行政院國家科學委員會專題研究計畫 期末報告

財金管理之風險與波動率估計及預測(第 3 年)

計 畫 類 別 ： 個別型
計 畫 編 號 ： NSC 99-2410-H-004-101-MY3
執 行 期 間 ： 101 年 08 月 01 日至 102 年 07 月 31 日
執 行 單 位 ： 國立政治大學應用數學學系

計 畫 主 持 人 ： 吳柏林

計畫參與人員： 碩士班研究生-兼任助理人員：許書瑜
碩士班研究生-兼任助理人員：李志陞
碩士班研究生-兼任助理人員：廖育琳
博士班研究生-兼任助理人員：楊志清
博士班研究生-兼任助理人員：張天財
博士班研究生-兼任助理人員：陳正傑

報 告 附 件 ： 出席國際會議研究心得報告及發表論文

公 開 資 訊 ： 本計畫涉及專利或其他智慧財產權，2 年後可公開查詢

中 華 民 國 102 年 09 月 21 日

中 文 摘 要 ： 本研究計畫探討亞洲主要市場各型財金經濟時間數列財金中，風險與波動率估計及預測的統計程序。波動率是測量標的物價格變動速度的指標。如何根據個別選擇權目前市價，反推算出的標的物價格變動的速度，並進行風險管理是目前相當重要之課題。在實務上，相同現貨但不同月份、不同履約價的隱含波動率，經常出現極大的差異。本研究主要重點在於發展基於 Choquet 積分形式的一系列風險測度。並將此應用在量化風險相關之價格波動率，投資組合配置與精確風險評價及管理。

中文關鍵詞： 關鍵字:風險，波動率估計，模糊比率熵，區間預測

英 文 摘 要 ： This research project aims at investigating practical statistical procedures, based upon various types of economic/financial time series data, for risk estimation and volatility forecasting. Volatility a measure of the risk in a financial instrument, refers to the standard deviation of the continuously compounded returns with a specific time horizon. The main focus will be on developing the current theory of coherent risk measures, based on fuzzy ratio entropy integral formulation, with applications to assessing quantitative risks related to volatility of stock prices for option pricing, efficient portfolio allocation and accurate risk assessment and management.

英文關鍵詞： Keywords: Risk, volatility estimation, fuzzy ratio entropy, forecasting

行政院國家科學委員會補助專題研究計畫 □期中進度報告
✓ 期末報告

財金管理之風險與波動率估計及預測

計畫類別：個別型計畫
計畫編號：NSC 99-2410-H-004-101-MY3
執行期間：2010 年 8 月 1 日至 2013 年 7 月 31 日

執行機構及系所：國立政治大學應用數學系

計畫主持人：吳柏林
共同主持人：
計畫參與人員：

本計畫除繳交成果報告外，另含下列出國報告，共 2 份：
□移地研究心得報告
✓ 出席國際學術會議心得報告
□國際合作研究計畫國外研究報告

處理方式：除列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，X 一年□二年後可公開查詢

中 華 民 國 101 年 8 月 1 日

Practical Estimation and Forecasting of Risk and Volatility in Financial Management. 財金管理之風險與波動率估計及預測

摘要

本研究計畫探討亞洲主要市場各型財金經濟時間數列財金中，風險與波動率估計及預測的統計程序。波動率是測量標的物價格變動速度的指標。如何根據個別選擇權目前市價，反推算出的標的物價格變動的速度，並進行風險管理是目前相當重要之課題。在實務上，相同現貨但不同月份、不同履約價的隱含波動率，經常出現極大的差異。本研究主要重點在於發展基於Choquet積分形式的一系列風險測度。並將此應用在量化風險相關之價格波動率，投資組合配置與精確風險評價及管理。

關鍵字:風險，波動率估計，模糊比率熵，區間預測

Abstract

This research project aims at investigating practical statistical procedures, based upon various types of economic/financial time series data, for risk estimation and volatility forecasting. Volatility a measure of the risk in a financial instrument, refers to the standard deviation of the continuously compounded returns with a specific time horizon. The main focus will be on developing the current theory of coherent risk measures, based on *fuzzy ratio entropy* integral formulation, with applications to assessing quantitative risks related to volatility of stock prices for option pricing, efficient portfolio allocation and accurate risk assessment and management.

Keywords: Risk, volatility estimation, fuzzy ratio entropy, forecasting

1. **Introduction** 研究計畫之背景及目的。

This research project aims at investigating practical statistical procedures, based upon various types of economic/financial time series data, for risk estimation and volatility forecasting in Asia leading markets. The main focus will be on developing the current theory of coherent risk measures, based on *Choquet* integral formulation, with applications to assessing quantitative risks related to volatility of stock prices for option pricing, efficient portfolio allocation and accurate risk assessment and management.

In financial mathematics and financial risk management, Value at Risk (VaR) is a widely used risk measure of the risk of loss on a specific portfolio of financial assets. For a given portfolio, probability and time horizon, VaR is defined as a threshold value such that the probability that the mark-to-market loss on the portfolio over the given time horizon exceeds this value (assuming normal markets and no trading in the portfolio) is the given probability level. It is a category of risk metrics that describe probabilistically the market risk of a trading portfolio. It has five main uses in finance: risk management, risk measurement, financial control, financial reporting and computing regulatory capital. Important related ideas are economic capital, backtesting, stress testing and expected shortfall. Unlike retrospective risk metrics, such as historical volatility, VaR is prospective. It quantifies market risk while it is being taken.

During the late 1980s and early 1990s, a number of institutions implemented VaR measures to support capital allocation or market risk limits. The Group of 30 (1993) published a groundbreaking report on derivatives practices. It was influential and helped shape the emerging field of financial risk management. It promoted the use of value-at-risk by derivatives dealers and appears to be the first publication to use the phrase VaR. In 1994, JP Morgan launched its free RiskMetrics service. This was intended to promote the use of VaR among the firm's institutional clients. The service comprised a technical document describing how to implement a VaR measure and a covariance matrix for several hundred key factors updated daily on the internet. These were based upon a crude VaR measure, but the committee also approved, as an alternative, the use of banks' own proprietary VaR measures in certain circumstances.

The following statements bring out the research problems we will investigate in this proposal. In a sense, this research proposal is in line with the spirit of the Nobel Lecture " *Risk and Volatility: Econometric Models and Financial Practice*" delivered by Robert F. Engle III, on December 8, 2003. Specifically, while modern quantitative theory of financial management started with the theory of portfolio selection [13], [14], and the Capital Asset Pricing Model (CAPM) [15] [16], it was the foundational work and which put financial management on a firm footing for further quantitative analysis.

On the other hand, financial market activities can be analyzed by examining the ways in which complex financial instruments are priced via the now well known Black-Schole model. When practitioners want to implement financial strategies, they require, first of all, estimates of the fluctuation (called volatility) of the underlying, say, stock prices. As variance (or standard deviation) of the random variable representing the future stock prices, the volatility is in fact changing over time. The main contribution of Engle (1982) is to model the volatility as a stochastic process, that is a collection of random variables indexed by time, in a specific form, now known as the Autoregressive Conditional Heteroskedasticity (ARCH model). Clearly financial behavior analysis depends essentially on the knowledge of volatility involved. Note
that it is really the volatility over a future period that should be considered as "risk". Thus, in fact, we need a *forecast of volatility*, in addition to *estimation of volatility*.

Now, the volatility in stock prices is a useful index for financial analysis as it is an indication of some sort of *risk*. However, risk management science is based also on other quantitative risk assessment such as the popular *VaR*. In a sense, volatility in financial markets is the likelihhood of fluctuations, say, in the exchange rate of currencies. Therefore, it could be modeled as the probability of the threat that an exchange rate movement poses to an investor's portfolio in a foreign currency. An extension of the concept of standard deviation to *VaR* will help to determine the actual risk exposure to a protfolio of several currencies. And yet, current research on Coherent Risk measures extends *VaR* further into more appropriate risk measures.

