ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

A THESIS PRESENTED TO THE
ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

IN PARTIAL FULFILLMENT OF THE THESIS REQUIREMENT
FOR THE DEGREE OF
PHILOSOPHY DOCTOR IN ENGINEERING
Ph.D.

BY
PU YUN MA

A FRESH ENGINEERING APPROACH FOR THE FORECAST OF FINANCIAL INDEX VOLATILITY AND HEDGING STRATEGIES

MONTRÉAL, JULY 28, 2006

© copyright reserved by Pu Yun Ma

THIS THESIS WAS EVALUATED

BY THE COMMITTEE COMPOSED BY:

Dr. Tony Wong, Thesis supervisor
Département de génie de la production automatisée à l'École de technologie supérieure

Dr. Thiagas Sankar, Thesis co-supervisor
Département de génie de la production automatisée à l'École de technologie supérieure

Dr. Edmond T. Miresco, President
Département de génie de la construction à l'École de technologie supérieure

Dr. Bruno de Kelper, Examiner
Département de génie électrique à l'École de technologie supérieure

Dr. Virendra Jha, Examiner
Vice-president, Science, Technology and Programs, Canadian Space Agency

THE THESIS WAS DEFENDED IN FRONT OF THE EXAMINATION

COMMITTEE AND THE PUBLIC

ON JULY 28, 2006

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

UNE NOUVELLE APPROCHE POUR LA PRÉVISION DE LA VOLATILITÉ FINANCIÈRE ET DES STRATÉGIES DE COUVERTURE DE RISQUE

Pu Yun Ma

SOMMAIRE

Cette thèse vise à apporter une nouvelle perspective sur un problème important en ingénierie financière – la prévision de la volatilité des indices boursiers. La volatilité des indices boursiers est une mesure du risque dans le domaine de l'investissement. De plus, la volatilité journalière est un facteur déterminant dans l'évaluation des prix et dans l'application des stratégies de couverture de risque. Or, le succès dans l'estimation et la prévision de la volatilité est encore loin d'un niveau acceptable.

En intégrant d'une façon judicieuse les méthodes analytiques (la transformée par ondelettes, l'optimisation stochastiques basée sur les chaînes de Markov) et les méthodes de l'intelligence computationnelle (les algorithmes évolutionnaires dans un contexte de l'exploration de données), les travaux de cette thèse ont mené à une stratégie systématique permettant la caractérisation et la prévision de la volatilité dont le niveau de succès est supérieur aux méthodes standards de l'industrie.

L'aspect saillant de la stratégie proposée est « la transformation et le traitement séquentiel ». En considérant la variation des indices boursiers comme un signal temporel et stochastique, cette thèse préconise une stratégie en cinq étapes : i) transformation des données variables en volatilité intégrée, c'est-à-dire passage d'un phénomène non observable à un qui est observable; ii) détermination de l'horizon temporel propice à la prévision par la transformée par ondelettes; iii) formulation des coefficients d'ondelettes en ensembles de données récursives; iv) extraction, par algorithmes génétiques, des règles de type « IF – THEN » caractérisant les structures cachées ou inhérentes des données; v) effectuer la prévision de la volatilité à court terme en appliquant la programmation génétique et en utilisant les règles dégagées dans l'étape précédente. Ces étapes et la stratégie proposée sont appuyées sur le cadre théorique de l'optimisation stochastique, ce qui constitue une contribution majeure de cette thèse.

Pour appliquer concrètement cette approche, des données historiques de S&P 100 s'échelonnant sur 5 ans ont été utilisées pour générer les règles de caractérisation. Ces règles sont ensuite appliquées dans la prévision de la volatilité. L'analyse statistique des résultats ont permis de constater un taux de succès moyen supérieur à 75%. Dépassant ainsi la performance des méthodes de prévision publiées dans la littérature. Il s'agit là une autre contribution de ce travail de recherche.

A FRESH ENGINEERING APPROACH FOR THE FORECAST OF FINANCIAL INDEX VOLATILITY AND HEDGING STRATEGIES

Pu Yun Ma

ABSTRACT

This thesis attempts a new light on a problem of importance in Financial Engineering. Volatility is a commonly accepted measure of risk in the investment field. The daily volatility is the determining factor in evaluating option prices and in conducting different hedging strategies. The volatility estimation and forecast are still far from successfully complete for industry acceptance, judged by their generally lower than 50% forecasting accuracy.

By judiciously coordinating the current engineering theory and analytical techniques such as wavelet transform, evolutionary algorithms in a Time Series Data Mining framework, and the Markov chain based discrete stochastic optimization methods, this work formulates a systematic strategy to characterize and forecast crucial as well as critical financial time series. Typical forecast features have been extracted from different index volatility data sets which exhibit abrupt drops, jumps and other embedded non-linear characteristics so that accuracy of forecasting can be markedly improved in comparison with those of the currently prevalent methods adopted in the industry.

The key aspect of the presented approach is "transformation and sequential deployment": *i*) transform the data from being non-observable to observable *i.e.*, from variance into integrated volatility; *ii*) conduct the wavelet transform to determine the optimal forecasting horizon; *iii*) transform the wavelet coefficients into 4-lag recursive data sets or viewed differently as a Markov chain; *iv*) apply certain genetic algorithms to extract a group of rules that characterize different patterns embedded or hidden in the data and attempt to forecast the directions/ranges of the one-step ahead events; and *v*) apply genetic programming to forecast the values of the one-step ahead events. By following such a step by step approach, complicated problems of time series forecasting become less complex and readily resolvable for industry application.

To implement such an approach, the one year, two year and five year S&P100 historical data are used as training sets to derive a group of 100 rules that best describe their respective signal characteristics. These rules are then used to forecast the subsequent out-of-sample time series data. This set of tests produces an average of over 75% of correct forecasting rate that surpasses any other publicly available forecast results on any type of financial indices. Genetic programming was then applied on the out of sample data set to forecast the actual value of the one step-ahead event.

The forecasting accuracy reaches an average of 70%, which is a marked improvement over other current forecasts. To validate the proposed approach, indices of S&P500 as well as S&P100 data are tested with the discrete stochastic optimization method, which is based on Markov chain theory and involves genetic algorithms. Results are further validated by the bootstrapping operation. All these trials showed a good reliability of the proposed methodology in this research work. Finally, the thus established methodology has been shown to have broad applications in option pricing, hedging, risk management, VaR determination, *etc*.

UNE NOUVELLE APPROCHE POUR LA PRÉVISION DE LA VOLATILITÉ FINANCIÈRE ET DES STRATÉGIES DE COUVERTURE DE RISQUE

Pu Yun Ma

RÉSUMÉ

La volatilité d'un actif, défini par l'écart-type du taux de rentabilité, est souvent utilisée pour déterminer le prix du marché. La volatilité quotidienne est une variable cruciale dans l'évaluation du prix des options et dans le choix d'une stratégie de couverture du risque surtout dans le contexte où l'industrie des fonds de placement à risque connaît une forte croissance. L'importance de la volatilité a été dûment reconnue en 1997 et en 2003 par le décernement de prix Nobel aux chercheurs travaillant sur ce sujet. Dans le domaine de l'investissement, on peut négocier la volatilité directement par des contrats à terme ou indirectement en misant sur différents arrangements d'options (différent entre l'offre et la demande, différence entre présent et le futur, etc.). À la différence du taux de rentabilité, la volatilité a été confirmée pour être raisonnablement prévisible (Alexandre, 2001) et est devenue, pour un grand nombre d'analystes, un sujet d'intérêt d'exploration et d'exploitation.

La volatilité ponctuelle est latente tandis que la volatilité intégrée (IV) est une mesure efficace de la volatilité quotidienne de l'actif. Elle est obtenue à partir des transactions de haute fréquence (Andersen, 1998, 2001b) enregistrées. La recherche contemporaine sur le comportement de la volatilité est dominée par des méthodes statistiques paramétriques de type GARCH (Generalized Autoregressive Conditional Heteroscedastic) (Davidson & MacKinnon, 1993). D'une manière générale, les méthodes statistiques utilisent des modèles discrets pour traiter les données historiques du rendement d'un actif. La variance, par exemple, est une mesure convenable de la volatilité d'une telle série temporelle. L'objectif de ces modèles statistiques consiste à obtenir la meilleure correspondance possible entre la série temporelle produite par un modèle et celle des données historiques. Cependant, le pouvoir de prévision résultant n'est pas toujours convaincant (Harvey Dec 1999, Peters 2001, Chong et al. 1999, Park 2002, Brooks 1998, 2003, Christofferson 2004) et les modèles n'ont pas pu démontrer leur efficacité dans un contexte réel. De nos jours, les analystes en investissement ont développé des produits commerciaux qui sont basés sur des méthodes ad-hoc non divulguées. Le mieux connu de ces produits, est le « Volatility Report™ » permet aux investisseurs d'identifier les moments propices à des transactions de la volatilité en se basant sur un indice de volatilité propre et qui mesure la volatilité sous-jacente plus exactement et efficacement que des méthodes traditionnelles.

Selon les développeurs du Volatility Report™, le système est capables de prévoir correctement la direction future de la volatilité à une moyenne de 72 % - 75 % du temps dans l'univers des titres et des actions. Compte tenu que la méthodologie employée par ce système (et d'autres projets de recherche à caractère confidentiel) est du domaine privé, notre objectif de recherche est donc de réaliser une meilleure prévision à l'aide de techniques d'ingénierie connues et de produire ainsi des méthodologies systématiques et reproductibles.

Dans cette recherche, nous établirons des relations entre l'ingénierie et des méthodes du domaine des finances. Nous avons ainsi développé une approche systématique dans la prévision de la volatilité d'un ensemble d'indices boursiers. L'approche préconisée implique les algorithmes évolutionnaires (EA) dans un contexte d'exploration de données temporelles (TSDM). Cette approche est appuyée sur la théorie de l'optimisation stochastique et les chaînes de Markov.

Un concept clef emprunté de l'ingénierie consiste à appliquer des transformations judicieuses des données avant d'effectuer le traitement et l'analyse. Nous avons donc converti les données historiques des indices sous forme de la volatilité intégrée (IV) et puis en coefficients d'ondelette. Par la suite, le signal transformé est traité par l'algorithme génétique (GA) et la programmation génétique (GP) afin d'identifier les structures inhérentes qui sont favorables à la prévision de la volatilité.

L'exploration de données (*Data Mining*) est le terme employé pour représenter le processus d'analyse de données dans le but de découvrir des modèles cachés. Povinelli (1999) a formulé le concept TSDM qui révèle les modèles temporels cachés qui sont caractéristiques d'événements d'une série temporelle. Pour trouver ces modèles, la série de temps est incorporée dans un espace de phase reconstruit avec un retard unitaire et d'une dimension *D*. C'est-à-dire, la série temporelle est formée en créant le chevauchement des observations :

$$\theta = \{\theta_t, t = j, \ldots, j + N\}.$$

La série de prévision est alors formée à partir d'une seule observation :

$$\eta = \{\theta_t, t = j + N + 1\}$$

où θ_j est la valeur de la série au temps $t = j$, alors que N est la taille de la fenêtre d'observation. Les structures temporelles intéressantes sont alors identifiées par la fouille utilisant les techniques GA/GP. Ces structures sont déterminées par les 4 points précédents θ_t de la série IV pour permettre la prédiction de η_t dans le prochain intervalle de temps.

L'optimisation stochastique discrète (DSO) procure un appui théorique à l'approche proposée dans cette recherche. En employant la méthode DSO présentée par Andradottir (1995, 1999), nous démontrons la faisabilité et la convergence d'un GA pour le cas de la non-homogénéité temporelle. Andradottir a proposé d'utiliser toutes les valeurs de fonction d'objectif observées produites comme la trajectoire de la fouille aléatoire autour d'une région faisable pour obtenir des évaluations de plus en plus précises de la fonction d'objectif. À n'importe quel temps donné, la solution faisable qui a la meilleure valeur d'objectif évaluée, par exemple le plus grand pour des problèmes de maximisation est utilisé comme l'évaluation de la solution optimale.

Parallèment, Andradottir a spécifié le taux de convergence de cette méthode et a montré que la convergence est garantie presque assurément à l'ensemble des solutions optimales globales. La convergence numérique présentée par Andradottir (1996b) et par Alrefaei et Andradottir (1996a, c) suggère que cette technique pour évaluer la solution optimale est supérieure en comparaison avec d'autres techniques parue dans la littérature.

La condition d'application de l'analyse DSO est que la série temporelle soit une chaîne de Markov. Cette condition est satisfaite par notre proposition. Rappelons que les données historiques sont d'abord transformée en une série IV puis en une série de coefficients d'ondelette. La fenêtre mouvante de la IV avec $\theta = \{\theta_{t+j}, j = 0, 1, 2, 3\}$ convertira ces coefficients en un ensemble de symboles {1, 2, 3, 4 *} en se basant sur des intervalles tels $\{(-\infty, -a], (-a, b], (b, c], (d, \infty), *\}$, où « * » signifie « peu importe ». Ainsi, les données seront transformées en une série de nombres.

Incidemment, on peut démontrer que ce processus de transformation produit une chaine de Markov. Les règles de traitement utilisées par les algorithmes évolutionnaires ont la forme <IF [(($\theta_t = I$) AND/OR ($\theta_{t+1} = J$) AND/OR (($\theta_{t+2} = K$)) AND/OR ($\theta_{t+3} = L$))], THEN ($\theta_{t+4} = M$)> où la fonction de caractérisation $g(t) = \theta_{t+4}$ aura comme valeur, un entier dans l'intervalle [1, 4] prévoyant la direction et l'étendue de la valeur subséquente de la série IV. Dans chacune des itérations, 100 règles par groupe pour un totale de 100 groupes sont générées d'une façon aléatoire et indépendante. Ainsi, nous pouvons définir une fonction stochastique :

$$X(\theta) = \begin{cases} 1, \text{ if } \{\theta_t, \theta_{t+1}, \theta_{t+2}, \theta_{t+3}, \theta_{t+4}\} \text{ correspond à la séquence de données;} \\ 0, \quad \text{autrement.} \end{cases}$$

où $\{\theta_t, \theta_{t+1}, \theta_{t+2}, \theta_{t+3}, \theta_{t+4}$, AND, OR, $*\}$ représente les règles. Conséquement, $f(\theta)$ dans l'expression suivante:

$$\max_{\theta \in \Theta} f(\theta), \quad \text{where } f(\theta) = E\{X(\theta)\}, \forall \theta \in \Theta.$$

serait l'espérance mathématique de l'exactitude de la prédiction $E\{X(\theta)\}$ pour une règle θ. Après l'optimisation de $f(\theta)$, les 100 meilleures règles seront utilisées dans le processus de la prédiction.

Les données quotidiennes de S&P100 et ensuite S&P500 ont été employées pour évaluer la validité de l'approche proposée. Dans chaque cas, la première partie du jeu de données est utilisée comme le jeu d'apprentissage, tandis que la partie suivante est utilisée pour calculer l'exactitude de prédiction des règles. Dans tous les essais, l'exactitude de prédiction a été supérieure à 60 %, tandis que les résultats de obtenus par application des concepts de la DSO sont supérieurs à 70 %. De plus, la vitesse de traitement a été 10 fois plus rapide que celle d'un algorithme génétique conventionnel en raison de l'utilisation vaste de mémoire pour tenir les règles évaluées durant l'exécution de l'algorithme.

ACKNOWLEDGEMENT

Resuming formal study after a pause of nearly a decade is indeed a challenge beyond my initial anticipation. Throughout the entire process however, I am blessed to attain the generous support from many professors, colleagues, friends, family members and particularly from my thesis co-supervisor Dr. T. S. Sankar who encouraged me to do this Ph.D. at ETS, helped me beyond a professor's capacity to overcome many rough moments, mentored me at those critical stages and provided the financial support to complete the journey. My other thesis co-supervisor Dr. T. Wong guided me through the research especially during the literature publication process and replenished the funding whenever necessary to ensure a smooth voyage. Raymond Sui of Positron Inc. helped me in the complex computer programming to ensure the proposed method is being implemented effectively and efficiently. Dr. H. F. Li of Concordia University provided valuable inputs regarding the methodology developed in the current research.

"The ocean of knowledge has no shore", such a Chinese idiom has duly reflected the pursuance of my Ph.D. voyage during the past few years. My Master Hellyer Chan jump-started the expedition and tirelessly cheerleads along the route. And my wife Grace Yeung strived as usual to keep the boat afloat. If the completion of this phase of study is like a boat reaching a dock at a midway island, it has been an exciting yet rewarding journey that showcased all participants' efforts.

I feel extremely privileged to have this opportunity to associate with such a group of intellectually celebrated and/or practically empowered individuals. The experience will undoubtedly enlighten my future endeavor in pursuance of further knowledge expansion.

TABLE OF CONTENTS

Page

LISTE OF TABLES

LISTE OF FIGURES

LIST OF SYMBOLS

$b_{jo,k}$	Coarse scale coefficients of wavelet transform
$c_{j,k}$	Detailed (fine scale) coefficients of wavelet transform
d	The first-order autocorrelation coefficient between the lagged squared return and the current squared return
$\overline{d}$	Expected dividend
db_i	i^{th} class of the Daubechies wavelet family
δ	The distance between a random point(s) and a selected pattern in TSMD
$E(\sigma)$	Expected volatility
$E(\sigma^2)$	Expected variance
ε_t	IID variables representing the model approximation error or the dynamic noise
ϕ	Scaling function (or farther wavelet) such that its dilates and translates constitute orthogonal bases
ψ	Scaling function (mother wavelet) such that its dilates and translates constitute orthogonal bases
g_l	Unit scale wavelet scaling low-pass filters
$G(f)$	Transfer functions (Fourier Transforms) for the filters $\{g_l\}$
$g(t)$	Event characterizing function in TSDM
$H(f)$	Transfer functions (Fourier Transforms) for the filters $\{h_l\}$
h_l	Unit scale wavelet high- pass filters
$K_m(\Theta)$	The total number of matches between θ and the data sequence
n_{1t} , n_{2t}	Fractions of fundamentalists and technical analysts
$p(t)$	Asset price at time t

p^*	Fundamental price
$p_{bid}(t_i)$ and $p_{ask}(t_i)$	The bid and ask prices of the underlying asset at time t_i.
Θ	Discrete feasible region containing at least two states, *e.g.* θ;
θ_j	One of the data point in the 4-lag recursive data set, which is converted from the IV time series
θ	An IV data set or the 4-lag recursive data set (*e.g.*, θ_t, θ_{t+1}, θ_{t+2}, and θ_{t+3})
P	Vector representing hidden patterns in a time series while conducting TSDM
Q	Dimension of hidden patterns in a time series captured by a window while conducting TSDM
r	Discount rate
R_t or r_t	Rate of return at time t
ρ	The first-order autocorrelation coefficient between lagged returns and current squared returns
s	An intensity factor that optimizes the past successful strategies
σ	Estimated integrated volatility
$U_{h,t-1}$	Performance measure, which is the evolutionary fitness of predictor h in period $t-1$ given by utilities of realized past profits
u_{i-1}^2	The square of the continuously compounded return during day $i-1$
v	Fundamentalists' belief that tomorrow's price will move in the direction of the fundamental price p^* by a factor ν
$\nu(t_i)$	The moment of the rate of return distribution
V	Long term volatility, which is often a constant for a given asset
ω	The integrated volatility data set
x	An individual value in a time series *e.g.*, a homogeneous sequence of logarithmic prices of S&P100 index
$X(\theta)$	Random variable indicating if a complete rule $\{\theta_t, \theta_{t+1}, \theta_{t+2}, \theta_{t+3}, \theta_{t+4}\}$ matches the data sequence
Z_t	Normalization factor such that the fractions add up to one

LIST OF ACRONYMS

ARCH	Autoregressive Conditional Heteroscedasticity
CBOE	Chicago Board Options Exchange
CFE	Chicago Futures Exchange
CI	Computational Intelligence
DGP	Data Generating Process
DSOM	Discrete Stochastic Optimization Method
EA	Evolutionary Algorithm
EMA	Exponential Moving Averages
FX	Foreign Exchange
GA	Genetic Algorithm
GAB	Guggenheim-Anderson-De Boer model
GARCH	Generalized Autoregressive Conditional Heteroscedasity
GP	Genetic Programming
IV	Integrated Volatility
MAE	Mean Absolute Error
MSE	Mean Square Error
NIKKEI 225	Tokyo Stock Exchange Index with 225 selected issues
S&P 100	Standard and Poor 100 Index
S&P 500	Standard and Poor 500 Index
TSDM	Time Series Data Mining

TSDM*e*2	Time Series Data Mining evolving temporal pattern method
TSE	Toronto Stock Exchange
VIX	CBOE Market Volatility Index
VXB	Futures of Volatility Index

INTRODUCTION

Over the years, the field of engineering has evolved so rapidly that many powerful, practical and effective methods have been well established and widely applied to solve a wide array of problems of different scales. However, some complicated and/or difficult problems would demand a systematic approach that synthesizes different methods and creates the synergy in order to overcome seemingly insurmountable obstacles. The subject of Financial Engineering or more specifically Applied Engineering in Derivatives is a relatively young field in both academic and scientific sense, and is full of newly emerged problems with many critically important ones. The following statement exemplifies one such typical problem:

> *"The benchmark model for option pricing is the classic Black-Merton-Scholes (BMS) model published in 1973. This colossal achievement was recognized when two of the three surviving founders of the model were awarded the Nobel Prize in Economics in 1997. Although the BMS model assumes volatility to be constant, its creators knew that volatility is itself volatile -- an observation underlined by the Nobel Prize awarded last year to professor Engle for his pioneering work on modeling volatility dynamics." (Carr, 2004)*

That is essentially the motivating force behind this research thesis to address the important problem of volatility dynamics inherent in financial systems through development of solutions using judicious combination of engineering methodologies. Volatility is a commonly accepted measure of *risk* in the investment field. Volatility of an asset is used to determine the market price for insuring against future events. It is usually defined as the standard deviation of the rate of return distributions, whereas the commonly understood mean of the rate of return distribution is the first statistical moment of the corresponding asset price. In basic finance, *risk* is always associated with rate of return. Higher returns are usually obtained by taking higher risks.

Spot volatility is latent. One can only approximate it by short-windowed standard deviation of the return rate. Barndorff-Nielsen *et al.* (2001) and Meddahi (2002) showed that high frequency realized volatility converges to integrated volatility. Integrated volatility (IV) is an effective measure of the daily volatility of the selected asset and is derived from the high frequency transaction prices (Andersen, 1998, 2001b). In a strict sense, high frequency prices refer to tick-by-tick market prices, but for convenience purpose any intra day data such as minute by minute data are considered high frequency. The daily volatility is a crucial variable in evaluating option prices and in conducting different hedging strategies. It can also be traded indirectly for its own sake by either configuring different option schemes such as spreads and straddles or traded directly through related futures contracts. Unlike rate of returns, volatility has been confirmed to be reasonably predictable (Alexander, 2001) and therefore, is worth while for analysts to explore for prediction techniques. As defined by Anderson *et al.* (1998), IV at time t is expressed by the equation

$$\sigma(t) = \sqrt{\frac{1}{n}\sum_{j=1}^{n} r^2(\Delta t, t_{i-n+j})} \tag{1}$$

where $\sigma(t)$ represents the standard deviation of the time series or the volatility, r is the rate of return, Δt is the time interval of the data in which integration is done, n is the total time length of the integration, and i is the total number of data. Refer to APPENDIX 1 and 2 for further details on this IV representation. The volatility estimation and forecast are still far from successfully complete, judged by their lower than desirable accuracy. It would be appropriate to link some well-established engineering analytical formulations and methods to the field of finance to help solve this critically challenging problem. In this thesis, a general overview is provided about the data mining and forecasting of time series, which include the related concepts, available methods, potential applications and challenges that confront the contemporary researchers. The philosophy adhered in this research is to apply well proven engineering approaches to solve difficult financial problems especially in the derivative asset analysis – a relatively young field. Such an approach happens to have a long history in a closely related field – economics. We believe that not just one but a judicious combination of a multiple of engineering tools could be deployed to help understand, represent and eventually solve the problems in econometrics and computational finance. And this is the objective of this thesis. Financial volatility is inherently a time series, thus potentially suitable for us to use different engineering methods such as those applied in a typical vibration test of some mechanical signals. For exemple:

a) Record the time-varying signal of *g*-force in the field from the corresponding accelerometers;
b) In a lab, such data are reduced to some manageable digital forms which become a time series for further study;
c) Analyzers equipped with different analytical tools such as FFT, Hilbert transform, orbit analysis, wavelet transform, *etc.* are applied to analyze the data to obtain the essential signal information for applications, and
d) By synthesizing the observed patterns, forecast could be made by comparing with the historical baseline.

Based on existing methods, the current research intends to develop and establish a systematic engineering approach to forecast effectively the volatility of selected financial indices *i.e.,* to improve the accuracy of the forecast. The approach that is taken will revolve around evolutionary algorithms (EA) within a time series data mining framework, which is supported by a Markov chain based random search discrete stochastic optimization method. A key concept borrowed from engineering practice is to enforce a proper data transformation prior to any in depth analysis. As such, time series of equity index is first converted into IV and then into a set of wavelet coefficients. By converting the latent variance of a financial index into an effectively observable IV, we arrive at a typical time-series forecasting problem, which could be defined in an appropriate framework. Instead of going the classical econometric route that has been put forward by many, a wavelet transform is used to pre-process the time series of IV data and then an EA approach that comprises genetic algorithms (GA) and/or genetic programming (GP) is applied to explore the repetitive patterns inherent in the dynamic signal for forecasting purpose.

This thesis consists of three parts, Chapters 1 to 3 provide background information regarding volatility forecasting and the methodology that is formulated and adopted, followed by the general plan of deployment of the current research effort. The latter part of Chapter 3 and the body of Chapter 4 lay down the analytical foundation for the subsequent systematic evolutionary approach, whereas the last part describes the computational procedure to implement and to verify the proposed IV-wavelet-EA method including a complete analysis on market generated real data, and then the corresponding results are presented for critical discussion to enable application in practice.

More specifically, Chapter 1 defines the challenge of the problem in a form of answering such basic questions as “what”, “why”, “when”, “where”, and “how”, needed to define the precise scientific framework, whereas Chapter 2 reviews volatility related literature which tried to employ the EA techniques and/or attempted to address the inherent non-linearity problems associated with the volatility time series. Based on this review of the previous works, we try to establish the rationale to take the EA route in approaching the volatility forecasting problem. In Chapter 3 the general methodology to be developed and adopted in this research thesis is sequentially outlined. Here the four-step recursive data conversion is elucidated in the Time-Series-Data-Mining and Evolutionary-Algorithms (TSDM-EA) framework in search of practical means to forecast the value of a time series. Chapter 4 substantiates the GA approach by making use of a Markov chain based discrete time series optimization method. Chapter 5 describes data pre-processing procedure where the equity index data is transformed into IV time series and then wavelet coefficients by applying the wavelet packet technique. It forecasts the one-step-ahead direction and range of the transformed data sets. And it attempts to forecast the one-step-ahead value of an IV series whereas Chapter 6 analyzes the IV series of another index for validation purpose. Results derived from this research effort are interpreted and critically discussed, and then we attempt to conclude what has been achieved during this Ph.D. investigation and recommend some future research directions for further study. Finally, relevant concepts and technical terms related to this research thesis are given in APPENDIX 1 through 6 to enable easy understanding of the financial and some engineering terms often employed in this manuscript as well as in the research solutions proposed here.

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 $\tau = 2$, *i.e.* σ_{t-2} has only 0.81 contribution to the current volatility σ_t if $\lambda = 0.9$. Second, the model only contains one unknown parameter, namely, λ. RiskMetrics actually sets $\lambda = 0.94$ 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.

CHAPITRE 2

RELEVANT LITERATURE SURVEY

As indicated in the previous chapter, the research efforts on the behavior of rate of return of equities or indices, specifically their volatility is recently dominated by the parametric statistical approaches, such as a variety of GARCH models. Contributions of Andersen *et al.* (1998, 2001, 2001b, 2003) gave a jump start to the theoretical and experimental research on the estimation and forecasting of the realized/integrated volatility and produced some decent results. However, there is need for further research to be done in order to improve the forecasting power of their methodology. For example, the one-day ahead forecasting of IV in a 25-day period generally lags the actual data by roughly two days, even though their results are considered to be already far superior to the GARCH results. To improve the forecasting accuracy, some researchers started to venture into different non-traditional exotic techniques however non-systematic the attempt might be at the current time. As indicated by Arino (1996) and Li (2003), the wavelet methodology, though it has not been used extensively, began to play a more important role in economic and/or financial time series studies because of its many unique features. A few researchers do make use of wavelets to analyze individual stock returns, high frequency stock index returns (Arino, 1996) and the foreign exchange rates (Kaboudan, 2005). Among the publications that apply wavelet analysis in either economics or finance applications (Li, 2003), there are still very few publications that deal with integrated/realized volatility regardless of the types of asset (Hog, 2003, Wang *et al.* 2005). As background information, Lee (1998) reviewed many of the applications of wavelets to provide general estimations of the output.

By applying GA, Fong and Szeto (2001) trained a group of rules based on predetermined format in order to extract patterns from a time series data set that is artificially generated with a short memory.

Those trained rules are first tested on one data set and then used to predict another similarly generated data set. With 100 simple 'IF/THEN' rules on a 4-lag recursive memory, the forecasting accuracy is shown to approach 50% ~ 60%. One could expect better result if more elaborate GA strategy is employed. Since Szeto's method was designed to forecast the rate of return of securities, it inherently conflicts with the well documented fact that rates of return usually do not exhibit straightforward autocorrelation (LeBaron, 1992). On the other hand, the corresponding equity, commodity and foreign exchange markets often exhibit volatility clustering, *i.e.* a strong autocorrelation in squared returns thus, the volatility variable. As a result, Szeto's algorithm could be considered as more useful in IV forecasting, and is used here as a basis for the deployment of a pattern recognition tool for the current research problem to find methods to forecast IV in different time horizons. Both wavelet transform and GA are just two of the critical components in the proposed integral approach presented in this thesis. In this chapter, we will concentrate on reviewing only those relevant published papers that attempt to deal firstly with volatility forecasting by applying EA's, and secondly with other type of methods which mainly deal with the non-linearity issues in volatility. Their strengths as well as shortcomings are indicated at the end of each subsection along with a commentary on the concepts that is brought in to play in this investigation. The title of each paper will be used as the heading of each subsection. In the subsequent chapters, other related literature will be further reviewed prior to the detailed discussion in introducing the IV forecasting methodology for the current investigation.

