

藉由馬可夫轉換模型解決風險值計測過程高峰、厚尾
與偏態問題—國內主要股市指數實證結果
**Mitigating Tail-fatness, Lepto Kurtic and Skewness
Problems in VaR Estimation via Markov Switching
Settings – An Empirical Study on Major TAIEX Index
Returns**

林修葺 Hsiou-Wei Lin¹
National Taiwan University
國立台灣大學

饒秀華 Hsiu-Hua Rau

黎明淵 Ming-Yuan Li
National Cheng-Chi University
國立政治大學

摘 要

本論文以馬可夫模型估算風險值，分析討論與實證應用或屬原創。風險值估算是財金學術與實務領域很重視課題，我們應用雙重狀態期望值與波動設定，採馬可夫過程控制狀態間切換，也藉國內主要股價指數報酬樣本，以貝氏混合分配與GARCH測試結果與馬可夫互對照。在各類模型降低報酬衝擊分配非常態問題效果比較，馬可夫表現最佳，貝氏混合分配與GARCH則分呈過度與不足。就實務應用角度言，馬可夫優勢對5%臨界機率風險值不顯著，在1%與2.5%臨界機率情境則頗明顯。在學習窗期的選擇，為掌握殺傷力大但不常發生的特殊事件，描述報酬損失較利得需更長的歷史資料。

Abstract

This paper serves as one of the first studies that estimate the value at risk (VaR) via a Markov-switching (MS) model. Specifically, we use a two-regime MS specification, a MS setting with two sets of regime mean and regime variance, on TAIEX as well as Taiwan's major industrial group stock index returns. We demonstrate that MS effectively correct non-normality problems and outshine both GARCH and the mixing normal models, with the former (latter) alternative being subject to over- (under-)estimating the persistence of stock return volatility (hereafter volatility). As for estimating the 5% VaR, MS appears to be equally effective as Bayesian mixing normal and GARCH. In contrast, MS significantly outperforms the two non-linear alternatives for estimating VaR with 1% or 2.5% tail probabilities. Furthermore, as for the window of learning period on rare events, we find that one need to go much farther back to effectively depict the left as opposed to the right tail.

關鍵詞：馬可夫轉換模型、風險值

Keywords: Markov-switching models, value at risk

¹ 本文撰寫過程中，承Darrel Duffie、James D. Hamilton、暨本刊兩位匿名評審提供多項富建設性的建議，對於Rual Susmel與Chang-Jin Kim對本文電腦程式的撰寫提供了寶貴的意見，暨The *Eighth Conference on the Theories and Practices of Security and Financial Markets*中諸先進的意見，亦一併感謝。惟本文若有任何謬誤，當屬三位作者之責。

Mitigating Tail-fatness, Lepto Kurtic and Skewness Problems in VaR Estimation via Markov Switching Settings – An Empirical Study on Major TAIEX Index Returns

1. Introduction

This paper serves as one of the first studies that adopt Markov-switching (MS) models to control regime switches and gain efficiency in estimating the value at risk (VaR). Specifically, we use a two-regime MS setting for both mean and variance parameters of Taiwan Stock Exchange market index (hereafter TAIEX) returns as well as major industrial group stock index returns including TAIEX-Construction (hereafter TAIEX-CONS), TAIEX-Finance (hereafter TAIEX-FIN), TAIEX-Electronic (hereafter TAIEX-ELEC), and TAIEX-Electric&Machinery (hereafter TAIEX-EL&MACH). Our empirical findings support the notion that the MS models significantly help mitigate typical problems such as fat tails, Lepto kurtic and skewness in VaR estimation.

Most practitioners turned their attention to VaR no earlier than the mid 1990s. The Basle Committee on Banking Supervision of Bank for International Settlements (BIS) presented Amendment to Capital accord to Incorporate Market Risks in January 1996. Most financial institutions now diversify their operations out of their original businesses and actively trade capital market securities, foreign currencies and derivative instruments. These institutions' risks keep increasing because of increased market return volatility, new competition, and deregulation. In April 1996, The Basle Committee further endorsed the use of the value at risk to measure the banks' capital adequacy ratios. By the end of 1997, the G10 bank regulators took market risk into account for determining the risk-based assts.² Since January 1999, financial institutions in Taiwan have been instructed to use VaR to measure their capital adequacy ratios. Financial institutions have two alternatives. The first alternative, or standardized risk measurement proposal, appears to be ad hoc and subject to strong assumption of homogeneity among the banks. The second alternative, also named the value at

² The G10 group includes Belgium, Canada, France, Germany, Italy, Japan, Netherlands, Sweden, Switzerland, United Kingdom, United States and Luxembourg.

mogeneity among the banks. The second alternative, also named the value at risk approach, in contrast, allows the users to gain forecast accuracy for estimating market risks via tailored VaR methods³.

