CHAPITRE 1

CURRENT STATE OF THE ART OF VOLATILITY FORECAST

1.1 Need to Forecast Financial Index Volatility

At the heart of financial risk modeling, the estimation and forecasting of volatility are critical for institutional investors like banks, mutual fund companies, credit unions, *etc.* as well as for borrowers *e.g.,* governments, corporations and individuals. Specifically, asset managers require a reasonable volatility estimate of standard measures of the markets such as S&P100 index, TSE30, Nikkei 225 index, *etc.* in order to estimate and to be able to forecast the price of the related security as required in the prevalent model of price development (Dupacova, 2002). Based on the forecasts, the concerned parties could hedge their investment portfolios according to the level of their risk tolerance. They could either adjust the contents of the portfolios themselves or use a combination of the corresponding put/call options – contracts that gives the holder the right to sell/buy the underlying security at a specified price for a certain period of time. For example, with the forecasted direction and range of the volatility movement, one could buy/sell selected options to complement his/her investment strategy. If the forecast of volatility is up, the values of the relevant call and put options would go up given that other factors remain reasonably constant. The investor could decide to set up a straddle and expect the rate of return of options to go up, just like the approach associated with the rate of return of equity investment.

The hedging strategies are important for corporations in risk management to maintain their credit and borrowing power under the scrutiny and monitoring of the public and credit unions such as banks. This aspect is more important than before given the serious accounting allegations reported in recent news and the SEC tightening the regulatory controls.

When financial institutions design hedging strategies in portfolio management, they need to determine option prices, including index option, foreign exchange option, equity option, interest rate forwards, variance swaps and so on. All these investment vehicles are available in one or more financial markets in North America and in Europe, except in Asian financial markets where one may have fewer selections. Any forward looking investment vehicle that needs an estimate of risk premium will need to have a forecast of the volatility of the underlying security. To an equity option trader, volatility is a measurement of an underlying stock's price fluctuation. It is a critical factor in calculating an option's current theoretical value (The Option Industry Council, 2004). Volatility is the only major unknown entity in calculating prices of derivatives such as equity and index options and futures and likely to be the only unknown entity to quantify and understand in establishing a variety of trading strategies. Therefore, the accuracy of the forecast has a direct impact on the current price of the security and eventually affects the rate of return of the investment. It is the most vital factor in the outcome – success or failure.

Recently investors could also use the volatility itself as an investment vehicle by directly trading the variance swaps (Wang, 2005) or the Chicago Board Options Exchange (CBOE) Market Volatility Index (VIX) futures. By doing so, one could use VIX futures (VXB) to hedge equity option in the U.S. equity market. As summed up by Poon and Granger (2003), volatility forecasting is recognized as an important task in financial markets, especially in asset allocation, risk management, security valuation, pricing derivatives and monetary policy making.

At the current time, researchers are only beginning to address the fundamental questions of what "risk" or "volatility" mean in precise terms, and how best to model it – mathematically or otherwise. What we do know about volatility from the empirical research reported so far is very encouraging in this regard: volatility process is eminently more persistent and forecastable than the typical asset return process. This indicates an investment opportunity *i.e.,* the potential for generating abnormal returns is more likely to be found in the relatively uncharted territory of volatility arbitrage. And volatility arbitrage is therefore, considered the most fruitful investment opportunity of the next decade, and has provided adequate means and descriptors that can be found to describe and model the underlying processes (Kinlay *et al.* 2001). As Kinlay (2005) recently claimed that his Caissa Capital Fund with $170 million market capitalization has achieved a compounded rate of return 382.91% over the period between Oct 2002 and Jul. 2004. The fund's operation is mainly based on a non-directional market neutral strategy – making money from arbitrage opportunities rather than from directional trades. The core of the strategy hinges on volatility forecasting – making use of genetic algorithms to construct long/short volatility portfolios. Please refer to APPENDIX 1 and B for the definition of arbitrage and other related financial terms.