2.1 Volatility Forecasting with Genetic Programming – Key Publications

2.1.1 "Volatility forecasting using genetic programming" (Pictet et al., 2001)

This study uses GP to discover new types of volatility forecasting models for financial time series. The authors improve on the standard GP approach by introducing types in the GP trees, and by optimizing the program constants with a gradient search. These two modifications improve significantly the convergence properties of the algorithm. Moreover, the typing (type-casting) is used to impose a well-defined parity of the solutions so that only meaningful volatility models are built from the price time series. The volatility models are searched with data sampled at hourly frequency, and the optimization criterion is based on the in-sample forecasting quality of the average daily volatility. The heterogeneity of the financial markets is introduced into the models by price change information measured at different frequencies. Finally, the results are compared to standard models like GARCH(1,1). This research was further extended in the paper shown in the next subsection and thus shares the similar shortcomings.

2.1.2 "Genetic programming with syntactic restrictions applied to financial volatility forecasting" (Zumbach et al., 2001)

The use of data at higher frequency opens new avenues in volatility forecasting as the statistical uncertainty is decreased and intra-day effects must be taken into account. Undoubtedly, the use of high frequency data permits better short-term volatility estimations, but also implies more complexity in data treatment and volatility modeling. The main problems can be summarized as a direct question: what are the most important stylized properties that must be taken into account and brought in to play in any new methodology in order to obtain a good volatility forecast? In this referenced study, the authors use GP to discover new types of volatility forecasting models for foreign exchange (FX) rates at hourly frequency.

In the popular application to foreign exchange financial time series, there exists an important exact symmetry induced by the exchange of the two currencies, and this symmetry must be respected by any of the solutions proposed. For this purpose, the research investigators have used a strongly typed GP approach, where the typing system keeps track of the parity of the GP trees. In this way, they have been able to reduce their search space from all possible GP trees down to the subspace of trees that have the proper symmetry. However, it is critical to note that no such properties exist in equity index, so for our current research problem, we would need to look for other properties to be incorporated in the GP process, similar to ones used in the foreign exchange.

From a general viewpoint, volatility forecast is a function-fitting problem, where the realized value for the volatility is the function to be discovered using a causal information set. The time series of the realized volatility is dominated by randomness, and the actual amount of information contained in the information set about the future evolution is rather low. This is measured for example by the lagged correlation for the volatility, which is in the order of 3% to 15%, depending on the actual definition of volatility that is accepted. In short, this essentially means that the volatility forecast becomes a very difficult challenge for GP, and the algorithms need to be very efficient. In order to develop and test special tools, the authors have applied GP to a similar problem, namely the discovery of the transcendental function cosine, using polynomials. As for the volatility, the cosine function obeys a symmetry of parity since $\cos(-x) = \cos(x)$. The study with the cosine makes clear that the conventional GP cannot tackle the volatility forecast problem, and the main difficulty lies in the discovery of good workable values for the constants included in the GP trees. In order to speed up convergence to the optimal values of the constants, they have to use a local search algorithm, like a conjugate gradient. Only when using a mixed algorithm, are they able to obtain good solutions for the cosine problem, and to find volatility forecast that can compete with the standard GARCH(1,1) model. In this work the gradient of the cost function needs to be evaluated numerically, and they recommend use of the Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS) – a one dimension search quasi Newton method, for the local search (Press *et al.,* 1986).

Overall, the remarkable aspect of the result is that the GP solutions are consistently better than the benchmarks, including out-of-sample forecasting. This clearly shows that

GP, with the syntactic restrictions and the local optimization of the constants, can be considered as an efficient tool for discovering new forecasting models, without over fitting the data sets. Thus, we could in the future use such local optimization algorithms to improve the efficiency in predicting IV of equity index. Furthermore, the following facts can be considered for deployment when selecting the initial functions. The exponential moving averages (EMA) function is always used at least once in each tree, and most of the time an EMA operator is the root node. Clearly, a good forecast needs to have enough memory of the past historical behavior, and this is achieved through the use of EMAs. The EMAs have mostly a constant range z, and not a variable range given by the value of a subtree. All the good solutions do have to contain product of returns at different time horizons.

The key shortcoming of this recent research paper is that it did not explicitly account for the ever-present non-linearity that is inevitably encountered in every problem dealing with the analysis leading to the IV estimation and forecasting, and hence hampers the accuracy of the forecasting. Yet non-linearity is known to be inherent in every volatility time sequence, particularly when jumps, discontinuity or other structural changes in data flow arise. Moreover, the referenced paper's methodology focuses mainly on the characteristics of foreign exchanges, while our research focus lies mainly on handling the equity index.

2.1.3 "Using genetic programming to model volatility in financial time series: the cases of Nikkei 225 and S&P 500" (Chen and Yeh, 1997)

This paper proposes a time-variant and non-parametric approach to estimate volatility. This approach is based on the so-called recursive genetic programming (RGP). In simple terms, RGP is a method that applies Koza's basic GP concepts (Koza, 1998) to sliding windows of a time series and generates an improved sequence based on the evaluation of the average fitness. Such an approach can estimate volatility by simultaneously detecting and adapting to structural changes – which can, in a way, deal with non-linearity. Non-linearity is a key concern in estimating and predicting volatility, because high volatility is usually associated with discontinuity, abrupt breaks and other disruptive structural changes. It may be difficult to model the non-linearity, but they may possess certain distinct patterns that can be recognized. The main contribution of this paper is to employ a model-less adaptive approach to detect structural changes. Therefore, when the underlying structure experiences a certain change, RGP can detect it and, in the mean time, generates a population of volatility estimates under the new structure. As a result, one can avoid firstly, reliance on out-of-date knowledge and secondly the problem of overestimation of volatility.

The key shortcomings of this contribution are that the method given does not employ the most up-to-date estimation techniques such as those methods that are now based on the more accurate estimation of current volatility – IV. Please refer to Section 5.1 for more details. Another important shortcoming is that this technique does not forecast the future outcomes, which is important for equity index applications.

2.1.4 "Predicting exchange rate volatility: genetic programming vs. GARCH and RiskMetrics" (Neely and Weller, 2001)

It has been well-established that the volatility of asset prices displays considerable persistence. That is, large movements in prices tend to be followed by more large moves, producing positive serial correlation in squared returns. Because of this characteristic, the current and past volatility can then be used to predict future volatility. This fact is important to both financial market practitioners and regulators.

This article investigates the performance of a GP applied to the problem of forecasting volatility in the foreign exchange market. GP application here is a computer search and problem-solving methodology that can be adapted for use in non-parametric estimation.

It has been shown to detect patterns in the conditional mean of foreign exchange and equity returns that are not accounted for by standard statistical models (Neely and Weller 2001). This suggests that a GP may also be considered as a powerful tool for generating predictions of asset price volatility for developing our solutions in this thesis. The authors compare the performance of a GP in forecasting daily exchange rate volatility for the dollar-deutschemark and dollar-yen exchange rates with that of a GARCH(1, 1) model and a related RiskMetrics volatility forecast. These models are widely used both by academics and industry practitioners and thus are good benchmarks to which to compare the GP forecasts. While the overall forecast performance of the two methods is broadly similar, on some dimensions the GP produces significantly superior results. This is an encouraging finding, and suggests that more detailed investigation of this methodology applied to volatility forecasting would be warranted.

A core component of the RiskMetrics system is a statistical model — a member of the large ARCH/GARCH family as described in Section 1.2 and 1.3 — that forecasts volatility. Such 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. The key problems in this paper are that the functional forms do not explicitly account for non-linearity and structural change of the time series and IV has not been used. This drawback will make on a universal basis, its volatility applications less attractive.

2.1.5 "Extended daily exchange rates forecasts using wavelet temporal resolutions" (Kaboudan, 2005)

In this research investigation, by employing the natural computational intelligence (CI) strategies such as GP and Artificial Neural Networks (ANN), three exchange rate series were fitted and trained for the forecasts of one-step as well as 16-step-ahead exchange rates. Results seem to be generally more successful than the random walk predictions.

As concluded by the author, not all exchange rate series could be forecasted by this or any other method, and this conclusion agrees with the established efficient market theory. The work cited is among the first few that attempted to incorporate wavelet and CI methods such as GP to approach systematically the time series forecast problems. However, in using the Haar wavelet instead of Daubechies transformation, flexibility of finding the most suitable shape to match the data set is forfeited. Further work, therefore, is needed to extend the method to forecast volatility series.

2.1.6 "Forecasting high-Frequency financial data volatility via nonparametric algorithms -evidence from Taiwan financial market" (Lee, 2005)

This paper uses two CI algorithms ANN and GP for forecasting financial data volatilities of four residents in the Taiwan Stock Exchange (TAIEX) at high frequency with different horizons and compares the output sample forecasting performances with certain parametric volatility models such as HISVOL, GARCH(1,1), EGRACH(1,1) and GJRGARCH(1,1). Their results reveal that algorithms based on nonparametric CI are powerful for modeling high-frequency intraday financial data volatility and establishes the fact that this is the way for future research.

To compare the forecasting performance, several well known parametric volatility models, *e.g.* HISVOL, EGARCH, GARCH and GJR-GARCH model are used. The reason of taking parametric models as benchmark is that parametric model such as GARCH-type are easy to estimate and readily interpretable. It is also the least complex model that describes volatility clustering. Furthermore, it is widely used by both academic researchers and market practitioners and proves to be a good benchmark volatility model in majority of contemporary research works.

To sum up, the authors cited here find GP volatility forecasting model is clearly superior to all other models when judged by kernel density plot in most of different time horizons. One potential drawback of the approach is the calculation efficiency. The problem would become more acute if optimization is needed to determine the best forecasting horizons. The approach would be limited for medium to longer term applications and not for, say, day traders, if hours of CPU time in typical PC's are a constraint to complete one set of forecast.

2.2 Volatility Forecasting with Genetic Algorithms – Key Publications

2.2.1 "Explaining exchange rate volatility with a genetic algorithm" (Lawrenz and Westerhoff, 2000)

Traders evaluate and update their mix of rules from time to time. To be able to understand the dynamics, the authors referred here concentrate only on a limited number of trading rules. The agents are assumed to have the choice, between three technical and three fundamental trading rules. The selection process is modeled by GA's, which has proven to be a useful tool for describing learning behavior in a variety of earlier papers (Dawid 1999).

The chartists-fundamentalists approach is another research direction, which focuses on explaining speculative transactions. Of crucial importance in this class of models is the behavior of the so-called chartists and fundamentalists because the interaction between the two groups has the potential of generating the interesting non-linear dynamics into the problem at hand. Chartists are those who base their investment decisions purely on analyzing historical price data, while fundamentalists focus on the fundamental aspects of the underlying assets, *e.g.* company's cash flow, equity's price to earning ratio, *etc.* By analyzing only two groups, strong structural dynamic relations among the variables still remain. To overcome this problem the models have to introduce some stochastic features. More recently, some multi-agent models in the spirit of the chartists-fundamentalists models have emerged. See LeBaron (2000) for a full survey. Since these models allow for many interacting heterogeneous agents, the structure in the data declines endogenously.

The aim of the cited paper is to develop a realistic, yet simple exchange rate model to get a deeper understanding of the driving forces underlying the foreign exchange dynamics. Rather than deriving the results from a well-defined utility maximization problem, details from the market microstructure and psychological evidence are used to motivate the framework. They construct a model where heterogeneous boundedly rational market participants rely on a mix of technical and fundamental trading rules. The rules are applied according to a clever weighting scheme. Traders evaluate and update their mix of rules from time to time. To be able to understand the dynamics they concentrate only on a limited number of trading rules. The agents have the choice between three technical and three fundamental trading rules. The selection process is modeled by a GA, which has proven to be a useful tool for describing learning behavior in a variety of papers. For an overview, please refer to Dawid (1999). Thus, the authors here derive the dynamics endogenously from learning processes on individual level rather than imposing random disturbances.

In case one wants to derive a model that includes some driving factors that underlie the system of volatility, this could be a good starting point. In other words, this is a traditional approach such that a model is built to explain a phenomenon and to forecast the future event based on the built model. However, market driving forces are more than two groups of investors. Moreover, chartists and fundamentalists behave in ways that are far more complex than three trading rules can explain. As a result, this approach suffers a bias problem just like those parametric models.

2.2.2 "Investment analytics volatility report" (Kinlay et al. 2001)

A proprietary asset allocation and optimization model reported in this work enables the construction of a portfolio comprising different weightings in each long or short option positions that optimize the investment objectives. These weightings are computed using a **proprietary non-linear GA**. Not much public information is available regarding the internal composition of this GA method. In simplified terms, the authors of the paper find, from empirical data, that realized variances tend to be log-normally distributed and asset returns standardized by realized standard deviations tend to be normally distributed. This suggests that a lognormal-normal mixture may be considered as a good model for asset return forecasts. Based on such a model, the authors could estimate and forecast the volatility values that are crucial in calculating the price of the index options. As a result, one can be more confident about the strategies designed to capitalize on the corresponding option trading.

In essence, the report referred here may be characterized by the following investment strategy:

a) Asset class: Equity options.
b) Strategy: Volatility arbitrage.
c) Methodology: Statistical modeling.
d) Style: Market neutral.

Equity options are among the most popular derivative investment vehicles and provide investors with high level of liquidity. As indicated in Section 1.1, volatility arbitrage is more promising than concentrating on the rate of return investments, because of the more predictable nature of volatility. Moreover, market neutral investment strategies enable investors to make profits without taking significant directional risks, *i.e.* less relying on the forecast of the direction of the market. Up to now, forecasting the movement of the market itself is far more difficult than forecasting that of volatility of the market. Kinlay *et al's* proprietary asset allocation and optimization model enables the construction of a portfolio comprising different weightings in each long or short option positions that optimize the investment objectives.

As a result, the subject matter of this research dissemination is exactly the area where our research should target. The results reported in this reference are encouraging as it claims up to 75% of accuracy in forecasting the future direction of volatility of the stock and equity indices. The fact that their methodology is completely proprietary makes our research effort on this thesis subject even more compelling.

2.2.3 "Rules extraction in short memory time series using genetic algorithms" (Fong and Szeto, 2001)

On an artificially generated time series data with short memory, rules are extracted and tested in order to form a basis for the predictions on the test set. A simple GA based on a fixed format of rules is introduced to do the forecasting. The results are markedly improved over those derived based on the traditional approaches, *e.g.* random walk and random guess.

There are different methods to estimate and forecast volatility, as summarized in the current work. However, the methodology demonstrated in this paper, *i.e.* attaining a forecasting accuracy of 50% – 60% by using 100 simple IF/THEN rules on a 4-lag recursive memory (sliding window), seems to be more straightforward and feasible. Therefore, it would be interesting to build upon this method to analyze some time series data in the real world. This approach could thus be considered in our research when applying the GA to find the rules that can best forecast IV in different time horizons.

2.3 Volatility Forecasting with other Methods – Key Publications

2.3.1 "Nonlinear features of realized FX volatility" (Maheu, 1999)

In order to investigate the time series properties of the FX (foreign exchange) volatility, this paper implements a non-parametric measure of daily volatility that is estimated by using the sum of intra-day squared returns. The specification of a functional relationship between this estimate of *ex-post* (realized) volatility and a latent data generating process (DGP) for daily volatility, allows the latter to be parameterized in terms of realized volatility, other variables in the information set, and an error term. The author explores nonlinear departures from a linear specification using a doubly stochastic process under duration-dependent mixing. For example, the author evaluates the importance of time varying parameters and persistence. Furthermore, the author also finds the structure that parameterizes the conditional variance of volatility and can capture large abrupt changes in the level of volatility. The importance of non-linear effects in volatility is gauged by in-sample statistical tests and by out-of-sample forecasts. The volatility forecasts are also evaluated using a simulated trading exercise involving FX straddles. These results have implications for forecast precision, hedging, and pricing of derivatives. The author indicates that stochastic jumps in the conditional mean of the price process have a long tradition in the finance literature but that further research trying to establish the presence of jumps in the conditional variance was just getting underway (Maheu, 1999). This paper states that it intends to solve a similar problem as ours here, but with different type of approaches. Since it still employed stochastic analysis, the question of "which type of distribution should be used?" still remains. Therefore accuracy of their forecasts would become questionable. However, results of this paper could be used to compare with results from any subsequent research such as ours, for establishing thresholds of accuracy of a prediction.

2.3.2 "A Nonlinear structural model for volatility clustering", (Gaunersdorfer, 2000)

Gaunersdorfer's paper referred here attempts to explain the source of the volatility clustering, *i.e.* the interaction between two types of traders: fundamentalists and technical analysts. The beliefs of these two types of traders are driven by an adaptive, evolutionary dynamics according to the reported success of the prediction strategies in the recent past, conditional upon price deviations from the rational expected fundamental price. Such beliefs cause asset price to switch irregularly between *a*) the fundamental price fluctuations with low volatility and *b*) persistent deviations from fundamentals, which are triggered by technical trading, and thus create higher volatility.

The key feature of the nonlinear structural model is therefore, given by the following expression:

$$(1+r)p_t = n_{1t}\left(p^* + \nu\left(p_{t-1} - p^*\right)\right) + n_{2t}\left(p_{t-1} + s\left(p_{t-1} - p_{t-2}\right)\right) + \bar{d} + \varepsilon_t, \qquad (4)$$

where $p^* = \bar{d}/r$ with r being the discount rate and $\bar{d}$ the expected dividend, whereas p_t is the asset price at time t, n_{1t}, n_{2t} are the fractions corresponding to fundamentalists and technical analysts, ν is fundamentalists' belief that tomorrow's price will move in the direction of the fundamental price p^* by a factor ν, s is an intensity factor that is specified for optimizing the past successful strategies, and ε_t is some IID variables representing the model approximation error or the dynamic noise. The modeling of n_{1t} and n_{2t} is where the non-linearity is introduced, *i.e.* they are estimated as $\tilde{n}_{h,t}(h,t)$ given by the relation:

$$\tilde{n}_{h,t}(h,t) = \exp\left(\beta U_{h,t-1}\right)/Z_t, \quad h \in \{1,2\}; \qquad (5)$$

where β is the intensity of choice, measuring how fast the mass of traders will switch to the optimal prediction strategy. $U_{h,\ t-1}$ is a performance measure, which is the evolutionary fitness of predictor h in period $t-1$ given by utilities of realized past profits. Z_t is a normalization factor such that the fractions do add up to unity or one.

This paper goes on to use the model to calculate the autocorrelations of the rate of returns, absolute rate of returns, return squares for an assumed asset to simulate the S&P 500 index during the last 40 years. It concludes that the rate of returns have the non predictability just like the real world data, while the absolute rate of return and return square (*i.e.* variance) do show to follow the trend closely with the index. However, the paper does not mention about the model's forecasting capability and did not intent to deal with it in the future. Moreover, the parametric nature of the model deems to exhibit a lack of flexibility and is difficult to test for validity. Please refer to APPENDIX 2 for more discussion about the parametric model of volatility forecasting.

2.3.3 "Forecasting and trading currency volatility: An application of recurrent neural regression and model combination" (Dunis, 2002)

This paper uses non-parametric Neural Network Regression (NNR) and Recurrent Neural Network (RNN) regression models to forecast and trade FX. Results based on this NNR model depend crucially on the choice of the number of hidden layers, the number of nodes and the type of non-linear transfer function retained. The use of NNR models enlarges the analysts' toolbox of available techniques by adding convenient models where no specific functional form is *a priori* assumed. RNN models are different from NNR models in that they include a loop back from one layer, either the output or the intermediate layer, to the input layer. The following factors have been applied to explain the exchange rate volatility: exchange rate volatilities (including the one to be modeled), the evolution of important stock and commodity prices, and the evolution of the yield curve as a measure of macro-economic and monetary policy expectations. Allowing for transaction costs, most of the trading strategies that are constructed based on the two neural network models produced positive returns. In other words, the volatility forecasting accuracy is in average, slightly over 50%. One might argue that a more elaborate model could produce better results. The contrary has been demonstrated by the authors when they combined the NNR with RNN models. Other recent work in the NNR field might shed more light in this regard. However, it is out of the scope of the current research, as we propose to concentrate mainly on analytical techniques in the EA area.

2.3.4 "An Empirical approach toward realistic modeling of capital market volatility" (Wang et al. 2005)

The authors referred here find that an all-over-the-time stationary generalized constant elasticity of variance (CEV) model will mismatch the mean reverting level θ as well as ignore the jump phenomenon. So they propose a jump-decaying CEV model to depict the realized IV process. The CEV model assumes the volatility process with a mean reverting level. Wavelet technique is introduced to verify that the volatility jumps are not natural to CEV modeling. Therefore, they find that the volatility series is essentially a stochastic jump-decay process rather than being all-over-the-time stationary in each market. These findings strengthen the theoretical foundation of our non-linear evolutionary approach towards IV forecasting.

2.3.5 "Detection and prediction of relative clustered volatility in financial markets" (Hovsepian *et al.* 2005)

This paper presents a methodology for detection and prediction of periods of relatively increased volatility in the time series data. It uses a synthesis of three concepts and methods derived from the field of computer science – *support vector classifiers* (SVC), statistics – GARCH, and signal analysis – *periodogram*. However, it still is GARCH based and only predicts a period of either volatile or non-volatile without specifying how volatile the future reading is. Moreover, it is verified with only simulated real-time cases and hence does not give the reliability expected for market applications.

2.3.6 "Volatility forecasting with sparse Bayesian kernel models" (Tino, *et al.* 2005)

These authors find that models built on quantized sequences *i.e.*, symbols such as signs + and - representing the sign of daily volatility differences, give superior results when compared to those constructed on the original real-valued time series. More specifically, their paper finds that quantization technique significantly improves the overall profit and quantization into just two symbols gives the best results. By trading option straddles on indices and achieving positive profit daily, it confirms the benefit of converting the real value time series into quantized sequence. It also finds that volatility patterns in durations of five and ten days could be detected and used to forecast the one-day-ahead volatility. The authors claim to achieve profitability by applying sparse Bayesian Kernel models based on neural networks. However, it works on daily volatility instead of IV, which has been demonstrated to be less accurate and appropriate. Moreover, this method is based on the premises that the price of the option straddle is strictly positively proportional to volatility. In reality many factors could affect their relationship. In cases where volatility switches signs, *e.g.* slightly negative to slightly positive, option prices will not change substantially and return of buying a straddle option could be negative after deducting the transaction cost. This is the other disadvantage of using just the sign instead of interval to classify the data.

2.3.7 "Boosting frameworks in financial applications: from volatility forecasting to portfolio strategy optimization" (Gavrishchaka1, 2005)

Based on several past experiences and trials, it is believed that accurate volatility modeling does not always warranty optimal decision making that leads to acceptable performance of a portfolio strategy. In this work, a boosting-based framework for a direct trading strategy and portfolio optimization has been introduced to strengthen such optimization procedure.

The author attempts to construct a model from the well-known market indicators for short-to-medium time horizons (from several months to 1-2 years). By calculating strategy returns on a series of intervals of length τ shifted with a step $\Delta\tau$ and encoding them as +1 (for $r >= r_c$) and -1 (for $r < r_c$), it is shown that one can obtain symbolically an encoded time series (distribution) of strategy returns. This is an incremental search algorithm without the aid of computation accelerating tools such as EA. Therefore, its efficiency is questionable. We can only conclude that it would be interesting to combine this approach with GP, in which different formulae could be evolved to find the ones best fit the time series. Moreover, this method could only distinguish positive from negative volatility, which has limitations similar to what was described earlier in section 2.3.6.

2.4 Summary and Discussion

We begin this subsection by presenting a table to summarize the important contributions and analytical characteristics of the reviewed papers discussed in the sections preceding this page. The information contained in the following table will provide us with important guidelines to further our discussion on the approach that we plan to take in this thesis research.

Based on the following summarized literature survey, one can only conclude that the contemporary research of volatility forecast has just begun to venture into the non traditional domain particularly in the CI area such as GA/GP in an attempt to seek better solutions on a backdrop of active IV research; each approach has its own advantages and weakness. IV provides a good starting estimation of the current volatility, and as indicated in previous sections, we as researchers could certainly apply a variety of powerful techniques including stochastic analysis to forecast future volatility more accurately. GA/GP on the other hand, could deal with non-linearity in an effective and progressively efficient manner, which opens up alternative application avenues besides the rigorous exercise in the traditional academic sense.

Table I

Summary of the papers reviewed in details

Author(s)	Approach	Goal	Comments
Pictet et al., 2001	GP	Discover new FX volatility models using "typed" GP trees.	Does not take into account non-linearity. Use FX symmetry to reduce considerably the search space.
Zumbach *et al.*, 2001	GP+LS	Use hybrid GP to forecast FX volatility.	Does not take into account non-linearity. Use FX symmetry to reduce considerably the search space.
Chen & Yeh, 1997	GP	Use a recursive GP to detect and adapt to structural changes of market volatility.	Explicit recognition of non-linearity but does not attempt to forecast. Integrated Volatility was not used.
Neely & Weller, 2001	GARCH, GP, RiskMetrics	Compare three approaches: Parametric, generalized parametric and non-parametric in FX forecasting.	In many instances, GP out-performed the other approaches. They were tested on FX volatility only. Non-linearity is not accounted for. No IV.
Kaboudan, 2005	GP, wavelet, NN	Apply an integrated approach to forecast one-step as well as 16-step-ahead exchange rate forecasting.	Does not deal with volatility. Other type of wavelet might improve the effectiveness.
Lee, 2005	ANN + GP, GARCH	Compare the computation intelligence method with GARCH models.	Better results are achieved at questionable calculation efficiency for medium to long forecasting horizons.
Lawrenz & Westerhoff, 2001	GA	Explore how trading rules can explain market volatility. Use GA to combine six simple trading rules using the chartists – fundamentalists point of view.	In real world there are more than two players (*i.e.* chartists and fundamentalists) and trading rules are much more complex.
Kinlay *et al.*, 2001	GA, SM	Asset allocation and optimization system based on a weighted sum technique. The weights are determined by statistical inference and aided by a GA.	Proprietary techniques with many undisclosed details. Best published results with 72% – 75% prediction accuracy.
Fong & Szeto, 2001	GA	Use GA to determine simple if – then – else rules in order to predict the behaviour of artificially generated time series.	Obtained 50% - 60% accuracy using only 100 simple if – then – else rules. Demonstrated the search power of GA applied to stochastic series.

Table I (continued)

Summary of the papers reviewed in details

Maheu, 1999	SM	Explores the nonlinear features of FX integrated volatility.	Found that stochastic jumps (structural changes) are a very determinant feature in IV.
Gaunersdorfer, 2000	SM	Attempt to define a nonlinear model that explains the volatility clustering phenomenon.	It concludes that the rate of return have non predictable behaviour while the variance does show trend that is close to the index measured. Thus confirming the usefulness of the integrated volatility approach.
Dunis & Huang, 2002	NN	Applied a non-linear non-parametric approach to forecast and trade FX	Achieved slightly above 50% of forecast accuracy. But elaborate models produced poor results.
Wang *et al.* 2005	CEV	Account for the non-linearity in volatility with a stochastic jump-decay process	Provides further theoretical foundation for the current research to deal with non-linearity.
Hovspian, *et al.* 2005	SVC, periodogram, GARCH	Detect and predict periods of relatively increased volatility by a synthesizing method.	It is still a GARCH based approach, *i.e.* parametric models and lacks verification with real data sets.
Tino *et al.* 2005	Sparse Bayesian Kernel	By quantizing real value time series, forecast the one-day-ahead volatility to generate profit.	Only forecasts the directions of volatility.
Gavrishchaka, 2005	Boosting framework	Make optimal investment decisions by forecasting the directions of volatility.	Other calculation methods could help improve efficiency. Only forecasts the directions of volatility.

GA – Genetic Algorithm, GP(+LS) – Genetic Programming (with local search), SM – Statistical Methods, NN – Neural Network, CEV – Constant Elasticity of Variance, SVC – Support Vector Classifier.

CHAPITRE 3

PROPOSED RESEARCH METHODOLOGY

As indicated in the earlier sections, in deriving IV the daily volatility is converted from a latent variable into an observable one. Traditional financial analytical methods based on parametric models such as the GARCH model family, have been shown to have difficulty to improve the accuracy in volatility forecasting due to their rigid as well as linear structure (Harvey, 1999, Christoffersen *et al.*, 2004). The requirement of a particular distribution assumption further hinders the accuracy of forecasting (Andersen *et al.*, 2001). At the current time, there is just very rare number of publicly available, effective and systematical analytical method in the open literature to deal with the non-linearity inherent in the volatility series of financial indices (Kinlay, 2001, Christoffersen *et al.,* 2000, Kaboudan, 2005, Lee, 2005, Gavrishchaka, 2005). In this regard, we believe that the recent development in financial time series analysis could be beneficial to the forecasting problem, which is the focus of this thesis. Work conducted in (Ma *et al.*, 2004b) has clearly established that by using a GA method, the one-step-ahead moving direction/range of volatility of selected underlying securities could be forecasted at an average accuracy of up to 75%. See references cited in APPENDIX 7 of our own publication list about this work.

3.1 Research Objectives

Given the importance and market opportunity associated with the volatility forecasting, the research efforts covered in this thesis intend to establish a fresh systematic approach and eventually propose an analysis based software tool for financial analysts to forecast more precisely the direction, range and eventually real value of future volatility of financial indices as well as different equities.