Confectionary we choose the best fitness model from the appropriated model family, like ARIMA model family, GARCH model family or nonlinear modes (such as threshold model family) when we are doing the analysis work. Yet, due to the inaccuracy or incomplete information, time lag, or the interwork among variables, it seems very difficult to find an appropriate/accurate model in the reality life. For example, which is the exact number of registered students every year-at the beginning of the semester, in the middle of the semester, at the end of the semester or an average of the above numbers? Different sampling time often results in different model. Furthermore, which is the weighted stock price index--the opening quotation, the closing quotation, or an average of the highest and the lowest stock price?

Zadeh's (1965) proposed of Fuzzy Set Theory, the theory has incorporated the property of linguistic variables, which is capable of reducing possible trouble in dealing with uncertain problems as well as providing a more reliable way in processing complex, diverse and uncertain phenomena. Lately, the application of fuzzy set theory in time series has been increasing gradually, e.g. Wu and Hung (1999) proposed a fuzzy identification procedure for nonlinear time series: with example on ARCH and bilinear models.. Wu and Chen (1999) utilized fuzzy clustering method to check the data structure's transitional span in time series. Tseng and Tzeng, etc. (2001) proposed a fuzzy ARIMA model to predict the NT dollar exchange rate to US dollar by combining traditional time series ARIMA model with fuzzy regression model. Tseng and Tzeng (2002) also combined fuzzy theory with seasonal ARIMA to establish a prediction model. Kumar and Wu (2001) discovered that the concept of fuzzy logic can be used to effectively determine structural changes in non-linear time sequence. Zhou (2005) introduced a new observation in structural change – the Integrating Bayesian Structural Break Model and the change point detection methods. Zeng and Li (2006).Relationship between similarity measure and entropy of interval valued fuzzy sets.

The fuzzy time series analysis and forecasting are recent hot research topics. Conventional studies in related literature emphasize on performing anti-fuzziness on data, and then classify them. So-called anti-fuzziness of data, is to describe the nature of fuzziness of subject matters, or to find out the fuzzy relationship between every subject matter. "Classification", on the other hand, is the grouping of data that share the same nature of fuzziness or relationship. However, literature discussing the structural change of data was rare. That was why (Wu, 1999) combined these two kinds of knowledge to construct a series of procedures that is effective in finding out the structural change within data using the fuzzy classification.

This proposed technique is applied to a financial time series forecasting problem for demonstration. Stock price forecasting is considered in the empirical application. We also propose an appropriate method to measure the forecasting performance for the interval data. Finally, comparing the performance with the traditional ARIMA model and , it has shown that the proposed method demonstrates a promising performance on predicting future values for interval time series.

Our research project aims at addressing practical aspects of supplying estimates and predictions for volatility and risk from available economic data.

(i) Estimation of volatility.

Recall that the classic Black-Schole option pricing formula (providing a way to price options so that no arbitrage is possible) requires, as inputs, five parameters : the initial stock price, the option exercise time, the option strike price, the interest rate, and the volatility. All these parameters are known (observable, obtained from the market) except the last one, namely the volatility (which is not directly observable) which must be forecasted.

Even in the case where the volatility is assumed to be non stochastic and time-invariant (i.e. an unknown constant), its estimation efficiency depends on the type of data used. For example, the volatility can be estimated by the sample standard deviation of returns over a short period of time. But, what is the right period of time to use? In other words, how to get the best estimate of the "risk" from daily, weekly or monthly data? as well as how to take the length of the data into account in the estimation process ? Thus, we are facing the problem of historical data should be used in predicting volatility.

In the context of volatility estimation, we will address current estimation procedures based on conditional volatility, stochastic volatility, realized volatility, implied volatility as well as liquidity risk and variance of Black-Schole option pricing.

Stochastic volatility models, such as ARCH processes, are heavily used in financial economics to capture the impact of time-varying volatility on financial markets and decision-making. The studies of inference in such models will produce methods that aid our understanding of option pricing, efficient portfolio allocation and accurate risk assessment and management. The statistical inference we intend to pursue is based upon return data. Practical methods to be considered include likelihood or quasi-likelihood, generalized method of moments, Kalman filter, and possibly simulated-based inference procedures such as Markov chain Monte Carlo techniques.

Under the general heading of realized volatility, an important research issue is how to use high-frequency data to estimate, to predict future levels of volatility. Here, we intend to examine the recent work of Andersen et al (2001) on using predictions of the future daily quadratic return variation as a key input for forecasting the volatility of future asset returns. Of course, this research will be based on stochastic calculus (Ito calculus) involving stochastic differential equations driven by semi-martingales.

(ii) Volatility forecasting.

Volatility forecasting is vital for derivatives trading, but it remains an art rather than a science, particularly among derivatives traders.

Forecasting of quantities of interest gives important information during decision-making processes, especially on economic developments, population policies, management planning or financial control. In our context of volatility forecasting, several drawbacks exist such as : Inefficiency in market applications due to the inacurate forecasting, the model constructed only by the closing price may not illustrate the whole process of daily or monthly trend.

One of the reasons for the above drawbacks is that the business marketing is full of uncertainty and human being's manipulation. To try to overcome these drawbacks, we will consider several avenues, including interval forecasting, such as interval ARIMA time series models.

(iii) Modeling of quantitative risk measures.

For both investment and actuarial sciences, decision-making can be in principle carried out in the context of Von Neumann utility maximization theory. However, the difficulty of specifying individuals' utility functions has led to the recent development of the *stochastic dominance theory*, in which comparisons of loss variables are performed via stochastic orders based upon distribution functions (see e.g. the recent research monograph " Stochastic Dominance and Applications to finances, Risk and Economics" by Songsak Sriboonchitta et al, 2009). While this theory avoids the specific knowledge of utility functions (only use their shapes as risk attitudes), its stochastic orders are partial order relations and as such, cannot handle situations where loss variables are not comparable. A "follow-up" decision-making criterion is needed and that should be a consistent total order relation. This is achieved by asking the fundamental question " Can we assign a numerical value to a loss variable, representing its quantitative risk ?"

Note that it has been recognized that the variance (or standard deviation), while representing volatility, is not an appropriate risk measure in view of its symmetric property. Risk measures are rather down side risks, and as such should be assymmetric. That was precisely the reason why the *VaR* was proposed. With the modern research on coherent risk measures, started with Artzner et al (1999), the state-of-the-art is this. The *VaR* is not coherent since it violates the diversification principle in investment (namely, diversification should reduce risk), and should be modified into another risk measure, namely the Tail-*VaR*. But this modified risk measure is just one candidate among many others for modeling risk.

The list of desirable properties of coherent risk measures suggests that the most general form for coherent risk measures is built up by a mathematical integral known as Choquet integral. The successful risk measure proposed for premium calculations in actuarial science by Wang (2002) is in fact a special case of a Choquet integral risk measure. Thus, in studying Choquet integral risk measures, we will establish a unifying pricing theory for insurance and financial risks towards applications to a unified risk management.

While the Choquet integral risk measures satisfy desirable properties for quantifying financial risks, its practical implementation remains an important issue of concern for applications. The main problem is the subjective choice of a "distortion function" in the construction of the Choquet integral. We intend to address this problem as follows.

First, it is possible to transform a Choquet integral risk measure (based on distortion functions) to another type of risk measures, known as spectral risk measures, in which the distortion function is mapped into a weighting function (the sprectrum). A systematic analysis of spectrums could reveal that they are built from utility functions in some canonical fashion, reflecting risk attitudes of investors.

The research should bring out at least two important facts : *How to choose risk measures consistently with investors' risk aversion ? and how to estimate such risk measures, say, from historical economic data ?* This second point will be relied upon the well established asymptotic theory of statistical inference of linear functions of order statistics, since the spectral risk measures are operating on quantile functions (of the loss variables) rather than directly on their distribution functions.

(iv) Prediction and Risk

In the standard approach to risk analysis of options, the volatility of the underlying asset is used to assess its risk in the sense that " the greater the volatility of the underlying asset, the greater is the risk attached to the option". Now, the *risk* of an option is measured by the variance of the rate of return on the option. As stated before, coherent risk measures have been shown to be more appropriate to quantify risks. Thus, while we estimate or forecast volatility for option pricing, we will develop a new approach to risk assessment in the context of financial markets by replacing standard deviations by more general and coherent risk measures, such as Choquet integral risk measures.

References.

[1]. P. Jorion (2006), *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd ed. McGraw-Hill.