VaR is also at the center of the recent interest in the risk management field.⁴ It serves to measure the level of market risks and accordingly, the capital adequacy. Duffie and Pan (1997) further suggested that one could measure default risk, credit risk, operation risks and liquidity risk via the algorithm of VaR.

There has been an extensive literature on VaR estimation. Estrella et al. (1994), Kupiec (1995), Jackson, Maude and Peerraudin (1997) explored banks' capital adequacy ratios and VaR. Hendricks (1994), Bender (1995), Simons (1996), Fong and Vasicek (1997) investigated the strengths and weaknesses of VaR estimation measures including parametric, historical data simulation and Monte Carlo simulation methods. Empirical studies on estimating VaR for portfolios of TAIEX securities include Wang and Chen (1998), Liao and Lin (1998), Uong and Chen (1998), Ho and Lin (1999).

Many investment risk studies, nevertheless, documented that the histograms for historical returns most financial assets deviate from the normal distributions. Typical non-normality properties include Lepto kurtic, tail-fatness and skewness. Furthermore, there exist problems of twin peak, the result often found in the New Taiwan dollar price per U.S. dollar histograms.

The purpose of this study is to examine the extent to which competing non-linear settings including Markov Switching, GARCH and the mixing normal models help mitigating non-normality problems in VaR estimation. We consider our research question a non-trivial issue because (1) VaR is at the center of the recent interest for both researchers and practitioners, and (2) as demonstrated by studies in prior studies as well as the empirical section of this study, different VaR models generate significantly different VaR estimates. We recommend the use of Markov Switching models since we conjecture that structural changes of financial and economic variables serve as a driver to the non-normality problems. Statistically, an increase in sample size can eliminate measurement errors but inevitably gain exposure to structural changes. Transitional variability in financial variables may

³ This statement is especially descriptive for banks with large trading accounts.

⁴ One of the most frequently adopted VaR models is the RiskMetrics of J. P. Morgan Financial Service Co. Incorporated. Its web site is: <http://www.jpmorgan.com>. RiskMetrics, which is publicly available on the Internet, is a representative parametric VaR model.

dwarf most typical scenario analyses, stress tests, and risk limit guidelines. Unfortunately, the traditional linear models do not allow their parameter sets to adjust for structural changes.

This study serves as one of the first papers to demonstrate the effectiveness of MS models in filtering structural changes in return volatility and thus correct the downward bias in estimated losses encountered by the popular linear models due to tail-fatness.⁵

This paper also contributes to the contemporary literature in the following aspects. First, we document the relative strengths of Markov-switching, the mixing normal and GARCH models in mitigating non-normality problems. Namely, all these three alternatives help reduce Lepto kurtic and fat-tail problems for return distributions. Nevertheless, the mixing normal (GARCH) models appear to over- (under-) react to Lepto kurtic and tail fatness. The explanation is that the general mixing normal settings disregard information for regime persistence, whereas GARCH models over-estimate the persistence of return volatility when there exist structural changes during the estimation periods. Accordingly, the prevalent normal setting for the error term can benefit more from incorporating the MS model to control the structural changes.

Second, we show that the two- variance-regime specification helps incorporate Lepto kurtic and tail-fatness, and that the two-mean-regime setting helps filtering skewness in return distributions. Namely, via endogenizing the state variable for each period, one may more effectively assign weights for the two normal distributions with different variance and mean parameters. Our finding that the MS models help solve the non-normality problems and achieves higher fitness than competing models supports the notion that structural changes are a major variable to non-normality.