In other words, this volatility forecasting method should be capable of dealing with complex signal attributes such as non-linearity and also possess the following advantages, *i.e.* to be:

a) assumption free – no need to assume normal or any other statistical distributions associated with the time series and its estimation errors;
b) more flexible – not limited by the parametric structure;
c) more accurate on the current and hence the forecasted volatility – the conversion of IV transforms volatility from a latent variable into an observable variable;
d) more efficient – data pre-processed by means of wavelet transform.

More specifically, the goals to be attempted would be that

a) to establish, from an EA perspective the rationale for the conversion of a typical time series into a four-lag recursive data set by employing a recently available data mining technique;
b) to establish the rationale for the use of GA in forecasting time series by applying a Markov chain based discrete stochastic optimization method;
c) by using wavelet transform to verify the hypothesis that volatility clustering also occurs at the scale levels;
d) to verify that wavelet transform can improve the efficiency of GA/GP in search of different patterns hidden in noise;
e) to verify that EA can effectively simulate/forecast the volatility time series by minimizing the inaccuracy caused by non-linearity;
f) to demonstrate that the current IV-wavelet-EA method is at least as accurate as the proprietary approach shown in Section 2.2.2 in the volatility forecasting.

3.2 Research Plan and Procedure

As the volatility time series being pre-processed into IV and then applying the wavelet transformation, by treating the wavelet coefficients as a 4-lag recursive data set, we can then apply simple IF/THEN rules with GA's in order to capture the typical patterns most frequently found in the resulting data set. Since wavelet coefficients are far smaller in size compared with the original time series, certain degree of calculation efficiency could be achieved. A hundred rules in the wavelet domain would cover a much broader range than the same number of rules in the time domain alone. Moreover, regularities, if any in the frequency domain could be detected, which adds considerably to the prediction power.

To account for non-linearity in the time series *i.e.,* to represent any abrupt changes or discontinuities, appropriate non-linear functions can be introduced when applying GP to forecast the one-step-ahead values of volatility. Thus, the general strategy is to judiciously use a combined EA approach to obtain a better forecasting volatility model by incorporating non-linearity and abrupt structural changes in the time series. The block diagram in Fig. 1 summarizes the proposed forecasting architecture to solve the volatility problem.

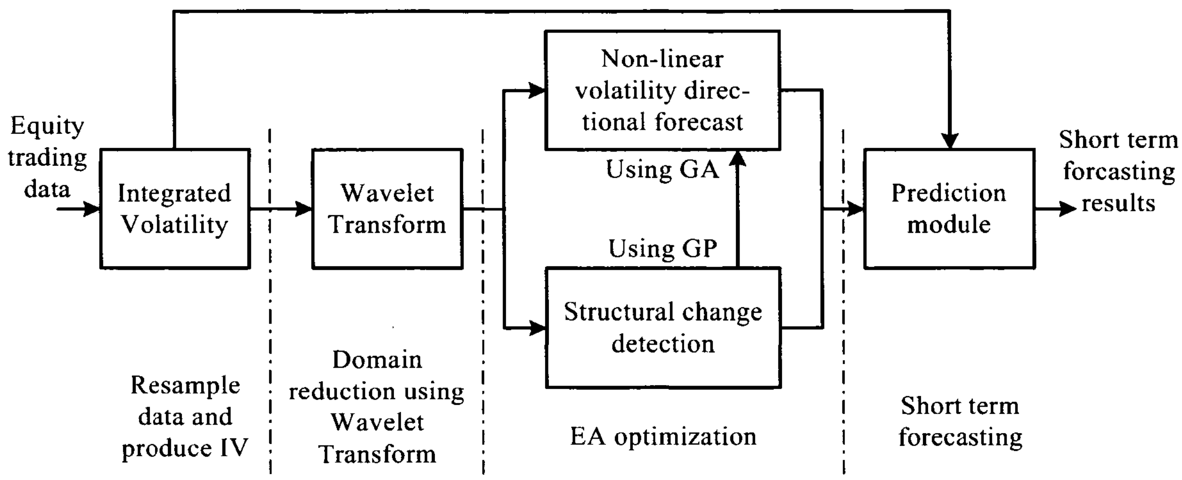

Figure 1 Proposed volatility directional forecasting system architecture

More specifically, the overall methodology for this thesis research can be represented in the form of Fig. 2:

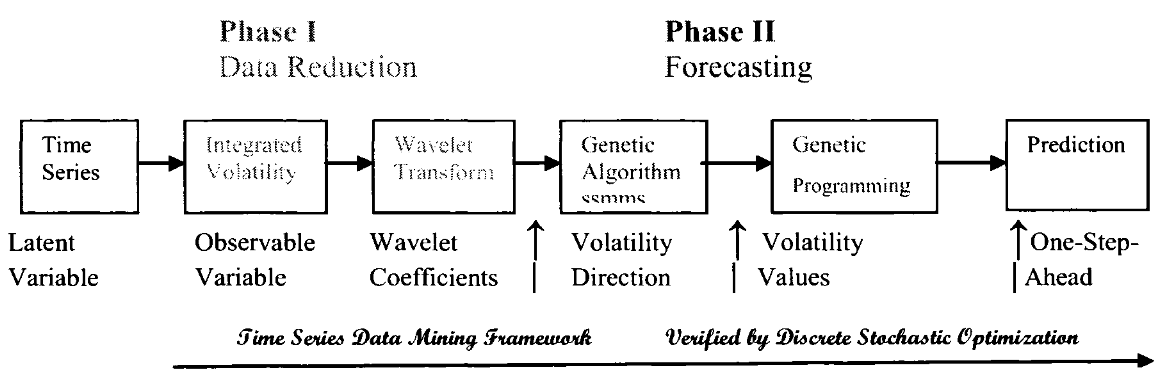

Figure 2 Research plan for this thesis work

The above figure characterizes the rationale of the method implemented throughout the current research work *i.e.*, the application of a combination of data reduction and analysis techniques which converts a difficult volatility forecasting problem into a classical signal analysis problem. And this represents one of the main contributions made in this thesis research effort.

It is apparent that most of the contemporary researchers use EA in time series analysis simply as a tool without much rigorous analytical substantiation. EA, be it GA or GP, typically lacks rigorous mathematical proof and justification even though it is a powerful tool for optimization applications. In order to lay the foundation for applying EA to construct the IV-wavelet-EA time series forecasting approach, the IV time series is first converted into a four-lag recursive series in the TSDM framework. The resultant series is in fact a Markov chain and could be analyzed with a discrete stochastic optimization method in conjunction with GA. Since GP is introduced in the TSDM framework, its operation is better illustrated, whereas its use to forecast IV is also built on a more solid analytical foundation.

The S&P100 and S&P500 indices are selected here for the investigation, not only because of the availability of VIX, hence the convenience of a direct reference, but also because of the popularity of equity and the associated option applications which affords the ease of and usefulness in future portfolio design and trading (Kinlay, 2005). One of the unique characteristics of equity trading is that they are traded only during the prescribed working hours of the working days. An alternative would be to use S&P100 futures (VXB), which is traded both day and night. Martens (2002) actually found that volatility models perform better when index futures instead of indices themselves are used due to the relatively more continuous nature of the data set. However, in this research work, we are more interested in finding ways for dealing with jumps and drops in time series. As a result, it is preferable to use the daily estimation of IV based on indices themselves. In other words, we will calculate IV based on the intra-day data and then use wavelet transform and the EA approach to conduct forecast estimates for the next time step. In the future, Barndorff-Nielsen's working paper (2004) could help select sampling frequencies to avoid market microstructure problems. The research covered in this thesis intends to forecast not only directions and ranges, but as well as values of the volatility movement. For validation purposes, the S&P500 index is obtained from a different source and analyzed using a similar approach. The results are compared with those derived with the S&P100 index and conclusions are then drawn for suitability of application of the developed methods.

3.3 Data Mining of the IV Time Series

Data mining is the term often employed to represent the process of analysis of data with the goal of uncovering hidden patterns especially those complex relationships in large data sets (Povinelli, 1999). Weiss *et al.* (2004) define data mining as *"the search for valuable information in large volumes of data. Predictive data mining is a search for very strong patterns in big data that can generalize to accurate future decisions."* Similarly, Cabena *et al.* (1994) define it as *"the process of extracting previously unknown, valid, and actionable information from large databases and then using the information to make crucial business decisions."* Data mining evolves from several scientific and computational fields, including machine learning, statistics, and database design. It uses techniques such as clustering, association rules, visualization, decision trees, nonlinear regression, and probabilistic graphical dependency models to identify novel, hidden, and useful structures in large databases (Weiss *et al.,* 2004).

The approach adopted in contemporary literature often requires *a priori* knowledge of the types of structures or temporal patterns to be discovered and represents these temporal patterns as a set of templates. Their use of predefined templates completely prevents the achievement of the basic data mining goal of discovering useful, novel, and hidden temporal patterns. Prompted by shortcomings of parametric models, new class of methods has been created so that they do not rely on pre-assumed models but instead try to uncover/induce the model, or a process of computing values from vast quantities of historic data. Many of them utilize learning methods of CI. Non-parametric approaches are particularly useful when parametric solution either lead to bias, or are too complex to use, or do not exist at all. Many recent publications attempted to use GA's and/or GP's to forecast financial time series such as stocks, indices and options. Refer to Table 2.1 for a summary of important related works (Chen and Lee, 1998, Goldber, 1998, Chen, 2003, Iba, 1999).

In search of a relatively more general approach for data mining of financial time series, Szpiro's (1997) method permits the discovery of equations of the data-generating process in a *symbolic* form. The GA that is described there uses parts of equations as building blocks to breed ever better formulae. Apart from furnishing a deeper understanding of the dynamics of a process, the method also permits global predictions and forecasts.

Povinelli and Feng (1999) took the subject one step further by introducing the TSDM framework, which differs fundamentally from most of the approaches mentioned above. Povinelli (1999) formulated the TSDM framework that reveals hidden temporal patterns that are characteristic and predictive of time series events. This contrasts with other time series analysis techniques, which characterize and predict all observations.

From a critical review of the above literature, instead of using just one single formula to represent the entire time series, a better idea would be to use multiple GA's and GP's exploring sequentially in the search space to obtain an overall estimation represented by a set of formula rules. For example, the GP's try to find the best fitting rules based on the input time series. The best formula rules are then combined as suggested by Chan and Stolfo (1996) and later used to forecast the future IV values. The same can be applied to rules in the GA approach.

3.3.1 Rule-based Evolutionary Algorithm Forecasting Method

In this section, the necessary background information regarding Povinelli's (1999) TSDM methods is introduced to evoke a good understanding of our research direction.

The TSDM methods create a new structure for analyzing time series by adapting concepts from data mining, time series analysis, EA, and nonlinear dynamics system (Iba, 1999). They are tactfully designed to predict non-stationary, non-periodic and irregular time series, and not restricted by the use of predefined templates. More specifically, they help discover hidden temporal structures predictive of sharp movements in time series, using a time-delay embedding process that reconstructs the time series into a phase space that is topologically equivalent to the original system under certain assumptions (Iba, 1999). The TSDM methods are developed and applicable to yield one-step predictions for time series data sets (Povinelli, 2000). In order to extract non-stationary temporal patterns, a specific TSDM method could be used to address quasi-stationary temporal patterns *i.e.*, temporal patterns that are characteristic and predictive of events for a limited time window Q. It is called the Time Series Data Mining evolving temporal pattern (TSDMe2) method, which uses a fixed training window and a single period prediction window. The TSDMe2 method differs from the other TSDM methods in how the observed and testing time series are formed. The TSDMe2 method creates the overlapping observed time series :

$$\theta = \{\theta_t, t = j, \ldots, j + N\}. \tag{6}$$

The testing time series is formed from a single observation:

$$\eta = \{\theta_t, t = j + N + 1\}, \tag{7}$$

Where θ_j is the time series value at time $t=j$, while N is the size of the window. The TSDM*e2* method was created for discovering multiple temporal pattern clusters in a time series (Povinelli, 1999).

In characterizing different embedded patterns hidden in the time series, there are two key factors to consider, number of pattern types and size of the patterns (or windows Q). By parsimony, the simplest characterization of events possible is desired *i.e.*, as small a dimensional phase space Q as possible and as few characterization patterns as necessary. However, the following modifications have been made to the typical TSDM in order to implement the proposed data mining procedure,

a) to increase the pattern characterizations by involving as many as 100 different arithmetic expressions to describe a windowed time series;
b) to use a 4-lag recursive memory as the size of the patterns Q. For definitions of some related concepts, refer to APPENDIX 4 at the end of this thesis and the paper by Povinelli (1999).

There will be many different patterns present in financial time series, both linear and nonlinear types. A financial series is a dynamic entity which is affected by many variables, economic, financial, politics, psychological, legal, *etc.* It is philosophically unwise to use one fixed model, linear or nonlinear to estimate such a process, let alone for forecasting. In using statistical models to estimate time series, wherever abrupt structural changes the model will need to adjust and change its parameters. Furthermore, the application of volatility estimation in option trading deem necessary to extract also the non-event, so that one could capitalize on the time value of the option. In our case, we used 100 different formula/rules to match the frequently appearing events and to extract different patterns buried in noise. Therefore, there is a high probability to extract the patterns and further to forecast the one step-ahead activity. Note that in each of the 100 rules used to characterize different patterns, the value of δ could be considered as the margin of accuracy the rules match the points in the window. In case of GA's, since the data will be divided into four ranges δ is not applicable. Details could be found in the following paragraphs as well as in Section 5.4.

The patterns are determined by the four previous points to make a prediction for the value at the next time interval. The 4-lag recursive system is used due to the potential weekly seasonality of the IV time series as well as the convenience of weekly option trading. Using 4-lag approach could save time and memory and is particularly useful in dealing with volatility forecasting because of past research showing that most of the information is contained in the most recent lags, resulting the popularity of GARCH(1,1) or other short memory models. One of the contributions of the current research is in dealing with characterizing the non-events. This is especially important in the forecast of volatility. For example, let us assume that one could accurately predict that no major change in the volatility of S&P100 index in the following week. The investor could sell a number of contracts on the strike or near the strike price of the S&P100 option with the shortest expiration available in order to earn the time value.

The current method involves matching clustering patterns, which include sharp fluctuations in the IV series. To find these temporal patterns the time series is embedded into a reconstructed phase space with a time delay of one and a dimension of four (Povinelli, 1999). Once the data is embedded, temporal structures are located using a GA/GP search. Clusters are made of points within a fixed distance of the temporal structures δ. In case of using GP, the event characterization function $g(t) = \theta_{t+1}$, determines the value given to the prediction made from the clustering using the temporal structures. This value is the IV value for the next time interval. The temporal structures are next ordered by how well each predicts the IV movements. A ranking function is defined as the average value within a temporal structure, and it is used to order the structures for optimization. The optimization is a search to find the best set of temporal structures and is done with EA that finds fitness value parameters that maximize the ranking function – the frequency of the correctly guessed patterns in case of GA or minimize the fitness function $f(P)$ – the difference compared with the guessed patterns in case of GP. This leads to the Eq. (8) given below :

$$\min_{P,\delta} f(P)$$
$$f(P) = \sum_{t=5}^{N} (\sigma_t - \theta_t), \quad t = 5, \ldots, N \tag{8}$$

where σ_t is the forecasted volatility and θ_t is a value in the series to be forecasted. The EA uses a combination of Monte Carlo search for population initialization with a fixed percentage selection, crossover and mutation to find the optimal P^*, and a limited number of generations to halt the genetic programming (Povinelli, 1999, 2000).

When GA is used, the IV window $\theta = \{ \theta_{t+j}, j=0, 1, 2, 3\}$ will be converted into a set of numbers $\{1, 2, 3, 4, *\}$ by classifying the range as $\{(-\infty, -a], (-a, b], (b, c], (d, \infty), *\}$, where '*' means "`don't care`". Therefore, all data will become a sequence of numbers. The rules will take the form of <IF [(($\theta_t = I$) AND/OR ($\theta_{t+1} = J$) AND/OR (($\theta_{t+2} = K$)) AND/OR ($\theta_{t+3} = L$))], THEN ($\theta_{t+4} = M$)>, where the event characterization function $g(t) = \theta_{t+4}$ will be a number that predicts the range of the subsequent IV value. And δ will become obsolete. The key difference between using GP and GA is the form of the rule. More details could be found in Chapter 5 where both techniques are applied respectively to the current problem.

In general, such combination can potentially eliminate the erroneous predictions caused by noise in the data. During the rule learning process, independent trials of the GA/GP allow many rules simultaneously to explore different parts of the search space, thereby learning different types of patterns for yielding a prediction. As a result, at a given time some rules generate better predictors than others, thus making them ideal candidates as base predictors to achieve increased predictive accuracy.

CHAPITRE 4

DSO FOR IV FORECASTING

As shown in Section 2.2, GA's have been used to extract patterns from the volatility time series and enable analysts to achieve a high forecasting accuracy. In this chapter we attempt to demonstrate that a Markov chain based discrete stochastic optimization (DSO) method could provide the theoretical support for applying GA's to forecast IV, if the IV time series is properly converted into a Markov chain. By employing this DSO method introduced by Andradottir (1995, 1999), we demonstrate the feasibility and convergence of GA's in case of time non-homogeneity. Viewed differently, the current work demonstrates the efficiency improvement which GA brings to the application of the stochastic optimization process in forecasting IV.

4.1 Literature Review

In order to model and forecast volatility, a wide variety of methods have been attempted in the last decade. Among literatures that use the traditional approaches, Barndorff-Nielsen & Shephard (2002) formed a general stochastic volatility model to estimate IV so that model based approaches can potentially lead to significant reductions of the mean square error. Working from the concepts of realized power variation and realized volatility, they provided a limiting distribution theory to strengthen the consistency results (Barndorff-Nielsen & Shephard 2003). They went on to provide a systematic study of kernel-based estimators of the integrated variance in the presence of market microstructure noise by deriving the optimal kernel-based estimator under an assumption that the noise is without memory and independent of the efficient price (Barndorff-Nielsen & Shephard 2004). Refer to Ma *et al.* (2004*a*, *b*) for a list of the related literature on this subject.

As discussed in the earlier chapters, it was concluded that GA's could prove to be the more practical and effective approach at present in tackling the stochastic optimization problems. Its advantage lies in the ease of variable coding and its inherent parallelism. The use of genotypes instead of phenotypes to travel in the search space makes them less likely to get stuck in local maxima. The GA approach employed in Ma *et al.* (2004*b*) satisfied some stringent criteria and yielded forecasting accuracy that is higher than those derived from other publicly available research. GA methods have, however certain drawbacks, *e.g.* GA's are not guaranteed to give an optimal solution and they lack convergence proof. In comparison to other stochastic optimization techniques such as simulated annealing, GA's lack rigorous supporting mathematical theory such as the one based on the principle of Markov chain (Pinto, 2000).

The recent advancements in discrete stochastic optimization methods provide the theoretical foundation to solidify the GA approach. For example, Andradottir (1995, 1999) demonstrated the feasibility of applying the Markov chain method when the transitional matrix is initially non-time homogeneous and asymptotically approaches time homogeneous state, unlike Duan *et al.* (2003) and most other work in the field, which are confined to time-homogeneous cases. However, the main difficulty while applying Markov chain theory to solve time series problems is that data in time series problems are typically correlated, while Markov chain by definition does not concern about the historical states prior to the current one. This is exemplified by the application of Markov chain method on the non-linear asymmetric GARCH(1,1) process, as done in Duan *et al.'s* research publication (2003). Therefore, one needs to transform a time series into a Markov chain while maintaining the necessary characteristics of the original data, in order to make use of the rigorous mathematical theory to substantiate the stochastic GA operation.

Duan *et al.* (2003) introduced the use of a time-homogenous Markov chain for the valuation of options, in which volatility determination is a key. The Markov chain approach allows one to decouple the partitioning of time and state. In other words, one can use time steps suitable for a particular contingent claim without being unduly constrained to have a particular set of state values, unlike other option valuation methods such as binomial tree or lattice and finite difference methods. Such a characteristic motivates the current IV data conversion into the overlapping four-lag recursive data groups, thus enabling the joining force of both Markov chain and GA's for the purpose of optimization. And this is another key contribution of this work, *i.e.* to substantiate the GA operation with Markov chain when applied to optimize the forecast of a volatility time series.

In the following sections we attempt to apply a Markov chain based Discrete Stochastic Optimization (DSO) method to substantiate the use of GA's in Ma *et al.*'s (2004*b*) published paper, since GA based DSO typically lacks rigorous mathematical proof. The key in such a process is to transform the IV time series data set into a cross-sectional one. It turns out that the 4-lag recursive transformation in the TSDM framework as described in Chapter 3 fulfils precisely such a purpose. In Section 4.2, Andradottir's (1999) global search DSO method is introduced, whereas Section 4.3, the method is applied to substantiate the use of GA's for volatility forecast.

4.2 Discrete Stochastic Optimization Method

In this section, a typical form of DSO or Discrete Stochastic Optimization problems is outlined while the basic concept, procedure as well as advantages and disadvantages of DSO are introduced. As shown in Ma *et al.* (2004*b*), the calculation of IV converts the volatility from a latent variable into an "observable" one. Upon certain conversion as shown in the next sections, the IV time series could become the random variable of a DSO process, thus allowing us to make use of the DSO method. This is demonstrated in the following sections.

4.2.1 Typical Markov Chain Approach

The following is the general form of a DSO problem for which the current approach we take, needs to determine global optimal solutions:

$$\max_{\theta \in \Theta} f(\theta), \quad \text{where } f(\theta) = E\{X(\theta)\}, \forall \theta \in \Theta. \tag{9}$$

a) Here, $f: \Theta \rightarrow \Re$ is the objective function, where $\Re$ is the domain of real numbers.

b) Θ is the discrete feasible region containing at least two states; in the current case, for a finite feasible set, $\Theta^* \neq 0$, where $\Theta^* = \{ \theta \in \Theta : f(\theta) \geq f(\theta') \text{ for all } \theta \in \Theta\}$ $(\theta' \in \Theta \backslash \{\theta\})$ is the set of global optimal solutions to the optimization problem; since f: $\Theta \rightarrow \Re$, the optimal value $f^* = \max_{\theta \in \Theta} f(\theta)$ is finite and can be achieved.

c) $\{X(\theta) : \theta \in \Theta\}$ is a collection of random variables having the property that $E\{X(\theta)\}$ cannot be evaluated analytically but estimated or measured.

d) θ is a random variable in a stochastic process.

e) And $X_l(\theta), \ldots, X_L(\theta)$ are independent and identically distributed (IID) observations of $X(\theta)$ for all $\theta \in \Theta$.

In seeking the solution for the next step, many traditional random search algorithms estimate the optimal solution by using either the feasible solution the method is currently exploring or the feasible solution visited most often so far. A feasible solution is one that corresponds to the state within $\tilde{\Theta}$. On the other hand, Andradottir (1996*b*) believe that their performance can be improved and proposed an alternative approach. Further details are given in the following section.

4.2.2 More Contemporary Approch

Andradottir proposed using *all the observed objective function values* generated as the random search method moves around the feasible region to obtain increasingly more accurate estimates of the objective function values at different points. At any given time, the feasible solution that has the best estimated objective function value, *e.g.* the largest one for maximization problems is used as the estimate of the optimal solution. At the same time, Andradottir specified the rate of convergence of this method and proved that it is guaranteed to converge almost surely to the set of global optimal solutions. Numerical evidence presented by Andradottir (1996*b*) and by Alrefaei and Andradottir (1996*a*, *c*) suggests that this approach for estimating the optimal solution appears to help yield improved performance relative to other approaches for estimating the optimal solution.

Andradottir's (1999) Lemma 3.1 assumes that P_m, $m = 0, 1, 2, \ldots$ and P are Markov matrices on the state space Θ such that P is irreducible and aperiodic and $P_m \rightarrow P$ as $m \rightarrow \infty$. If q: $\Theta \rightarrow \Re$, then as $M \rightarrow \infty$

$$\frac{1}{M}\sum_{m=1}^{M} q(\theta_m) \rightarrow \sum_{d=1}^{J} \pi_d q(d), \quad as\ M \rightarrow \infty, \tag{10}$$

where $\pi^T = (\pi_1, \ldots, \pi_J)$ is the steady-state distribution corresponding to P, while $\{X_m\}$ is a non-homogeneous Markov chain with transition probabilities

$$P\{\theta_{m+1} = d \mid \theta_0, \mathrm{K}\ \theta_m\} = P_m(\theta_m, d) \\ \forall d \in \Theta \wedge m = 0, 1, 2, \mathrm{K} \tag{11}$$

In other words, at iteration $m+1$, θ_{m+1} has $d=J$ possible states. Here the number of states is countable and limited. At the limit, the transitional matrix becomes time-homogeneous, *e.g.* stable. Andradottir's (1999) preferable approach would involve maintaining two variables for each point $\theta \in \Theta$. One of these, say $K_m(\theta)$ would count how many estimates of $f(\theta)$ have been generated in the first m iterations for the respective θ, while the other one $\sum_m(\theta)$ would contain the sum of all $K_m(\theta)$ estimates of $f(\theta)$ that have been generated in the first m iterations. The specific procedure is outlined in Algorithme 1.

Algorithm 1 – Modified Global Search Method

Step 0: Select a starting point $\theta_0 \in \Theta$. Let $K_{-1}(\theta) = \Sigma_{-1}(\theta) = 0 \ \forall \ \theta \in \Theta$. Let $m = 0$ and $\theta^*_m = \theta_0$ and go to Step 1.

Step 1: Given the value of θ_m, generate a uniform random variable θ'_m on $N(\theta_m)$ independently of the past (so that $\forall \ \theta \in \Theta$, $\theta \neq \theta_m$, we have that $\theta'_m = \theta$ with probability $1 / (|\Theta| - 1)$). Go to Step 2.

Step 2: Given the value of θ_m and θ'_m, generate observations $X_{m,l}(\theta)$ of $X(\theta)$, for $l = 1, \ldots, L$ and $\theta = \theta_m$, θ'_m independently of the past. Let $R_m = \Sigma^L_{l=1} (X_{m,l}(\theta_m) - X_{m,l}(\theta'_m)/L$. if $R_m > 0$, then let $\theta_{m+1} = \theta_m$. Otherwise let $\theta_{m+1} = \theta'_m$. Go to Step 3.

Step 3: Let $K_m(\theta) = K_{m-1}(\theta) + L$ for $\theta = \theta_m, \theta'_m$, and $K_m(\theta) = K_{m-1}(\theta) \ \forall \theta \in \Theta \backslash \{\theta_m, \theta'_m\}$. Moreover, let $\sum_m(\theta) = \sum_{m-1}(\theta) + \sum^L_{l=l} X_{m,l}(\theta)$ for $\theta = \theta_m$, θ'_m, and $\sum_m(\theta) = \sum_{m-1}(\theta) \ \forall \theta \in \Theta \backslash \{\theta_m, \theta'_m\}$. Let $\theta^*_m \in \arg\max_{\theta \in \Upsilon_m} \sum_m(\theta)/K_m(\theta)$, where $\Upsilon_m = \{\theta \in \Theta : K_m(\theta) > 0\}$. Let $m = m + 1$ and go to Step 1.

There is no particular requirement how θ, the solution should behave. On the other hand, $\{X(\theta) : \theta \in \Theta\}$ should be a collection of random variables having the feature that $E\{X(\theta)\}$ is the unbiased and consistent estimation of $f(\theta)$. Details regarding the rationale of unbiased and consistent estimation of $f(\theta)$ are given in Section 4.4. The main issue in applying Algorithm 1 will be the way to use the state data generated by a random search method in order to obtain an estimate of the optimal solution. Here, solutions are first compared against each other pair by pair. Those solutions that have higher averaged $X(\theta)$ will be retained for the next generation, *i.e.* the selection of θ_{m+1} based on the value of R_m. Andradottir's (1999) approach requires the search of optimized solution to be identified in Step 3, where $K_m(\theta)$ and $\sum_m(\theta)$ for each $\theta \in \Theta$ are stored, accumulated and compared for maximization. At the last generation, among thousands of solutions in the memory the optimization is performed with $\theta_m^* \in \arg\max_{\theta \in \Upsilon_m} \sum_m(\theta)/K_m(\theta)$. The top solutions could be selected and used for validation by testing against other set of data. The fact that all values of θ_{m+1} are kept in memory while the optimization is ongoing makes it the key difference between Andradottir's method and others including the one used in Ma *et al.* (2004*b*). The detailed procedure to incorporate Andradottir's method with GA will be given in the next section, which can further accelerate the optimization process.

4.3 Proposed Methodology

To apply Algorithm 1, we classify the respective values in the IV time series into four ranges, *e.g.* $(-\infty, -a]$, $(-a, b]$, $(b, c]$ and (d, ∞). Therefore, all data will become a sequence of numbers, *e.g.* 1, 2, 3, & 4. We then define a set $\theta_j = \{\theta_t, \theta_{t+1}, \theta_{t+2}, \theta_{t+3}, \theta_{t+4}\}$, where $\theta \in \{1, 2, 3, 4, *\}$ is a state in Θ with * = *don't care* and $j = 1, \ldots, J$, where J is the total number of states as indicated in Eq. (11). In other words, θ could be defined as the successively overlapped 4-lag recursive data set that has been converted from the original IV time series. We then generate rules randomly in the form of <IF [((θ_t = I) AND/OR ($\theta_{t+1} = K$) AND/OR (($\theta_{t+2} = L$)) AND/OR ($\theta_{t+3} = M$))], THEN ($\theta_{t+4} = N$)>, where the "IF" part $\{\theta_t, \theta_{t+1}, \theta_{t+2}, \theta_{t+3}$, AND, OR, *} is used as the qualifying criteria and the "THEN" part $\{\theta_{t+4}$, *} is for predicting the subsequent IV value. Each rule will work as a sliding window to pass across the entire IV time series point by point. In light of Andradottier's Lemma 3.1 as shown in Eq. (9) and (10), such a plan is sound in our GA operation, because in practical sense it is acceptable to assume that there is a limited number of patterns existing in the IV time series, *e.g.* J. And we are looking for the rules that most frequently match with the overlapped 4-lag IV data. Those patterns that appear more often tend to be caught by rules derived from crossover and/or mutation, and will gradually lead to more successful estimates. Moreover, Theorem 1 as shown in Section 4.4 demonstrates when each estimate is obtained from a single trial (in our case, each generation in GA generates 100 rules per group of total 100 groups in which each rule is independent of each other) the random search method is of first order convergence (Andradottir, 1999). Upon satisfying the two key conditions, we could define the stochastic function as :

$$X(\theta) = \begin{cases} 1, & \text{if } \{\theta_t, \theta_{t+1}, \theta_{t+2}, \theta_{t+3}, \theta_{t+4}\} \text{ matches the data sequence;} \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

where $\{\theta_t, \theta_{t+1}, \theta_{t+2}, \theta_{t+3}, \theta_{t+4}$, AND, OR, *} represents the rules. Consequently, $f(\theta)$ would be the expected prediction accuracy E{X(θ)} for a rule θ and could be any real value in [0, 1]. Each rule will be independent of any other rules or at least treated so in view of GA, thus satisfies the requirement of Markov chain operation, and the nature of X(θ) makes it IID as required in Eq. (9). The problem is therefore, converted into a search of rules that best fit the four-point patterns in the IV data set so that the immediate fifth IV value could be forecasted upon knowing the previous four points. In a more general sense, a time series problem is thus converted into a set of random data that could be approached with the Markov chain method.

