# Multi-Level Time Series Clustering: Issues with Traditional Risk Management Frameworks

Michael Kotarinos[1], Doo Young Kim[2], Chris P. Tsokos[1]
[1]University of South Florida, 4202 East Fowler Ave. Tampa, FL 33620
[2]Sam Houston State University, Lee Drain Building 439D, Huntsville, Texas 77341

**Abstract**
Multi-Level Time Series Clustering (MLTC) is a distance based technique which allows for efficient clustering of time series data. While MLTC has previously been used to study the characteristics of different sectors and form a diversified portfolio it has not been used in context of broader problems in asset allocation. In particular, there is no automated procedure for the generation of portfolios based on preferences over risk. In this paper, the authors present an overview of some of the issues associated with traditional risk management and portfolio techniques. The authors introduce the need for methods that address deficiencies in asset selection, and how MLTC Time Series Clustering is the first step in a new possible framework for addressing these issues. This paper sets criteria, methods, and techniques for evaluating proposed frameworks and new ways of thinking about asset allocation.

**Key Words:** Suitability, Multi-Level Time Series Clustering, Modern Portfolio Theory, Risk-Adjusted Return, Preferences over Risk

## 1. Issues with Traditional Asset Selection Techniques

Traditional asset allocation techniques use a variety of statistical tools to make trade-offs between risk and return. While other risk factors are often considered important factors of consideration, risk and return remain the dominant aspects used to evaluate a broad portfolio of assets[1]. This section aims to investigate issues that arise in traditional asset selection that should be addressed in a comprehensive alternative format.

### 1.1 Suitability is taken as an absolute constraint, with risk-return trade-offs occurring over allowable asset classes

In traditional finance, suitability is taken as an absolute condition above all else. For example, a younger investor would be unlikely to invest in fixed income instruments and would be heavily invested in assets with equity-like characteristics, while individuals more concerned with needing funds immediately would be unlikely to be placed in illiquid assets[2]. In traditional finance, a financial planner would have an initial consultation with a client where investment goals and time horizons are discussed. These are then used to discuss different kinds of asset classes and the risks and benefits associated with each class.

The authors of this paper are by no means against discussing investment goals and risk factors with clients. To the contrary, the authors are fully in support of open, honest, transparent communication with clients and investors. The issue is often the way this communication takes place. First of all, these risk factors are often discussed in isolation rather than as part of an entire portfolio. One example of this are derivative contracts: in a vacuum, derivatives are risky tools that allow a client to become highly leveraged quickly for low levels of capital investment[3]. On its own, this asset class is very risky. If this asset class is so risky, then why is it the case that derivatives are often also considered "insurance" contracts that can allow large farms and corporations cheaper access to capital, portfolio managers a way to hedge risk and cover positions, and a way to earn additional portfolio income on lots held in long positions?

The answer is that derivatives can serve as a risk accelerator or as a risk mitigator, depending on the other assets held in a portfolio. Similarly, commodities, convertible debt, and many other instruments can appear to have one impact on a portfolio but have radically different results when mixed with other assets. While modern portfolio theory recognizes that assets are related, the level of discussion on the nature of these interactions is often limited to some Monte Carlo simulations over risk with dubious at best assumptions[4].

In complex, rapidly changing global markets it can be difficult for asset managers to assess the relationships between asset classes[5]. While there are theoretical relationships that hold and some beliefs held by financiers that can establish some base-line for analysis, many portfolios quickly deviate from this base-line and take varied and complex characteristics. If asset managers cannot agree on the nature of the market and its implication for a varied basket of assets, how can these risks be adequately communicated and addressed to clients? If it is difficult to communicate these ideas to clients, how are clients supposed to communicate to asset managers their concerns? In an ideal framework, clients and asset managers would be able to find a technique, approach, or method to aid in the communication and discussion of the overall goals of the client's portfolio, with a way to assess different allocations not just in terms of risk and return but how well these portfolios met investment goals and objectives.

### 1.2 Volatility is difficult to measure and often means many different things

In statistics, variance is known as an inherent "evil" that one must address when assessing models. Perhaps most fitting is the argument that without variance, statisticians would not be as in demand as they currently are in the "data analytics" revolution. While many concepts such as variance-bias trade-offs are quite clear in statistics, oddly enough finance has a much more muddled and esoteric treatment on the topic. Rather than variance, we have the notion of "volatility." What is volatility? Variance is often treated by statisticians working in finance as equivalent to volatility (it is not). Rather than being defined, volatility is often best thought of as an elephant: one knows it when it is seen, and at any rate all of the common ways of describing volatility tend to measure very similar things that tend to be related.

Some take volatility to be related to VIX-related financial instruments which measure a modified standard deviation. Other analysts use GARCH or other time series techniques to measure volatility. Mathematical financiers often use volatility as the "drift" or "non-deterministic integral" component of a stochastic differential equation. Still others simply use the standard deviation of a fixed amount of time (typically quarters, but even this is not universally set) as the "volatility" of an asset. Bearing these many ideas in mind, this leads to many issues:

1. No one actually agrees what volatility really is, and all of the "camps" on this issue are interested in vastly different objectives.