[2]. Songsak Sriboonchitta et al (2009), *Stochastic Dominance with Applications to Finance, Risk and Economics*, Chapman and Hall/ CRC Press.

[3]. C. Acerbi (2002) Spectral measures of risk: A coherent representation of subjective risk aversion, *J. Banking and Finance* (7) 1505-1518.

[4]. P. Artzner et al (1999) Coherent measures of risk*, Mathematical Finance* ,9(3) 203-228.

[5]. D. Bosq and H.T. Nguyen (1996) *A Course in Stochastic Processes and Statistical Inference*, Kluwer Adacemic Press.

[6]. R. F. Engle (1982) Autoregressive conditional heteroscedasticity with estimates of variance of UK inflation, *Econometrica* (50) 987-1008.

[7]. H. Follmer and A. Schied (2004) *Stochastic Finance*, De Gruyer.

[8]. R. Dana and M. Jeanblanc (2003) *Financial Markets in Continuous Time*, springer-Verlag.

[9]. A. Shiryaev (2008) *Essentials of Stochastic Finance*, World Scientific.

[10].van Deventer, R. Donald, K. Imai and M. Mesler (2004). *Advanced Financial Risk Management: Tools and Techniques for Integrated Credit Risk and Interest Rate Risk Management*. John Wiley.

3. Research methodology and approach.

We spell out here the basic background of our research framework along with our research methodology in order to achieve our desirable goals in this research proposal.

(i) Quantitative coherent financial risk measures.

Since the variance of a future return is a symmetric measure of deviation about the expected value, it is not appropriate for a downside risk measure. As such, the *VaR* of , say, a loss variable XF_α , with (unknown) distribution function Fx, at confidence level _ 2 (0; 1), was proposed as a risk measure, where

$$VaR_\alpha(X) = F_X^{-1}(\alpha) = inf\{x: F(x) \geq \alpha\}$$

While $VaR_\alpha(X)$ has a transparent meaning of the maximum possible loss with confidence, it violates the reduction of risk by diversification. In mathematical terms, the operator $VaR_\alpha(\cdot)$
Is not subadditive:

$$VaR_\alpha(X+Y) \nleq VaR_\alpha(X) + VaR_\alpha(Y)$$

A modified type of quantile-based risk measure is the Tail *VaR*

$$TVaR_\alpha(X) = \frac{1}{1-\alpha}\int_0^\infty F_X^{-1}(t)\,dt$$

which is subadditive.

This is just one possible candiate for a risk measure, i.e. a way to assign a numerical value to a distribution function, representing its risk. Now, economists have listed a collection of desirable properties for any risk operator to be qualified as a reasonable risk measure. They are called coherent risk measures. It turns out that the most general form of coherent risk measures is obtained by an integral called Choquet integral. Specifically, a Choquet integral risk measure is of the form

$$\rho(X) = \int_0^\infty g(1 - F_X(t))\,dt$$

where *g*:[0; 1]→ [0; 1] , nondecreasing function. The above VaR_α and $TVaR_\alpha$ are special cases. Moreover if the "distortion" function g is concave, the associated Choquet integral risk measure is coherent.

Choquet integral risk measures thus surface as desirable candidates for measuring financial risks. For applications, we need to pursue two more issues. The first one is how to supply the knowledge of the unknown distribution function F_X? Clearly this is a statistical estimation problem which, as usual, requires the attention to two things : the model and the data.

We will investigate details in each case of model, namely parametric, semiparametric and nonparametric. The second thing is the type of available data, such as high frequency data. Accuracy of estimators depends on these two ingredients. Our approach will consist of using the theory of linear functions of order statistics.

The second issue in implementing a Choquet integral risk measure is the choice of the distortion function *g*. Our approach is this. It is possible to convert a Choquet integral risk measure into a "spectral" risk measure, i.e. a risk measure of the form

$$\rho(X) = \int_0^\infty \varphi(t) F_X^{-1}(t)\, dt$$

where the spectrum φ:[0; 1] → R^+, nondecreasing and $\int_0^\infty \varphi(t)\, dt$ =1

Now it can be shown that the spectrum spectrum φ reflects risk aversion attitude of investors, and hence a reverse transformation back to distortion function g should yield a reasonable choice.

(ii) Using coherent risk measures in relation with volatility.

First of all, recent advances in coherent risk measures have shown that Choquet integral risk measures are appropriate for risk assessment both in investment and actuarial sciences. Their relation with option pricing seems promising. We will investigate the methods from insurance risk pricing and their relationship to financial theory.

Financial market volatility is indispensable for asset and derivative pricing, asset allocation, and risk management. As volatility is not directly observable, research efforts have largely devoted to how to obtain information about it.

The volatility in the classic Black-Schole option pricing model is the constant variance σ^2 of the underlying stock price S_t at time t : Assuming that S_t follows a geometric Brownian motion W_t, the price C of an option in a market with interest rate r, with strike price K, is a function $C(S_0; K; t; r; \sigma^2)$, where S_0 is the initial stock price. Pricing according to this formula avoids arbitrage. The volatility in the Black-Schole model is called the implied volatility (of an option), obtained by solving for _2, giving the market price. Note that, in general, the volatility depends on the strike price, and hence forms a curve called "volatility smile".

In general, the volatility is time-varying and stochastic, i.e. it is a stochastic process σ_t^2 which can be modeled as a Generalized Autoregressive

$$\sigma_t^2 = \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_i \sigma_{t-j}^2$$

Where ε_i is a white noise. The above view of volatility is clearly model-dependent.

We will also look at model-free measure of volatility where volatility can be estimated by sample variance of returns.

In the context of volatility, we will investigate its relation with risk measures expressed as Choquet integral.

(iii) Estimation of Choquet integral risk measures.

For practical applications, we will consider various statistical models and data in the search for efficient statistical estimators. Several venues will be considered :

(a) Estimating $\int_0^\infty g(1 - F(t))\, dt$ by a plug-in estimator $\int_0^\infty g(1 - F_n(t))\, dt$ where $F_n(\cdot)$ could be some consistent estimator of $F(\cdot)$ based on data sample of size n, such as the empirical distribution function

(b) $$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{(X_i \le x)}$$

or, when F is absolutely continuous, with density function $f(\cdot)$, by some smooth estimator using nonparametric kernel estimation

$$\hat{F}_n(x) = \int_{-\infty}^{x} f_n(t)\, dt$$

where $f_n(t) = \frac{1}{nh_n}\sum_{i=1}^{n}\frac{K(x-X_i))}{h_n}$ a kernel density estimate of $f(\cdot)$

Of course, statistical properties of the above estimators should be investigated. For example, large sample properties could be examined within the framework of V- statistics.

(b) Nonparametric estimation of quantile functions could be used since, as noted above, Choquet risk measures can be written in terms of quantile functions.

Given a sample $X_1;X_2; :::;X_n$, the order statistics are denoted as $X_{(k)}$. The quantile function

$$F_n^{-1}(\alpha) = X_{(k)} \text{ for } \frac{k-1}{n} < \alpha \leq \frac{k}{n}$$

So that $\int_0^{\infty} \varphi(t) F_X^{-1}(t)\, dt$ is estimated by $\int_0^{\infty} \varphi(t) F_X^{-1}(t)\, dt = \sum_{i=1}^{n}[\int_{i-1/n}^{i/n} \varphi(t) dt]\, X_{(i)}$

From that, asymptotic properties of these estimators can be derived by using the theory of linear functions of order statistics.

(iv) Volatility forecasting.

For both estimation and forecasting of volatility, we will examine various practical models as well as data. for example, we address the problem of estimating and predicting volatility by using implied and historical volatilities.

Two cases of interests are : when σ_t^2 is an unknown non-random function, and when it is a random process, independent of the driving Brownian motion noise in the stock. We consider both the situations where future volatility could be calculated from time series economic data. Also, we will investigate the problem from the use of realized volatility (the realized volatility measures what actually happened in the past).