Volatility is typically an unobservable random variable. It is thus not easy to estimate or predict where a pattern heads towards in the next time step; let alone what value it is going to take. Since it is the only unknown so to speak in determining option values, forecasting volatility becomes a major challenge in the financial engineering field that is rapidly evolving in many recent research undertakings.

1.2 Parametric Volatility Forecasting

Statistical methods typically use discrete time models to process the historical data of the underlying asset returns in order to calculate the variance, hence has a measure of the volatility of a time series. The contemporary research about the behavior of rate of return of equities/indices, specifically their volatility is dominated by the parametric statistical approaches, such as a variety of Generalized Autoregressive Conditional Heteroscedastic (GARCH) models (Davidson & MacKinnon, 1993). The objectives of these models are to obtain the most efficient as well as the best fit of the time series, and only lately the effectiveness of these models is improving gradually. Volumes of literature in this field have documented the use of a wide array of variant GARCH models, such as stationary GARCH, unconstrained GARCH, non-negative GARCH, multivariate GARCH, exponential GARCH, Glosten, Jagannathan, and Runkle GARCH, integrated GARCH, GARCH with normal distribution, asymmetric GARCH with (skewed) student-*t* densities, GARCH with measures of lagged volume *etc.* However, the resulting predicting power is still lagging (Harvey, Dec. 1999, Peters, 2001, Chong et al. 1999, Park, 2002, Brooks, 1998, 2003) and has not been known to be reliable in application to real practice.

JP Morgan's RiskMetrics (Neely & Weller, 2001) system is among the most popular models for market risk management, where weights on past squared returns decline exponentially as moving backwards in time. The RiskMetrics volatility model or the exponential smoother is written in the following form as

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)u_{t-1}^2 \qquad (2)$$

where σ_t^2 is the variance of the rate of return at time t, u_{t-1} is the rate of return at time $t-1$ and $|\lambda| < 1$ is a scaling factor. The RiskMetrics model has some clear advantages. First, it tracks volatility changes in a way, which is broadly consistent with observed returns. Recent returns matter more for tomorrow's volatility than distant returns as λ is less than one and therefore, gets smaller when the lag τ becomes bigger, where τ is the discrete time difference between the past and the current volatility. This is essentially a Markovian approach, at least in basic concept. For example, when τ =2, *i.e.* σ_{t-2} has only 0.81 contribution to the current volatility σ_t if λ = 0.9. Second, the model only contains one unknown parameter, namely, λ. RiskMetrics actually sets λ = 0.94 for every assets for daily volatility forecasting. In this case, no estimation is necessary, which is a huge advantage in dealing with large portfolios. However, the RiskMetrics model does have certain serious shortcomings. For example, it does not allow for a leverage effect, *i.e.* when the equity price falls, the debt/equity ratio increases. As a result, the debt becomes more highly leveraged and the general risk level is elevated. The model also provides counterfactual longer-horizon forecasts. Please refer to APPENDIX 1 and 2 for more details.

The GARCH model helps resolve some of the above mentioned problems. From the analysis of empirical data, many stylized properties related to volatility are available. The most important of these properties is the long memory of the volatility, as measured by a lagged autocorrelation function that decays as a power law. This property is also called volatility clustering, and the slow decay of the autocorrelation means that this clustering is present at all time horizons (Alexander, 2001). A simple model that describes volatility clustering is the GARCH(1,1) model described by Eq. (3). The GARCH(1,1) model is a GARCH model with the variance at time t completely defined by the data at time t-1 :

$$\sigma_i^2 = \omega + \alpha u_{i-1}^2 + \beta \sigma_{i-1}^2. \tag{3}$$

where σ_i^2 and σ_{i-1}^2 are the variances at time i and $i-1$, u_{i-1}^2 is the square of the continuously compounded return during day $i-1$; while ω, α and β are constants (Hall, 2000). Please refer to APPENDIX 2 for a more detailed mathematical treatment of the parametric forecasting.