To extend Andratottir's strategy of comparing rules pair by pair, we make use of the GA technique such as the tournament/elitist selection criterion to improve the chance of reaching the optimal objective function. As defined by Langdon (1996), tournament selection is

> *"a mechanism for choosing individuals from a population. A group (typically between 2 and 7 individuals) are selected at random from the population and the best (normally only one, but possibly more) is chosen. An elitist GA is one that always retains in the population the best individual found so far. Tournament selection is naturally elitist."*

In every generation, new rules in the groups that have been derived from crossover and mutation in the previous generation will be put back into the pool to be compared with those retained from the last generation. Only those new ones that have higher prediction rates will replace the respectively selected peers for the next generation. Either accepted or rejected they are recorded in memory together with other existing rules. Here, L is the smaller number of the possible matches derived by comparing θ_m and θ'_m and is at maximum equals the number of data points in the IV time series minus four, while m is the number of generations to perform GA. One important feature GA incorporates in

Step 2 is the way of generating $X_{m,l}(\theta)$ of $X(\theta)$, for $l = 1, \ldots, L$ and $\theta = \theta_m, \theta'_m$ independently of the past. By applying GA, θ'_m are generated through crossover or mutation, while $X(\theta)$ depend on whether the qualified rule predicts correctly. With the value of R_m we could choose from θ_m and θ'_m, to make one of them go through further GA manipulation such as crossover or mutation. At the last generation, we could retain θ^*_m as the optimal solution for the m^{th} generation by carrying out the optimization process. Note that the calculation of $K_m(\theta)$ could be modified as

$$K_m(\theta) = \begin{cases} K_{m-1}(\theta) + L, & \text{if } \{\theta_t, \theta_{t+1}, \theta_{t+2}, \theta_{t+3}, \theta_{t+4}\} \text{ matches the data sequence;} \\ K_{m-1}(\theta), & \text{otherwise.} \end{cases} \quad (13)$$

where $\{\theta_t, \theta_{t+1}, \theta_{t+2}, \theta_{t+3}\}$ is again the qualifying part of the rule.

4.4 Proposed Procedure to Apply DSO with GA

When applying Algorithm 1 to solve the current discrete stochastic optimization problem, we obtain the following Algorithm 2:

Algorithm 2 – Discrete Stochastic Optimization with GA

Step 0: Randomly assign any one value of {1, 2, 3, 4, *} to the first four fields in $\theta = (\theta_t, \theta_{t+1}, \theta_{t+2}, \theta_{t+3}, \theta_{t+4})$, randomly assign operators "AND" and "OR" to join these four fields and then assign $\theta_{t+4} = 1$ for the first 25 rules. Repeat the same process with $\theta_{t+4} = 2$, 3 and 4 respectively to form a total of 100 rules. Repeat the operation to generate another 99 such groups. Then randomly select 50 rules in each group as θ_0's. Set all counters to zeros.

Step 1: The rest of 50 rules in each group that have been generated in Step 0 will become θ'_m's. Or when $m > 0$ θ'_m are derived by applying crossover or mutation on the first four points and the three joining operators of rules in those ones rejected in Step 2 during the previous generation.

Step 2: Generate the random variable $X_{m,l}(\theta)$ by running the pair of rules respectively selected from θ_m and θ'_m sequentially through the entire IV data set. L would be the smaller of the two corresponding total matches for each θ_m and θ'_m. $X_{m,l}(\theta) = 1$ when predict correctly, 0 otherwise. Let $R_m = \sum^{L}_{l=1}(X_{m,l}(\theta_m) - X_{m,l}(\theta'_m))/L$. If $R_m > 0$, then let $\theta_{m+1} = \theta_m$. Otherwise let $\theta_{m+1} = \theta'_m$. Select another pair rules from θ_m and θ'_m and repeat the comparison procedure until obtaining 50 θ_{m+1} rules. 25 of the rejected rules will be used for crossover and the other 25 mutation at Step 1 in the next generation. Repeat the entire process for the rest of the 99 groups.

Step 3: $K_m(\theta)$ would be the total number of matches in the qualifying part of rules θ_m and θ'_m up to generation m, while $\sum_m(\theta)$ is the number of correct predictions for the corresponding rules. Increase the counter by 1 until reaching the preset limit. At the last generation, optimize among all rules stored in the memory based on the given criteria and retain the top 100 θ^*_m that could best forecast in the given data set, i.e. maximize the percentage of correct forecast by letting $\theta^*_m \in \arg\max_{\theta\in\Upsilon_m} \sum_m(\theta)/K_m(\theta)$, where $\Upsilon_m = \{\theta \in \Theta : K_m(\theta) > 0\}$. In ranking all stored rules, among those rules that are numerically identical, qualified and predicting correctly only the one has minimum "*don't care*" fields and "OR" operators will be retained.

At Step 0 generation 0, first rule is generated to take a value of θ_0 and the success rate of prediction to be zero. For whatever value of θ_0 we generate a different rule based on criteria given in Step 1. At Step 1 we apply the GA techniques such as tournament/elitist selection criterion, crossover and mutation to generate rules for comparison. At Step 2, we generate the expected outcome $X_m(\theta)$ for both rules by comparing each rule with all data points in the IV series. In carrying on the same process to the next point in the data set till completion, we find the respective L. For generation m>1, we only need to go through this process for θ'_m while values of $X_m(\theta)$ and L for θ_m have been derived in the previous generation. If θ'_m have higher rates of success, replace the current rules with the more successful ones and keep them in memory as θ_{m+1}. In such an operation, the same θ_{m+1} from different groups could appear more than once as indicated in Step 1, and it will yield the same $X(\theta)$ as before due to the nature of the data set. But only one of them should be registered when they predict better than the current best θ_m. In order to comply with Algorithm 1, we could incorporate a screening mechanism firstly to reject rules that are the same as those currently exist in the memory and secondly to reject rules that are identically qualified and correctly predicting in the current generation. This is necessary because in Andratottir's algorithm (1995), θ'_m that is the same as previous θ's will be rejected in Step 2. This process is repeated in parallel for all 100 groups. At step 3, we calculate for the optimal solution θ^*_{m+1} at the last generation based on the corresponding number of correct predictions, *i.e.* determine the rules that maximizes the prediction among all retained rules. Once the top 100 rules are derived, we could use them to predict another set of IV data especially those at an immediately subsequent time period in order to confirm the validity of the approach.

4.5 Rate of Convergence

Yan and Mukai (1992) defined the rate of convergence of the algorithm to be the rate at which the distribution of θ_m in Algorithm 1 converges to an optimal distribution, *i.e.* only puts a positive mass on elements of Θ^*. In other words, rate of convergence of a random search method for DSO is the rate at which the estimated value of the objective function at the estimated optimal solution converges to the optimal values of the objective function.

Theorem 1: Rate of convergence of Random Search Methods [1]. Assume that

a) $\Theta^* \neq 0$ and is finite;

b) The estimate of the optimal solution $\theta^*_{m+1} \in \arg\max_{\theta \in \Upsilon_m} \sum_m(\theta)/K_m(\theta)$ in Algorithm 2 converges almost surely to the set Θ^* as $m \to \infty$. Since $\sum_m(\theta)$ is the number of correct predictions while $K_m(\theta)$ is the number of hits, i.e. the number of matches between the first four points of the rule and the 4-lag recursive points in the IV data set, as $m \to \infty$, $K_m(\theta) \to \infty$. From the Strong Law of Large Numbers, consistent and unbiased solutions exist [2];

c) For all $\theta \in \Theta^*$, the estimate of $f(\theta)$ (obtained from single trials, i.e. at a certain value of m) are independent and identically distributed with mean $-\infty < f(\theta) < \infty$ and variance $0 < \sigma^2 < \infty$; If Θ is finite and for all $\theta \in \Theta$ we have $|f(\theta)| < \infty$, the estimates of $f(\theta)$ here are IID [2]. Since the rules are initially randomly generated, and each rule is independent of each other; rules after randomly crossover and mutated are also independent. Moreover, they are generated in a similarly random fashion, therefore it is understandable that the rate of correct prediction for all rules at each iteration is IID.

d) The estimates $f(\theta)$ are independent of the estimates of $f(\theta')$ for all $\theta' \in \Theta^* \backslash \{\theta\}$ (when each estimate is obtained from a single trial); and there exists a constant $0 < c(\theta) < \infty$ and a sequence $\{a_m\}$ of constants such that as $m \to \infty$, $a_m \to \infty$ and $K_m(\theta) / a_m \to c(\theta)$. (i.e. $K_m(\theta)$ can be tracked so that it is feasible for each θ to have a distinguishable value of $K_m(\theta)$.) We then have

$$\sqrt{a_m}\left(\frac{\sum_m\left(\theta^*_{m+1}\right)}{K_m\left(\theta^*_{m+1}\right)} - \min_{\theta\in\Theta} f(\theta)\right) \Rightarrow \min_{\theta\in\Theta^*} Z(\theta), \text{ as } m \to \infty \tag{14}$$

where $\forall\theta \in \Theta^*$, the random variables $Z(\theta)$ are independent and

$$Z(\theta) \sim N(0, \frac{\sigma^2(\theta)}{c(\theta)}). \tag{15}$$

The recently available real time data base for equity, indices, foreign exchanges or even fixed incomes makes it sensible to assume $m \to \infty$. However, one may need to take into account properties associated with the nature of financial markets. For example, the micro-structure of the equity bid-ask prices makes it difficult to use data that have higher frequency than say one reading in every 15 minutes for the purpose of volatility evaluation (Andersen, 1998 & 2001, Barndorff-Nielsen, 2004). Moreover, data patterns that occurred more than one or two years earlier may have little influence on the recent data, thus may not be applicable in the current volatility forecasting process. Further evidence and discussion on this issue is given in Chapter 6.

CHAPITRE 5

FORECASTING WITH CI TECHNIQUES

IV estimation enables one to forecast volatility directly at the horizons of interest, without making assumptions about the nature of the volatility process. Christoffersen (2002) has further shown that the short-term volatility could be relatively accurately forecast *e.g.,* for a time horizon of five to ten working days. On the other hand, long term volatility forecastability seems to subject to some debate. Christoffersen and Diebold (1998) found that volatility forecastibilty declines quickly with horizon, and seems to have largely vanished beyond horizons of ten or fifteen trading days, whereas Ding, Granger and Engle (1993), Baillie *et al.* (1996) and Andersen *et al.* (2001, 2001b) claimed that volatility displays long memory.

The possibility of market timing focuses, not on prediction of returns directly, but rather on prediction of *signs* of returns, on the grounds that profitable trading strategies may result if one is successful at forecasting return signs, quite apart from whether one is successful at forecasting the mean of returns. A well-known and classic example involves foreign exchange trading. If the Yen/$ exchange rate is expected to increase, reflecting expected depreciation of the Yen relative to the dollar and hence a negative expected "return" on the Yen, one would sell Yen for Dollars, whether in the spot or derivatives markets (Christoffersen and Diebold, 2002). From the viewpoint of trading VIX and the associated futures, such an approach is also applicable for the volatility forecasting. In this chapter and in the following one, we attempt to forecast both the direction and range of the volatility by accounting for the non-linearity of the IV in a relatively short time span, *e.g.* within five working days. More specifically in the last part of this chapter, the actual value of the volatility will be forecasted based on a similar approach.

5.1 IV Calculation

The first step is to calculate the IV of the S&P100 index between Jan. 2, 1998 and Aug. 29, 2003, a series containing 37,995 entries. The return of an asset r at time t_i is defined as

$$r(t_i) = r(\Delta t; t_i) = x(t_i) - x(t_i - \Delta t) \tag{16}$$

where $x(t_i)$ is a homogeneous sequence of logarithmic prices as defined by

$$x(t_i) = x(\Delta t, t_i) = \tfrac{1}{2}\left(\log p_{\text{bid}}(t_i) + \log p_{\text{ask}}(t_i)\right) \tag{17}$$

where $p_{\text{bid}}\,(t_i)$ and $p_{\text{ask}}\,(t_i)$ are the bid and ask prices of the underlying asset at time t_i.The IV is evaluated based on intra-day historical data and is a more accurate approximation of the daily volatility (Andersen and Bollerslev 1998) :

$$\text{v}(t_m) = \text{v}(\Delta t, n; i) = \left[\frac{1}{n}\sum_{j=1}^{n}\left|r(\Delta t; t_{i-n+j})\right|^2\right]^{1/2}, \tag{18}$$

where $\text{v}(t_m)$ is the moment rate of return distribution, Δt is the time interval of the data in which integration is done, n is the total time length of the integration, i is the total number of data and p defined the $(1/p)^{th}$ moment.

In the GA case detailed in Section 5.4, the S&P100 index 15-minute interpolated time series, $\Delta t = 15$, $n = 28$, $1/p = 2$, and $m = i/n$ (where $m = 1, 2, \ldots, 1350$). In other words, one deals with m days of data with 6.5-hour usual trading time per day. The hence derived IV series could be de-trended through different ways of normalizations *e.g.*, take logarithm of the data and subtract all by the mean (Kinlay, 2001). In the current study, the main interest is to determine the short term abrupt changes in the IV series, which is the key difficulty confronted the contemporary researchers. Volatility term structures are normally mean-revert, with short-term volatility lying either above or below the long term mean, depending on whether current conditions are high or low volatility. Therefore, the IV series is normalized by its past 21 day moving average, which roughly encompassing the immediate past month trading. The one-month time horizon is for the convenience of trading index options.

5.2 Wavelet Transform

Prior to analysis, the data representation and/or pre-processing are often necessary before actual data mining operations can take place. The representation problem is especially important when dealing with time series, since direct manipulation of continuous, high frequency data in an efficient way is extremely difficult, although the current case is a one-dimension series. This problem can be addressed in different ways. One possible solution is to use windowing and piecewise linear approximations to obtain manageable sub-sequences. The main idea of transformation based representations is to transform the initial sequences from time to another domain, and then to use a point in this new domain to represent each original sequence. Wavelet analysis is a form of non-parametric regression analysis. It extracts both low and high frequency components of a given signal. It has found much success in applications such as image processing, geological testing and many other engineering applications.

Specifically, the Discrete Wavelet Transform (DWT) translates each sequence from the time domain into the time/frequency domain. The DWT is a linear transformation, which decomposes the original sequence into different frequency components, without loosing the information about the instant of the elements occurrence. A wavelet transform is a scaling function used to convert a signal into father and mother wavelets. Father wavelets are representations of a signal's smooth or low-frequency components. Mother wavelets represent the details or high-frequency components in the signal (Kaboudan, 2005). The sequence is then represented by its features, expressed as the wavelet coefficients. Again, only a few coefficients are needed to approximately represent the sequence. With these kinds of representations, time series becomes a more manageable object, which permits the definition of efficient similarity measures and an easier application to common data mining operations (Antunes and Oliveira, 2002).

As indicated in Chapter 2 however, the wavelet methodology has not been used widely in economic time series studies. So far only a few researchers make use of wavelets to analyze individual stock returns, high frequency stock index returns (Capobianco, 1999, and Arino, 2000) and the foreign exchange rates (Los and Karuppiah, 2000, Kaboudan, 2005). Hog and Lunde (2003) made used of wavelet transform to estimate IV of the foreign exchanges, but did not touch on the topic of forecasting.

In this research work, the proposed methodology uses wavelets to decompose the series into superposed spectral components. The nonlinear time-varying underlying trends in noisy series, such as the clustering phenomenon of the current volatility series, can be identified. After the process, the coefficient of each scale can provide clues for the pattern of the volatility in the corresponding time horizon. For example, long-term trends can be assessed from the coarse-scale behaviour of the series. Shorter-term trends and noise can be estimated from fine-scale behaviour. In a decomposed wavelet plot, if the volatility at a certain scale is high, we might hypothesize that high volatility is also likely to appear in the next few components at and/or near the same scale, because of the clustering effect. It will be interesting to confirm such a hypothesis by finding out that the high values indeed appear at the *same scale* in the subsequent time intervals. If so, when a high volatility appears, one could reasonably expect that high volatility would also occur in a time horizon that is associated with the found scale. Therefore, a qualitative forecasting on a component basis will be of practical value in the process of building a quantitative forecasting algorithm.

Once the time series is converted to the normalized IV's, it will be analyzed with the DWT, which decomposes the original sequence V into different frequency components without loosing the information about the instant of the elements occurrence. The sequence is then represented by its features, expressed as the wavelet coefficients $b_{j,k}$ and $c_{j,k}$. Refer to APPENDIX 3 for details. Since only a few coefficients are needed to approximately represent the sequence, time series becomes a more manageable object, which permits the definition of efficient similarity measures and an easier application to common data mining operations (Antunes and Oliveira, 2002).

The main advantage when using the wavelet method is its robustness due to the absence of any potentially erroneous assumption or parametric testing procedure. Another advantage is that wavelet variance decomposition allows one to study different investing behaviour in different time scales (horizons) independently. Different investing styles may cluster into different time horizons. In wavelet packet trees, those IV values that occur often *i.e.,* at higher probability correspond to nodes at a higher level in the tree (Mallat, 1999). In the best tree, only those IV patterns that are uniquely decodable get to the bottom of the tree. Other repetitive patterns get located closer to the root. The root node represents the original data sequence, which would mean daily forecast and thus provide no calculation economy.

Another property attribute is the de-correlation, or so called whitening, which means correlated signals in time domain become almost uncorrelated coefficients in the time-scale domain. And due to the inherent characteristic property that wavelets are very good at compressing a wide range of signals into a small number of large wavelet coefficients, a very large proportion of the coefficients can be set to zero without any loss of important embedded information. This property allows wavelets to deal quite well with heterogeneous and intermittent behaviors (Silverman 2000). Therefore, wavelet transform can be utilized to filter out "noise" traders (wavelets compressing) and separation of short term and long run performances (time-scale decomposition). In summary, wavelet method has the following property attributes for our purpose:

a) Perfect reconstruction if and when needed

b) Locality in time and in scale
c) Whitening
d) Dis-balance energy
e) Filtering
f) Detect self-similarity
g) Efficient algorithm capability (Li, 2003)

5.3 Wavelet Packets

Details regarding the concepts and advantages of applying wavelet packets, thresholding and filtering to analyze nonlinear time series are provided in the following statements. Wavelet packet analysis provides richer details compared with simple DWT. In wavelet packet trees, those IV patterns that occur often, *i.e.* at higher probability correspond to nodes at a higher level in the tree (Mallat, 1998).

The wavelet packets can help differentiate the signal in a wider range of scales and it could identify the coefficients where patterns repeat the most so that the analysis could be more focused. By employing the *besttree* function, one could pinpoint the best way to dissect the original series and find out the best forecasting frequency. For example, the *besttree* function returns a tree with high repetition rate at node [2, 0], so that one only needs to focus on the eight-day forecasting range instead of daily, an eight-time economics in computation time. Moreover, a better forecasting accuracy could be expected due to the focus at the range where higher concentration of energy locates. The general criteria in selecting the forecasting range are listed below as a sequence of steps:

Run the wavelet packet analysis and find the best tree by selecting the number of levels based on analyst's interest and by selecting the order of the filters in order to eliminate maximum amount of data without affecting the reconstruction while matching most patterns of the original data;

Find the entropy values associated with all terminal nodes on the best tree;

Concentrate on the terminal nodes with the low entropy values, knowing that the original node carries the highest value.

In this research thesis, the number of level of the wavelet packet as well as the order of filter is determined through a trial and error process. Once the best tree is selected based on the entropy value, the corresponding wavelet coefficients could be analyzed with GA programs and are described in the next section.

5.3.1 Strategic Deployment of some Wavelet Packet Techniques

The simplest wavelet non-linear compression technique is thresholding (Donoho 1995). Some recent work can be found in (Donoho and Johnstone 1998, Donoho and Yu 1998, and Donoho and Johnstone 1999), where statistical optimality of wavelet compression was explored, and Bayesian approaches are incorporated in selections of the compressing thresholding rules. The commonly used criterion for choosing the most efficient or best basis (pattern) for a given signal is the minimum entropy criterion (Coifman and Wickerhauser, 1992 and Wickerhauser, 1994). That is, let $\{p_i\}$ be the decomposition coefficients of a signal for a particular choice of the wavelet packet basis. For each set of decomposition coefficients $\{p_i\}$ we associate a nonnegative quantity $\eta^2(\{p_i\})$ called entropy defined by

$$\eta^2(\{p_i\}) = -\sum_i \frac{p_i^2}{\|p\|^2} \log_2 \frac{p_i^2}{\|p\|^2} \tag{19}$$

where $\|p\|^2 = \sum_i p_i^2$. The best basis is the one which produces the least entropy. Intuitively, the entropy defined above gives a measure of how many effective components are needed to represent the signal on a specific basis. For example, if in a particular basis the decomposition produces all zero coefficients except one *i.e.,* the signal coincides with a wave form, then the entropy reaches its minimum value of zero. On the other hand, if in some basis the decomposition coefficients are all equally important, say $p_i = 1/\sqrt{N}$ where N is the length of the data, the entropy in this case is maximum, $\log_2 N$. Any other decomposition will fall in between these two extreme cases. In general, the smaller the entropy the fewer significant coefficients would be needed to represent the signal.

In using the one-dimensional wavelet packet compression approach, the current research effort employs the global soft thresholding method (Mallat, 1998). Predefined thresholding strategy is based mostly on empirical methods – to strike a balance between sparseness and details. For example, in retaining 90% of the entropy, a 70-80% of data would have been removed which in turn results in the data compression.

The type of filter to be deployed could be selected based on the nature of the data set *e.g.,* for time series with sharp jumps or steps, one would choose a boxcar-like function such as Harr (Torrence and Compo, 1998). Daubechies wavelets filters are optimal in the sense that they have a minimum size support for a given number of vanishing moments. When choosing a particular wavelet, one thus faces a trade-off between the number of vanishing moments and the support size. If V has few isolated singularities and is very regular between singularities, one must choose a wavelet with many vanishing moments to produce a large number of small wavelet coefficients $<\theta, \theta_{j,n}>$. If the density of singularity increases, it might be better to decrease the size of its support at the cost of reducing the number of vanishing moments. Indeed, wavelets that overlap the singularities create high amplitude coefficients (Mallat, 1998).

5.4 Forecasting by Genetic Algorithms

Review of technical literature on applying a variety of techniques including the GA/GP to forecast volatility reveals that except those based on proprietary methodology which is not available in the public domain, researchers still rarely have found any effective and systematic method to deal with non-linearity. Moreover, very little research has been attempted to take advantage of the recent advances in the estimation of volatility, namely to couple the calculation of IV with those dealing on non-linear analytical techniques such as GA/GP and wavelet transforms. By applying the GA approach, the nonlinear time-varying underlying trends in noisy series, such as the clustering phenomenon as well as sudden jumps and drops in the volatility series can be properly identified. Since no pre-assumption is needed to be made and since both wavelet transformation and GA are non parametric, the current IV-wavelet-GA method is therefore inherently more robust, accurate and efficient to yield an effective solution for use in forecasting financial index volatility compared with any of the prevalent methods of today. In the following sections, the proposed IV-wavelet-GA approach is explained and is followed by a description of the investigation to forecast the short term moving direction and range of IV of the S&P100 index.

As indicated earlier, GA's can optimize some arbitrary function with straightforward representation better than many other procedures. Simplicity of operation and power of effect are two of the main attractions of the GA approach. GA's efficiently build new solutions from the best solutions of the previous trials, regardless of the linearity conditions of the problem on hand (Bauer, 1994). Moreover, GA's may help uncover those hidden rules, either general or specific for the asset besides the well documented clustering effect for the volatility such as crucial non-linear regularities.

5.4.1 Problem Formulation

As indicated in Chapter 1 and APPENDIX 2, a classical linear expression or factor format expression of GARCH explains mainly the clustering effect, but inherently lacks in accuracy when dealing with non-linearity. Andersen *et al.* (1998) introduced the concept of IV and improved the forecasting confidence of the one-day-ahead volatility. They made use of the diffusion limit of the weak GARCH(1,1) process to construct the continuous time model of daily volatility σ_t:

$$d\sigma_t^2 = \xi\left(\omega - \sigma_t^2\right)dt + \left(2\lambda\xi\right)^{1/2}\sigma_t dW_{\sigma,t} \tag{20}$$

where $W_{\sigma,t}$ is the generalized Wiener process and ω, λ and ζ can be expressed in terms of the discrete-time weak GARCH(1,1) parameters. Such a weak GARCH(1,1) model converges to an IGARCH(1,1) as the sampling frequency increases. Andersen (2001) ruled out any sizable improvement of predictability by means of higher order discrete time ARCH approximations or by using more complicated stochastic differential equations, due to the weak GARCH(1,1) interpretation of diffusion approximation in Eq. (20). However, the derivation of IV converts the volatility from a latent variable into an "observable" one. As a result, any volatility forecasting could be performed based on more realistic historical data.

5.4.2 Proposed Methodology

The research attempted in this part of the thesis work may be represented in terms of the process modules as shown in Fig. 3.

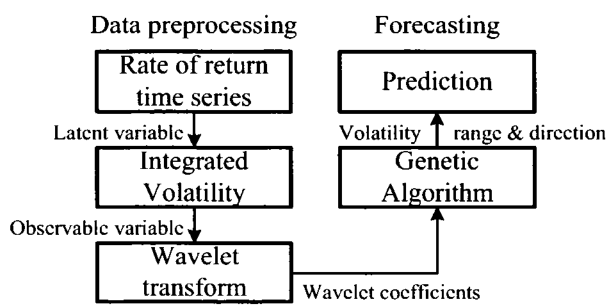

Figure 3 Proposed volatility forecasting methodology

Details about the implementation of this proposed methodology are described in the following sections.

5.4.3 Parameters Selection

One advantage of combining wavelet analysis with GA is the flexibility that they bring in. When applying the GA technique in the following sections, as mentioned before, we will adopt a 4-lag recursive forecasting approach. By selecting the corresponding wavelet coefficient series, this 4-lag configuration could help focus on different time ranges depending on nodes on the wavelet tree as given by the following expression:

$$\frac{f_{j,n}}{\Delta t} = \left[\frac{n}{2^{j+1}\Delta t}; \frac{n+1}{2^{j+1}\Delta t} \right], \tag{21}$$

where j is the level on the tree, n is the location on the tree, Δt is the sampling period; in the current case, it is one day and $f_{j,n}$ is the nominal frequency band.

For example, wavelet packet node (1, 0) gives information up to four days ahead; while (5, 5) is between 11 and 13 days. As a result, one could concentrate on uncovering trading patterns of different horizons.

For a data set of N samples of IV's, level 2 of the wavelet packets has $N/2$ number of coefficients, representing a saving of 50% calculation for the subsequent GA processing. A maximum of level 5 has been selected as the analysis scale here, because as mentioned earlier reliability of forecasting accuracy drops as time horizon expands. Filters such as the db2 wavelet, which has two vanishing moments, were used for the current analysis in order to maximize the match of the reconstructed data with the original data while retain the minimal amount of data. Note that a wavelet of the Daubechies family with fewer vanishing moments may fail to suppress the higher order polynomial signal. This has been confirmed in analyzing the current S&P100 series when db1, db2 and db3 occasionally fail to generate the wavelet coefficients based on the best tree that is created from the wavelet packet. On the other hand, higher order wavelet tends to generate smoother decomposed plots, which may loose some desirable details from the original series. Different combinations of orders and levels of the db wavelets could be tried to obtain the best tree. Analysis could focus on the node with the lowest entropy. For example, db3 with level 5 in the best tree, the lowest entropy occurs in packet (4, 0), where $j = 4$ and $n = 0$. Packet (4, 0) corresponds to the frequency of $(n + 1)/(2^{j+1}\,\Delta t) = 1/(32)$=> 32 days (in case of $j = 1$, $n = 2$, $(2 + 1)/(\,4 \times 1) \geq 1.3$ days). In general, there are five parameters to be determined before conducting the analysis *i.e.*, number of levels of the wavelet packet tree, order of the filter, number of generations, number of groups of rules and the training period. In this thesis research study, the effect of each variable is investigated in a sequence of comparative analysis by holding others temporarily constant. Since the main difficulty that many contemporary researchers were facing is the forecast of abrupt changes, the short term patterns in the volatility are the primary focus of this research investigation.

5.4.4 Implementation of the GA Forecasting Process

The premise of the GA approach described in this chapter is based on the concepts derived from Fong and Szeto's method (2001). First, we de-trend the wavelet coefficients with a 21-day-moving-average operation, convert them into integers of one to four corresponding to the four selected ranges, and then generate the rules randomly in the form as was shown in Section 3.3.1 earlier. For the '`THEN`' part of the rules, there are four different classes, 1, 2, 3 and 4. We randomly generate 25 rules for each class to have a total 100 rules. Then we repeat the process to generate 100 groups of such rules.