The analyst using basic standard deviation to measure volatility is probably a very different kind of analyst from the mathematician interested in a stochastic differential equation[6]. Neither one of these analysts would probably consider it to be the implied volatility from Black-Scholes. It is unclear how these measures relate to VIX-related characterizations[7].

2. Volatility is difficult to measure, and data is not always reliable.

If one cannot agree what volatility actually is, then how would one even measure it? One approach is to use some sort of GARCH method to estimate volatility. This first involves determining which of the many varied and diverse GARCH methods are most appropriate for a given problem, then figuring out what the best GARCH paramaterization is, and then finally convincing third party regulators that the paramaterization was set in a neutral, fair way and not in a way to generate the desired results[8].

Rather than trying to go through all the different GARCH approaches and figure out the best possible approach, most quantitative analysts simply use GARCH(1,1) and call it a day[9]. The logic is while this may not be the most appropriate paramaterization, most of GARCH techniques are highly correlated anyways and it is widely viewed that this is the "safest" default paramaterization for use. One of the main philosophical differences is that the estimates produced are not considered a be all to end all but simply a decent working tool that gives some degree of insight into a problem. This is based on the general philosophy of valuation in finance: one does not need to have an exact valuation, but generally be in the right ballpark. Informally, simply being in the right neighborhood of the price or volatility is "good enough" because of how heavily mispriced an asset might be.

Suppose a team does finally agree to use, say, a GARCH(1,1) method to model the volatility of an asset (say, a stock). How exactly should one implement this measure? Does one use the daily closing prices of the stock? What about intra-day trading? Does one consider the bid, the ask, the spread between the two, or the midpoint of these two measures? All of these have issues. For example, the closing trade is just a single trade. The bid and the ask are often typically just values being quoted by the market maker on the NYSE. Does a NYSE market maker's quotes reflect the market smaller trades see on a daily basis? What about dark pooled trade orders and other off-the-market activity? Casual observers might believe that the density of data in finance makes it ideally suited to "plug and chug" implementations, but there are often very serious questions about how a method is implemented and what it means about the larger market[10]. In this case, even if one can agree on a method, the implementation and usage of the method is often far from settled.

### 1.3 Traditional methods make unrealistic assumptions that are difficult to verify with possibly unknown consequences

Traditional techniques such as optimization under the Sharpe ratio rely and the Capital Asset Pricing Model rely on assumptions that are often difficult to verify with unknown implications. Consider first the Sharpe ratio: the ratio requires the use of the risk-free rate, which is assumed to be known. What, exactly, is it known to be? Is it the overnight treasury bill rate? Is it the 30 year treasury rate? Is it some sort of LIBOR rate? If it is a LIBOR rate, what does it say if our "risk-free rate" was for years arbitrarily set by bankers over tea in London to maximize the gains on swap contracts[11]?

Many analysts simply used either LIBOR or the treasury rate, because they were usually very close. Often the choice of a risk-free rate would depend on one's investment horizon, which is also assumed to be known (but in practice is often not, at least not with certainty). The capital asset pricing model has numerous issues, including complaints about the way it treats access to information to computational issues in the way the curves are constructed as well as documentation of poor empirical performance and validity[12]. Fama and French proposed an alternative model with additional information on the relationship between the size of companies and the way the market values the books of companies. This model was later revised to include additional factors, leading one to wonder whether the model will continue to need to be revised to reflect new and differing changes in the market[13].

The idea that the models we have need to keep being revised and changes is perhaps not shocking: markets themselves continue to change and adapt over time, and different business cycles and changing economies mean the data we collect may not convey information the same way as it did in the past. Another way of saying this idea addresses a very dangerous idea: perhaps it is the case that the data we have from the past to describe risk, return, volatility, and preferences at some point no longer becomes useful in the future to describe these same ideas.

To imagine this idea, imagine a company before the 2008 financial crisis. The company may appear to be well capitalized and a relatively safe investment in possession of numerous AAA-rated bonds. These bonds later turn out to be toxic assets, and the company collapses. The data we had before 2008 ("all AAA rated fixed income assets are extremely safe") was, for a brief period of time, no longer useful information for the current financial period. This is not merely a story of volatility, but of the fundamental way analysts interpret, report, and analyze data relating to companies and performance. This is perhaps one of the biggest threats facing the advocates of machine learning and automated trading: what happens when the data we used to make decisions in the past is not a reliable predictor of the future? What happens when markets suddenly change and the information we collected either is no longer valid or is greatly called into question?