The concept of entropy originates from machine engineering, where in a process involving heat it is a measure of the portion of heat becoming unavailable for doing work. In information theory, Entropy is a measure of *unpredictability* or *information content* in a random variable. Since entropy can be measured in price values or range of prices, in this context, the term will be referred to the *ratio entropy*, which ratio quantifies of the time series. The role of entropy rate for a dynamic process is one bit per value. For exmple, if the process demonstrates upward trend, and hence the entropy rate, is lower. This is because, if asked to predict the next outcome, we could choose the most frequent result and be right more often than wrong. The entropy of a message multiplied by the length of that message is a measure of how much information the message contains. If some messages come out smaller, at least one must come out larger. In practical use, this is generally not a problem, because we are usually only interested in compressing certain types of messages, for example English documents as opposed to gibberish text, or digital photographs rather than noise, and it is unimportant if our compression algorithm makes certain kinds of sequences larger. However the problem can still arise even in every day use when applying a compression algorithm to an already compressed data

Wu (1999) suggested that, the use of fuzzy entropy is effective in identifying whether a structural change happens in a time series. Besides, it can also be used together with the mean cumulated fuzzy entropy of t times, to observe the change in message of fuzzy entropy, based upon which a standard for the classification of change model can be established.

$$\begin{cases} x_{t+1} = a_1 + \phi x_t + \varepsilon_{t+1} & \varepsilon_{t+1} \sim WN(0, \sigma_\varepsilon^2) \\ l_{t+1} = a_1 + l_t + \delta_{t+1} & \delta_{t+1} \sim WN(0, \sigma_\delta^2) \end{cases} \tag{3.1}$$

Definition 3.1 Forecasting for Interval time series model

Let $\hat{x}_t(i)$ and $\hat{l}_t(i)$ be the one step forecasting value of the interval time series as in equation (2.1), $xf(i)$ and $lf(i)$ We denote the revised forecasting value as

$$\begin{cases} xf(1) = \hat{x}_t(1) + sign(l_t - l_{t-1})\frac{l_t}{l_{t-1}} \\ lf(1) = \hat{l}_t(1) + sign(x_t - x_{t-1})\frac{x_t}{x_{t-1}} \end{cases} \tag{3.2}$$

$$\begin{cases} xf(i) = \hat{x}_t(i) + sign(\hat{l}_t(i) - \hat{l}_t(i-1))\frac{\hat{l}_t(i)}{\hat{l}_t(i-1)} \\ lf(i) = \hat{l}_t(i) + sign(\hat{x}_t(i) - \hat{x}_t(i-1))\frac{\hat{x}_t(i)}{\hat{x}_t(i-1)} \end{cases} \tag{2.3}$$

An integrated decision system for fuzzy time series analysis and forecasting

Here we wish to find two cluster centers. This is determined based on common experience of empirical analysis and trend of time series. The procedures are as follows:

Step 1: Use the k-means method to find out two cluster centers C_1 and C_2 in time series $\{x_t\}$, and determine the membership degree μ_{it}, i=1,2 of $\{x_t\}$ to the two cluster centers

Step 2: Compute the fuzzy entropy $\delta(x_t)$, mean cumulated fuzzy entropy $MS\delta(x_t) = \frac{1}{t}\sum_{i=1}^{t}\delta(x_i)$, and Median($MS\delta(x_t)$) of this series, that correspond to x_t.

Step 3: Take a suitable threshold value λ, classify the mean cumulated fuzzy entropy $MS\delta(x_t)$ series that correspond with x_t. If the mean cumulated fuzzy entropy $MS\delta(x_t)$ falls into the interval [0,Median($MS\delta(x_t)$)−λ), we will use 1 to represent Group 1; if $MS\delta(x_t)$ falls into the interval [Median($MS\delta(x_t)$)−λ,Median($MS\delta(x_t)$)+λ), we use 2 to represent Group 2; and if $MS\delta(x_t)$ falls into the interval [Median($MS\delta(x_t)$)+λ,1], 3 will be used to represent Group 3.

Step 4: If the result of classification is inconsistent, we then make adjustment to the result. If it is consistent, go to Step 5.

Step 5: Select an appropriate determination level α. If the number of consecutive samples is greater than $[\alpha N]$, then these consecutive samples belong to the same group. During classification, if more than *one group is found, we know that structural change happens in this time series. Thereafter, find the change interval.*

Step 7. Construct a system of AR(1) model for the center and length of the dynamic process.

Step 8.forecasting the time series as in the equation (2.2).

Comparison with forecast results

After constructing the model, we look at the core interest of this study – the forecasting ability. Table 3.1 is the comparison results for the forecast of TSMC between the best ARIMA model and fuzzy classification ARIMA. We can see that the forecast result of the TSMC using the best fuzzy classification Threshold model haa better forecasting performance than the traditional ARIMA model.

Figure 3.2 illustrates the forecast result. In the process of correlation, the fuzzy classification is also an important procedure, because this reduces the number of samples required. The forecast result with the ratio entropy technique for TSMC provides a better forecasting performance than threshold ARIMA model. We also find that, for a time series, if the change period of its structural change can be determined, better results on the model construction and forecasting ability can be produced.

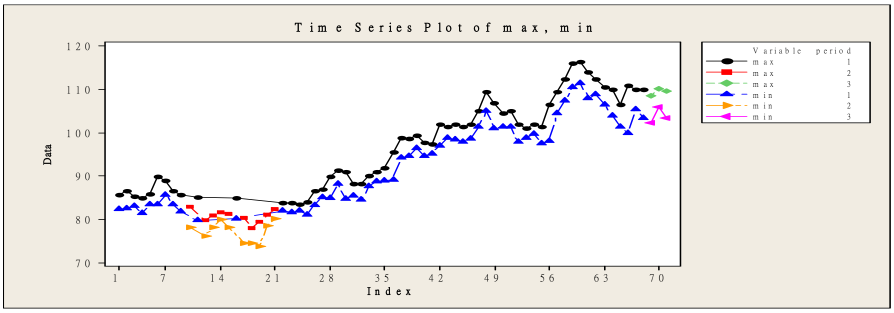

Figure 3.2 Trend and forecasting for the interval TSMC data

Finally, from Table 3.1 we can see the *IMSE* is very small, and this indicates that the result is quite appreciated.

Table 3.1 Comparison between the forecasts of the TSMC

Time/week	*Actual value*	*Interval forecasting(110x111)*	*Revised Interval forecasting*
102/06/24	[100,111]	[104.3,109.5];(106.9;5.2)	[102.3,108.5];(105.4;6.2)
102/07/01	[105.5,110]	[104.8,109.9];(107.3;5.1)	[106.0,110.2];(108.1;4.1)
102/07/08	[103.5,110]	[105.0,110.2];(107.6;5.1)	[103.5,109.7];(106.6;3.1)
IMSE		*6.49* $(=2.24^2+0.55^2+1.07^2)$	*1.04* $(=0.75^2+0.42^2+0.55^2)$

Conclusion and sugestions

Economic and financial analysts often need to know when changes occur in a time series. In this research we formalize the concept of change periods in contrast with traditional change points as more realistic structural features of certain time series. We present an approach to detect change periods by partial cumulative sums of fuzzy statistics, allowing us to identify the beginnings and ends of trends. Through the use of fuzzy statistics, our proposed change period detection approach is able to systematically address fuzziness in the data. As a consequence, its results are more meaningful in financial/economic time series analysis.

The key contribution of this paper is that we provide a new method to prediction as well as the decision making

In comparison with conventional methods, our approach offers several advantages:

1. Initial knowledge about the structure in the data is not required, so we can take full advantage of the model-free approach.
2. We can select standards for change periods by controlling the parameters to detect change periods at the scale desired and filter noise in a time series.
3. The fuzzy data can be handled.
4. Practical aspects of volatility forecasting. The problem of building optimal predictors for volatility will be investigated on financial/economic time series data. Developing models for analyzing financial risk measures and volatility. Essentially, we will investigate practical ways to choose coherent risk measures, reflecting risk aversion , and models for stochastic volatility.

Although the simulation and empirical results show that our approach of change period detection is visually satisfactory and can be generally applied, there remain several points to note and problems to be solved:

Because the change periods we consider are defined as intervals where trends change, the stationary part of a time series may be viewed as a change period. Future study should address sensitivity of its results to the parameter choice.