1.3 Difficulties with Parametric Approach

An inconvenience shared by both RiskMetrics and GARCH models is that the multi-period distribution is unknown even if the one-day-ahead distribution is assumed to be Gaussian. Thus while it is easy to forecast longer-horizon variance in these models, it is not as easy to forecast the entire conditional distribution (Christoffersen, 2002). In practice, it is quite desirable to forecast volatility at more than just one time horizon.

The GARCH(1,1) model has an exponential autocorrelation function for the volatility, meaning that it captures the volatility clustering only at one time horizon. In order to remedy the shortcomings of the simplest GARCH(1,1) model, a large number of variations in the popular Autoregressive Conditional Heteroscedastic (ARCH) class of models have been studied, mostly using daily data (Zumbach *et al.*, 2001). However, all ARCH/GARCH models are parametric. That is, they make specific assumptions about the functional form of the data generation process and the distribution of error terms, which may not be amenable to be generalized. More importantly, both models lack the capability in dealing with non-linearity because of the rigid structure of the parametric modeling. Stated differently, one explanation to the poor forecasting performance of the entire GARCH model family could be that although the rate of returns are posited by some type of stochastic process, the volatility is formulated to be entirely deterministic in nature (Brooks, 1998), thus introducing inherent incompatibility in the basic hypothesis itself. These models all ignore the observation that the volatility of volatility is itself stochastic in character and behavior (Carr, 2004). These limitations are also present even in the more recent work by P. Christofferson, K. Jacob and Y. Wang (2004), in which better Root-Mean Square Error (RMSE) estimations were accounted for in comparison with those in the benchmark GARCH(1,1) model. Refer to APPENDIX 2 for more details.

Partially due to these fundamental problems, a recent study (Christoffersen, 2002) found that while the risk forecasts on average tended to be overly conservative, at certain times the realized losses far exceeded the risk forecasts. More importantly, the excessive losses tended to occur mostly on consecutive days. Thus, looking back at the data on the *a priori* risk forecasts and the *ex ante* loss realization, one would have been able to forecast an excessive loss for tomorrow based on the observation of an excessive loss of today. This serial dependence unveils a potentially serious flaw in current financial sector risk management practices, and it motivates the development and implementation of new tools such as those attempted and presented here.

1.4 Currently Best Known Results

In recent years, high frequency data have become available to the research teams and to the general public, particularly the hence derived data on the rate of returns that are based on minute-by-minute transaction prices. Volatility can be measured arbitrarily from return series that are sampled sufficiently more frequently. For example, one could estimate daily or weekly volatility by integrating the realized volatility obtained in shorter time spans, *e.g.* say 15 minutes.

In modeling the volatility, emerging opinion on its theory suggests that the estimation of typical volatility in terms of IV possesses some specific advantages. Under the usual diffusion assumptions, Andersen and Bollerslev (1998) have shown that IV computed with high frequency intra-day returns could effectively be an error-free volatility measure. Diffusion is a phenomenon that in any one direction at a unit time the net flux of molecule movement equals zero. As a result, we can treat volatility as observable in analyzing and forecasting by much simpler methods than the complex econometric models that treat volatility as latent (Kinlay *et al.* 2001).

Although IV provides improved estimation accuracy, it has been shown that only about 50% of the variance in the one-day-ahead volatility factor could be accounted for. Beyond the 50%, this approach is hindered by the parametric nature of the GARCH model, which has difficulty in handling abrupt changes and discontinuities, where volatility is most significant. Therefore, the forecasting results are still far from being reliable and not useful for prediction purposes. Parametric models are easy to estimate and readily interpretable, but they possess major difficulty in dealing with non-linearity such as jumps, abrupt discontinuities, structural changes in data flow, *etc.* As a result, these type of approaches lack accuracy in forecast. For instance, one of their problems would be that IV measures do not distinguish between variability originating from continuous price movements and from jumps. As noted by Christoffersen and Diebold (2002), the dynamic impact may differ across the two sources of variability.

