearly invariant to time scale (e.g. thirty-second, one-minute, or five-minute returns), to stock (e.g. IBM or Citigroup), and to year (e.g. 2001 or 2007).
2. The waiting-time to large returns is of obvious interest to investors, and much easier to study if, and to the extent that, it is invariant across time scale, stock, and year.
3. The particular waiting-time distribution found in the data and its invariance to time scale have implications for models of price and volatility movement. For instance, Lévy processes, “market-time” models based on volume or trades, and GARCH models are each one way or another inconsistent with the empirical data.
4. Overwhelmingly, the evidence for self-similarity comes from studies of the univariate (marginal) return distributions (e.g. evidence for a stable-law distribution), but marginal distributions leave data models underspecified. Waiting-time distributions provide additional, explicitly temporal, constraints, and these appear to be nearly universal.

Larger returns can be studied by using more extreme percentiles. Although we have not experimented extensively, the empirical results we will report on appear to be qualitatively robust to the chosen percentiles and hence the definition of “large return.” In general, the upper and lower percentiles index a family of waiting-time distributions that might prove useful to systematically constrain the dynamics of price and volatility models.

In §2, we study the invariance of the empirical waiting-time distribution. Starting with the Lévy type models, we first make a connection between the model-based distribution and the geometric distribution. To be concrete, let  $S(t)$  follow the “Black-Scholes model” (geometric Brownian motion) as an example:  $d \log S(t) = \mu dt + \sigma dw(t)$ , where  $w(t)$  is a standard Brownian motion. Because of the independent increments property of Brownian motion  $w(t)$ , the return sequence under this model is exchangeable (i.e. the distribution of any permutation remains the same). Therefore, the empirical waiting-time distribution under this model is provably invariant to time scale and to time period. More specifically, the probability of getting a “large” return, with  $l = 10$ ’th percentile and  $u = 90$ ’th percentile, is exactly 0.2 at each return interval and the empirical waiting-time distribution is, therefore, nearly a geometric distribution with parameter 0.2 (see §2.1 for more detail). We emphasize these considerations apply without modification not just to the geometric Brownian motion but to all of its popular generalizations as geometric Lévy processes.

Not surprisingly (cf. “stochastic volatility”), the *actual* (i.e. empirical) waiting-time distribution is different from geometric. But what is surprising is the invariance of this distribution across time scale, stock, and year. In §2.2 we make an exhaustive comparison of empirical waiting-time distributions, using trading prices of approximately 300 stocks from the S&P 500 observed over the eight years from 2001 through 2008. Invariance to timescale is strong in all eight years; invariance to stock is strong in years 2001–2007 and less strong in 2008; and invariance across years is stronger for pairs of years that do not include 2008. (We have not studied the years since 2008.) In §2.3, we will connect waiting-time invariance to self-similarity, being careful to distinguish a self-similar *process* from a process having self-similar increments (i.e. distinguish dynamics from marginal distributions).

Which of the state-of-the-art models of price dynamics are consistent with the empirical distribution of the excursion process? The existence of a nearly invariant waiting-time distribution between excursions provides a new tool for evaluating these models, through

which questions of consistency with the data can be addressed using statistical measures of fit and hypothesis tests. In general, we will advocate for permutation and other combinatorial statistical approaches that robustly and efficiently exploit symmetries shared by large classes of models, supporting exact hypothesis tests as well as exploratory data analysis. In §3 we introduce some combinatorial tools for hypothesis testing and explore the implications of waiting-time distributions to the time scale of volatility clustering. We continue with this approach, in §4, with a discussion of stochastic volatility modeling, as well as “market-time” and other stochastic time-change models. We conclude, in §5, with a summary and some proposals for price and volatility modeling.