Fitness value of each rule is calculated as described in the following paragraphs. In each training step, the rules for class k are trained by comparing the patterns of the randomly generated rules with the S&P100 IV historical data. Three possible cases can arise. They are:

Case 1: the '`IF`' part of the rule does not match the data point pattern. So, no prediction can be made.

Case 2: the 'IF' part of the rule matches the data points in the training set. Prediction can be made. When the 'THEN' part of the rule also matches the class of the corresponding data point, it is counted as a correct guess otherwise a wrong guess. The fitness value of rule *i*, *i.e.* **the forecasting accuracy** will be

$$F_i = N_c / N_g = N_c / (N_c + N_w). \tag{22}$$

Here N_c is the number of correct guess and N_w is the number of wrong guess, so that

$$N_g = N_c + N_w. \tag{23}$$

We apply each rule to all training data and find the accumulated N_c.

Case 3: there are more than one rule with the 'IF' part, which matches the data points in the training set. The most specific rule, which does not have "don't care" and all logical operators are 'AND', is chosen.

These rules will be ranked based on their fitness for the subsequent self-reproduction process. We then repeat these steps sequentially throughout the training data set for other 99 groups. Out of the 50 groups with F_i's below its medium, we randomly choose 25 groups for crossover, in which each group goes through the following process :

a) From the pool of 100 rules, we randomly select 2 rules to conduct crossover;
b) We register the rules in a memory;
c) From the second round of selection onward, we compare selected the rules with those stored in the memory;
d) If both rules have been selected as a pair before, then we repeat the selection process till picking a different pair. We repeat the process until we have formed 100 new rules.

The other 25 groups undergo mutation with at a rate of 4%, which means 1% overall in each generation. The same process is repeated for a preset number of generations, *e.g.*

1000. In each generation, only 50% of rules need to be evaluated for F_i's, therefore resulting in a 50% of CPU time saving. At the end, the best group of rules is selected for the purpose of testing of their forecasting accuracy with the subsequent part of the data.

5.4.5 Testing of the Methodology and Results

The intraday data of S&P100 between 1987 and Aug. 2003 was acquired from TickData Inc. Part of the data set, the 15-minute high-low prices between 1998 and 2002 is taken for the training purpose. The second part *e.g.*, between Jan. 2 and Aug. 29, 2003 is used to test the validity of the rules. The data are imported into a Matlab environment to calculate the corresponding normalized IV's. The IV values are then wavelet transformed to find the best tree. The GA programs are then applied to forecast the IV values at the selected time ranges ahead. In applying the GA programs, all rules are initially assigned to have zero fitness. The data range. $(-\infty,-a]$, $(-a, b]$, $(b, c]$, (c, ∞) are preset at $a = -0.3$, $b = 0$ and $c = 0.3$ to evenly distribute data into four ranges and to meet analyst's risk requirement. The data is then processed with the GA programs and the best group of rules is found. Upon completion of the training process, the best group of rules will be used to test the subsequent part of the S&P100 IV data to assess the forecasting accuracy, *i.e.* if each of the forecasted one-step-ahead point is at the same range as the actual data. To achieve calculation economy, the programs that involve GA are written in Java while the wavelet transformation is performed with Matlab. It could be observed from Fig. 4 and 5, (while it may be less obvious in Fig. 6 that

a) the forecasting accuracy is generally above 60% shown at the Y-axis in the respective figures, which is better than the traditional methods and matches those derived from the proprietary methods (Kinlay, 2001);
b) the forecasting accuracy is higher for the wavelet transformed series with higher scales compared with those derived on the original series *i.e.,* the non-transformed

ones. This may be attributed to the fact that variance of the original series is the sum of variances of its spectral components.

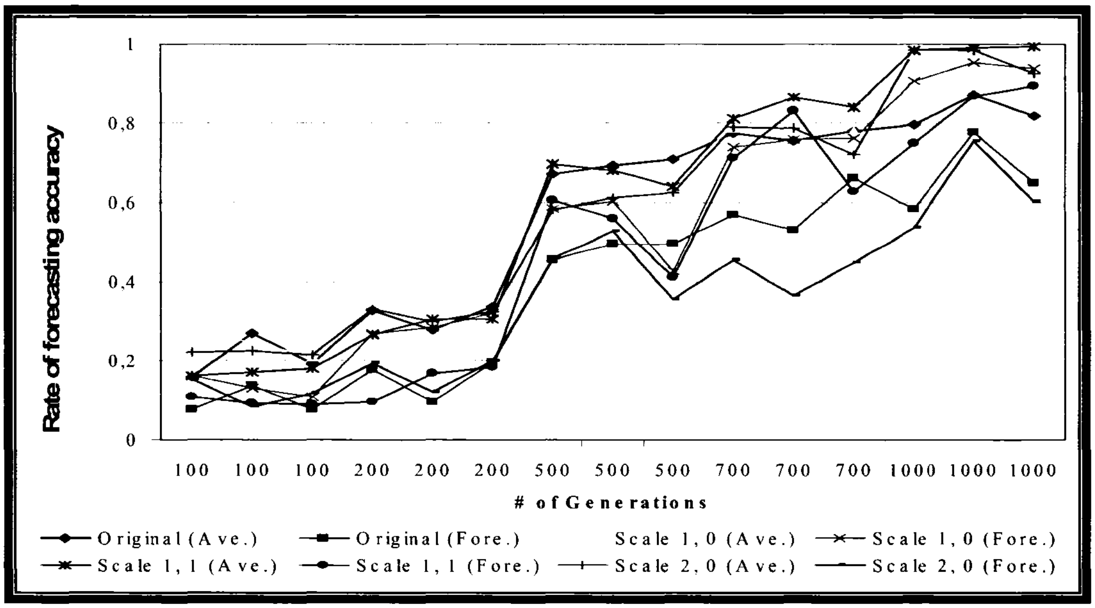

Figure 4 Daily 2003 S&P100 forecasting accuracy based on 2002 S&P100 data

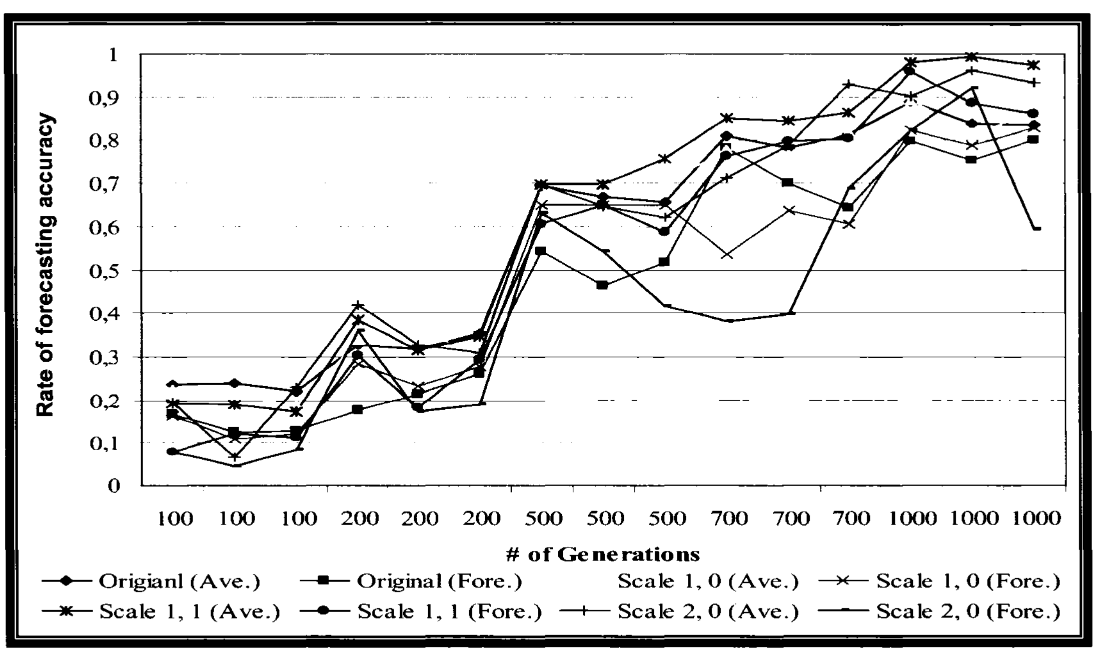

Figure 5 Daily 2003 S&P100 forecasting accuracy based on 2001-2 S&P100 data

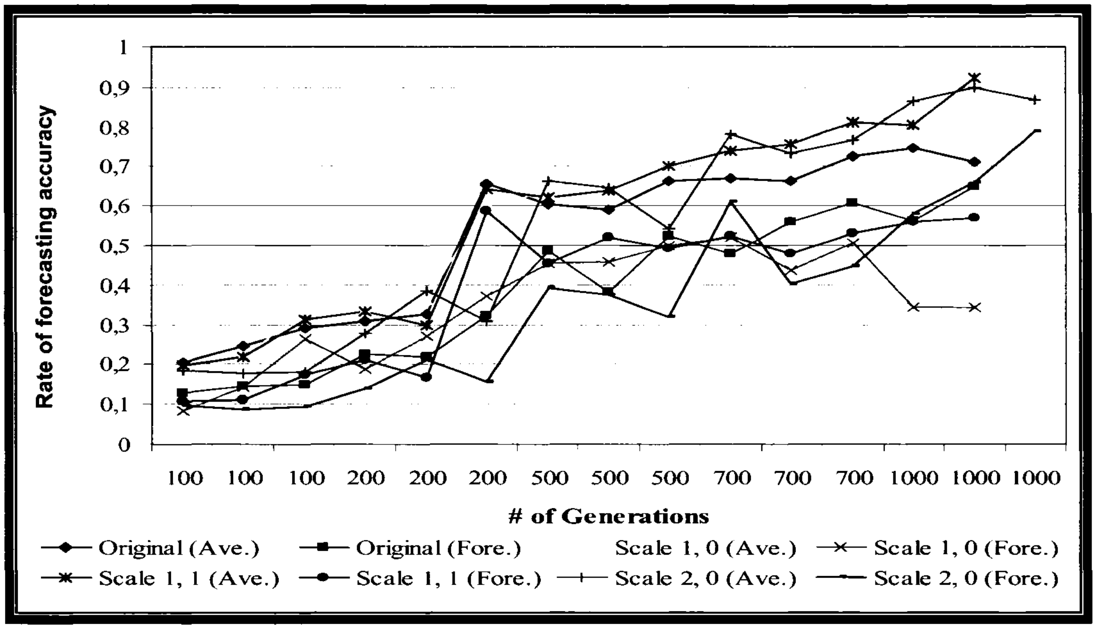

Figure 6 Hourly 2003 S&P100 forecasting accuracy based on 2002 S&P100 data

The same data in the same one-year, two-year and five-year time horizons as shown above in the current research are also processed with the GARCH(1,1) model. The forecasting values from GARCH(1,1) are first normalized to the respective logarithmic means and are then converted into values of 1, 2, 3 and 4 according to their amplitudes, similar to the preprocessing described in the previous section where GA's are used. The accuracy of forecasting is calculated based on the comparison between the converted data and the realized volatility in the validity-testing period *i.e.*, from Jan. 03 to Aug. 29, 2003. The one day ahead forecasting accuracy for the 2003 S&P100 daily data based on training sets at the selected periods are summarized as shown in Table II.

Table II

The one-day-ahead GARCH forecasting accuracy

Period	% Accuracy
2002	0.485
2001 – 2002	0.491
1998 – 2002	0.503

They agree well with the results derived in many contemporary GARCH as well as IV studies (Andersen, 1998, 2001), but are markedly lower than those achieved by using the GA method proposed and developed in this research thesis. It may be argued the GA method is superior to the GARCH approach simply because it takes more historical patterns linear or nonlinear, into consideration for forecasting purposes. Our current method takes the past four data points joined with the logical arguments 'AND' to form 1024 possible combinations in a training data set to find a relatively optimal group of rules. By taking the 'OR' argument and "don't care" class, even more combinations have been included.

In many contemporary research publications dealing on volatility estimation and forecasting, the dominant evaluation criterion is the coefficient of determination R^2 (multiple correlation coefficient), which simply represents the fraction of variability in y (the linear regression function) that can be explained by the variability in x (regressor) (Taylor, 1999). In other words, it explains how much of the variability in the y's can be explained by the fact that they are related to x *i.e.*, how close the points are to the line. Since we are not using a linear model in the current approach and since Taylor (1999) has demonstrated some of the drawbacks of using R^2, we adopt instead the percentage of correct forecast as the measure of accuracy.

5.4.6 Discussions on the Present Methodology

The use of the GA four-lag forecasting algorithm is based on the stylized fact of volatility clustering that has been the foundation of the GARCH approach. This method helps analysts identify different patterns in the time domain even if those patterns are abrupt jumps or drops. Since it is non parametric and free of any pre-assumption, it is more flexible and robust to deal with non-linearity. The wavelet transform enables analysts to study the volatility patterns in different frequencies (time horizons). The combination of these two techniques opens up a broader field for analysts to explore different properties hidden in the volatility series. The fact that for some time horizons the forecasting accuracy is higher for the wavelet transformed series with longer time ranges implies that it could be beneficial for analysts to focus on those time ranges with lower entropy.

With such a speedy processing, one could afford to increase the number of lags from four to five or even higher, and increase number of the amplitude ranges from the current four to six to be more precise in forecasting the volatility values. In such a case, the need to de-trend the data by normalizing the data with the 21 day moving average could be diminished.

5.5 Forecasting by Genetic Programming

In this Section, we intend to take the forecast process one step further – instead of using GA's to forecast the moving direction and range of IV, we will apply GP to forecast the value of IV. GP was first developed as an extension to GA's. The most important feature of GP is their representation of individual solution structure. Unlike traditional GA's, which usually represent individuals as vectors of fixed length, GP individual is represented as hierarchical composition of tree-like structure with variable length from basic building blocks called functions and terminals. The function set is composed of the statements, operators, and functions available to the GP system. The terminal set is comprised of the inputs and constants to the GP program. GP possesses no inherent limitations on the types of functions, as long as the closure property is satisfied, that is each function should be able to handle gracefully all values it might receive as inputs.

Traditionally, GP uses a generational EA. In this approach, there exist well-defined and distinct generations. Each generation holds a complete population of individuals. The newer population is created from the genetic operation on the older population and then replaces it (Chen, 2003). GP-resulting specifications may be viewed as coincidental equations that may capture the dynamics of a process. Coefficients in GP-evolved equations are not computed but randomly generated. It is therefore, advantageous over the conventional statistical regression, *e.g.* robust against problems of multicolinearity, autocorrelation and heteroscedasticity. There are also no degrees of freedom lost to compute the coefficients. However, because variables and operators selected to assemble the equations are random, while attempting to breed the fittest individual equation, the program occasionally gets trapped in local optima rather than a global one within the search space. It is therefore, necessary to generate a large number of equations and then select the best ones for forecasting as suggested by (Kaboudan, 2005).

GP is fundamentally a computer search and problem-solving methodology that can be easily adapted for use in non-parametric estimation. It has been shown to detect patterns in the conditional mean of foreign exchange and equity returns that are not accounted for by standard statistical models as shown in the corresponding references listed in Table 2.1 and others (Neely, Weller, and Dittmar 1997; Neely and Weller 1999; Neely 2000). This suggests that a GP based method as presented here may also serve to be a powerful tool for generating predictions of asset price volatility. For a summary of the basic concepts of GP, please refer to APPENDIX 6 at the end of this thesis.

5.5.1 Forecast of IV of a Financial Time Series

As indicated by Radzikowski (2000), modern parametric option-pricing models, where volatility is often the only stochastic variable, were expected by many researchers as well as end-users to :

a) Be well-specified,
b) Consistently outperform other models,
c) Be statistically consistent with underlying asset return dynamics,
d) Provide a statistical theory of option pricing error, and
e) Be elegant and not difficult to estimate

But they have uniformly failed to deliver against these expectations, as they either are too complex, have poor out-of-sample performance, make unrealistic distribution assumptions, and/or use implausible and/or inconsistent implied parameters. While parametric models provide internal consistency, they do not out-perform simplistic approaches out-of-sample. Even the most complex modern parametric models are imperfect and are outperformed by simple, less general models. Jackwerth and Rubinstein (2001) applied series of tests to a variety of models and concluded that naïve approaches are consistently the best, stochastic deviation models are the next best, then there are deterministic volatility models that follow and then comes finally the traditional parametric models.

5.5.2 Problem Statement of IV Forecast

In this Section, we will attempt to forecast the numerical values of the volatility by formulating a nonlinear and non parametric approach based on GP in the TSDM framework. Different patterns, linear or non-linear including the stylized clustering effect of volatility may repeat in different time intervals. This is true when dealing with different types of financial securities or dealing with different historical periods for the same underlying security. By making use of the stylized characteristics of financial volatility, we extend the TSDM method with GP to forecast as many events/non-events as practically feasible in the IV time series in order to guide option trading.

5.5.3 Data Analysis and Final Results

The intraday data of S&P100 index between 1987 and Aug. 2003 is acquired from TickData Inc., and the 15-minute high-low prices between Dec. 3, 2001 and Dec. 31, 2002 are taken for our training purpose. The second part, *e.g.* between Jan. 2 and Aug. 29, 2003 is utilized to test the validity of the rules. The first 21 days of both sets of data are used to prepare for the 21-day moving average, in order to take the monthly effect into consideration, to de-trend (or normalize) and to improve the forecasting accuracy. The corresponding normalized IV's were then calculated and fed into the GA's to forecast the moving directions and to find the best 100 rules by maximizing the value f (Ma *et al.*, 2004b). The GP programs are then applied to forecast the IV values at the selected time ranges ahead, *e.g.,* one period ahead.

The execution cycle of the generational GP algorithm consists of the following steps:

1). First, we initialize the population. An initial population of 100 is created randomly from the basic building blocks.
2). We then evaluate the individual programs in the existing population. A value for fitness, *e.g.* the absolute difference between the individual and the desired one is assigned to each solution depending on how close it actually is to solving the problem (thus arriving to the answer of the desired problem).
3). Until the new population is fully populated, we repeat the following steps:
    a. We select an individual or individuals in the population using the selection algorithm

b. We perform genetic operations (crossover & mutation) on the selected individuals

c. We insert the result of the genetic operations into the new population.

4). If the termination criterion is fulfilled, then we continue. Otherwise, we replace the existing population with the new population and repeat steps 2-4

5). We can then present the best individuals in the population as the output from the algorithm.

Table III

GP Configuration

Parameter	Values
Generations:	25/50/100
Populations:	100
Function set:	+, –, %, *, sin, exp, sqrt, ln
Terminal set:	$\{x(t\text{-}1), \ldots x(t\text{-}4)\}$
Fitness:	difference between actual and desired
Max depth of new individual:	6
Max depth of new subtrees for mutation:	6
Max depth of individuals after crossover:	17
Muataion rate:	0.05
Generation method (selection):	50%

Table III lists the parameters used in this research. Note that based on the findings in Neely and Weller's research publication (2001), the fitness of the GP operation in the current investigation is derived from the Mean Absolute Error (MAE) between the generated individual and the actual IV value. As found by Park (2002) while testing an MAE based GARCH model (so called robust GARCH), MAE tends to generate results superior to GARCH models that are MSE based. Two separate tests have been conducted on the 2002 training data set by pre-processing them with the GA's (Ma *et al.*, 2004*b*), one for 500 generations and the other 1000. The intermediate results are then passed through GP programs and the final results of percentage accuracy are randomly selected and shown in Table IV. For example, an initial population of 100 rules is generated and 25 generations of GP are performed with a maximum depth of six of new individuals.

Table IV

Random samples of forecasting accuracy for 2003 IV data (2002 training data set)

GP Parameters	2002 data set (500 generations GA)	2002 data set (1000 generations GA)
[25, 100, 6]	72.65, 73.21, 74.40	74.23, 66.00, 68.77
[50, 100, 6]	71.46, 75.36, 74.46	76.77, 69.13, 67.54
[100, 100, 6]	71.44, 69.60, 68.42	68.09, 67.49, 67.17

The 2001-2002 training data set was then pre-processed using GA's (Ma *et al.*, 2004*b*) and 1000 generation GP was implemented to obtain the results as shown in Table V. These results are randomly selected from the respective test groups for sake of brevity.

Table V

Random samples of forecasting accuracy for 2003 IV data (2001 – 2002 training data sets)

GP Parameters	2001/2002 data (based on 1000 generations GA)
[25, 100, 6]	78.25, 77.66, 78.25
[50, 100, 6]	80.20, 79.51, 79.54
[100, 100, 6]	79.10, 78.86, 78.98

An interesting phenomenon could be observed that the forecasting accuracy in the current tests is not positively related to the number of generations used in either GA's or the subsequent GP operations. This may be caused by the early convergence to the local minima in the search process. Further investigation and appropriate search strategy may be necessary to resolve the issue.

5.5.4 Discussions

By working in the novel TSDM framework with the associated methods, tests conducted on the proposed methodology developed in this research have made use of GP to find optimal temporal pattern clusters that both characterize and predict time series events. Additionally, a time series windowing techniques is adapted to allow prediction of non-stationary events. Results of the tests have demonstrated that the modified TSDM framework successfully characterizes and can be relied on to predict complex, non-periodic and irregular IV time series. This was done through testing the S&P100 index of different years. The one step ahead forecasting accuracy reaches an average of 74% with standard deviation of 4.6%. This accuracy as well as reliability of forecast in market applications is considered well above average.

The forecasting accuracy achieved with GP is, however, somewhat lower than those derived by GA's not only in terms of absolute level but also on the trend of convergence. As shown in Table V, accuracy did not improve with the increase of the number of generations. One reason may be that the forecasting accuracy of GP in the current case is built upon results of the GA processing. Errors from the GA part may affect the subsequent GP operation. This simply demonstrates the high level of difficulty of pinpointing a value of future IV. And it leads us to believe that some more work needs to be carried out before this part of the program can be successfully used in practice. Local search algorithms such as conjugate gradient technique proposed by Zumbach *et al.* (2001) described in Chapter 2 could be helpful in tackling problems such as this one. On the other hand, the GA part seems to be more robust and reliable, therefore it is recommended for immediate application. Future work in this regard may include, *a*) incorporating the modified TSDM*e2* based GP directly with wavelet transform, which might lead to improved forecasting accuracy, *b*) using parallel processing techniques to accelerate GP process, and *c*) testing other financial indices over a wider time span, *etc.*

These are just a few suggestions that may fill the missing links still left in the complex problem of market index forecasting.

CHAPITRE 6

VALIDATION WITH THE DSO METHOD

The primary objective of this chapter is intended to make use of DSO Algorithm 2 that combines GA's and the discrete stochastic optimization theory in order to compare and further to validate the IV-wavelet-GA approach which has been established in the previous chapters, but would need a rigorous mathematical proof in order to bridge such a traditional short fall. The key element in such a validation process is to transform the IV time series data set into a cross-sectional one. The validation process would then involve the forecast of the one-step-ahead directions and ranges of the IV series derived from the S&P500 index. The results hence obtained are expected to have a similar nature to what have been achieved in Chapter 5 *i.e.,* the accuracy in forecasting the S&P500 volatility is expected to be in the same order as that was derived in forecasting the S&P100. In Section 6.1, 6.2 and 6.3, we have presented the forecast for the S&P500 2004 volatility by means of GARCH(1,1), compared critically the hence obtained results with those derived from applying the DSO method and then attempted to validate those DSO results with the bootstrapping tests. In Section 6.4 and 6.5 a general discussion on the approach taken is presented with respect to its limitations and potential applications.

6.1 Forecast with GARCH(1,1) method

The following section describes the computational phases involved in this GARCH forecast. The data set S&P500 index from Olsen & Associates Inc. is processed employing the Matlab applications.

The 2002-2003 and 2002 index data are used to train the algorithm, whereas the Jan. 2-Oct. 21, 2004 data are used as the out of sample data to test the derived rules. The training data sets are first processed to derive the parameters for the GARCH(1,1) model. The data in 2004 are then fed to the resulting GARCH(1,1) model in order to obtain the one-day-ahead forecasted IV values. The same data are then used to compute the (realized) IV values for the 2004 data set. Both groups of results are normalized to the respective logarithmic means and then converted into values of 1, 2, 3 and 4 according to their amplitudes, similar to the pre-processing described in the previous sections where GA's are used. The accuracy of forecasting is computed based on the comparison between the two groups of the converted data. The one-day-ahead forecasting accuracy of 2004 based on training sets corresponding to 2002-2003 and 2003 alone is thus obtained and is shown in the following table:

Table VI

The one-day-ahead GARCH forecasting accuracy for the 2004 S&P500 daily data

Training Period	% Accuracy
2003	43.9
2002—2003	35.6

They agree moderately well with those results derived in a number of contemporary GARCH as well as IV studies (Andersen, 1998, 2001), and they validated our data pre-processing procedure, but are markedly lower than those achieved by using the GA-DSO method proposed and developed in this research work. The GA-DSO method can thus be shown to be superior to the GARCH approach simply because it takes more historical patterns, linear or nonlinear, into consideration for forecasting purposes. More pertinent details regarding the comparison are given in the following sections.

6.2 Data Processing with S&P500

To apply the DSO method for the validation purpose, the same IV time series S&P100 derived in Section 5.1 is converted into wavelet coefficients by following the same procedure as those discussed in Sections 5.2 and 5.3. Based on the entropy level, coefficients at node (2, 0) in the wavelet tree are selected for the Algorithm 2 to process with Matlab applications. To form the initial population, 100 groups of rules are randomly generated with each consisting of 100 rules. GA operations such as selection, cross-over and mutation are then applied to find ultimately the best 100 rules among all rules regardless of which groups they belong. Initially, up to 500 generations have been imposed as the stopping criterion according to the experience gained in previous tests, but it was soon found that results tend to converge after 75 generations. Fourteen runs of tests each with 75 generations have been performed. For the sake of brevity, results of the last five are plotted in Fig. 7 as shown in the following.

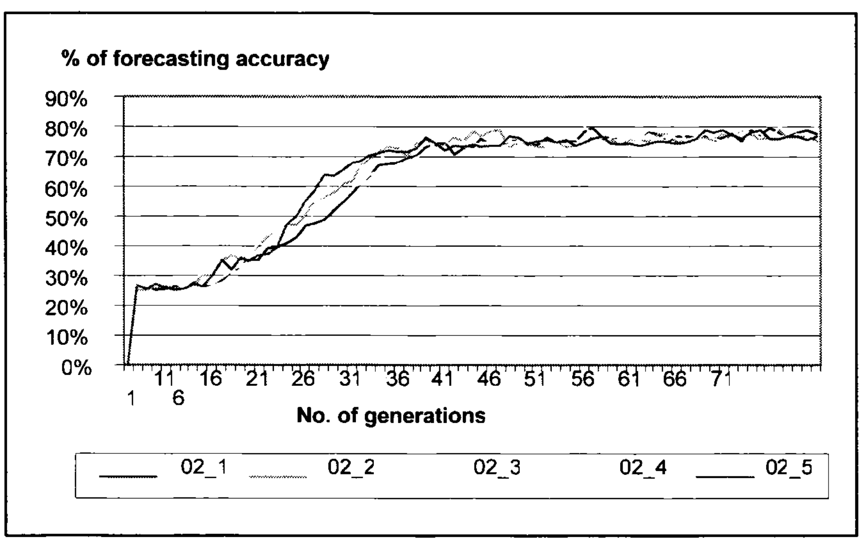

Figure 7 Daily 2003 S&P100 forecasting accuracy with 2002 data as training set

The data set S&P500 index from Olson & Associates Inc. is then processed. The price data in 2003 is used for training purpose, while the 2004 data are used to test the optimized rules. The same procedure is then repeated with the 2002 – 2003 data as the training set. Again, tests with up to 500 generations have shown that results tend to converge after 75 generations. Please refer to Fig. 8 for a display of selected results of the last few runs for the two sets of tests.

One could easily observe that the forecasting accuracy reaches well above 70% for both sets of tests. One advantage of using the DSOM algorithm 2 is its fast convergence due mainly to the usage of far larger amount of memory to record all the rule history. In contrast, simple GA's discussed in Chapter 5 made used of great deal of loops generation over generation. It is interesting however, the forecasting accuracy for the 2003 data set is generally higher than those derived based on 2002-2003 data sets *e.g.*, above 80% versus above 70%. Such a phenomenon actually agrees well with the widely documented observation in volatility time series – the lesson being that the more recent events carry more weight to the current reading than those further back in the past.