Investment analysts have developed proprietary research products based on undisclosed mathematical models. The best known of these products, the Volatility Report™, enables investors to identify opportunities to trade asset volatility at times of favorable market conditions, based on a proprietary volatility index that measures underlying volatility more accurately and efficiently than traditional methods. Using statistical techniques, they claim to be able to anticipate correctly the future direction of volatility an average of 72% - 75% of the time in the universe of stock and equity indices that are analyzed. They are able to identify regimes of unsustainably high or low levels of volatility with a high degree of accuracy (Kinlay 2001). More details could be found in Section 2.2.2. But the basis of these results are not known and hence the justification for this research to advance a fresh approach to the volatility problem.

Since the methodology employed in the Volatility Report™ (and other classified research projects) is proprietary, our research objective here is to achieve the same or better forecasting accuracy based on novel engineering–based techniques and thus generate methodologies for financial applications that can be made publicly available to the financial analysts, traders and the research community.

1.5 Alternative Forecasting Methods

In the domain of parametric estimation, there are many different forms of nonlinear regression models that are statistically well behaved, *i.e.* different algebraic forms, such as the Oswin model, Smith model, Guggenheim-Anderson-De Boer model (GAB), *etc.* All these are mathematical models used for fitting data, which is of nonlinear nature (Stencle, 1999). Unlike engineering, biochemistry or other scientific disciplines where much research has been done to establish the corresponding models for a variety of phenomena, applications of nonlinear analysis is far from being systematic in their research efforts and has not been well-established in the financial time series analysis field. It is therefore, necessary to observe in-depth the behavior of the volatility of different underlying securities before one could choose the appropriate model. That is why the solution for the problem at hand has to go beyond a model-based computational approach.

In the extreme case such as the far-from-linear type of nonlinear model, even if a model is selected and the respective parameters have been reasonably well-estimated, comparing different sets of data, in order to test the validity of the model will be difficult, because there usually exist following constraints and obstacles :

a) The parameter estimates may be very biased;
b) The standard errors of the parameter estimates may be grossly underestimated.

Other, perhaps much more complex models, may give better representations of the underlying data generation process. If so, then procedures designed to identify these alternative models have an obvious payoff. Such procedures are described as non-parametric. Instead of specifying a particular functional form for the data generation process and making distributional assumptions (*e.g.* Gaussian distribution) about the error terms, a non-parametric procedure will search for the best fit over a large set of alternative functional forms (Neely, 2001).

As observed from the relevant literature survey presented in the next chapter, there is an obvious lack of systematic approach to simplify the problem in a procedure similar to solving engineering problems such as machinery vibration analysis, which involve data acquisition, reduction, analysis, synthesis (base line comparison) and forecast. By employing the wavelet transform on the IV data, by working in the time series data mining framework to implement EA, and by substantiating the use of GA with the discrete stochastic optimization method we will demonstrate a practical way to resolve the IV forecasting problem.

1.6 Engineering Perspective with Wavelet Transforms

In order to deal with the difficulty in forecasting volatility caused by non linearity in the underlying security transactions, some ingenious EA maneuvers could be employed to extract the unknown hidden patterns in the time series. However, it takes much computation time to compile the data each time. Wavelet analysis is a form of nonparametric regression analysis that decomposes a time function into a number of components, each one of which can be associated with a particular scale at a particular time (Hog, 2003). Thus it could be used to analyze time series that contain non-stationary components at many different frequencies (Daubechies, 1990). The proposed methodology uses wavelets packets to decompose the time series into a set of coefficients representing superposed spectral components ranging from low to high scales (frequencies). The wavelet coefficient of each scale can provide clues for patterns of the volatility in the corresponding time horizons. As a result, one could have a better understanding on behavioral patterns of institutional investors versus day traders *i.e.,* long term versus short term transactions. This is particularly important because by employing the wavelet packet transform, analysts could focus on the forecasting time horizons where entropy value is the lowest or activity patterns repeat the most.