## References

- [1] T. G. Andersen and T. Bollerslev. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4:115–158, 1997.
- [2] T.G. Andersen. Return volatility and trading volume: An information flow interpretation of stochastic volatility. *Journal of Finance*, 51:169–204, 1996.
- [3] T. Ané and H. Geman. Order flow, transaction clock, and normality of asset returns. *Journal of Finance*, 55(5):2259–2284, 2000.
- [4] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.
- [5] J.-P. Bouchaud. Power laws in economics and finance: some ideas from physics. *Quantitative Finance*, 1:105–112, 2001.
- [6] L. Calvet and A. Fisher. Multifractality in asset returns: theory and evidence. *The Review of Economics and Statistics*, 84:381–406, 2002.
- [7] L.-B. Chang and S. Geman. Empirical scaling laws and the aggregation of non-stationary data. *Physica A*, to appear, 2013.
- [8] L.-B. Chang, A. Goswami, F. Hsieh, and C.-R. Hwang. An invariance for the large-sample empirical distribution of waiting time between successive extremes. *Bulletin of the Institute of Mathematics, Academia Sinica*, 8:31–48, 2013.

- [9] P.K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1):135–155, 1973.
- [10] P. Diaconis and D. Freedman. Finite exchangeable sequences. *Annals of Probability*, 8(4):745–764, 1980.
- [11] D. Easley, M.M. López de Prado, and M. O’Hara. The volume clock: insights into the high frequency paradigm. *Journal of Portfolio Management*, 39:19–29, 2012.
- [12] P. Embrechts and M. Maejima. *Selfsimilar Processes*. Princeton University Press, Princeton, NJ, 2002.
- [13] R.F. Engle. Autoregressive conditional heteroscedasticity with estimates of variance of united kingdom inflation. *Econometrica*, 50:987–1008, 1982.
- [14] X. Gabaix, P. Gopikrishnan, V. Plerou, and H.E. Stanley. A theory of power-law distributions in financial market fluctuations. *Nature*, 423:267–270, 2003.
- [15] H. Geman. From measure changes to time changes in asset pricing. *Journal of Banking & Finance*, 29:2701–2722, 2005.
- [16] H. Geman, D.B. Madan, and M. Yor. Time changes for Lévy processes. *Mathematical Finance*, 11(1):79–96, 2001.
- [17] R. Gencay, F. Selcuk, and B. Whitcher. Scaling properties of foreign exchange volatility. *Physica A*, 289:249–266, 2001.
- [18] J.B. Glattfelder, A. Dupuis, and R.B. Olsen. Patterns in high-frequency fx data: discovery of 12 empirical scaling laws. *Quantitative Finance*, 11:599–614, 2011.
- [19] S.L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.
- [20] C.C. Heyde. A risky asset model with strong dependence through fractal activity time. *J. Appl. Prob.*, 36:1234–1239, 1999.
- [21] F. Hsieh, S.-C. Chen, and C.-R. Hwang. Discovering stock dynamics through multi-dimensional volatility-phases. *Quantitative Finance*, 12:213–230, 2012.

- [22] J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42:281–300, 1987.
- [23] B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36:394–419, 1963.
- [24] B. Mandelbrot and H. Taylor. On the distribution of stock price differences. *Operations Research*, 15:1057–1062, 1967.
- [25] R.N. Mantegna and H.E. Stanley. Scaling behaviour in the dynamics of an economic index. *Nature*, 376:46–49, 1995.
- [26] R.N. Mantegna and H.E. Stanley. *An Introduction to econophysics*. Cambridge University Press, 2000.
- [27] T.D. Matteo. Multi-scaling in finance. *Quantitative Finance*, 7:21–36, 2007.
- [28] U.A. Muller, M.M. Dacorogna, R.B. Olsen, O.V. Pictet, M. Schwarz, and C. Morgenegg. Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. *Journal of Banking and Finance*, 14:1189–1208, 1990.
- [29] N. Shephard. *Stochastic Volatility*. Oxford University Press, Oxford, 2005.
- [30] A.E. Veraart and M. Winkel. Time change. In Rama Cont, editor, *Encyclopedia of Quantitative Finance*. Wiley Online Library, 2010.
- [31] B.H. Wang and P.M. Hui. The distribution and scaling of fluctuations for hang seng index in hong kong stock market. *The European Physical Journal B*, 20:573–579, 2001.
- [32] F. Wang, K. Yamasaki, S. Havlin, and H.E. Stanley. Scaling and memory of intraday volatility return intervals in stock markets. *Physical Review E*, 73:026117–1–023117–8, 2006.
- [33] Z. Xu and R. Gencay. Scaling, self-similarity and multifractality in fx markets. *Physica A*, 323:578–590, 2003.