When data is corrupted, unreliable, or in the cases of dark pools sometimes missing outright one can venture into frightening territory: the realm of uncertainty. Statisticians typically operate in the domain of risk, characterizing the possible scenarios and probabilities of events occurring. Despite what some analysts might suggest, uncertainty is a terrifying environment. An investment in uncertainty not only faces risks, but the risk characterization of these risks is poor, and there are many risks that are also not being considered or characterized. While markets are generally good at operating in environments with risk (where appropriate controls can be taken), it is extremely difficult to operate in an environment of uncertainty where the appropriate hedges, controls, or suitable investments are unclear. Traditional methods typically always assume certain characteristics or parts of the market are known. What happens when not only is an estimator or part of our model unknown or non-trivial (when we assumed otherwise), but we do not fully understand the impact of this faulty assumption and what it means in terms of exposure and risk? The consequences for misuse and recklessness in this modeling aspect are often severely understated, as was the case with sub-prime mortgages in 2008.

## 2. MLTC Method: Towards a Unified Framework

The Multi-Level Time Series Clustering (MLTC) method has two roots that are Lag Target Time Series Clustering (LTTC) and Multi-Factor Time Series Clustering (MFTC) [14]. LTTC uses the sum of weighted lag distances, and MFTC allows us to employ more than a single factor in calculating lag distances. MLTC is an improved method which is especially for clustering financial time dependent information by putting emphasis on a pure lag time dependency.

The most applicable data for MLTC method is stock price data which have lots of local extrema in a plot for price fluctuations even within one day. The first step in MLTC is a clustering step. In this step, we calculate pairwise distances in each lag of interest and perform a clustering procedure using dissimilarity matrices that are created based on our lag of interest. The second step in MLTC is a portfolio selection step. We now choose the same number of stocks within the same cluster from each sector based on the trading interval which is the same the degree of lag of our interest. Then, we investigate neighborhood lag solutions in order to verify that our original solution is unique.

Because MLTC method is another way of assessing, categorizing, and managing risk it gives analysts another way of "controlling" possible sources of variation. While not perfect, this provides an additional level of control not seen in traditional asset management by providing for assets that have different movement patterns of behavior. This practice of "hedging" different moving streams to provide additional control of risk is a concept that is deployed in a less technical level by traditional hedge funds, which use mixes of long and short positions to produce returns uncorrelated to the market. When this method for the control of variation is paired with other techniques and methods, the end result is that investors gain an additional level of safety and control in the asset allocation process.

While MLTC on its own does not completely control for variation, it is a useful part in a broader framework for assessing risk and volatility. By combining MLTC clustering with other techniques, one could develop methods, ideas, and ways of controlling for and managing risk and unique and novel ways.

## 3. Closing Thoughts

While any method has its risks and limitations, the authors of this article find that MLTC method is an under-used and under-utilized tool in risk management in the financial literature. Perhaps it is due to many quantitative funds not disclosing in the literature the sources of their above-market performance or the novelty of the application, but it is certainly not for lack of opportunity. As investors look to develop tools that not only provide suitable returns but also meet the other needs of clients and have built-in diversification, multi-level time series clustering may prove to be a useful component of new and broad frameworks that look to redefine the way investors, analysts, and clients think about asset allocation in changing markets.

## References

[1] S. Aboura and N. Wagner., Extreme asymmetric volatility: Vix and s&p 500, Working Paper, (2017).

[2] P. Ashton and B. Christophers, On arbitration, arbitrage and arbi-

trariness in _nancial markets and their governance: Unpacking libor and the libor scandal, Economy and Society, (2015).

[3] H. K. Baker and J. A. Haslem, The impact of investor socioeconomic characteristics on risk and return preferences, Working Paper, (2015).

[4] F. M. Bandi and R. RenÃ2;Nonparametricstochasticvolatility;EconometricT heory; (2018):

[5] e. a. Bouchaud, Jean-Philippe, Trades, quotes and prices: _nancial markets under the microscope, Cambridge University Press, 2018.

[6] K. Bruce and A. D. Ahmed., Conceptions of professionalism: meaning-ful standards in nancial planning, Taylor and Francis, 2016.

[7] E. E. Deng, Saiying and C. X. Mao, Derivatives-hedging, risk allo-cation and the cost of debt: Evidence from bank holding companies., The Quarterly Review of Economics and Finance, (2017).

[8] E. F. Fama and K. R. French, Choosing factors, Journal of Finan-cial Economics, (2018).

[9] P. Fernandez, Capm: an absurd model, Business Valuation Review, (2015).

[10] F. E. A. Kaiser, Michael G. and F. Ahlemann., Successful project portfolio management beyond project selection techniques: Understanding the role of structural alignment., International Journal of Project Management, (2015).

[11] R. N. M. Kourtis, Apostolos and L. Symeonidis, An international comparison of implied, realized, and garch volatility forecasts, Journal of Futures Markets, (2016).

[12] e. a. Kroese, Dirk P., Why the monte carlo method is so important today, Wiley Interdisciplinary Reviews: Computational Statis-tics, (2014).

[13] C. L. Zhang, Rongmao and L. Peng, Inference for the tail index of a garch (1, 1) model and an ar (1) model with arch (1) errors, Econometric Reviews, (2016).

[14] D. Kim and C. P. Tsokos, Active and Dynamic Approaches for Clustering Time Dependent Information: LTTC and MFTC, Journal of Statistical Theory and Applications, 17 (3), (2018)