1. While our research efforts are interconnected, they can be focused as follows.
2. Developing models for analyzing financial risk measures and volatility. Essentially, we will investigate practical ways to choose coherent risk measures, reflecting risk aversion , and models for stochastic volatility.
3. Practical aspects of financial risk estimation. We will conduct research for estimating financial risk under various forms of economic data, as well as under various forms of distribution assumptions. In other words, both parametric and non-parametric statistical estimation (and testing) will be investigated.
4. Practical aspects of financial risk estimation. We will conduct research for estimating financial risk under various forms of economic data, as well as under various forms of distribution assumptions. In other words, both parametric and non-parametric statistical estimation (and testing) will be investigated.

附件二

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估 ● X 達成目標 □ 未達成目標（請說明，以 100 字為限） □ 實驗失敗 □ 因故實驗中斷 □ 其他原因
2. 研究成果在學術期刊發表或申請專利等情形： 論文：□已發表 □未發表之文稿 □撰寫中 □無 專利：□已獲得 □申請中 □無 技轉：□已技轉 □洽談中 □無 其他：（以 100 字為限）
3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值 In comparison with conventional methods, our approach offers several advantages: ● Initial knowledge about the structure in the data is not required, so we can take full advantage of the model-free approach. ● We can select standards for change periods by controlling the parameters to detect change periods at the scale desired and filter noise in a time series. ● The fuzzy data can be handled. ● Practical aspects of volatility forecasting. The problem of building optimal predictors for volatility will be investigated on financial/economic time series data. Developing models for analyzing financial risk measures and volatility. Essentially, we will investigate practical ways to choose coherent risk measures, reflecting risk aversion , and models for stochastic volatility.

行政院國家科學委員會補助國內專家學者出席國際學術會議報告

2013 年 1 月 20 日

報告人姓名	吳柏林	服務機構及職稱	國立政治大學應用數學系教授
時間 會議地點	2013/1/10-1/12 Chingmai,Thailand	本會核定 補助文號	計畫編號 NSC 99-2410-H-004-101-MY3
會議名稱	The 6th International Conference on Econometrics 2013		
發表論文題目	Model Construction and Residues Analysis with Interval Time Series		

一.參加會議經過

The 6th International Conference on Econometrics 2013, 於 2013 年 1 月 10 日至 1 月 12 日在 Chingmai,Thailand 舉行, 由 Chinagmai University 主辦,. 來自各國之學者專家約有 70 餘人參加, 包括地主國外, 韓國 台灣, 美國, 加拿大, 澳洲, 南非, 馬來西亞, 捷克等 10 餘國. This conference aims at bringing together researchers in econometrics for an opportunity to present and discuss theoretical and applied research problems as well as to foster research collaborations. The main theme of this Fifth International Meeting is ***Financial Econometrics*** and the conference is open to any topics in economics. The Fifth Conference will feature two keynote addresses: (1) Paul Embrechts, Swiss Federal Institute of Technology, Zurich, Switzerland(2) Christian Gourieroux, Paris Graduate School of Economics, Statistics and Finance, France

2 本次大會就以下幾項重點主題進行研討

1. The Fifth Conference will feature two keynote addresses: (1) Paul Embrechts, Swiss Federal Institute of Technology, Zurich, Switzerland (2) Christian Gourieroux, Paris Graduate School of Economics, Statistics and Finance, France and several invited addresses from prominent economists. 以及一些新觀念如: Perception measurement, Stochasticvalidity，創新管理與創新計算等新看法。

二. 攜回的資料：

1. The 6th International Conference on Econometrics 2013 研討會論文集 2.與 Professor Nguyen, Hung, Vladick Kreinovich 等學者討論有關最近著作與研究結果。

三. 建議與其他

目前筆者在政大應數研究所開設模糊統計與時間數列分析與預測課程多年，深感學術研究發展日新月異、一日千里。國科會能給予補助出國出席國際學術研討會，收穫相當大。希望將來能多利用課餘時間出國做短期研究，吸收國外新知、及研究方向。回國後繼續開設模糊時間數列分析與預測課程，指導博、碩士班研究生，籌辦國際學術研討會，推動國際財金與管理學術研究工作。並於國際著名學術論文期刊，發表學術論文。

Model Construction and Residues Analysis with Interval Time Series

Berlin Wu

Department of Mathematical Sciences

National Chengchi University, Taiwan

berlin@nccu.edu.tw

Abstract: This paper demonstrates how to use the fuzzy classification technique to perform an intensive research on the change periods detection and model construction for the interval time series. We use average of the sum of fuzzy entropies to find out interval of the structural changes. Focusing on the time series of intervals, we build a model and make prediction about it. The residues analysis for fuzzy time series model are proposed. Empirical studies show that the unemployment rate does significantly correlate with the population of singles.

Keyword: *fuzzy data classification* 、*average of the sum of fuzzy entropies* 、*change periods* 、*unemployment rate* 、*population of singles*

1. Introduction

As the trend of getting married late or not getting married emerges, Taiwan faces the problem with low birth rate and aging of population. The issues of the welfare for the aged and age structure of population receive increasing attention too. The heightening unemployment rate, on the other hand, has caused more people to marry late or not marry at all, and this increases the number of people who remain single. Therefore, a study like this that investigates the future trend of singlehood rate becomes more and more important. Conventionally, studies recorded in related literature emphasize on performing anti-fuzziness on data,and then classify them. This process is called “fuzzy classification”. So-called anti-fuzziness of data, is to describe the nature of fuzziness of subject matters, or to find out the fuzzy relationship between every subject matter. “Classification”, on the other hand, is the grouping of data that share the same nature of fuzziness or relationship. However, literature discussing the structural change of data was rare. That was why (Wu, 1999, Kumar and We (2001)) combined these two kinds of knowledge to construct a series of procedures that is effective in finding out the structural change within data using the fuzzy classification. Weina Wang (2007), Malay K (2005), Kuo-Lung Wu (2005), Dae Won Kim (2004) further constructed different methods of calculation, to investigate the effective way of using the Fuzzy C-Means to determine the cluster in a large number of sets.

Although there have been numerous researchers performing studies on structural change, and countless of methods used, the mathematical inference process is complex, and there still has not been a clear standard in defining change point. Therefore another purpose of this thesis is to find out the structural change in the unemployment rate and effectively analyze it, and then transform its structural change part with the singlehood rate, to discover a better forecasting skill.

In Taiwan, it becomes increasingly common for males before 25 to be still pursuing their study. For these reasons, we do research on males between the ages of 25 - 34 as the object of study, to investigate the relationship between singlehood rate and unemployment rate.

2. Model construction with fuzzy data

2.1 The role of fuzzy entropy

There are many analysis techniques for time series. The oft-used ones include exponential smoothing, ARIMA, log linear trend, linear trend with seasonal terms, etc. Transfer function model is the extension of the construction method of a univariate time series model to the analysis method of a multivariate time series. In many cases, it is possible to have a set of data whose current observed value is affected by the past observed value, and one or more other set of time series are correlated to the mentioned set of data. This implies that there will be impact transferred to the output series, when there are any changes in the input series.

Since the possibility of the singlehood rate being affected by unemployment rate of previous periods and the past singlehood rate itself, it is more accurate to use the transfer function mode to perform the conversion of the singlehood rate.

When the fuzzy theory is used to examine whether there is any change point in a time series, first we cluster the time series, find out the cluster center, and then use the fuzzy membership degree, fuzzy entropy and other relevant concepts to perform classification.

Definition 2.1 Fuzzy Entropy

Let a time series be $\{x_t, t=1,2,\ldots,N\}$*, with* μ_{it} *being the membership degree of* x_t *to the cluster centers Ci (*$i=1,2,\ldots,k$*), the fuzzy entropy is thus defined as:*

$$\delta(x_t) = -\left(\frac{1}{k}\right)\sum_{i=1}^{k}\left[\mu_{it}\ln(\mu_{it}) + (1-\mu_{it})\ln(1-\mu_{it})\right]$$

Entropy is a concept in the thermodynamics study, it illustrates the degree at which work can be transformed. The Probability Theory and Information Theory give it a more common definition: measure of the unboundedness of a random variable, or the measure of the amount of missing information. So fuzzy entropy is used to measure the uncertainty of fuzzy sets, and is an important tool for the processing of fuzzy data, while the membership degree is used to characterize elements that do not clearly belong to some particular sets.