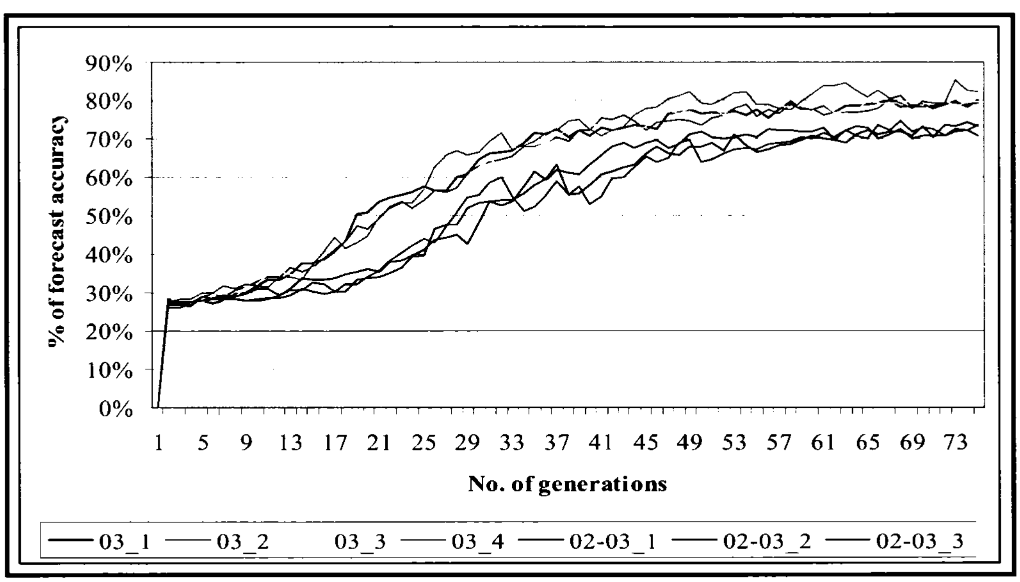

Figure 8 Daily 2004 S&P500 forecasting accuracy trained by 2002-3 and 2002 data

6.3 Statistical Validation

A confidence interval is usually needed to depict an accurate estimate of a random variable, which is supposed to include the true value of the variable with a specified probability. Bootstrapping provides a ready and reliable way to construct such a confidence interval that does not depend on the asymptotic normality assumption. This is important when the population consists of just a few observations. We have used Carpenter and Bithell's (2001) approach to process the two 14 forecasts that were respectively derived from the one-year data of S&P100 and S&P500. The bootstrapping samples are chosen to be 999, confidence level to be 0.05 and statistical function to be MEAN. The results are 0.7734 [0.7659, 0.7794] and 0.8144[0.8067, 0.8233] respectively. The 0.7734 forecasting accuracy of the S&P100 IV matches the result that was achieved with the GA alone as shown in Chapter 5. However, the current 75 generations versus then 1000 generations represent more than 10 folds reduction of <u>*CPU*</u> time.

6.4 Discussion of the Results

In comparison with the GA operation in Ma *et al.* (2004*b*) which repetitively loop through Step 1 to 3 in Algorithm1 without keeping all tested rules in memory, the current GA-DSO approach has the following advantages:

a) Evaluate fewer rules per iteration by avoiding re-tests of the rules stored in the memory, thus resulting in greater computation efficiency;
b) Use tournament selection method instead of the ranking method, hence minimizing the risk of early convergence;

c) Take all historical rules into account and not just those retained in the last generation while optimizing their respective final fitness values, hence improving forecasting accuracy.

It is therefore, obvious that results obtained in this chapter validate the following two claims:

a) The DSO procedure produces forecasting results at the same or higher accuracy levels in comparison with those obtained by using GA alone
b) The DSO procedure that incorporated GA produces results by processing IV's of S&P100, which are consistent with those by processing IV's of S&P500.

In other words, the current set of experimental tests coupled with the analysis developed in Chapter 4 confirm that the data conversion of a regular time series into the 4-lag recursive data set based on the TSDM framework could indeed be treated as a Markov chain. And the application of GA could indeed be substantiated by the DSO theory. For those who intend to try a test drive based on the approach established here, more numerical tests will definitely help them obtain a better grasp on the methodology and the claims made herein. However, the experimental nature of EA's thus dictates that no two tests would yield identical results, whereas slightly different results would not materially undermine the conclusions that we have reached here.

Another advantage of the current approach is characterized by the ease of assessment on the result quality, in which a direct comparison of percentage accuracy would suffice. This would have created a problem to most traditional statistical models. First, there are several quality measures to choose from *e.g.*, as simple as error measurement of Mean Absolute Error (MAE) or MSE. Second, bias of both realized volatility (the target variable) and volatility forecasts appears if the return intervals chosen are too small due to the microstructure of the market. Volatility forecast tests are affected by this bias. A treatment of the bias is almost inevitable when designing volatility forecasting models and tests based on high-frequency returns over intervals of less than an hour (Corsi *et al.* 2001). Due to these technical difficulties, there is not yet a comprehensive study of high-frequency volatility forecasts with particular focus on their qualities (Dacorogna et al. 2001).

One of the reasons for Andratottir's global optimization method adopted here to be superior to the local optimization method is attributable to the nature of the converted data sets and the objectives of the current tests (Andradottir, 1999). Assume that $\theta_M^{*(1)}$ and $\theta_M^{*(2)}$ are ranked top one and two rules among the top 100 rules after the terminal generation. Based on the initial assumption, $\theta_M^{*(1)}$ will be the global optimal solution of the problem, while $\theta_M^{*(2)}$ could be a solution that locates as a state next to $\theta_M^{*(1)}$. Even in such a case, the importance of $\theta_M^{*(2)}$ is not diminished. This is mainly because the prediction accuracy calculated here is based on those qualified rules and not the absolute number of data points in a data set. In fact, the discrete nature of the rules dictates that different rules characterize different patterns of data points. Our objective is to find those patterns that match the four lags in the IV data set for the forecast of the fifth point. Rules corresponding to better fitness values would naturally have better chance to achieve higher forecasting accuracy in the entire data set regardless of the relation among rules if there is any. Therefore, rules derived from the global optimization method have a better generality than those derived from the local optimization method. For further proof of such a claim, it may be worthwhile to determine the effect of using different number of lags in forming rules and different values of *a, b, c* and *d* in classifying the IV data ranges on the prediction accuracy. Upon implementing these procedures, we anticipate that the current methodology would become quite complete and ready to support a commercial software package for the use of general risk forecasting in a wider industrial fields. However, these tasks are beyond the objective of this thesis research and are recommended for future expansion of our concepts.

6.5 Application of the Results

6.5.1 Volatility – from an Option Trading Perspective

The Black-Scholes model indicates that the price of an option is a function of the stock price, exercise price, risk-free rate, time to expiration, volatility, and any dividends on the stock over the life of the option. Of these six variables, the stock price, exercise price, and time to expiration are easily observable. Hence, one could easily measure these without introducing any appreciable error. The risk-free rate is largely observable, and its impact is small. The dividends are not observable, but they are not too difficult to measure accurately. The volatility, however, is almost completely unobservable. And this explains why the implied volatility in the Black-Scholes model is more or less a catch-all term, capturing whatever variables are missing, as well as the possibility that the model is improperly specified or blatantly wrong.

The Black-Scholes model is a critical component in the modern option pricing theory, but it tells us nothing about why anyone would hold an option. The very fact that it ignores the stochastic nature of the volatility variable means that any option serves as well as any other option and it therefore, cannot motivate the holding of options. In reality some options are more desirable than others. Whatever factors that motivate the holding of options are simply not captured in the Black-Scholes model. Hence, these factors show up hidden within the implied volatility.

Other researchers believe that the volatility smile as defined in Appendix 1 reflects stochastic volatility. Volatility is surely not constant as assumed in the Black-Scholes model. If volatility is stochastic, researchers argue that the smile reflects the failure of the Black-Scholes model to capture the random nature of volatility. Others argue that the Black-Scholes model, which assumes that stock prices fluctuate in a smooth and continuous manner, fails to capture the true nature of stock price movements, which are observed to have discrete jumps. However, practitioners seem capable of operating in a world of volatility smiles. They even use the smile to simplify how they trade. For example, they oftentimes quote option prices not in terms of the actual price but in terms of the implied volatility. More specifically, a dealer might indicate an intention to sell the January 36 call by quoting a volatility of 25.92. This statement is interpreted to mean that the actual price is derived from the Black-Scholes model using a volatility of 25.92. Assuming agreement on the dividends and risk-free rate, such a quote for this option would lead to a price of $1.85. By quoting prices this way, traders can immediately see which options are truly more expensive, that is, after accounting for moneyness (how much the security price is in, at or out of the strike price), time to expiration, and whether the option is a call or a put. The phenomenon of different implied volatility values for the same security simply reflects another aspect of the stochastic nature of volatility. It is obvious that volatility has to be evaluated for individual cases in most of the practical applications. The methodology established in this thesis help analysts determine values of volatility for different time horizons at a higher confidence level and at a more efficient speed. Put differently, it helps confirm the stochastic nature of the volatility of equity indices – volatility of volatility. The proposed approach could deal with different volatility regardless of its distributions or other random nature such as the smile effect. The recently available option chain data enables analysts to compile the implied volatility series for a selected security. The IV-wavelet-EA method would process the historical time series as described above and would provide implied volatility forecasts as required. More details could be found in the subsequent sections in this chapter about the concept of volatility curve.

6.5.2 Potential Applications of the Proposed Methodology

Volatility estimation is in the heart of option valuation for securities such as equity, indices and fixed-income. As a result, a large and growing body of literature that emerges recently in this area has proposed a variety of approaches. Their two primary limitations are firstly the contemporary estimation techniques do not allow inference of the volatility of the underlying security's price movement in an arbitrary term. This restriction might not be a concern for applications targeted at real-time security trading. However, since one usually relies on a variable time-scale in the formulation of stochastic programming (*e.g.* short time-steps for the near terms and much longer time-steps for the longer terms) in strategic risk management applications, the prevalent estimation techniques may not be suitable for users to cope with wider time horizons *e.g.,* banks, hedge fund management firms and other major investment institutions. The second limitation is that the prevalent approaches require an *a priori* assumption of a particular functional form for the estimated volatility curve, which is undesirable under most practical circumstances (Marti, 2004).

For evaluating different types of securities particularly fixed-income securities and equity options that have expiration dates, besides the observable quantities such as the term structure of interest rates (yield curve), one would also need the term structure of volatility. Like the yield curve, the term structure of volatility describes the relationship between the term to expiration of the equity options and their corresponding volatility (*i.e.* the volatility of 3-day, 6-days, 1-month, *etc.*).

Extended from the discussion thus far, it becomes apparent that another advantage of current IV-wavelet-GA approach is that by incorporating the wavelet transform (Ma *et al.*, 2004*b*) one could easily construct a yield-curve-like volatility curve. Conversely, she could use it as a guideline to determine the forecasting horizon based on the values of entropy of the respective wavelet coefficients. This will be more reasonable than arbitrarily selecting the forecast horizon based on analysts' experience or requirements as used in other prevalent approaches. With an improved prediction of volatility, besides trading options we could also trade volatility itself. One way to do so would be to use the newly formed VIX Futures in the CFE (Chicago Futures Exchange), namely VXB. Refer to APPENDIX 2 for details.

6.5.3 Applications on a Winder Perspective

In a wider perspective, risk management as a field has evolved and expanded, more so in the recent years. Few years ago it covered credit, individual security and market risk. Today, it also covers operational risk, fraud risk, and several other dimensions. Based on a survey conducted in the SAS Institute, all of the interviewees note that the underlying growth in the number and volume of financial risk-management securities and contracts are the fundamental driving force for continued growth in this field. One of the most popular approaches to risk measurement is by calculating what is known as an institution's "Value at Risk" (VaR), an estimation of likely losses that could arise from changes in market prices. More specifically, it is the money-loss in a portfolio that is expected to occur over a pre-determined horizon and with a pre-determined degree of confidence. Brooks paper (2003) illustrated the directly proportional relationship between volatility forecast and financial risk management *e.g.*, the large role played by the time-varying volatilities in minimizing VaR. The application of the current forecasting strategy is therefore, not limited to financial data forecasting or security trading, but have broader implications. For example, banks could use the technique to detect credit card frauds by following the spending patterns of their customers. Lending organizations could evaluate the bankruptcy risk of certain types of firms based on the selected criteria. Insurance companies could be in a better position to determine the risk of certain group of individuals by better predicting their behaviour. In short, the strategy of employing multiple engineering techniques to simplify and further to analyze time series could have a great potential in the data mining field.

In summary, we have incorporated Povinelli's (1999) TSDM framework with carefully selected engineering techniques such as wavelet transforms and GA/GP, substantiated the operation with DSO and consequently formulated a systematic approach to achieve markedly improved effectiveness and efficiency in forecasting volatility, which was not available till now. Volatility modeling and forecasting are one of the major obstacles in contemporary computational finance and econometrics. Using a data mining/EA approach to help effectively solve such a problem should be considered a meaningful, timely and important contribution in the field.

CONCLUSION

The current thesis research was pursued after an extensive and exhaustive survey of the relevant literature, published and otherwise, that deal with the estimation and forecast of financial time series such as rates of return, foreign exchange rates, variances as well as volatility by means of various GARCH models, IV calculation, EA's, neural network, non-linear parametric modeling, *etc.* In analyzing the rationale and results of this set of scientific publications, it became clear that a more systematic and solid approach making use of a number of versatile engineering tools and related analytical methods could account for more types of patterns, and thus could yield higher forecasting accuracy. For this purpose, we have set our objectives to formulate a methodology that can be developed to make it applicable to forecast volatility of equity indices and potentially other investment vehicles. In other words, the proposed approach was specifically targeted to embody the following vital characteristics, *i.e.* to be:

a) assumption free – no need to assume normal or any other statistical distributions associated with the time series and its estimation errors;
b) absent from problems of bias/consistency/standard-error-estimation during modeling process;
c) more flexible – not limited by the parametric structure;
d) more accurate on the current and hence the forecasted volatility – the conversion of IV by transforming volatility from a latent variable into an observable one; and finally
e) more efficient – in both stages of data pre-processing and analysis.

We therefore, decided to establish for the first time, a methodology by judiciously incorporating IV estimation, wavelet transforms and EA's to estimate and forecast financial volatility series that are known to have abrupt changes at accuracy higher than those achieved by even the proprietary research methods.

At the stage of the completion of this thesis research, we believe hat we have generally achieved each of these goals while establishing a systematic approach characterized by robustness, flexibility, efficiency and accuracy and we are able to show that each of these characteristics has been attained at different stages of the current analysis process.

By applying the wavelet transformation, we have achieved economy in computation time and better flexibility in selecting the optimal forecasting horizons based on the values of entropy of the respective wavelet coefficients. And the transformation allows one to study different investing behaviour in different time scales (horizons) independently, because different investing styles may cluster into different time horizons. Another advantage when using the wavelet method is its robustness due to the absence of any potentially erroneous assumption or parametric testing procedure.

The robust nature of the approach, particularly in dealing with non-linearity is further strengthened by the adoption of EA's due to its four main associated strategies – direct use of coding, search from a population, blindness to auxiliary information, and randomized (stochastic) operators. The experimental tests on equity indices of the S&P100 and 500 have been shown to reach a forecasting accuracy level of above 60%, which is superior to those derived based on publicly available methods.

The implementation of EA's in the TSDM framework provides an analytical basis to better illustrate the data transformation operation, in which a sliding window converts a wavelet coefficient series into a four-lag recursive cross-sectional panel data set. We have further proved that such a data set is in fact Markovian, whereby such a proof of concept allows for the rigorous mathematical proof and justification for the application of EA's in the current research work. And such a substantiation operation, being one of the main contributions of this research thesis, is a unique proposition that characterizes and sets it apart from the majority of contemporary publications. The introduction of DSOM with EA's further improves the computational speed by a factor of greater than

10 times in comparison with that of just a simple application of EA's. Forecasting accuracy is also improved by evaluating more rules kept in the memory. The subsequent experimental tests on equity indices of the S&P100 and 500 at different time horizons demonstrate the merits of the proposed methodology not only in forecasting accuracy in terms of direction/ranges (> 74% higher than those based on proprietary methods) and actual values of IV (>70%), but also in flexibility of forecasting horizons as well as computational economy. The bootstrapping tests that follow further validate the results derived in these experimental tests, and thus complete the process of establishing the validity of our proposed fresh engineering approach in forecasting financial index volatility series.

The proposed methodology enables all practitioners, both scientific and market-based, to capitalize on the associated economical benefits related to the following four key aspects:

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

More specifically, based on our results, the 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, more cost effective hedging strategies could be designed and implemented and better portfolio performance could be achieved through optimizing risk management. This has been our objective in the pursuit of this doctoral research work in this hybrid interdisciplinary field.

RECOMMENDATIONS

In the following three sections, we will layout the guidelines for researchers to conduct future work on the proposed forecasting methodology to make it universally more valid, functionally more powerful and extensively more useful.

Recommendation 1: To alleviate the existing limitations

Certain limitations listed in this section were encountered during the derivation of the general approach, but they neither fundamentally affect its validity nor tangibly jeopardize its completeness as defined in the pre-set objectives. For the sake of pragmatic considerations, the work to alleviate them has been left to the future stages of research on this topic.

Besides other easily satisfied assumptions such as irreducible, aperiodic, *etc.*, Andradottir's (1999) local and global search methods are based on the assumptions of initially time non-homogeneous but asymptotically homogeneous Markov transition matrix as $m \to \infty$. We accepted such a principle of time averages for non-homogeneous Markov chains to be applicable to our case, because of the vast number of data points available as the historical time series for volatility forecasting study. Furthermore, the 100 rules found by GA's are derived from matching the 100 most popular patterns in the entire IV data set through an evolutionary process. However, the 100 popular patterns may not necessarily be found at a steady-state because of the ultimately limited number of generations and size of the available data set. In other words, the required conditions for applying Andradottir's method are stronger than what we actually possess. Yet, Andradottir's approach provides the basic mathematical foundation for the current research and the starting point for further development.

The second limitation is closely related to the one listed above. Andradottir assumed that unbiased estimates of the objective function values are available. In particular, if $X_1(\theta),\ldots,X_L(\theta)$ are independent observations of the random variable $X(\theta)$, then $\sum^{L}_{l=1}X_l(\theta)/L$ is an unbiased estimate of $f(\theta)$ – the expected prediction accuracy for all $L \in N$ and $\theta \in \Theta$. At current time we will need to accept tacitly (in order to move forward) such an assumption prior to the EA operation with a limited data set, which leads to limited *m,* thus limited *L*. Both of these afore-mentioned limitations may not affect materially the effectiveness of our proposed approach, because of the nature of EA's, *e.g.* the great number of rules generated over many generations, which covers far more ground than the traditional GARCH method. In the future, building upon the current system which relies on the central limit theorem, we could develop further mathematical proof in order to alleviate limitations such as the potential bias of $f(\theta)$ and the non-homogeneous Markov transition matrix at limited *m*. Furthermore, the current IV-Wavelet-GA method could be applied to test other financial indices, such as S&P/TSX composite index, foreign exchange rates, futures, *etc.* in order to validate further the high accuracy achieved with the current tests.

Another issue is that our current method has not accounted for the leverage effect, because we only focus on the volatility itself and not in relation to the absolute value of the underlying security. The same volatility value may have larger impact when the security value is small. To counter such an effect, practitioners could use the Sharpe-ratio-like volatility, *e.g.* value of the volatility normalized by the corresponding security value itself.

Recommendation 2: To extend the existing functionality of the method

There is little doubt that volatility is possible to forecast on a relatively high frequency basis, such as hourly or daily. Interestingly however, much less is known about volatility forecastability at longer horizons – the pattern and decaying rate as one moves from short to long horizons. Thus, open and key questions remain for risk management at all but the shortest horizons. How forecastable is volatility at various horizons? With what speed and pattern does forecastability decay as horizon lengthens, or is longer-horizon volatility approximately constant? By decomposing the IV time series into wavelet coefficients, one could study at the scale levels and clarify some of these doubts. For example, wavelet decomposition of IV time series could help construct a yield-curve-like volatility curve that guides the determination of the forecasting horizon, which could vastly advance researchers into the position to answer majority of these questions. One of the ways to construct such a volatility curve is to include simultaneously different lags in windowing, *e.g.* 3-lag, 5-lag, 6-lag or more in order to yield rules that cover even more variety of patterns. And these data sets including the original 4-lag one could be configured to deal with hourly, weekly or even monthly data.

At Step 2 of Algorithm 2 in Chapter 4, the selection criterion between two rules is based on the percentage of correct prediction rate when the rule is qualified, *i.e.* the first four fields of the rule match the four-lag recursive points in the data set. It might be interesting to find out what if the criterion is based on the percentage of correct prediction when each time the rule predicts whether the rule qualified or not. It would also be interesting to try different ranges in classifying the values of IV data points *e.g.*, different values of *a, b, c* and *d.*

In using the GA forecasting method, we have attempted to convert the Matlab applications into Java format. Preliminary study shows that as much as 20 times of *CPU* time saving could result from such a conversion. Over 10 times of efficiency improvement has also been achieved with the use of DSO algorithms. Better performance is expected if other more elaborate strategy is implemented inside the relevant programs. With such a faster processing, as pointed out in the previous paragraphs, one could afford to increase the number of lags from four to five or even higher and increase number of the amplitude ranges from the current four to six to be more precise in forecasting the volatility ranges. In such a case, the need to de-trend the data by normalizing the data with the 21 day moving average could be removed. A richer variety of rules should better characterize the different behaviors of a time series. In other words, we have laid the foundation to construct a software package that can automatically obtain suitable number of rules covering different lags in order to forecast the value of the *n*-step-ahead volatility.

Recommendation 3: To improve the performance of the method

Further testing and more trials would be unquestionably desirable to quantify the benefits brought about by the different wavelet packet techniques, such as compressing/thresholding. A more systematic method may be sought to replace the current trial and error approach in determining parameters in conducting the wavelet analysis, *e.g.* level of tree, order of filter, *etc.* More work may be needed to formulate an appropriate strategy in selecting the suitable types of filter, because different patterns of different types of data/indices could be better captured by suitable filters. By implementing a more organized system in managing the function selection process, performance of the general approach could be further improved.

The reliability of the GP part of the method could be improved even though matching the forecasting accuracy with that of the GA part of the method is indeed difficult because of the more stringent requirements. The application of local search algorithms such as conjugate gradient technique as cited in Section 2.1.2 could be helpful to improve both efficiency and accuracy of the forecasting process. Besides, as indicated in Fig. 1 in Chapter 3 and Section 2.1.3, techniques such as RGP could be employed to detect structural changes in the time series so that the capability of the proposed approach particularly for the GP part to recognize non-linear patterns could be enhanced.

In summary, forecasting volatility, or risk for that matter, is fundamentally and inherently a challenging problem. Making use of a combination of the certain engineering tools/techniques is as powerful as discovering brand new strategies for the sake of innovation. Engineering as a field has far longer and stronger scientific history in problem solving compared to the field of finance as a research subject. The current research takes one important step in an attempt to draw the immense power of the engineering analysis to benefit financial applications and to solve some of its pressing issues, which have been puzzling researchers in finance over the last two decades. The preliminary results shown here are very encouraging and further exploration is definitely worthwhile. We believe that we have opened up a new avenue and new thinking in the field of financial engineering.

APPENDIX 1

GLOSSARY OF VOLATILITY CONCEPTS

Arbitrage: A trading strategy that takes advantage of two or more securities being mispriced relative to each other. It is a zero-risk and zero – net investment strategy yet generates profits (Bodie, 1996).

Chartist vs. Fundamentalist: Chartist concentrates on the study of market action reflected on the trading price of assets, while fundamentalist focuses on the economic forces of supply and demand that cause prices to move/stay (Murphy, 86)

Conditional mean: In traditional stochastic analysis it is often assumed that conditional mean is a constant for the purpose of estimating and forecasting conditional volatility.

Correlation: It is related to the slope parameter of a linear regression model and thus is only a linear measure of association. In financial markets, where there is often a non-linear dependence between returns, correlation may not be an appropriate measure of co-dependency.

Delta hedging: Delta is the rate of change of the option price with respect to the price of the underlying asset. Delta hedging is to use the delta of the derivative position, *e.g.* option to offset the delta of the underlying security such as stock or vice versa.

Derivatives: An instrument whose price depends on or is derived from the price of another asset.

Diffusion assumptions: A phenomenon that in any one direction at a unit time the net flux of molecule movement equals zero.

Distributional assumptions: Assumptions that are imposed on the distribution of a random variable when conducting a stochastic analysis; *e.g.* normal distributions.

Equity: Ownership interest of common and preferred stockholders in a corporation (Ross, 1995).

Forwards: A contract that obligates the holder to buy or sell an asset at a predetermined delivery price at a predetermined future time period.

Futures: A contract that obligates the holder to buy or sell an asset at a predetermined delivery price during a specified future time period. The contract is marked to market daily.

FX Straddle: Foreign exchange option spread trading – simultaneously buying & selling two different contracts (Murphy, 1986).

Hedging: In order to remove the risk arising from market moves and leaves the hedger exposed only to the performance of the portfolio relative the market, a hedger uses derivatives of underlying securities, such as options and futures to offset the volatility movement of the market that is related to the portfolio.

Implied Volatility: Besides realized volatility that is estimated from the history of the asset price, the alternative approach is the volatility implied by an option price observed in the market.

Index Option: It is an option contract based on a stock index or other index.

Integrated Volatility: It is the standard deviation of rate of return distribution of an asset, which is evaluated based on intra-day historical data.

Leverage effect: Refer to APPENDIX 2 for details.

Multi-period distributions: Distributions that may vary at different time periods in a time series.

Oswin, Smith, GAB models: Mathematical models that are used to fit data of nonlinear nature.

Put / Call Options: A contract entitling the holder to sell/buy an asset for a certain price by a certain date.

Rate of return: Ratio between profit and capital investment or securities (Ross *et al.*, 95).

Realized Volatility: Realization of the stochastic volatility process that governs the movement of rate of returns (Alexander, 2001)

Risk: Uncertainty about future rates of return and can be quantified by using probability distributions.

Stationary series – A time series that has a finite "unconditional" variance σ^2.

Swaps: An agreement to exchange cash flows in the future according to a pre-arranged formula.

Time-varying volatility models: Models that describe a process for the conditional volatility.

Vega: Vega of a portfolio of derivatives is the rate of change of the value of the portfolio with respect to the volatility of the underlying asset.

Volatility: A measure of the uncertainty of the return realized on an asset.

Volatility Clustering: The property of the long memory of the volatility, as measured by a lagged autocorrelation function that decays as a power law (Alexander, 2001).

Volatility Smile: For equity options, it is the phenomenon that volatility decreases as the strike price increases. Volatility used to price a low strike price option *i.e.,* a deep-out-of-the–money put or deep-in-the-money call is significantly higher than that used to price a high-strike-price option *i.e.*, a deep-in-the-money put or a deep-out-of-money call. It is caused by the fat left tail and thin right tail of the implied distribution of the corresponding options in comparison with the lognormal distribution (Hull, 2000).

APPENDIX 2

VOLATILITY CONCEPTS

1 Basic of Volatility

Contrary to traditional beliefs, volatility is found to be positively correlated to market presence, activity, and volume. Karpoff (1987), Baillie and Bollerslev (1989) and Muller et al. (1990), emphasize the key role of volatility for understanding market structures. From risk management perspective, the only significant risks are the "irreducible" risk: those that cannot be reduced by hedging or diversification. Thus, the concerns of a portfolio manager focus not on the total volatility of a portfolio, but on the volatility that is collinear with the market. This volatility is represented as the portfolio '*beta*'. In a capital asset pricing model framework a high *beta* can be attributed to a positive correlation with the market index and a high relative volatility for the portfolio. In other words, it is important to estimate and forecast the volatility of indices such as S&P100 and S&P500 which represent the activity of the general market.

It is a fundamental question that how change of volatility associated with an underlying stock affects prices of the corresponding equity options. Higher volatility reflects more price fluctuations, leading to a higher probability that the underlying stock will perform either very well or very poorly. Such a fluctuation increases the profit potential for a long call when the stock performs well or for a long put position when the stock performs poorly. Therefore, call and put buyers are willing to pay more for the higher earning potential upon holding the corresponding options.

As volatility rises, call or put writers on the other hand, expect to receive more money because more fluctuations of the underlying stock price carries more risk. Since a call writer has assumed a bearish position, higher underlying stock volatility infers a greater chance that the stock might do very well over the lifetime of the contract, and assignment is more likely with a higher stock price. A bullish put writer incurs more risk from increasing volatility for the opposite reason as assignment is more likely from a lower stock price. When all other option pricing factors remain the same, a rule about the relation of volatility and option prices in the marketplace is as shown in Fig. 9.

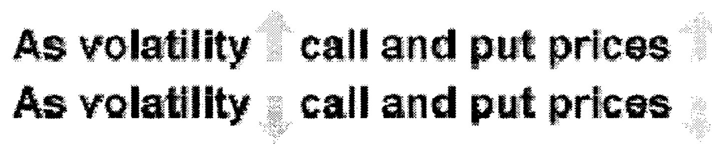

Figure 9 Rule about volatility in the option marketplace

Events that can result in changing stock volatility include factors such as :

- **Fundamental changes** in a company's balance sheet
- Uncertain **future earnings projections**
- **Economic factors** affecting the overall market or a stock's particular market sector, *e.g.*, pharmaceutical or retail
- A company's **move into uncharted waters**, such as a more speculative business
- **Changing market psychology**, or political and military situations possibly arising
- **Stock splits** can have an effect on volatility as well. In general, splits have a tendency to decrease stock volatility, while mere anticipation of a possible split can increase it.

Here are some important characteristics about volatility:

- Volatility represents underlying stock price **fluctuation**, not price trend.
- It can be measured as the annualized standard deviation of returns.
- Implied volatility represents a **market consensus** on a stock's future volatility.
- As implied volatility increases, call and put prices also increase, and vice versa.
- Implied volatility is dynamic. It can increase or decrease during an option's lifetime and lead to unexpected profit or loss.
- Buying options that are **overvalued** or selling those **undervalued** because of current implied volatility level by no means guarantees profits.

- For long option positions before expiration, increasing implied volatility has a positive effect on potential profits. For short positions, decreasing volatility has a positive effect (The Option Industry Council, 2004).

2 What is Volatility

Analysis of volatility of financial assets is a vast subject that is approached from two different technical perspectives. The option pricing school models the variation in asset prices in continuous time, while the statistical forecasting school models volatility from the perspective of discrete time series analysis. More specifically, in generating volatility forecasts, statistical methods use historical data on the underlying asset returns in a discrete time model for the variance of a time series. In other words, despite $E(\sigma) \neq \sqrt{E(\sigma^2)}$, most statistical models for 'forecasting volatility' are actually models for forecasting variance: they first forecast a variance, and then the volatility forecast is taken as the square root of the variance forecast.

One of the most important characteristics of volatility is clustering. Many financial time series display volatility clustering, that is, autoregressive conditional heteroscedasticity. Equity, commodity and foreign exchange markets often exhibit volatility clustering at the daily, even the weekly frequency, and volatility clustering becomes very pronounced in intra-day data. A typical example of a conditionally heteroscedastic return series is shown in Fig. 10 (Alexander, 2000). Note that two types of news events are apparent. The second volatility cluster shows an anticipated announcement, which turned out to be good news: the market was increasingly turbulent before the announcement, but the large positive return at that time shows that investors are pleased, and the volatility soon subsided.