Third, we examine the statistical properties of Taiwan' major indices including TAIEX, TAIEX-ELEC, TAIEX-FIN, TAIEX-CONS, TAIEX-EL&MACH. For the return distributions of these indices, we document serious (non-trivial) fat tails in the 1% (2.5%) regions for both left- and right-tails but insignificant tail-fatness in the 5% regions. This finding supports the notion of significant lumps towards both extremes in the return distributions and is consistent with the result of Ho and Lin (1999).

Fourth, this study compares the relative accuracy in estimating VaR for lin-

⁵ See Simons (1996) and Ho and Lin (1999).

ear and MS models in terms of violation rates. Our results on TAIEX indices show that, to conservative investors, who are typically the audience with 1% left-tailed critical probability, MS models are significantly more accurate. As to the 2.5% left-tailed region, Markov-switching model still performs better. But the differences in accuracy are much less than that corresponding the 1% region. Furthermore, when the left-tailed probability is set to be greater than 5%, the differences in accuracy become trivial. As to the learning window, the setting that uses the sample of 500 (250) pre-VaR-day observations appears to yield a more (less) accurate estimator. A potential explanation is that the former setting incorporates a richer information set.

Fifth, we avoid the strong assumption of symmetric return distributions in Venkataraman (1997) and alternately apply the VaR algorithm on each of the two tails with horse races for the competing models and learning windows. Our results for the sample index returns show that the forecast accuracy with MS exceeds that of the others in both 1% and 2.5% right-tailed regions in terms of violation rate. However, with 5% right-tailed probabilities, the difference in accuracy becomes insignificant.

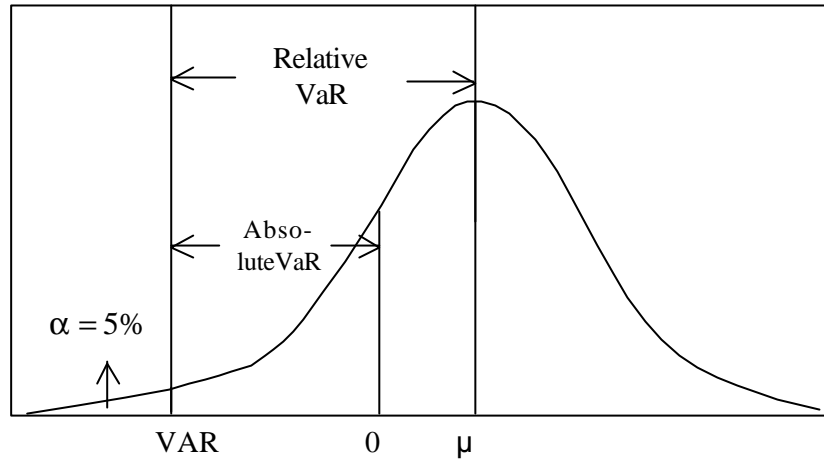
The second section presents the definition for VaR and evaluates the competing models. Section 3 presents our empirical results. Finally, Section 4 concludes the paper.

2. The Definition of the VaR and the Settings for Return Distributions

1. The Definition for the Value at Risk

Set the critical probability to be α , then VaR, the absolute value at risk, is the expected maximum (worst) loss associated with a portfolios over a target time horizon, say, one trading day. Namely, VaR is the lower bound for equation $\int_{VaR}^{+\infty} f(R_p) dR_p = 1 - \alpha$, where R_p denotes the daily returns of a portfolio with a probability distribution function $f(R_p)$. Taking the firms' opportunity costs into account, one can further derive the relative value at risk, which is defined as the difference between the absolute VaR and the expected returns. With a given information set, one can complete the probability distribution for R_p as shown in the following figure and thus identify VaR :

VaR for a Confidence Interval of 95%



2. Specifications for the return distributions

The key to VaR estimation is to estimate the return distribution based on our existing information set. Let R_t denote the returns for some specific portfolio for period t . There exist the following five competing specifications.

(a) Linear Model with Standard Gaussian Distribution (hereafter the Linear Model)

$$R_t = u_0 + \sigma_0 \cdot e_t$$

(1)

In this setting, u_0 and σ_0 represent mean and standard deviation of R_t , respectively. Also, e_t , which denotes the return shock, follows a standard Gaussian distribution.

(b) Linear Model with Student t Distribution (hereafter Student t Setting)

$$R_t = u_0 + \sigma_0 \cdot e_t$$

(2)

In this setting, return shock e_t follows a t distribution with standard error equal to unity.