The flexibility of selecting different forecasting horizons helps analysts deal with some inherent constraints of IV. For example, any attempt to estimate a time-varying volatility model if there is any, for daily or hourly variations using low-frequency data such as weekly or monthly would not give very meaningful results. This is because any lower frequency data will not capture the volatility clustering that is a characteristic of most financial markets. On the other hand, it is important to be consistent with the horizon of the forecast *i.e.,* to forecast a long-term average volatility it makes little sense to use a high-frequency time-varying volatility model. With wavelet transform, user could easily match the estimation horizon with the forecast horizon by choosing the corresponding node on the tree instead of re-sampling from the original data set. Economy of calculation time could thus be achieved and this factor is critical in many of the financial applications which require massive amounts of data to process and manipulate.

1.7 Finance Perspective with Evolutionary Algorithms

> *According to Chen and Wang(eds.)(2002): "Computation intelligence is a new development paradigm of intelligence systems which has resulted from a synergy between fuzzy sets, artificial neural networks, evolutionary computations, machine learning ...etc., broadening computer science, physics, engineering, mathematics, statistics, psychology, and the social sciences alike."*

Viewed from a traditional perspective, many general soft computing methods in the domain of computation intelligence such as the theory of GA, neural networks and rough sets are currently considered as a sort of exotic methods. They are applied to the evaluation of ordinary share value expectations for a variety of financial purposes, such as portfolio selection and optimization, classification of market states, forecasting of market states and for data mining towards algorithm developments. This is in contrast to the wide spectrum of work done on exotic financial instruments, wherein other advanced mathematics is used to construct financial instruments for hedging risks and for investment (Tay and Cao, 2003).

In business and in most walks of life, goodness is judged as a measure only relative to its competition, while convergence to the absolute best is not necessarily an issue, because we are only concerned about doing better relative to others. The most important practical goal of optimization is just improvement in performance. Meanwhile, designers of artificial systems particularly those dealing on business systems, can only marvel at the robustness, the efficiency and the flexibility of solutions pertaining to biological systems. Therefore, one can see why EA's are becoming ever more popular tools in attacking a wide variety of problems in the business world and the justification for our current approach.

EA's work from a rich database of points simultaneously (a population of strings), climbing many peaks in parallel; thus the probability of finding a false peak is reduced over methods that utilize point – to – point approach. By working from a population of well-adapted diversity instead of a single point, an EA adheres to the old adage that there is strength in numbers or in organized clusters; this parallel flavor contributes mainly to EA's robustness. The four main strategies – direct use of a coding, search from a population, blindness to auxiliary information, and randomized (stochastic) operators (one can use econometric models to guide, such as the clustering effect) – also contribute to an EA's robustness and thus resulting in clear advantages over other more commonly used optimization techniques.

GA, a member of the EA family can optimize a broad class of functions with straightforward binary-coded or real-coded strings, by using the survival of the fittest paradigm. To perform an effective search for better and superior structures, they only require payoff values (biologists call this function the fitness function or a kind of objective function values) associated with individual strings. These characteristics make GA a more canonical method than many search schemes. After all, every search problem has a metric (or metrics) relevant to the search; however, different search problems have vastly different forms of auxiliary information. Simplicity of operation and power of effect are two of the main attractions of the GA approach.

GA builds efficiently new solutions from the best solutions of the previous trials. It ruthlessly exploits the wealth of information by reproducing high-quality notions according to their performance (Lawrenz, 2000) and crossing these notions with many other high-performance notions from other strings (Bauer, 1994). Reproduction is a process in which individual strings are copied according to their objective function values. This function can be some measure of profit, utility, or goodness that we want to maximize. Thus the action of crossover with previous reproduction speculates on new ideas constructed from the high-performance building blocks (notions) of past trials. The notion here is not limited to simple linear combinations of single features or pairs of features. Biologists have long recognized that evolution must efficiently process the epitasis (position-wise non-linearity) that arises in nature. In a similar manner, GA must effectively process notions even when they depend on their component feature in highly nonlinear and complex ways. In other words, GA's are largely unconstrained by the limitations that hamper other methods such as continuity, derivative existence, unimodality and so on (Bauer, 1994). Finally, mutation in GA is the occasional (in other words, at a small probability) random alteration of the value of a string position. By itself, mutation is a random walk through the string space. When used sparingly with reproduction and crossover, it is an insurance policy against premature loss of important notions and to provide exploration capability to the search process. In conclusion, GA can be adopted as a suitable tool used to forecast IV even if non-linearity is to be considered.