2.2 Distance with interval data

When a sample of interval-valued fuzziness is available, we have to consider the calculation for intervals. However, there is still no complete definition for the measure of interval distance (see Wu 2010). How to define a well-defined interval distance? First we represent the interval with $(c_i; r_i)$ with c being the center, r being radius. This way, the interval distance can be considered as the difference of the center plus the difference of the radius. The difference of the center can be seen as the difference in location, and the difference of the radius can be seen as the difference in scale. However, in order to lower the impact of the scale difference on the location difference, we take the *ln* value of the scale difference, and then plus exp(1) to avoid the *ln* value becoming negative.

Definition 2.2 ***Defuzzyfication for a trapezoid fuzzy number on R***

Let $X = [a,b,c,d]$ *be a trapezoid fuzzy number on* U *with its centroid Then the defuzzification value* X_f *of* *X*=[a,b,c,d] *is defined as*

$$X_f = cx + \frac{\|X\|}{2\ln(e+|cx)|};$$

where, $\|X\|$*is the area of the trapezoid.*

Definition 2.3 ***Distance among fuzzy data***

Let $X_i = [a_i, b_i, c_i, d_i]$ *be a sequence of trapezoid fuzzy number on* U *with its centroid* (cx, cy). *Then the distance between the trapezoid fuzzy number* X_i *and* X_j *is defined as*

$$d(X_i, X_j) = \left|cx_i - cx_j\right| + \left|\frac{\|X_i\|}{2ln(e+|cx_i|)} - \frac{\|X_j\|}{2ln(e+|cx_j|)}\right|$$

Definition 2.4 I***nterval means square error (IMSE)***

Let $\{s_i = [a_i, b_i], i = 1, \ldots, N\}$ *be an interval time series, with prediction interval being* $\hat{s}_i = [\hat{a}_i, \hat{b}_i]$ *and* $\varepsilon_i = d(s_i, \hat{s}_i)$*beging the error between the prediction interval and the actual interval, thus:*

$$IMSE = \frac{1}{l}\sum_{i=N+1}^{N+l} \varepsilon_i^2$$

where l is the forecasted expectancy value.

Example 2.1 Let S={[3,6], [4,5], [2,6], [5,8], [3,8]} *and* $\hat{S}$= {[3,4], [2,6], [3,5], [5,7], [4,5]} *be two set of fuzzy sample about expected salary and real salary from a survey for graduated students. Then the distance from expected salary and actual salary are computed as follows:*

Table 2.1 Distance for the interval data

Sample	*Estimated Salary*	*Actual Salary*	*distance*
1	[3,6]	[3,4]	$\mathrm{d}(s_1, \hat{s}_1) = \|3.5 - 4.5\| + 2.05 = 3.05$
2	[4,5]	[2,6]	$\mathrm{d}(s_2, \hat{s}_2) = \|4 - 4.5\| + 2.82 = 3.32$
3	[2,6]	[3,5]	$\mathrm{d}(s_3, \hat{s}_3) = \|4 - 4\| + 1.90 = 1.90$
4	[5,8]	[5,7]	$\mathrm{d}(s_4, \hat{s}_4) = \|6 - 6.5\| + 1.17 = 1.67$
5	[3,8]	[4,5]	$\mathrm{d}(s_5, \hat{s}_5) = \|4.5 - 5.5\| + 4.28 = 5.28$
			IMSE=$\frac{1}{5} \times (3.05^2 + 3.32^2 + 1.90^2 + 1.67^2 + 5.28^2) = 10.9$

Definition 2.5 ***Cluster of interval-valued time series***

Let$\Psi = \{s_t$，t=1,2,……,N}be an interval time series,k ∈Nbe the number of clusters. If there exists a set $\mathcal{I} = \{\mathrm{I}_i \in$ interval; $i = 1,2,\ldots,k\}$,where by the distance square sum of an element s_t from Ψ and an element I_i from $\mathcal{I}$ is the least, then:

$$\mathrm{Min}\sum_{\mathrm{t}=1}^{\mathrm{N}}\sum_{i=1}^{\mathrm{k}} \mathrm{d}(e_{\mathrm{t}}, \mathrm{I}_i)^2$$

Set $\mathcal{I} = \{I_i \in \text{interval}; i = 1,2,\ldots,k\}$, is then known as the set of cluster interval for interval time series Ψ.

Example 2.2: The distribution of 27 sets of interval data of unemployment rate (unit: percent) is shown as Figure 2.1, as follows:

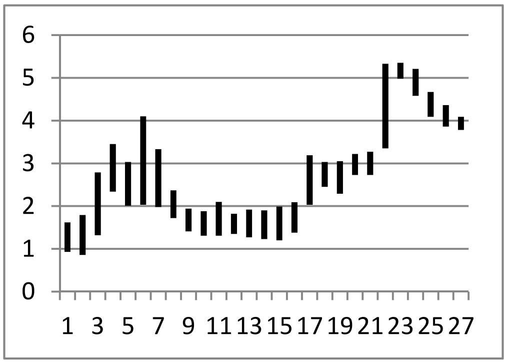

Figure 2.1 Trend of interval time series

If we wish to divide the data into two groups, using Definition 2.5, we can obtain two interval clusters $I_1 = (1.83, 2.46)$ and $I_2 = (3.71, 5.23)$The result of clustering is as follows:

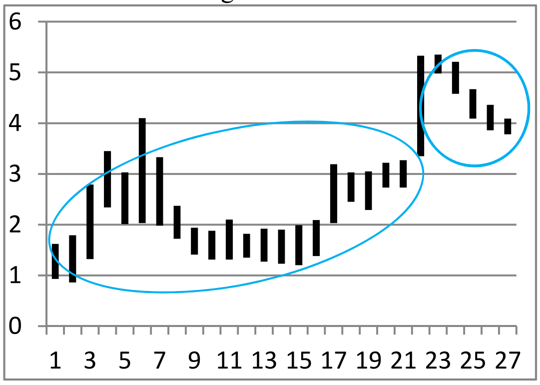

Figure 2.2 Result of interval clustering

2.3 Change periods

Conventionally, the detection of the structural change in a system takes the change point as the main factor of consideration. However, structural change should be mainly based on variable and not the time value, and it should be a gradually-emerging change interval and not a change point whereby the change happens abruptly at a certain point of time. So the change interval that studies variables, compared with the classical method of investigation for time series, has better descriptive power. Wu (1999) suggested that, the use of fuzzy entropy is effective in identifying whether a structural change happens in a time series. Besides, it can also be used together with the mean cumulated fuzzy entropy of *t* times, to observe the change in message of fuzzy entropy, based upon which a standard for the classification of change model can be established.

Definition 2.5 *Mean Cumulated Fuzzy Entropy*

Let a time series be $\{x_t\}$, *t* =1,2,..,*N* , *with*$\delta(x_t)$ *being its fuzzy entropy. The mean cumulated fuzzy entropy is thus defined as:*

$$MS\delta(x_t) = \frac{1}{t}\sum_{i=1}^{t} \delta(x_i)$$

There is usually a threshold level λ set up for fuzzy classification, because no matter it is for nature or humanities, the determination for classification is very subjective and often non-unanimous. Hence an objective measure is needed. According to empirical experience, λ cannot take a value too huge or too small, otherwise the classification cannot be done or too many classes will be created. So a value for λ between 0.001 and 0.1 will be ideal.

Here we wish to find two cluster centers. This is determined based on common experience of empirical analysis and trend of time series. The procedures are as follows:

Step 1: Use the k-means method to find out two cluster centers C_1 and C_2 in time series $\{x_t\}$*, and determine the membership degree* μ_{it}*,* i*=1,2 of* $\{x_t\}$ *to the two cluster centers*

Step 2: Compute the fuzzy entropy $\delta(x_t)$*, mean cumulated fuzzy entropy* $MS\delta(x_t) = \frac{1}{t}\sum_{i=1}^{t}\delta(x_i)$*, and Median(*$MS\delta(x_t)$*) of this series, that correspond to* x_t.