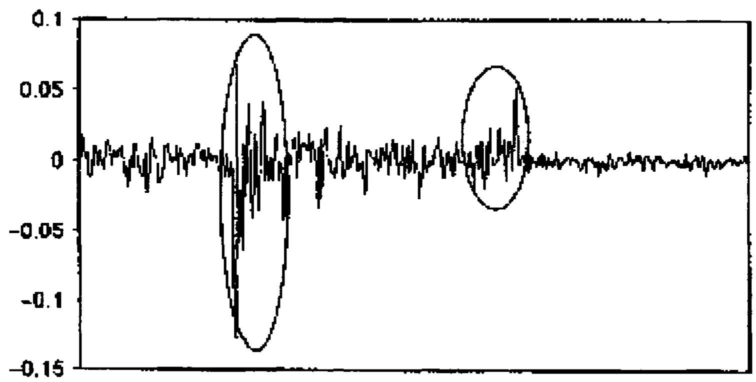

Figure 10 Volatility clustering

The first cluster of volatility indicates that there is turbulence in the market following an unanticipated piece of bad news. Volatility clustering implies a strong autocorrelation in squared returns, so a simple method for detecting volatility clustering is to calculate the first-order autocorrelation coefficient *d* in squared returns:

$$d = \frac{\sum_{t=2}^{T} r_t^2 r_{t-1}^2}{\sum_{t=2}^{T} r_t^4}, \tag{24}$$

where r_t and r_{t-1} are the rate of return at time t and $t-1$. A basic test for the significance of autocorrelation is the Box-Pierce LM test (Alexander, 2001), which is an econometric test for serial correlation between two random variables at different lags based on residual autocorrelations (Dacorogna, 2001).

3 Why Volatility Needs Forecast?

Volatility is a crucial element in a simple yet popular model for price development. Assume the following for an efficient market that

1) all the past history of the price development is reflected in the present price (*i.e.* a Markov property);

2) the response of the market on any new piece of information is immediate

Let $\Delta t > 0$ and denote $\Delta P := P_{t+\Delta t} - P_t := P_t$, where $\Delta P/P$ could be decomposed into a deterministic and a stochastic part in the following way:

$$\Delta P / P = \mu\Delta t + \sigma\Delta W, \tag{25}$$

where $\mu\,\Delta t$ is the deterministic part, μ is called *drift* or a *trend coefficient*; $\sigma\,\Delta W$ is a stochastic term with *volatility, standard error* or *diffusion* σ and $\Delta W := W(t+\Delta t) - W(t)$ is the increment of a standard Wiener process. Both μ and σ could be functions of P and t. Recall that the Wiener process $\{W(t), t \geq 0\}$ is a stochastic process with continuous trajectories such that $W(0) = 0$ with probability 1, for positive s and t the distribution of $W(t) - W(s)$ is normal $N(0, |t-s|)$, and for any $0 < t_0 < t_1 < \ldots < t_n < \infty$ the random variables $W(t_0)$, $W(t_1) - W(t_0)$, …, $W(t_n) - W(t_{n-1})$ (the increment) are independent. For $\Delta t \to 0$, the above equation becomes the stochastic differential equation :

$$\Delta P / P = \mu dt + \sigma dW \tag{26}$$

which describes the geometrical Brownian motion. Apply the Ito formula and convert P to logarithmic (Dupacova, 2002), hence the relationship of price vs. volatility:

$$d \ln P = (\mu - \sigma) dt + \sigma dW. \tag{27}$$

In other words, volatility is a crucial factor in modeling the price movement of the underlying asset.

4 How is Volatility Currently Characterized?

A simple linear regression can provide a model for the conditional mean of a return process. The classical linear regression model assumes that the unexpected return ε_t *i.e.*, the error process in the model is homoscedastic. In other words, the error process has a constant variance $V(\varepsilon_t) = \sigma^2$ whatever the value of the dependent variable. The fundamental idea in GARCH is to add a second equation to the standard regression model: the conditional variance equation. This equation will describe the evolution of the conditional variance of the unexpected return process $V_t(\varepsilon_t) = \sigma^2_t$. The full GARCH($p$, q) model takes the following form for the conditional variance :

$$
\begin{aligned}
&\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_p u_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2, \\
&\alpha_0 > 0, \\
&\alpha_1, \ldots, \alpha_p \text{ and } \beta_1, \ldots, \beta_q \geq 0, \\
&u_i = \frac{P_t - P_{t-1}}{P_{t-1}}.
\end{aligned}
\tag{28}
$$

Traditionally however, it is rarely necessary to use more than a GARCH(1,1) model, which has just one lagged error square and one autoregressive term. Using the standard notation for the GARCH constant ω, the GARCH error coefficient α and the GARCH lag coefficient β, the symmetric GARCH(1,1) model is as shown in Eq. 3 in Chapter 1.

Putting $\sigma^2_t = \sigma^2$ for all t gives an expression for the long-term steady-state variance in a GARCH(1,1) model (Alexander, 2001) :

$$\sigma^2 = \omega/(1 - \alpha - \beta) \tag{29}$$

When $\alpha+\beta = 1$, the unconditional variance is no longer defined and term structure forecasts do not converge. Since in this case the variance process is non-stationary, the model of integrated GARCH model is used :

$$\sigma_{t+1}^2 = \omega + \lambda\sigma_t^2 + (1-\lambda)u_t^2, \quad 0 \leq \lambda \leq 1. \tag{30}$$

5 Leverage Effect – A Problem Needs to be Resolved

In equity markets it is commonly observed that volatility is higher in a falling market than it is in a rising market. The reason may be that when the equity price falls the debt becomes more highly leveraged and so the debt/equity ratio increases. The firm becomes more highly leveraged and so the future of the firm becomes more uncertain. The equity price therefore, becomes more volatile. This implies an asymmetry in volatility clustering in equity markets. A very simple test of the leverage effect is to compute the first-order autocorrelation coefficient between lagged returns and current squared returns (Alexander, 2001) :

$$\rho = \frac{\sum_{t=2}^{T} r_t^2 r_{t-1}}{\sqrt{\sum_{t=2}^{T} r_t^4 \sum_{t=2}^{T} r_{t-1}^2}}. \tag{31}$$

If the coefficient ρ is negative and the corresponding Box-Pierce test is significantly different from zero, then there is an asymmetry in volatility clustering, which will not be captured by a symmetric GARCH model. Instead, one of the asymmetric (non-linear) GARCH models should be used (Alexander, 2001). There is a fundamental distinction between the symmetric GARCH models that are used to model ordinary volatility clustering and the asymmetric GARCH models that are required to capture leverage effects. The estimation of the conditional mean and conditional variance for the asymmetric GARCH is very complex compared with the estimation of symmetric GARCH (Alexander, 2001).

6 Limitations of the Integrated Volatility Model

Andersen *et al.* (2001) made use of the diffusion limit of the weak GARCH(1,1) process to construct the continuous time model of daily volatility, described by Eq. 20 in Section 5.4.1, where σ_t^2 is the variance of the rate of return distribution at time t, ω is the long term average variance. As indicated by Andersen, following the ARCH approach may not yield much higher predictability in volatility forecasting. Moreover, the problem of discontinuity and non-linearity do exist in the volatility data and need to be addressed in order to match the reality. These observations support the consideration stated in the first few chapters in this thesis *i.e.*, a drastically different mathematical framework may therefore, be necessary.

In Andersen′s 1998 paper, which discussed about nonlinear modeling of volatility of financial time series, GARCH model is used as a benchmark to measure the effectiveness of the integrated volatility calculation. By comparing results, *e.g.* the implied volatility whose returns are calculated from a stochastic model, versus the actual volatility calculated from continuously recorded quotations, it is obvious that R^2 values are strikingly similar. Since both processes are based on the same benchmark, the GARCH model, the paper concludes that the IV calculation is very closed to the reality. However, there may be a need to compare directly the two in order to determine exactly how closely the implied volatility resembles the actual volatility. Assuming this observation is correct, one may question that:

1) Why R^2 can only be at maximum 0.5 in both cases?
2) Is it true that estimation of IV can be done independent from any GARCH model?

We may even take the modeless approach. For example, if IV is taken as the input data, it may demonstrate certain types of pattern characteristics. As long as these patterns are recognized, forecasting will become far easier. This approach is like the rule-based diagnostics in mechanical vibration and the rule-based technical analysis in stock trading. Granted, the credibility of the technical analysis of the mean of rate of return may not be favourable in the academic field, the current work and many others cited here did provide some interesting arguments for the application of technical analysis on the second moment of rate of returns.

The problem could be viewed from another perspective. Most of the contemporary work about the volatility of rate of return conducts research in the time domain. Rarely does anyone start to work from the frequency domain, for reasons ranging from comparably less revealing to lack of appropriate tools. It will be interesting to work in both the time and frequency domain simultaneously, which leads to the use of wavelet analysis techniques as an alternative approach to model the volatility.

7 How is Volatility Traded Directly?

For those engaged in proprietary trading, the fact that volatility is uncertain has interesting implications. Recently constructing a delta-hedged option position has becomes a prevalent way to trade volatility. In particular, one goes long volatility by buying options and one goes short volatility by selling them *i.e.*, using delta hedging to limit exposure to movements in the price of the underlying asset. This approach regards the implied volatility as the cost of buying volatility when buying options, while the realized volatility captures the benefit from holding the options. Since implied volatility is typically well above subsequent realized volatility, this view of the world encourages option selling coupled with delta hedging.

There is a view that if one sells options and delta-hedge, one's revenues are basically implied volatility and the costs are historical volatility. Historically, implied volatilities are much higher than historical volatilities. The result is that delta-hedged option selling looks like a license to print money. Only when things go against the option-seller *e.g.*, short volatility trading is a high-variance strategy, which means that one might win small 99% of the time but looses big 1% of the time (Wood, 2004). Another noteworthy point is that a delta hedged option position still has exposure to movements in the price of the underlying asset for several reasons. Among these, the simplest reason is the well-known leverage effect – the empirical observation that volatilities tend to move in the opposite direction of the price (Carr, 2004).

Since two years ago, CBOE has offered a way to trade directly the volatility of an equity index. Based on the forecast of the volatility of the S&P100 index, one could trade the volatility itself by using the CBOE Volatility Index (VIX) futures. According to CBOE,

> *"Volatility Indexes provide investors with up-to-the-minute market estimates of expected near-term volatility of the prices of a broad-based group of stocks by extracting volatilities from real-time index option bid/ask quotes. Volatility Indexes are calculated using real-time quotes of the nearby and second nearby index puts and calls on established broad-based market indexes, referred to herein as a "Market Index." (Securities & Exchange Commission, 2004).*

The Final Settlement Price for the VIX Futures is a Special Opening Quotation (SOQ) of VIX calculated from the sequence of opening prices of the volatility of the S&P 500 Index (SPX) options used to calculate the VIX index on the settlement date. The opening price for any SPX series in which there is no trade will be the average of that option's bid price and ask price as determined at the opening of trading. The settlement date for VIX futures is the Wednesday preceding the third Friday of each calendar month. Listed below in Table VII are settlement dates for expiring contracts through January 2005.

Table VII

Schedule of the Settlement dates for VIX Futures

Expiring VIX Contract	Settlement Date	SPX Series Used in VIX SOQ
May 2004	May 19	Jun
Jun 2004	June 16	Jul
Jul 2004	July 14	Aug and Sep
Aug 2004	August 18	Sep
Sep 2004	September 15	Oct
Oct 2004	October 13	Nov and Dec
Nov 2004	November 17	Dec
Dec 2004	December 15	Jan and Feb
Jan 2005	January 19	Feb

CBOE will employ a modified Rapid Opening System (ROS) opening procedure on the settlement date for VIX futures for those SPX option contract months whose prices are used to calculate the VIX index. In other words, the one-month-ahead SPX option series is used to price the VIX Futures, because this VIX Futures are based on implied volatility of a portfolio originated from S&P100 (CFE, 2004).

APPENDIX 3

BASICS OF WAVELET TRANSFORM

1 Continuous wavelet transform versus discrete wavelet transform

What is "continuous" about the Continuous Wavelet Transform (CWT), and what distinguishes it from the discrete wavelet transform (to be discussed in the following section), are the set of scales and positions at which it operates. Unlike the discrete wavelet transform, the CWT can operate at every scale, from that of the original signal up to some maximum scale that one determines by trading off his/her need for detailed analysis with available computational horsepower. As shown in Fig. 11, the CWT is also continuous in terms of shifting: during computation, the analyzing wavelet is shifted smoothly over the full domain of the analyzed function.

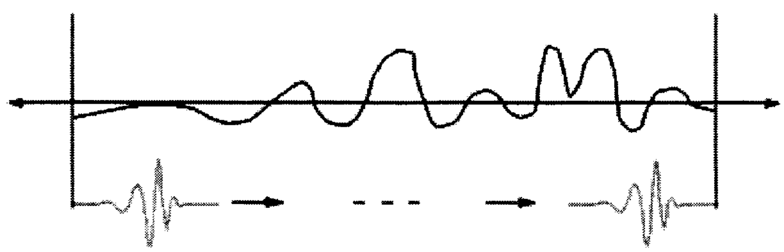

Figure 11 Continuous wavelet transform

Calculating wavelet coefficients at every possible scale is a fair amount of work, and it generates an awful lot of data. What if we choose only a subset of scales and positions at which to make our calculations? It turns out, rather remarkably, that if we choose scales and positions based on powers of two – so-called *dyadic* scales and positions – then our analysis will be much more efficient and just as accurate. We obtain such an analysis from the *discrete wavelet transform* (DWT). An efficient way to implement this scheme using filters was developed in 1988 by Mallat (1989). The Mallat algorithm is in fact a classical scheme known in the signal processing community as a *two-channel subband* (Strang, 1996). This very practical filtering algorithm yields a *fast wavelet transform* – a box into which a signal passes, and out of which wavelet coefficients quickly emerge.

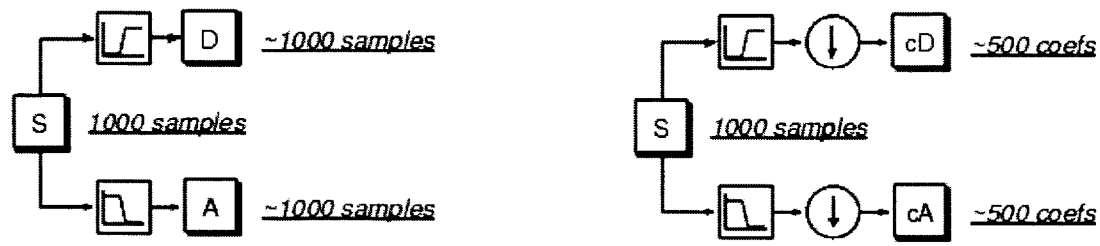

Figure 12 Decimation process in a wavelet filter

Consider a general function f belonging to a general functional space, for instance $L^2(R)$. Given a so-called *scaling function (or farther wavelet)* ϕ, such that its dilates and translates constitute orthogonal bases for all the V_j subspace, that are scaled versions of the subspace V_o, one can form a *Multiresolution Analysis* (MRA) of $L^2(R)$ once some properties are satisfied (Daubechies, 1992, and Meyers, 1993). With a DWT, one is constructing a map $\{f, \Psi_w\}$ from the signal domain to the wavelet coefficient domain. As shown in Eq. III-1, $\psi_{j,k}$ is the mother wavelet, while $\phi_{j,k}$ the father wavelet. The ϕ and ψ pair of functions generates the series of approximating spaces V_j. At a more specific level of analysis, the DWT algorithm is able to produce coefficients for fine scales, thus capturing the high frequency information, and for coarse scales, thus capturing low frequency information. A representation of the signal can be derived as

$$f(x) = \sum_k b_{jo,k}\phi_{jo,k}(x) + \sum_{j>jo}\sum_k c_{j,k}\psi_{jo,k}(x) \tag{32}$$

where

$$b_{j,k} = \frac{1}{n}\sum_{i=1}^{i=n}\phi_{j,k}(x_i), \quad c_{j,k} = \frac{1}{n}\sum_{i=1}^{i=n}\psi_{j,k}(x_i).$$

and $\phi_{jo,\,k}$ is a scaling function with the corresponding coarse scale coefficients $b_{jo,k}$, while $c_{j,k}$ are the detailed (fine scale) coefficients. The first term on the right hand side of Eq. III-1 is the projection of f onto the coarse approximating space V_{jo}, while the second term represents the details. A clear advantage of the orthogonal wavelet expansion is the resulting independence among coefficients. More specifically, let $\{h_l\}$ and $\{g_l\}$ denote the unit scale wavelet (high- pass) and scaling (low-pass) filters, respectively. Let $H(f)$ and $G(f)$ denote the transfer functions (Fourier Transforms) for the filters $\{h_l\}$ and $\{g_l\}$ respectively. Wavelet filters for higher scales are obtained through inverse Fourier Transform (Whitcher, 2000) :

$$H_j(f) = H(2^{j-1}f)\prod_{l=0}^{j-2} G(2^l f), \quad 0 \le f \le 1/2. \tag{33}$$

In summary, specific versions of DWT can be applied to a time series observed over a discrete set of times, say, t = 0, 1, ..., N-1. In a sense, the DWT can be formulated entirely in its own right without explicitly connecting it to any CWT, but it can also be regarded as an attempt to preserve the key features of the CWT in a succinct manner. From this point of view, the DWT can be thought of as a judicious subsampling of CWT(a, τ), in which one deals with just dyadic scales, *i.e.* a is picked to be of the form 2^{j-1} (j= 1,2, 3, ...) and then, within a given dyadic scale 2^{j-1} , one picks times t that are separated by multiples of 2^j.

2 Wavelet coefficients

While section 1 provides some operational details of the filters, this section is devoted to elaborate the relations among the number of samples, the time and the number of wavelet coefficients. For a data set of N samples, the DWT of these data also consists of N (*e.g.* 1024) values called "DWT coefficients". These coefficients can be organized into seven "wavelet coefficients" from scale 1, 2, 4, ..., 64. For scale 2^{j-1} *e.g.*, $Max\{j\}$ = 7 there are $N_j = N / 2^j = 8$ *wavelet coefficients*, and the time associated with these coefficients are taken to be $(2n + 1)\, 2^{j-1} - ½$, n = 0, 1, 2, ..., $N_j - 1$. For example, the second coefficient at scale 64 would be n = 1 and it is associated with time t = (2*1 +1)64 – 1/2 = 191.5. This coefficient reflects the difference between an average of 64 values before and after time 191.5. With appropriate normalization, the n^{th} wavelet coefficient at scale 2^{j-1} can be regarded as an approximation to CWT $[2^{j-1}, (2n + 1)\, 2^{j-1} - ½]$ (Percival, 2000).

Models for time series analysis usually should be built according to the *principle of parsimony* in the number of parameters characterizing the probability distribution involved. In wavelets a similar role is played by the concept of *sparcity;* in approximating a function by projecting the signal onto a sequence of subspace at different resolution levels, relatively few coefficients should be used in the representation. Thus, one would like to concentrate the information onto few coefficients able to give a high reconstruction power for the signal at hand. Since financial time series are noisy, once a wavelet decomposition of the signal has been applied, the resulting wavelet coefficients are heavily affected by the noise and thus care must be taken in selecting them for signal reconstruction purposes (Enrico, 1997).

3 Scale and frequency

In the wavelet coefficient plot, usually shown as y-axis labels, one could notice that the scales run from say 1 to 31. Recall that the higher scales correspond to the most "stretched" wavelets. The more stretched the wavelet, the longer the portion of the signal with which it is being compared, and thus the coarser the signal features being measured by the wavelet coefficients. On the other hand, the smaller the scale factor the more compressed the wavelet is. Thus, there is a correspondence between wavelet scales and frequency as revealed by wavelet analysis :

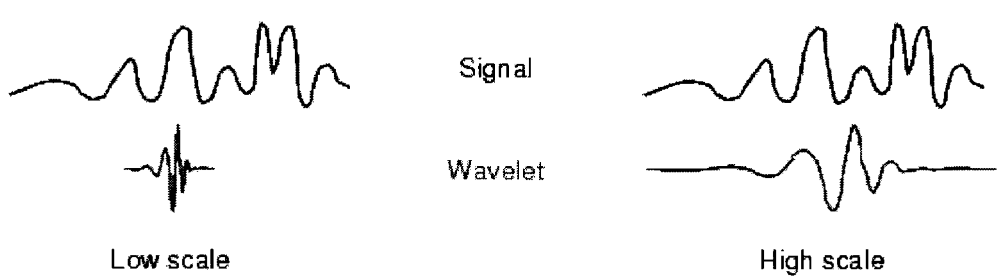

Figure 13 Different scales of the wavelet transform

- Low scale a compressed wavelet rapidly changing details High frequency
- High scale a stretched wavelet slowly changing, coarse features Low frequency

4 Number of levels

Since the analysis process is iterative, in theory it can be continued indefinitely. In reality, the decomposition can proceed only until the individual details consist of a single sample or pixel. In practice, one will select a suitable number of levels based on the nature of the signal, or on a suitable criterion such as *entropy* (Misiti, 1996).

5 Selection of a mother wavelet

Ingrid Daubechies, invented what are called compactly supported orthonormal wavelets – thus making discrete wavelet analysis practicable. The names of the Daubechies family wavelets are written `dbN`, where `N` is the order, and `db` the "surname" of the wavelet. The `db1` wavelet is the same as `Haar` wavelet. Here are the wavelet functions ψ(psi) of the next nine members of the db family :

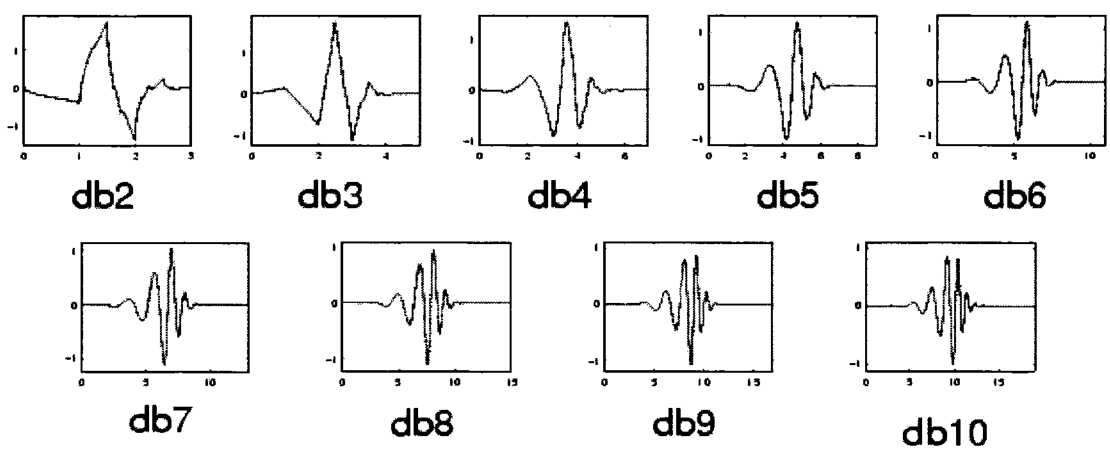

Figure 14 Different orders (levels) of the Dauberchies wavelets

And Eq. C-3 is the mathematical expression of the Dauberchies wavelets.

$$DWT(f) = 2\cos^{L}(\pi f)\sum_{l=0}^{L/2-1}\binom{\frac{L}{2}-1+l}{l}\sin^{2l}(\pi f) \qquad (34)$$

where L is a positive even integer and it is the width of the filter, *e.g.* $N = L/2$; while

$$\binom{a}{b} = \frac{a!}{b!(a-b)!} \quad \text{and} \qquad (35)$$
$$\sin^0(\pi f) = 1, \quad \forall f \text{ including } f = 0.$$

For more details on levels of the Dauberchies wavelet, refer to the multiple level decompositions outlined in Section 7.

6 De-noising

To further refine the data processing, the given data can be de-noised. With wavelets one basically adopts a flexible degree of smoothing, according to the resolution level. Thus by increasing the resolution level j, the analyst decreases smoothing and vice versa. However, care must be taken in analyzing the coefficients selected due to the highly noisy nature of these data; one thus needs reliable procedures to get rid of coefficients considered not useful for reconstructing the signal.

The wavelet shrinkage principle of Donoho and Johnstone (1990) applies a threshold based de-noising procedure to the observed data by shrinking wavelet coefficients toward 0 so that a limited number of them will be considered for reconstructing the signal. The algorithm that implements this principle can be dissected by the following three steps :

1) DWT is applied to the data to get the empirical wavelet smooth and detail coefficients;
2) the wavelet coefficients, in particular at the finest scales are shrunken toward zero by *thresholding*;
3) the inverse DWT is applied to the *thresholded* coefficients to reconstruct the signal

Two basic elements are to be determined: the shrinkage rule and the threshold value. Among several possible choices available, an adaptive procedure is chosen because given the nature of the problem it is the method that allows for a better signal/noise separation. The soft shrinkage rule selected is $\delta_c(x)$ = sign(x)($|x|$ - c), when $|x|>$ c; otherwise $\delta_c(x) = 0$. It keeps or shrinks values, compared to the keep-or-kill solution by the hard rule.

7 Multiple-level decomposition

The decomposition process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many *lower resolution components*. This is called the *wavelet decomposition tree* as shown in the following figure.

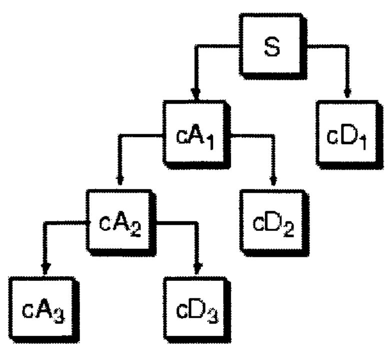

Figure 15 Wavelet decomposition tree

8 Orthogonality

The scale function ϕ and the wavelet function ψ are chosen to be orthogonal to each other *i.e.*, independent to each other therefore, in theory the signal could be divided indefinitely.

9 Compactly supported

A compactly supported function is non-zero only for a finite duration. A wavelet function that is compactly supported can be more efficient in a way that requires fewer coefficients to transform a function from the time domain to the wavelet (or scale) domain.

10 Vanishing moments

The ability of a wavelet to suppress a polynomial depends on a crucial mathematical characteristic of the wavelet called its number of vanishing moments. If the average value of $x^k \psi(x)$ is zero (whereas $\psi(x)$ the wavelet function), for $k = 0, \ldots, n$ then the wavelet has $n + 1$ vanishing moments and polynomials of degree n are suppressed by this wavelet.

The signal that is analyzed here is not known to be regular or symmetrical. Therefore, for a preliminary study, an asymmetric and irregular wavelet function like the Dauberchies can be a starting point for the practice purpose.

11 Wavelet Packets

In the orthogonal wavelet decomposition procedure, the generic step splits the approximation coefficients into two parts. After splitting, one obtains a vector of approximation coefficients and a vector of detail coefficients, both at a coarser scale. The information lost between two successive approximations is captured in the detail coefficients. Then the next step consists of splitting the new approximation coefficient vector; successive details are never re-analyzed.

In the corresponding wavelet packet situation, each detail coefficient vector is also decomposed into two parts using the same approach as in approximation vector splitting. This offers the richest analysis in which the complete binary tree is produced. The idea of this decomposition is to start from a scale-oriented decomposition, and then to analyze the obtained signals on frequency sub-bands.

APPENDIX 4

TIME SERIES DATA MINING

A group of TSDM definitions are given in this section, *e.g.* events (important occurrences), temporal patterns (vector of length Q), event characterization function ($g(t)$), temporal pattern cluster (P), time-delay embedding, phase space, augmented phase space, objective function (Povinelli, 1999).

An event may be defined as the sharp rise or fall of any selected time series *e.g.*, $H = \{x_t, t = 1,..., M\}$. A temporal pattern is a hidden structure in a time series that is characteristic and predictive of events. The temporal pattern P is a real vector of length Q. And it best characterizes the desired events *e.g.*, P is used to predict events in a testing time series. The temporal pattern is represented as a point in a Q dimensional real metric space, $P \in \Re^Q$. Because a temporal pattern may not perfectly match the time series observations that precede events, a temporal pattern cluster is defined as the set of all points within δ of the temporal pattern.

Let $\tau > 0$ be a positive integer. If t represents the present time index, then $t - \tau$ is a time index in the past, and $t + \tau$ is a time index in the future. A phase space is a Q dimensional metric space into which a time series is embedded. In our case, Q is chosen to be 4, representing the 4-lag recursive memory. To link a temporal pattern (past and present) with an event (future) the event characterization function $g(t)$ is introduced. The event characterization function represents the value of future "eventness" for the current time index. One possible event characterization function to address this goal is $g(t) = x_{t+1}$, which characterizes the one-step-ahead time series values. The concept of an augmented phase space follows from the definitions of the event characterization function and the phase space. The augmented phase space is a Q+1 dimensional space formed by extending the phase space with $g(.)$ as the extra dimension. Every augmented phase space point is a vector $< x_t, g(t) > \in \Re^{Q+1}$.

As shown in Fig. 16, the height of the leaf represents the significance of $g(.)$ for that time index. From this plot, the required temporal pattern and temporal pattern cluster are easily identified. The TSDM objective function represents the efficacy of a temporal pattern cluster to characterize events.