Some prior studies used Student t settings to approximate the security return distributions⁶, which are often subject to Lepto kurtic and tail-fatness. Our empirical results, nevertheless, indicate that the Student t setting, which is not descriptive for (1) distributions that are Lepto kurtic but not fat-tailed and (2) distributions that are fat tailed but not Lepto kurtic, does not fit most of Taiwan's industrial index returns. Note that all TAIEX industrial returns have significant skewness. TAIEX-ELEC and TAIEX-EL&MACH (TAIEX-FIN and TAIEX-CONS) returns have negative (positive) skewness⁷, consistent with the notion that large negative (positive) returns are more frequent than large positive (negative) returns. Nevertheless, most return series including TAIEX-ELEC have significant tail-fatness but insignificant kurtosis. Accordingly, this study excludes the measure of Student t settings.

(c) GARCH(p, q) Models with Standard Gaussian Distribution

Bollerslev (1986) first introduced the GARCH models. The setting for GARCH(p, q) is :

$$\begin{aligned} R_t &= u + \mathbf{s}_t e_t, \quad e_t \text{ iid} \sim ND(0,1) \\ \mathbf{e}_t &= \mathbf{s}_t e_t \\ \mathbf{s}_t &= a_0 + \sum_{i=1}^q a_i \mathbf{e}_{t-i}^2 + \sum_{i=1}^p b_i \mathbf{s}_{t-i}^2 \end{aligned} \quad (3)$$

The popular ARCH(q) model proposed by Engle (1982) can be viewed as a special case of a GARCH (p, q) with $p = 0$. In the GARCH, the conditional variance may be affected by prior-period conditional variance and error sum of squares. I , the volatility persistence measure, can be expressed as:

$$I = (a_1 + a_2 + \dots + a_q + b_1 + b_2 + \dots + b_p)$$

⁶ Please refer to Jorion (1997).

⁷ The kurtosis coefficient for t distributions is $3(-2)/(-4)$, > 4 . The skewness coefficient for t distributions is 0. The kurtosis of a t distribution decreases with the degree of freedom. As the degree of freedom approaches infinity, the t distribution approaches a standard normal distribution. With finite degree of freedom, a t distribution is *both* Lepto kurtic and fat-tailed. Namely, the t-distribution is not descriptive for distributions that are Lepto kurtic but not fat-tailed and distributions that are fat tailed but not Lepto kurtic.

(4)

(d) Mixture of Normal Distributions (Mixing-normal)

$$R_t = u_{s_t} + \mathbf{s}_{s_t} \cdot e_t \tag{5}$$

A two-regime setting incorporates an unobservable regime variable $s_t = 1, 2$, with corresponding mean returns of u_1 and u_2 , and corresponding standard deviations of σ_1 and σ_2 . Specifically, R_t is generated via two normal distributions with differential means and variances. Also, return shock e_t follows a standard Gaussian distribution.

In a mixing-normal setting, s_t follows a stochastic process. These models disregard the persistence of return volatility and assume that the unobservable regime variables s_t conforms the following Bernoulli distribution:

$$P(s_t = 1) = q, \quad P(s_t = 2) = 1 - q \tag{6}$$

where $0 < q < 1$.

Hamilton (1991) suggested that we introduce the prior constraint in Bayesian models and rewrite the ML function as the follows:

$$\begin{aligned} \log p(r_1, r_2, \dots, r_T | \mathbf{q}) &= \sum_{k=1}^{n_s} (a_k / 2) \log \mathbf{s}_k^2 - \sum_{k=1}^{n_s} b_k / (2\mathbf{s}_k^2) \\ &\quad - \sum_{k=1}^{n_s} (m_k - u_k)^2 / (2\mathbf{s}_k^2) \end{aligned} \tag{7}$$

where $\{a_k\}, \{b_k\}, \{c_k\}$ and $\{m_k\}$ are sets of constants summarizing an analyst's prior belief for u_k and σ_k for $k = 1, \dots, n_s$, with nonnegative a_k, b_k , and c_k for all k . The proposed estimator may degenerate to the MLE as a special case for the diffuse prior $a_k = b_k = c_k = 0$. In this study, we adopt the setting in Hamilton (1991) and set $a_i = b_i = 0.20, c_i = 0.10$ and $m_i = 0$, for $i = 1$ or 2 ⁸.