GP, also being a member of the EA family, is a tool to search the space of possible programs for an individual case (computer algorithm) that is fit for solving a given task or problem (Koza, 1992). It operates through a simulated evolution process on a population of solution structures that represent candidate solutions in the search space. The evolution occurs through different processes that will be described in the following sections.

A selection mechanism implements a survival of the fittest strategy by employing the genetic crossover and mutation of the selected solutions to produce offspring for the next generation. The generated programs are represented as trees, where nodes define functions with arguments given by the values of the related sub-trees, and where leaf nodes, or terminals, represent task related constants or input variables.

The selection mechanism allows random selection of parent trees for reproduction, with a bias for the trees that represent better solutions. Selected parents are either mutated, or used to generate two new offspring by a crossover operator. Crossovers and mutations are the two basic operators used to evolve a population of trees. The mutation operator creates random changes in a tree by randomly altering certain nodes or sub-trees, whereas the crossover operator is an exchange of sub-trees between two selected parents. The evolution of trees' population continues until a certain stopping criterion is reached. The initial population is composed of random trees, which are generated by randomly picking nodes from a given terminal set and function set. The only constraint is that the generated trees ought not to be too complex. A restriction on the maximum allowable depth or the maximum number of nodes is also frequently imposed for this approach.

Equipped with the EA methods, one could deal with the non-linearity in the process to search for either maxima or minima once the problem is properly formulated. Therefore, EA, incorporating both GA and GP features, becomes the key computational intelligence tool (CI) in constructing the proposed systematic approach in order to forecast IV. Such a judicious combination of best attributes of these powerful methods is attempted in this research thesis.

1.8 Impact of this Research

The main economical benefits resulting from the current research may be realized in four aspects:

a) To determine more accurately the option prices of the underlying securities;

b) To improve the rate of return of related derivative investments;
c) To achieve higher rate of return from direct investment in volatility securities such as futures based on VIX, and lastly
d) To perform risk management more confidently in various fields (Brooks, 2003).

More specifically, financial analysts could better identify buying and selling opportunities in equity options markets, select investment opportunities that offer the greatest risk-reward trade-off and generate specific buy or sell recommendations in selected stock and index options that are consistently profitable regardless of the direction of the overall market. As a result, better portfolio performance could be achieved through optimizing risk management.

1.9 Problem Statement

Different patterns, linear or non-linear, of the volatility time series of a security may repeat at different moments and intervals. This is true when dealing with different types of financial securities or dealing with different historical periods for the same underlying security. By capitalizing on the stylized clustering effect characteristics of financial volatility, we have attempted to formulate and apply different engineering analyses, solution techniques and proven computational transformation principles such as wavelet, the TSDM, Markov chain, GA, and GP methods to establish a fresh systematic approach in order to forecast as many events/non-events as practically feasible in the IV time series in order to optimize risk management (Brooks, 2003) and to guide derivative trading utilizing such tools.

Upon critically studying the background, history and basic concepts and methodologies in the financial analysis field, we have secured a good understanding regarding the necessity, difficulty as well as feasibility for the contemporary researchers to work on volatility forecast using advanced techniques from other disciplines. In the next chapter, a relevant literature survey will help set the basis for the work undertaken here and will guide both theoretical and experimental investigations in order to formulate an effective and efficient methodology that can be developed to make it applicable to forecast volatility of equity indices and potentially other investment vehicles.