Step 3: Take a suitable threshold value λ, classify the mean cumulated fuzzy entropy $MS\delta(x_t)$ *series that correspond with* x_t*. If the mean cumulated fuzzy entropy* $MS\delta(x_t)$ *falls into the interval [0,Median(*$MS\delta(x_t)$*)−λ), we will use 1 to represent Group 1; if* $MS\delta(x_t)$ *falls into the interval [Median(*$MS\delta(x_t)$*)−λ,Median(*$MS\delta(x_t)$*)+λ), we use 2 to represent Group 2; and if* $MS\delta(x_t)$ *falls into the interval [Median(*$MS\delta(x_t)$*)+λ,1], 3 will be used to represent Group 3.*

Step 4: If the result of classification is inconsistent, we then make adjustment to the result. If it is consistent, go to Step 5.

Step 5: Select an appropriate determination level α*. If the number of consecutive samples is greater than* $[\alpha N]$*, then these consecutive samples belong to the same group. During classification, if more than one group is found, we know that structural change happens in this time series. Thereafter, find the change interval.*

3. Empirical Studies

We use the populations of singles and unemployment rate to perform the fuzzy statistical analysis. The main source of information is the statistical data from the Department of Budget, Accounting and Statistics (DGBAS) of the Ministry of the Interior. The singlehood rate is calculated by dividing the population of single males in years 1980 to 2006, of ages 25- 34 years.

Firstly, by determining CCF of the rate of single men of ages 25 – 24 year old in years 1980 to 2006, and the unemployment rate of the same years. For such a comparison, the relationship we can obtain is as follows:

$$(1-B)Y_t = 0.00634 + (-0.148 - 0.083B + 0.492B^2)X_t \text{，} 1980 \le t \le 2006$$

in which, Y_t is the marriage rate at a point of time t, X_t is the employment rate at the point of time t, and the model of input series being:

$$(1-0.325B)(1-B)X_t = \varepsilon_t \text{，} 1980 \le t \le 2006$$

Take the threshold value λ=0.01, and set the change interval of the unemployment rate as years 1985 to 1988, we then consider this a transition period, and construct a new threshold model. But because the classification oscillation before year 1985 is too frequent, it is difficult to place the unemployment rate into any class. We will therefore deem it non-stationary, and only consider the unemployment rate after year 1988 to build the ARIMA model, as follows:

$$\begin{cases} \text{nonstationary}，1980 \le t \le 1987 \\ (1-0.0036B)(1-B)X_t = 0.0054 + \varepsilon_t，1988 \le t \le 2006 \end{cases}$$

and then construct the conversion model of the unemployment rate after the change interval to the singlehood rate, as follows:

$$(1-0.94B)Y_t = 0.0285 + (0.217 - 0.419B + 0.78B^2)X_t，1988 \le t \le 2006$$

When we take λ=0.01, change intervals are found in years 1983 to 1985 as well as years 1992 to 1994. This means structural change happens. So we need to consider the singlehood rate after year 1985. However, the sample size is less than 6, so a model cannot be built. Nonetheless, observing the change intervals, we know that the structural change in the unemployment rate happened earlier than that of the singlehood rate, and in the change period of the unemployment rate, structural change had also occurred on the singlehood rate. Therefore the unemployment rate has significant impact on the singlehood rate.

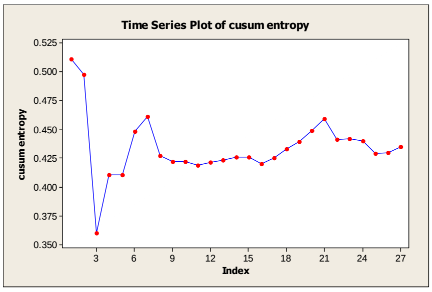

Figure 3.1 Chart of mean cumulated fuzzy

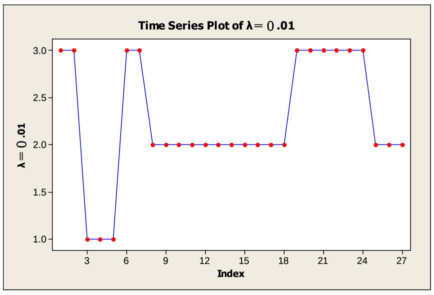

Figure 3.2 The classification of the unemployment rate

We will now perform the classification of fuzzy intervals using the fuzzy classification method, on the unemployment rate intervals. We will then obtain the cluster centers of the fuzzy intervals of the unemployment rate: $I_1 = (0.0183，0.0245)$, $I_2 = (0.0371，0.0523)$.

Also, with *Median*(MS$\delta(x_t)$)=0.2460, we know that the transition interval for the unemployment rate interval falls between years 1987 and 1988. We consider this a transition period, and construct a new threshold model. But because the classification oscillation before year 1987 is too frequent, it is difficult for us to place the unemployment rate in this period into any class. We will therefore deem it non-stationary, and take only the unemployment rate interval after year 1988 into account to build an ARIMA model, as follows:

High model of intervals:

$$\begin{cases} \text{nonstationary}，1980 \le t \le 1987 \\ (1-0.0005B)(1-B)X_t = 0.002 + \varepsilon_t，1988 \le t \le 2006 \end{cases}$$

Low model of intervals:

$$\begin{cases} \text{nonstationary}，1980 \le t \le 1987 \\ (1-0.147B)(1-B)X_t = 0.001 + \varepsilon_t，1988 \le t \le 2006 \end{cases}$$

The highest unemployment rate-to-singlehood rate conversion model:

$$(1-1.02B)Y_t = 0.0005 - (0.016 - 0.152B^2)X_t，1988 \le t \le 2006$$

The lowest unemployment rate-to-singlehood rate conversion model:

$$(1-0.945B)Y_t = 0.0292 + (-0.145 + 0.134B + 0.571B^2)X_t，1988 \le t \le 2006$$

Comparison with forecast results

Table 3.1 is the comparison results for the forecast of unemployment rate between the best ARIMA model and fuzzy classification ARIMA. We can see that the forecast result of the unemployment rate using the best ARIMA model performs bettern the best RIAMA model.

Table 3.1 Comparison between the forecasts of the unemployment rate

Retention phase	*Actual value*	*ARIMA(1,1,0)*	*Fuzzy classification ARIMA(1,1,0)*
2007	*0.0391*	*0.0390*	*0.0389*
2008	*0.0414*	*0.0397*	*0.0393*
2009	*0.0585*	*0.0406*	*0.04*
MSE		*0.000108*	*0.000116*

Table 3.2 shows the forecast of singlehood rate obtained by applying the threshold transformation on the intervals of the unemployment rate forecast. We know from Tables 3.5 and 3.6 that, no matter it is performing the fuzzy classification on the unemployment rate first, and then the transform the unemployment rate to the singlehood rate; or it is performing the fuzzy classification on the intervals of unemployment rate before transforming the result of classification to the singlehood rate to get the forecast or intervals of forecast of the singlehood rate, the differences in MSE is not great. This is because the sample size is not large. However, this provided an approach to effectively obtain the forecast intervals of the singlehood rate, and has demonstrated the uncertainty in singlehood rate.

Table 3.2 Forecast intervals of the transformation of the singlehood rate

Retention phase	Actual value	Forecast intervals of the transformation rate
2007	0.6549	(0.653,0.656)
2008	0.6619	(0.667,0.675)
2009	0.6819	(0.681,0.695)
IMSE		0.000129

Overall, the forecasting we made, whether on the unemployment rate or the singlehood rate, is reasonably accurate. The forecasting will improve if classification transformation is used – we only need to construct a model of the singlehood rate using the end result of classification of the unemployment rate. This way, the required number of data periods is lower (we only need the unemployment rate data after year 1998). In addition, the forecast interval of singlehood rate obtained from forecast of interval will provide a better forecasting ability for the uncertainty in forecasting singlehood rate.

At the end we also find that, for a time series, if the change period of its structural change can be determined, better results on the model construction and forecasting ability can be produced.

4.Conclusion

In the progress of scientific research and analysis, the uncertainty in the statistical numerical data is the crux of the problem that the traditional mathematical models are hard to be established. This paper proposes to use the interval data to avoid such risks happening. In fact, using interval data to establish a model and to predict, we can find that the forecasting in each step is carried out by means of intervals, so as to increase the objectiveness of the forecast results. In the general aspect, the ‘‘intervalization’’ seems to be a very normal phenomenon too.