⁸ Venkataraman (1997) adopted the same setting.

(e) Markov-switching (MS) Models with Standard Gaussian Distribution

$$R_t = u_{s_t} + \mathbf{S}_{s_t} \cdot e_t \quad (8)$$

A two-regime MS setting incorporates unobservable regime variable $s_t = 1, 2$. Specifically, R_t is generated via two normal distributions with differential means u_1 and u_2 and differential variances σ_1 and σ_2 . Also, return shock e_t follows a standard Gaussian distribution. MS models, in contrast with mixing-normal, incorporate Markov process for s_t and are effective in controlling regime switches and preserving the statistics for persistence in regimes. The two-regime Markov transition probabilities can be expressed as the follows:

$$\begin{aligned} p(s_t = 1 | s_{t-1} = 1) &= p_{11}, & p(s_t = 2 | s_{t-1} = 1) &= p_{12} \\ p(s_t = 2 | s_{t-1} = 2) &= p_{22}, & p(s_t = 1 | s_{t-1} = 2) &= p_{21} \end{aligned} \quad (9)$$

where $p_{11} + p_{12} = p_{21} + p_{22} = 1$. The stochastic process for s_t is strictly stationary if both p_{11} and p_{22} are less than unity and do not take on the value of 0 simultaneously. Because when $p_{11} = 1$, once the process enters state 1, it would never return to state 2.⁹

State variable s_t conforms to the following AR (1) setting:

$$s_t = (1 - p_{11}) + s_{t-1} + v_t, \quad v_t = -1 + p_{11} + p_{22} \quad (10)$$

The above equation is a special case of a typical AR(1) model with an unusual probability distribution of the innovation sequence $\{v_t\}$. As shown by Hamilton (1989), v_t can be interpreted as a measure of persistence in the forcing process¹⁰.

More general, an N-state Markov chain is said to be reducible if there exists a way to label the states¹¹ such that the transition matrix can be written in the form:

⁹ In such a case state 1 is an absorbing state and the Markov chain is reducible.

¹⁰ Please refer to Cai (1994).

¹¹ That is, a way to designate which state should be named state 1, which one should be named state 2, and so on.

$$P = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix} \quad (11)$$

where B denotes a $(K \times K)$ matrix, with $1 \leq K \leq N$. If P is upper block-triangular, then so is P^m for any m . Hence, once such a process enters a state j such that $j > K$, there exists no probability of its ever returning to any one of the states $K+1, K+2, \dots, N$.¹²

Note that, contrast with Li and Lin (1999), who adopted a three-regime SWARCH (Markov-switching ARCH) model,¹³ this study selects a more simplified setting for return volatility. The significance of the SWARCH model is that it incorporates MS and ARCH, with the Markov-switching specifications filtering out most return volatility and the traditional ARCH models and the t distribution for the error term controlling the residual return volatility. In contrast, this paper focuses on daily returns, which is less volatile in terms of magnitude as opposed to the weekly returns in Li and Lin (1999). Moreover, with limited number of prior observations for the learning process in the models, two- instead of three-regime settings may be sufficient. With Markov-switching models that help filtering return volatility, the contribution of the ARCH algorithm and the Student t setting would be less significant.

3. Why Do We Propose the Use of MS Models for VaR Estimation?

GARCH¹⁴, mixing-normal and MS¹⁵ models are stochastic volatility settings for non-constant volatility. A simple example would help illustrate how the stochastic volatility models help solving the non-normality problems. Assume that the return volatility for series X is non-constant and instead taking the value of either σ_1 or σ_2 . We may then sketch two normal distributions instead for X . As one incorporates the two distributions, he may resemble a return distribution with kurtosis and tail-fatness. Moreover, if the mean return can be either μ_1 or μ_2 , too, then one may thus mitigate skewness and twin peak.

¹² Please refer to Hamilton (1994).

¹³ SWARCH models incorporate the MS, ARCH and student t setting.

¹⁴ Duffie and Pan (1997) summarized the alternatives as *two* types of *stochastic volatility models*. However, our tests provide evidence as to the significance of the distinction among the three specifications.