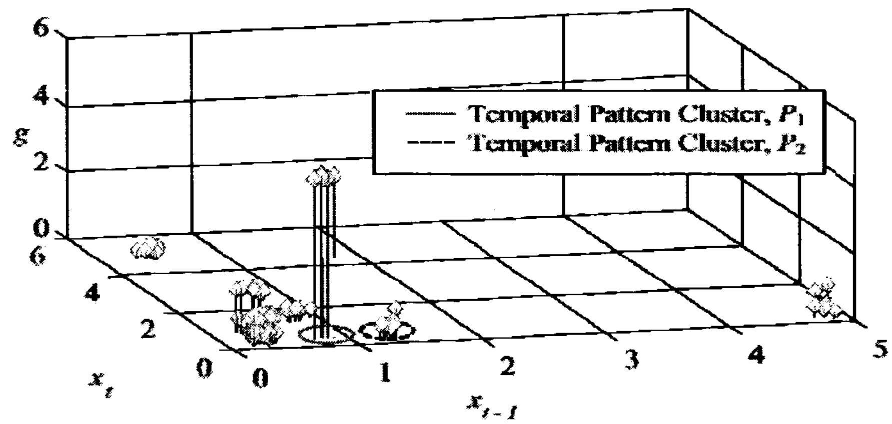

Figure 16 Augmented phase space for a time series (Povinelli, 1999)

The objective function f maps a temporal pattern cluster P onto the real line, which provides an ordering to temporal pattern clusters according to their ability to characterize events. The objective function is constructed in such a manner that its optimizer P^* meets the TSDM goal. The objective function must capture the accuracy with which a temporal pattern cluster predicts all events. Since it may be impossible for a single temporal pattern cluster to perfectly predict all events, a collection C of temporal pattern clusters is used for this objective function. The objective function $f(C)$ is the ratio of correct predictions to all predictions, *i.e.*

$$f(\mathrm{C}) = \frac{t_p + t_n}{t_p + t_n + f_p + f_n}, \tag{36}$$

where t_p is the correct prediction when there is temporal pattern cluster p; t_n is the correct prediction when there is no temporal pattern cluster found; f_p is the incorrect prediction when there is temporal pattern cluster p and f_n is the incorrect prediction when there is no temporal pattern cluster found.

This objective function would be used to achieve maximum event characterization and prediction accuracy for binary $g(t)$. The key concept of the TSDM framework is to find optimal temporal pattern clusters that characterize and predict events – thus, an optimization algorithm represented by $\max_{P,\delta} f(P)$.

APPENDIX 5

GLOSSARY OF MARKOV CHAIN RELATED CONCEPTS

Aperiodic: (A state *j* is said to be *periodic* with *period p*, if on leaving state *j* a return is possible only in a number of transitions that is a multiple the integer $p > 1$. A state whose period is p=1 is said to be aperiodic, *e.g. occurring without periodicity; irregular.*

Attribute: An attribute is any parameter that has significant impact on the behavior of a system that is being optimized and is being controlled in an attempt to achieve an optimum.

Ergodic: the states of a finite, aperiodic, irreducible Markov chain are ergodic.

Irreducible: A DTMC is said to be irreducible if every state can be reached from every other state, *i.e.* if there exists integer m from which $p_{ij}^{(m)} > 0$ from every pair of states i and j.

Markov chain: a stochastic process whose conditional probability distribution function satisfies the "Markov property":

$$\Pr\{\theta(t) < \theta \mid \theta(t_0) = \theta_0, \theta(t_1) = \theta_1, \ldots, \theta(t_n) = \theta_n\} = \Pr\{\theta(t) < \theta_0 \mid \theta(t_n) = \theta_n\} \qquad (37)$$

If the state space of a Markov process is discrete, it is referred as a Markov chain.

Objective Function: An objective function (also called a cost function) spells out the relationship between the outcome (objective) and the attributes mathematically.

Optimization: the process by which an optimum is achieved, which could include mixed integer programming, stochastic approximation, taboo search, neural networks, simulated annealing, *etc.*

Recurrent: The Discrete Time Markov Chain (DTMC) is guaranteed to return to these states infinitely often. Any finite state Markov chain must have at least one recurrent state.

Simulation in Management Science: It often has three features:

(1) There are one or more stochastic input processes that are specified via the language of probability and from which synthetic realizations can be generated.

(2) There is a logical model that completely describes how the system of interest reacts to realizations of the stochastic inputs. The logical model is usually an algorithm that updates the system state upon the occurrence of some discrete events (GA's typically don't need to have models.)

(1) There are output processes, typically ordered by some concept of time, that represent the system behaviour of interest to the analyst (Nelson, 2004).

Time homogenous: When the transitions out of state $\theta(t)$ depend on the time t, the Markov chain is said to be non homogenous, otherwise, homogeneous.

Transient: There is a nonzero probability that the DTMC will never return to such a state.

APPENDIX 6

GLOSSARY OF CONCEPTS OF GENETIC PROGRAMMING

A group of GP related definitions are provided in this section (Chen, 2003):

Crossover: The idea of crossover is to preserve "fit" segments and they will become larger and larger until an optimal solution is found. It combines the genetic attribute of the two parents by swapping part of their genes.

The operator first chooses two parents based on a certain mating selection policy (*e.g.* tournament selection). The top half depicts two chosen parents and they are two numerical expressions in tree structure. Then a sub-tree is selected form each parents with the same type. The selected sub-trees have the same type. Lastly, the sub-trees of the two parents are swapped and two children are born. One child is discarded randomly.

Evaluation of the Fitness: Fitness of a GP individual is the measurement of how well it is in relation to the optimal solution or comparing with other individuals. Fitness is application specific, and the measurement of fitness can be adjusted to provide convenient measurement for the completion among individuals in evolution process.

Genetic operators: Genetic operators control the evolving patterns of the population. It controls the ways, the speed and the level of "optimal" of the evolution. In GP, crossover, mutation and reproduction are three most common genetic operators.

Initialization: The first step in performing a GP run is initialization of the first generation of population. There are two different methods for initializing tree structures in common use namely, grow and full. Grow method will just randomly choose any kind of nodes to build an irregular tree until it ends the branch with a terminal. As for the full method, the builder will choose a function until the maximum depth of the tree is reached by ending with terminals. So every branch will reach the maximum depth.

Mutation: Mutation operates on only one individual to change part of its genes randomly. The probability of mutation is a parameter of the run. If one does not want the individual to differ much from its ancestor, small mutation should be selected.

There are many different kinds of mutation operators. The operator randomly selects a point in the tree of the ancestor, and replaces the sub-tree to that point with a newly generated sub-tree. The new sub-tree is generated the same way and is subject to the same restrictions with the initialization.

Another kind of mutation will choose randomly a point in the tree and mutate every constant in the sub-tree into a new value. Other different kinds of mutation include : point mutation, permutation, hoist, and expansion mutation, collapse sub-tree mutation, *etc.*

Reproduction: Reproduction operators simply select an individual, copy it and place the copy back into the population, so there will be two copies of the same individual in the population.

Selection: Selection mechanism uses a probabilistic survival rule on the population. According to the fitness measurement, there is a sequence of competition among the individuals of the population. Selection operator will decide which individual will survive and which will perish. Selection pressure places an important role in dealing with the trade-off between further exploiting promising regions of search and exploiting new regions at the same time.

APPENDIX 7

PUBLICATION LIST ARISING FROM THIS THESIS

1. Ma, I., Wong, T., Sankar, T., & L. Li, "An Engineering Approach to Forecast Volatility of Financial Indices", International Journal of Computational Intelligence, vol. 3, no. 1, pp. 23 – 35, 2006
2. Ma, I., Wong, T., Sankar, T., "Volatility Forecast by Discrete Stochastic Optimization and Genetic Algorithms", New Mathematics and Natural Computation, World Scientific Publishing Co., MA., U. S., Accepted in Oct. 2005.
3. Ma, I., Wong, T., Sankar, T., & Siu, R., "Volatility Forecasts of the S&P100 by Evolutionary Programming in a Modified Time Series Data Mining Framework", WAC2004, Seville, Jun. 2004 (peer reviewed).
4. Ma, I., Wong, T., Sankar, T., & Siu, R., "Forecasting the Volatility of a Financial Index by Wavelet Transform and Evolutionary Algorithm", IEEE SMC 2004 Conference, Hague, pp. 5824 – 5829, Oct. 2004 (peer reviewed).
5. Ma, I., Wong, T., Sankar, T., "Volatility Forecast by Discrete Stochastic Optimization and Genetic Algorithms", 4th International Conference on Computational Intelligence in Economics and Finance (CIEF), Utah, U. S. A., pp. 992 – 996, Jul. 2005 (peer reviewed).
6. Ma, I., Wong, T., Sankar, T., "A Novel Mathematical Approach to Forecast Volatility of Financial Indices", 2006 Financial Management Association (FMA) Annual Meeting, Salt Lake City, Utah, (p. TBD) October 11 - 14, 2006 (peer reviewed).

BIBLIOGRAPHY

Alexander, C., *Market Models – A Guide to Financial Data Analysis*, John-Wiley & Sons, Ltd., UK, 2001.

Andersen, T. G. and Bolersleve, T., "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Acceptable Forecasts," *International Economic Review*, 39, pp. 885-905, 1998.

Andersen, T., Bollerslev, T., Diebold F. and Labys P., "The Distribution of Exchange Rate Volatility," Manuscript, University of Pennsylvania, 1999.

Andersen, T., T. Bollerslev, F. Diebol, and H. Ebens, "The Distribution of Realized Stock Return volatility," *Journal of Financial Economics*, No. 61, pp.43-76, 2001.

Andersen, T. G., T. Bollerslev, F. X. Diebold. and P. Labys., "The distribution of realized exchange rate volatility," *Journal of the American Statistical Association*, No. 96, pp. 42-55, 2001b.

Andersen, T., T. Bollerslev, F. Diebol, and P. Labys, "Modeling and Forecasting realized volatility,", *Econometrica*, No. 71, 579-625, 2003.

Andradottir S., "A Method for Discrete Stochastic Optimization," *Management Science*, vol. 41, No. 12, Dec. 1995, p. 1953.

Andradottir, S., "A scaled stochastic approximation algorithm," *Management Science* 42:475–498, 1996a.

Andradottir, S., "Optimization of the transient and steady-state behavior of discrete event systems," *Management Science*, 42:717–737, 1996b.

Andradottir, S., "A global search method for discrete stochastic optimization," *SIAM Journal on Optimization*, 6:513–530, 1996c.

Andradottir, S., "Accelerating the Convergence of Random Search Methods for Discrete Stochastic Optimization," *ACM Transactions on Modeling and Computer Simulation*, Vol. 9, No. 4, p. 349-380, Oct. 1999.

Antunes C. M., and Arlindo L. Oliveira , *Proceedings KDD Workshop on Temporal Data Mining*, pages 1-13, 2001.

Arino, M. A., "Forecasting Time Series via Discrete Wavelet Transform," *Second International Conference on Computing in Economics and Finance*, Geneva, Switzerland, 26-28 June 1996

Baillie, R. T., and Bollerslev, T., "The Message in Daily Exchange Rates: A Conditional-Variance Tale," *Journal of Business and Economic Statistics,* vol. 7, pp. 297-305, 1989.

Barndorff-Nielsen and N. Shephard, "Estimating Quadratic Variation using Realized Variance," *Journal of Applied Econometrics,* 17: pp. 457– 477, 2001.

Barndorff-Nielsen, O. E. and N. Shephard, "Econometric analysis of realised volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society*, Series B 64, 253-280, 2002.

Barndorff-Nielsen, O. E. and N. Shephard, "Realised power variation and stochastic volatility", *Bernoulli* 9, 243-265. Correction published in pages 1109-1111, 2003.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard, "Regular and modified kernel-based estimators of integrated variance: the case with independent noise," Unpublished paper: Nuffield College, Oxford, 2004.

Bauer, R. J., *Genetic Algorithms and Investment Strategies*, by John Wiley & Sons, Inc., 1994.

Bodie, Z., Kane, A. and Marcus, A., *Investments*, 3rd edition, Irwin McGraw-Hill, Boston MA., 1996.

Brooks, C., "Predicting Stock Index Volatility: Can Market Volume Help?," *Journal of Forecasting*, 17, 59-80, 1998.

Brooks, C. and Persand, G, "Volatility Forecasting for Risk Management," *Journal of Forecasting*, 22, 1-22, Oct. 2003.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. and Zanasi A., *Discovering data mining: from concept to implementation*, Prentice-Hall, Inc. Upper Saddle River, NJ, USA , 1998.

Capobianco, E., "Wavelets for High Frequency Financial Time Series," *Interface 1999 Conference Proceedings*, pp. 373-378.

Carpenter, J. and J. Bithell, "Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians," *Statistics in Medicine*, 19:1141 – 1164, 2000.

Carr, P., "The Volatility of Volatility," *Financial Engineering News*, October 26th, 2004.

CFE Information Circular IC04-06, from http://www.cboe.com/CFE/general/CFEIC04-006.pdf: CFE Trading Permit Holders, Re: Settlement Procedure for VIX Futures, on May 6, 2004

Chan, P. K. and S.J. Stolfo, "Scaling learning by meta-learning over disjoint and partially replicated data," *Proceedings of the 1996 Florida Artificial Intelligence Society*, pp. 151-155, 1996.

Chan, Y. T., *Wavelet Basics*, Kluwer Academic Press Inc., 1995.

Chen, H., "Equity Premium Prediction and Investment Strategy Searching with Genetic Programming", M. Eng. Dissertation, Universiteit Leiden, Leiden, Netherlands, 2003.

Chen, S. H and Paul P. Wang (eds.), *Computational Intelligence in Economics and Finance*, Springer.2002.

Chen, Shu-Heng and Yeh, Chia-Hsuan, "Using Genetic Programming to Model Volatility in Financial Time Series: The Cases of Nikkei 225 and S&P 500", AI-Econ Research Group, Dept. of Economics, National Chengchi University, Taipei, Taiwan 1623, *The 4th JAFEE International Conference on Investments and Derivatives (JIC'97)*, Royal Park Hotel, Tokyo, Japan, Jul. 31-Aug. 1, 1997.

Chen, S. Y., C. Yeh, and W. Lee, "Option Pricing with Genetic Programming," in J. Koza, W. Banzhaf, K. Chellapilla, K. Deb, M. Dorigo, D. Fogel, M. Garzon, D., 1998.

Christoffersen, P., "Elements of Financial Risk Management Chapter 1:Volatility Forecasting", McGill University and CIRANO, August 5, 2002.

Christoffersen, P., Diebold, F. and Schuermann, T., "Horizon Problems and Extreme Events in Financial Risk Management", *FRBNY Economic Policy Review*, pp. 109, Oct. 1998.

Christoffersen P. and F. Diebold, "How relevant is volatility forecasting for financial risk management?," *Review of Economics and Statistics*, vol. 82, pp. 12-22, 2000.

Christoffersen, P., Diebold, F., "Financial Asset Returns, Market Timing, and Volatility Dynamics", Manualscript, McGill University, University of Pennsylvania, October 2, 2002.

Christoffersen, P., K. Jacob and Y. Wang, "Option Valuation with Volatility Components", working paper, McGill University, Sep. 24, 2004.

Chui, C., *An Introduction to Wavelets*, Academic Press, Inc., 1995.
Cohen, L., "Time-Frequency Distributions – A Review", *Proc. IEEE*, Vol. 77, No. 7, pp. 941 – 981, 1989.

Coifman, R., R. and Wickerhauser, M. V., "Entropy Based Algorithms for Best Basis Selection", *IEEE Transactions on Information Theory*, 32:712--718, March 1992.

Conrad J. & Kaul, "Time-Variation in Expected Returns", *Journal of Business,* 61, pp. 409-25, Oct. 1988.

Corsi, F., Zumbach, G. et al., "Consistent High-Precision Volatility from High-Frequency Data", *Economic Notes*, 2001.

Dacorogna, M., Gencay, R. et al, *An Introduction to High-Frequency Finance*, Academic Press, San Diego, 2001.

Daubechies, I., "Ten Lectures on Wavelets", Philadelphia: *SIAM*, 1992.

Daubechies, I., "The Wavelet Transform time-frequency localization and Signal Analysis", *IEEE Trans. Inform. Theory*, 36, 961-1004.

Davidson, R., and MacKinnon, J., *Estimation and Inference in Econometrics*, Oxford University Press, New York Oxford, 1993.

Dawid, H., *Adaptive Learning by Genetic Algorithms*, 2nd edition, Springer: Berlin, 1999.

Diggs, D. H. and R. J. Povinelli, "A Temporal Pattern Approach for Predicting Weekly Financial Time Series," *Artificial Neural Networks in Engineering*, St. Louis, Missouri, 707-712, 2003.

Donoho, D. L. & Johnstone, I. M., "Wavelets and the theory of nonparametric function estimation," *Phil. Trans. Roy. Soc. London*, 357, 2475–2494, 1999.

Duan, J. C., Gautier, G. and Simonato, J. G, "A Markov Chain Method foe Pricing Contingent Claims", *Stochastic Modeling and Optimization – with applications in queues, finance, and supply chains*, Springer-Verlag, New York, Inc., (Edited by Yao D. D., Zhang, H. and Zhou, X. Y., 2003.

Dupacova, J., Hurt, J. and Stepan, J., *Stochastic Modeling in Economics and Finance*, Kluwer Academic Publishers, Dordrecht, 2002.

Enrico, C., "On the Multiresolution Analysis of Financial Time Series Models", *6th International PASE' 97 Workshop*, Computers in Finance and Economics, Marianske` Lazne, Czech Republic, on November 9-12, 1997

Fong, L. Y. and Szeto, K. Y., "Rules Extraction in Short Memory Time Series Using Genetic Algorithms", *The European Physical Journal B,* 20, pp. 569-572, 2001

Gavrishchaka1, G. G, "Boosting Frameworks in Financial Applications: From Volatility Forecasting to Portfolio Strategy Optimization", *4th International Conference on Computational Intelligence in Economics and Finance* (CIEF), Salt Lake City, Utah, p. 1052 – 1058, Jul. 2005.

Gaunersdorfer C. and Hommes Cars, "A Nonlinear Structural Model for Volatility Clustering", Center for Nonlinear Dynamics in Economics and Finance (CeNDEF) Faculteit Economie, Universiteit van Amsterdam, CeNDEF Working Papers with number 00-02, Sept. 2000.

Goldberg, H., Iba, and R. Riolo, eds., *Genetic programming 1998: Proceedings of the Third Annual Conference*, San Francisco, CA: Morgan Kaufmann, 32-37.

Gregoire, A.& McKeague, I., "Wavelet Methods for Curve Estimation", *Journal of the American Statistical Association*, Vol. 89, Issue 428, Dec. 1994.

Hall, J., "Options, Futures and Other Derivatives", 4th edition, Prentice-Hall, NJ., 2000
Harvey C., & Siddique A., "Autoregressive Conditional Skewness", *Journal of Financial and Quantitative Analysis*, Vol. 34, No. 4, pp. 465-487, Dec. 1999.

Hayin, S., Editor, *Advances in Spectrum Analysis and Array Processing*, Vol. I, Prentice Hal Englewood, Cliffs, NJ., 1991.

Hog, E. and Lunde A., "Wavelet Estimation of Integrated Volatility", unpublished manuscript, *The Aarhus School of Business*, Denmark, Jan. 2003.

Hong, Y. and Lee, J., "Wavelet-based Estimation of Heteroskedasticity and Autocorrelation Consistent Covariance Matrices", Cornell University, Oct. 1999.

Hovsepian, K, Anselmo P. C., and Mazumdar, S., "Detection and Prediction of Relative Clustered Volatility in Financial Markets", *4th International Conference on Computational Intelligence in Economics and Finance (CIEF)*, Salt Lake City, Utah, p. 1052 – 1058, Jul. 2005.

Hubbard, B., *The World According to Wavelets*, A. K. Peters, Second Edition, 1998.

Hull, J., *Options, Futures, & other Derivatives*, 4th edition, Prentice Hall, Upper Saddle River, NJ, 07458, 2000.

Iba, H. and T. Sasaki, "Using genetic programming to predict financial data", *Proceedings of the Congress on Evolutionary Computation*, Vol.1, 244-251, 1999.

Jackwerth, J. C. and M. Rubinstein, "Recovering Stochastic Processes from Option Prices", working paper, London Business School, 2001.

Kaboudan, M, "Extended Daily Exchange Rates Forecasts Using Wavelet Temporal Resolutions", *New Mathematics and Natural Computation*, Vol. 1, No. 1, by World Scientic Publishing Co., p. 79- 107, 2005.

Karpoff, J. M., "The Relation between Price Changes and Trading Volume: A Survey", *Journal of Financial and Quantitative Analysis,* 22, 109-126, 1987.

Kinlay, H. R., Neftci, J. and Wilmott, P. S., "Investment Analytics Volatility Report", *Investment Analytics*, N.Y. 2001.

Kinlay, J., "Investment Analytics -- F1000 Volatility Opportunity Strategy", Investment Analytics (Bermuda) Ltd, extracted from http://www.investment-analytics.com/files/vol%20arb%20program/strategies/F1000/Investment%20Analytics%20Volatility%20Arbitrage%20Program%20Strategy%20Presentation%20F1000%20VOF%20with%20Notes%20Sept%2004.pdf, on Oct. 16, 2005.

Koza, J., *Genetic Programming – On the programming of computers by means of natural selection*, The MIT Press, Cambridge, Massachusetts, 1998.

Langdon, W. B., *Genetic Programming and Data Structures*, 1996, (retrieved from http://www.cs.bham.ac.uk/ ~wbl/thesis.html#toc on Dec. 15, 2004).

Lawrenz, C. and Westerhoff F., "Explaining Exchange Rate Volatility with a Genetic Algorithm", *6th International conference of the Society for computational Economics on Computing in Economics and Finance*, Barcelona, paper 325, Jul. 2000.

LeBaron, B., "Forecasting Improvements Using Volatility Index", *Journal of Applied Econometrics*, Vol. 7, pp. 137-149, 1992.

LeBaron, B., "Evolution and Time Horizons in an Agent Based Stock Market", Working paper,Brandeis University, June 2000.

Lee, G., "Wavelets and Wavelet Estimation: A Review", *Journal of Economic Theory and Econometrics*, 4, p. 123 – 157, 1998.

Lee, W. C., "Forecasting High-Frequency Financial Data Volatility Via Nonparametric Algorithms -Evidence from Taiwan Financial Market", *4th International Conference on Computational Intelligence in Economics and Finance (CIEF)*, Salt Lake City, Utah, p. 1031 – 1034, Jul. 2005.

Li, D., "Empirical Study of Investment Behavior in Equity Markets Using Wavelet Methods", Ph. D. Thesis, Rensselaer Polytechnic Institute, 2003.

Los, J., Karuppia. C., "Wavelet Multiresolution Analysis of High-Frequency Asian FX Rates", AFA 2001 New Orleans; Adelaide University, School of Economics Working Paper No. 00-6.

Ma, I., Wong, T., Sankar, T., & Siu, R., "Volatility Forecasts of the S&P100 by Evolutionary Programming in a Modified Time Series Data Mining Framework", *WAC2004*, Seville, Jun. 2004a.

Ma, I., Wong, T., Sankar, T., & Siu, R., "Forecasting the Volatility of a Financial Index by Wavelet Transform and Evolutionary Algorithm", *IEEE SMC 2004 Conference*, Hague, Oct. 2004b.

Maheu, J. and McCurdy, T., "Nonlinear Features of Realized FX Volatility", Department of Economics, University of Alberta and Rotman School of Management, University of Toronto.

Mallat, S. G., "A theory for Multiresolution signal decomposition: The wavelet representation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674-693, July 1989.

Mallat, S., G., *A Wavelet Tour of signal Processing*, Academic Press, San Diego, 1999.

Martens, M., "Measuring and Forecasting S&P500 Index-Futures volatility Using High-Frequency Data", *The Journal of Futures Markets*, Vol. 22, No. 6, 497-518.

Martens, M., "Measuring and Forecasting S&P500 Index-Futures Volatility Using High-Frequency Data", *The Journal of Futures Markets*, Vol. 22, No. 6, pp. 497-518, 2002.

Marti, K., Ermoliev, Y. and Pflung, G., "Dynamic Stochastic Optimization -- Lecture notes in economics and mathematical systems", Springer Verlag, Berlin Heidelberg, New York, p. 100, 2004.

Meddahi, N., "A Theoretical Comparison between Integrated and Realized Volatility," *Journal of Applied Econometrics* 17: pp. 479- 508, 2002.

Meyers, S. D., "Wavelets: Algorithms and Applications", *SIAM*, 1993.

Misiti, M., Misiti, Y. et al., "Wavelet Toolbox User's Gruide, Version 1", Mathworks Inc., Natick, M. A., 1996.

Muller, U. A. et al., "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.

Murphy, J., "Technical Analysis of the Futures Markets", *New York Institute of Finance*, NY 10270, 1986.

Nelson, B. L., "50th Anniversary Article Stochastic Simulation Research in Management Science, *Management Science*,Vol. 50, No. 7, pp. 855–868, July 2004.

Neely, C. J. and Weller, P., "Predicting Exchange Rate Volatility: Genetic Programming vs. GARCH and RiskMetrics", Research Department, Federal Reserve Bank of St. Louis and Department of Finance, Henry B. Tippie College of Business, Sept. 2001.

Oppenheim, A. V. and Schafer, R. W., *Discrete-Time Signal Processing*, Prentice Hall, Inc. Englewood Cliffs, NJ., 1989.

The Option Industry Council, retrieved from http://www.888options.com/oou/content, on Nov. 22, 2004.

Papoulis, A., *Signal Analysis*, McGraw-Hill, New York, NY, 1977.

Park, B. J., "An Outliner Robust GARCH Model and Forecasting Volatility of Exchange Rate Returns", *Journal of Forecasting*, 21, 381-393, Apr. 2002.

Peters, J. P., "Estimating and Forecasting Volatility of Stock Indices Using Asymmetric GARCH Models and (Skewed) Student-t Densities), unpublished manuscript, Ecole d'Administration des Affaires, Universite de Liege, Mar. 2001.

Percival, D. B., Walden, A. T., *Wavelet Methods for Time Series Analysis*, Cambridge University Press, Oct. 2000.

Pictet, O. V., Masutti, O. and Zumbach, G., "Volatility forecasting using genetic programming", *Dynamic Asset Management*, Geneva, Switzerland and Olsen & Associates, Zurich, Switzerland, 2000.

Pinto, V. D., Pottenger, W. M., Thompkins, W., "A Survey Of Optimization Techniques Being Used In The Field", IMRL 2000, *Third International Meeting for Research in Logistics*, Trois-Rivières, May 9, 10 and 11, 2000.

Povinelli, R. J., "Time Series Data Mining: Identifying Temporal Patterns for Characterization and Prediction of Time Series Events", Marquette University, Ph. D. dissertation, 1999.

Povinelli, R., "Identifying Temporal Patterns for Characterization and Prediction of Financial Time Series Events," Temporal, Spatial and Spatio-Temporal Data Mining: *First International Workshop*; revised papers / TSDM2000, 46-61, 2000.

Povinelli, R. J. and Xin Feng, "Data Mining of Multiple Nonstationary Time Series" *Artificial Neural Networks in Engineering*, Proceedings, 511-516, 1999.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T., *Numerical Recipes in C*, 2nd Edition, Cambridge Press, Cambridge, 1986.

Radzikowski, P. Bertelsmann, "Non-parametric methods of option pricing", *Informs and Korms* (Conference), Seoul, 2000.

Ross, S., Westerfield, R., Jaffe, J. and Roberts, G., *Corporate Finance*, 1st Canadian edition, McGraw-Hill Ryerson, Toronto, 1995.

Securities & Exchange Commission, "Release No. 34-49798; File No. SR-CBOE-2004-23", Jun. 3, 2004.

Simons, H., "Nothing to Fear", Futures: *The Magazine of Commodities & Options*, Vol. 28, Issue 5, P62-64, May 1999.

Szpiro, G., "A Search for Hidden Relationships: Data Mining with Genetic Algorithms", Israel, *Computational Economics*, 10: 267–277, Kluwer Academic Publishers,1997.

Strang, G., & Nguyen, T., *Wavelets and Filter Banks,* Wellesley Cambridge, 1996.

Tay, F., Shen, L. and Cao, L., "Ordinary shares, exotic methods: Financial forecasting using data mining techniques", *World Scientific*, 2003.

Taylor, J. W., "Evaluating Volatility and Interval Forecast", *Journal of Forecast*, 18, 111-128, 1999.

Tino, P., Nikolaev, N. and Yao, X., "Volatility Forecasting with Sparse Bayesian Kernel Models", *4th International Conference on Computational Intelligence in Economics and Finance (CIEF)*, Salt Lake City, Utah, p. 1052 – 1058, Jul. 2005.

Torrence, C. and Compo, G., "A Practical Guide to Wavelet Analysis", *Bulletin of the American Meteorological Society*, Vol. 79, No. 1, Jan. 1998.

Wang, N. J., K. Wang and J. Jeng, "An Empirical Approach toward Realistic Modeling of Capital Market Volatility", *4th International Conference on Computational Intelligence in Economics and Finance (CIEF)*, Salt Lake City, Utah, p. 1052 – 1058, Jul. 2005.

Weiss, S.M., Indurkhya, N., Zhang, T., Damerau, F, *Text MiningPredictive Methods for Analyzing Unstructured Information*, Springer, 1st edition, 2004

Whitcher, B. "Wavelet-Based Estimation Procedures for Seasonal Long-Memory Models", 5600 MB Eindhoven, The Netherlands, May, 2000.

Wickerhauser, M, V., "Two Fast Approximate Wavelet Algorithms for Image Processing, Classification, and Recognition", *Optical Engineering*, 33(7):2225--2235, July 1994.

Wood, D., "A Link between Option Selling and Rogue Trading?" *Financial Engineering News*, 2004.

Yan, D., and Mukai, H., "Stochastic Discrete Optimization", *SIAM J. Control Optimization*, 30, 3, p. 594-612, 1992.

Zumbach, G., Pictet, O. V. and Masutti O., "Genetic Programming with Syntactic Restrictions applied to Financial Volatility Forecasting", Olsen & Associates Research Institute for Applied Economics Seefeldstrasse 233, 8008 Zurich, Switzerland, Dynamic Asset Management Chemin des Tulipiers 9, 1208, Geneva, Switzerland, 2001.