This paper discusses the quality of the forecast result through evaluating forecasting performance, It is noteworthy that if we can establish a good efficiency process, we can make a superior interval forecasting for the interval time series. We also hope that with the experience in analyzing the structural change in time series, we can produce effective procedures to detect change interval, and this can be used to do forecasting, in order to fulfill human beings' thirst to grasp the future changes.

In recent years, the unemployment rate in Taiwan is on the rise. The challenges that the Taiwan society faces are not only limited to the disorderly situation caused by the political reconstruction after the first shift of political power in the millennium year, as well as the “Democratic Labor’s Pain”. As the international economic environment and the cross-Strait economic and trade environment change, the investment and employment markets in Taiwan becomes increasingly challenging, and this created an impact to the supply-and-demand balance in the local employment market.

5.Reference

[1]. Wu,B (1999).Use of fuzzy statistical technique in change period detection of nonlinear time series. *Applied Mathematics and Computation*,99,241-254.

[2]. A.F.Gómez-Skarmeta,M.Delgado and M.A.Vila,(1999).About the use of fuzzy clustering techniques for fuzzy model identification ,*Fuzzy Sets and Systems*,106,2,179-188 .

[3]. Bai, J. and Perron, P. (1998), Estimating and testing linear models with multiple structural changes, *Econometrica*, 66, 47–78.

[4]. Kumar, K. and Wu, B. (2001), Detection of change points in time series analysis with fuzzy statistics, *International Journal of Systems Science*, 32(9), 1185-1192.

[5]. Chow,G.C. (1960). Testing for equality between sets of coefficients in two linear regression. Econometrics,28,291-260

[6]. San, O.M., Huynh, V., and Nakamori, Y. (2004), An alternative extension of the K-means algorithm for clustering categorical data, Int. J. Appl. Math. *Comput. Sci*,14:2, 241-247.

[7]. Weina Wang, Yunjie Zhanga(2007). On fuzzy cluster validity indices, *Fuzzy Sets and Systems* ,58,2095–2117.

[8]. Dae-Won Kim, Kwang H. Lee, Doheon Lee(2004).On cluster validity index for estimation of the optimal number of fuzzy clusters.*Pattern Recognition*, 37: 10, 2009-2025

[9]. Kuo-Lung Wu, Miin-Shen Yang(2005). A cluster validity index for fuzzy clustering.*Pattern Recognition Letters,*26:9,1275-1291

[10]. Gin-Shuh Liang, Tsung-Yu Chou, Tzeu-Chen Han(2005).Cluster analysis based on fuzzy equivalence relation, *European Journal of Operational Research*, 166: 1,160-171

[11]. Xie, J., Wu, B. and Scriboonchita, S. (2010). Fuzzy Estimation Methods and their application in the Real estate Evaluation. *International Journal of Intelligent Technologies and Applied Statistics.* **3(2)** ,187-202

國科會補助計畫衍生研發成果推廣資料表

日期:2013/09/21

<table>
<tr><td rowspan="3">國科會補助計畫</td><td>計畫名稱: 財金管理之風險與波動率估計及預測</td></tr>
<tr><td>計畫主持人: 吳柏林</td></tr>
<tr><td>計畫編號: 99-2410-H-004-101-MY3　　　學門領域: 作業研究／數量方法</td></tr>
<tr><td colspan="2">無研發成果推廣資料</td></tr>
</table>

99 年度專題研究計畫研究成果彙整表

<table>
<tr><td colspan="3">計畫主持人：吳柏林</td><td colspan="5">計畫編號：99-2410-H-004-101-MY3</td></tr>
<tr><td colspan="8">計畫名稱：財金管理之風險與波動率估計及預測</td></tr>
<tr><td colspan="3" rowspan="2">成果項目</td><td colspan="3">量化</td><td rowspan="2">單位</td><td rowspan="2">備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）</td></tr>
<tr><td>實際已達成數（被接受或已發表）</td><td>預期總達成數(含實際已達成數)</td><td>本計畫實際貢獻百分比</td></tr>
<tr><td rowspan="13">國內</td><td rowspan="4">論文著作</td><td>期刊論文</td><td>2</td><td>2</td><td>100%</td><td rowspan="3">篇</td><td></td></tr>
<tr><td>研究報告/技術報告</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>研討會論文</td><td>6</td><td>6</td><td>100%</td><td></td></tr>
<tr><td>專書</td><td>3</td><td>3</td><td>100%</td><td></td><td></td></tr>
<tr><td rowspan="2">專利</td><td>申請中件數</td><td>0</td><td>0</td><td>100%</td><td rowspan="2">件</td><td></td></tr>
<tr><td>已獲得件數</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td rowspan="2">技術移轉</td><td>件數</td><td>0</td><td>0</td><td>100%</td><td>件</td><td></td></tr>
<tr><td>權利金</td><td>0</td><td>0</td><td>100%</td><td>千元</td><td></td></tr>
<tr><td rowspan="4">參與計畫人力
（本國籍）</td><td>碩士生</td><td>0</td><td>0</td><td>100%</td><td rowspan="4">人次</td><td></td></tr>
<tr><td>博士生</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>博士後研究員</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>專任助理</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr style="display:none"><td></td></tr>
<tr><td rowspan="12">國外</td><td rowspan="4">論文著作</td><td>期刊論文</td><td>6</td><td>6</td><td>100%</td><td rowspan="3">篇</td><td></td></tr>
<tr><td>研究報告/技術報告</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>研討會論文</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>專書</td><td>4</td><td>4</td><td>100%</td><td>章/本</td><td></td></tr>
<tr><td rowspan="2">專利</td><td>申請中件數</td><td>0</td><td>0</td><td>100%</td><td rowspan="2">件</td><td></td></tr>
<tr><td>已獲得件數</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td rowspan="2">技術移轉</td><td>件數</td><td>0</td><td>0</td><td>100%</td><td>件</td><td></td></tr>
<tr><td>權利金</td><td>0</td><td>0</td><td>100%</td><td>千元</td><td></td></tr>
<tr><td rowspan="4">參與計畫人力
（外國籍）</td><td>碩士生</td><td>0</td><td>0</td><td>100%</td><td rowspan="4">人次</td><td></td></tr>
<tr><td>博士生</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>博士後研究員</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>專任助理</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
</table>

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	近五年內主要研究著重於結合機率統計與資訊科學如知識經濟，人工智慧，等高科技領域與創新觀念。申請者帶領他的博碩士生研究團隊一直進行開創性的研究計劃。在時間數列分析與預測、模糊統計和智慧計算兩個領域有相當貢獻。傳統時間數列分析，很少離開穩定過程與線性模式的假設。但是在實證分析領域裡，時間數列的走勢往往呈非線形且不穩定的型態。因此在建構時間數列模式之前， 應先考慮其數列走勢的特性，如圖形識別與認定的問題。古典 ARMA、狀態空間， 時空模式等已無法滿足嚴格的科技要求。因此申請者考慮以非線性時間數列和目前計算機科學中發展迅速之神經網路理論為基礎，並結合隨機過程中的收斂理論，應用在實證資料的特徵分析與相關測度上。前幾年申請者提出的模式庫(models base)建構流程與認定法則，即由模式庫中找出一合適模式族，繼而建立一最穩健模式的程序。對於非線性的時間數列分析方面，申請者所提出遺傳模式觀念與建構演算則，在數學與統計領域上更是一大創新與突破。 2011 Life Fellow of International Society of Management Engineers, Japan

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

■達成目標

□未達成目標（請說明，以 100 字為限）

□實驗失敗

□因故實驗中斷

□其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文：■已發表 □未發表之文稿 □撰寫中 □無

專利：□已獲得 □申請中 ■無

技轉：□已技轉 □洽談中 ■無

其他：（以 100 字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

This research project aims at investigating practical statistical procedures, based upon various types of economic/financial time series data, for risk estimation and volatility forecasting. Volatility a measure of the risk in a financial instrument, refers to the standard deviation of the continuously compounded returns with a specific time horizon. The main focus will be on developing the current theory of coherent risk measures, based on fuzzy ratio entropy integral formulation, with applications to assessing quantitative risks related to volatility of stock prices for option pricing, efficient portfolio allocation and accurate risk assessment and management.