¹⁵ Duffie and Pan (1997) discussed the potential to apply MS models to VaR estimation without any analysis or empirical tests to discriminate amongst the competing models. But Li and Lin (1999) documented that, for TAIEX returns, MS effectively filters out structural changes during the estimation period and thus mitigates the tail-fatness problems. Li and Lin (1999) also suggested that MS models be applied to VaR estimation.

Now, let us evaluate the three competing models. First, mixing normal and MS are both mixing two normal distributions to approximate return distributions with kurtosis and tail-fatness. The mixing normal setting is a recipe for drawing fat-tailed returns by mixing two normal distributions¹⁶. In contrast, thanks to the Markov-chain property, MS models use the transition probabilities as the mechanism to incorporate the persistence of return volatility. The notion of volatility persistence is consistent with a large probability for a stage with low (high) volatility in returns to be followed by a low (high) volatility stage. Take the sunspots and earthquakes activities as our examples. They both have persistence in energy absorption and dissipation processes. Usually it is scientifically foreseeable that an earthquake would trigger after-shocks in the subsequent periods. Specifically, crush motions in the neighboring periods are serially correlated.

MS and mixing-normal both use a discrete state variable to indicate which distribution the returns are drawn from and are both nonlinear models. The difference between the two measures is that the mixing normal models use Bernoulli distributions to control the regime switching process, whereas MS models use Markov-chain to control it. In contrast with MS and GARCH, accordingly, mixing-normal specifications do not model the stochastic volatility with persistence. The mixing-normal settings assume the state variable for neighboring periods to be independent of one another and may not fit most business cycle variables and perhaps some financial market variables.

If there exists persistence, when the period t return is drawn from the first (the more volatile) distribution, then it is more (less) likely that the period $t+1$ return is drawn from the first (the second) distribution. Namely, $p_{11} > 0.5 > p_{21}$, $p_{22} > 0.5 > p_{12}$. During our empirical process, all estimates for transition probabilities (p_{11} and p_{22}) exceed 50%.¹⁷ The observed statistical property of our TAIEX samples appears to lend supports to the notion that MS outperforms mixing-normal. Furthermore, consistent with the notion that the Markov chain of

¹⁶ Duffie and Pan (1997) brought forward the jump diffusion models, which are akin to the mixing normal model. Instead of the traditional Bernoulli distribution, they adopted Poisson distribution to control the probability of jump diffusions. With simulated data, they further demonstrated the impact of λ , the expected jump frequency per unit time interval, and σ , the expected jump standard deviation (the magnitude), on VaR estimation. They conclude that, when the product of per period expected frequency and jump standard deviation is held constant, the lower the expected jump frequency per period, or the greater the jump standard deviation, the more significant is the tail-fatness in return distributions. Namely, exclusion of rare but extreme events may fuel the underestimation of VaR.

¹⁷ For example, p_{11} and p_{22} are 0.97 and 0.94, respectively, for TAIEX, and are 0.96 and 0.96, respectively, for TAIEX-ELEC.

return process is not reducible, in each and every phase in our tests the transition probability (p_{11} and p_{22}) is less than unity.

To remedy the general mixing normal models' omitting the persistence of regimes, Venkataraman (1997) incorporated quasi-Bayesian ML estimation of Hamilton (1991) and used Monte Carlo simulations to measure VaR¹⁸. Hamilton (1991) first proposed the use of quasi-Bayesian ML estimation, stating that via introducing prior constraints, analysts can mitigate the singularity problems in the converging process of ML function and thus eliminate overflows or underflows in ML estimation. A singularity arises whenever (1) the analysts over-react to an outlier and treat that single observation as a prevalent state, or (2) one of the distribution is imputed to have a mean exactly equal to one of the observations with no standard deviation ($\sigma = 0$, say.) The result of our rolling estimation in this study, nevertheless, indicates that the prior constraints in Bayesian, instead of eliminating overflows or underflows, serve at best to lower the frequency of this unpleasantness. More importantly, this alternative is with limited strength for our sample of TAIEX stocks, which are subject to a price limit of 7%.

Furthermore, our findings suggest that incorporating prior probabilities would add little to empirical studies for TAIEX series. Specifically, we document that the difference in regime volatility for TAIEX returns significantly contrasts with that in the setting in Hamilton (1991), who showed that with trivial difference between the parameters in two-regime models, additional prior constraints help enhance forecast accuracy and eliminate the multiple local maxima with EM-Algorithm. Nevertheless, we document that the standard deviation for TAIEX's high-volatility regime is almost 2.5 times as great as that for the low-volatility regime. Finally, the subjective prior constraints are essential in Bayesian models but are difficult to be relayed by risk managers or regulators for implementation within their specific economy or market.

GARCH, the fourth alternative, specifies the period t volatility as a function of prior period volatility and return shock squares. α , the volatility persistence measure, is equal to a_i plus b_i as described in Section 2.2. With parameters a_i and b_i unable to adjust for any structure changes in the estimation periods, the GARCH models are a linear volatility setting. Thus, contrast with Mixing-normal as well as MS, GARCH models are subject to over-estimating the persistence of return volatility.

¹⁸ See Hamilton (1991).

To sum up, the MS model is the least biased choice. First, the student t setting lacks the generalization for series such as TAIEX returns. Second, MS models effectively adjust for structural changes in the sample period and may thus outperform GARCH specifications. Third, MS models use Markov chain process to incorporate the persistence of volatility regimes and may thus outperform the mixture of normal models, which virtually disregard the persistence.¹⁹

3. Empirical Design and Test Results

This section first contrasts both kurtosis and skewness coefficients as well as the critical values of TAIEX return distributions with structural changes being filtered via MS with the (unfiltered) measures corresponding to the linear models. We then compare the effectiveness of MS, mixing normal and GARCH in mitigating Lepto kurtic and tail fatness. Furthermore, we examine the violation rates corresponding to linear and MS (non-linear) models for both tailed regions in return distributions. Consistent with our prior, after our filtering structural changes for the index returns via MS, each originally non-normal error term series appears to conform to standard normal distributions.

The sample data is provided by Taiwan Economic Journal (TEJ) and include Taiwan Stock Exchange market (TAIEX) returns as well as major industrial index returns. The latter test group includes TAIEX-CONS, TAIEX-FIN, TAIEX-EL&MACH, and TAIEX-ELEC. For TAIEX and the first four industrial indices, the sample period is between January 3, 1991 and May 28, 1998, whereas the sample period for TAIEX-ELEC only dated back to the beginning of 1995. Accordingly, we have 2,382 observations for the former indices and 1,236 observations for the electronics industry.²⁰

1. Kurtosis, Skewness and Critical Values of TAIEX Return Distributions with Structural Changes Being Filtered or Unfiltered

Table 1 presents statistics for return shocks including skewness and kurtosis as well as the 1%, 2.5% and 5% critical values for both tails in TAIEX return distributions. Without regarding any structural changes in mean and variance, as

¹⁹ Please refer to Li and Lin (1999) for further discussions.

²⁰ This study adopts the Basle Committee's proposal that the learning window should at least include 250 pre-VaR trading days. However, for TAIEX there are virtually approximately 281 daily observations of close prices per calendar year.

opposed to the standard normal distributions, the distributions for TAIEX-ELEC, TAIEX-FIN, TAIEX-CONS and TAIEX-EL&MACH return shocks are substantially skewed²¹. The kurtosis coefficient for TAIEX-ELEC and TAIEX-EL&MACH are 3.81 and 4.51, respectively. Moreover, the kurtosis coefficients for TAIEX-FIN and TAIEX-CONS returns are both greater than 5.

We further explore the significance of tail-fatness in the distributions of index return shocks. Table 1 shows that the realized absolute 1% critical values on both tails are significantly greater than that for a standard normal distribution. Moreover, the realized absolute 2.5% critical values for both tails are greater than that for a standard normal setting. The deviation, nevertheless, appear to decrease as we move towards the center. Up to the 5% critical probability, there appears no significant difference²². This finding shows that there exist small lumps on both tails in market return distributions.

Table 1 Skewness, Kurtosis, and 1%, 2.5%, 5% Critical Values for TAIEX and Taiwan's Major Industrial Index Return Shocks

	TAIEX	TAIEX -ELEC	TAIEX - FIN	TAIEX - CONS	TAIEX - EL&MACH
Skewness Coefficients (N=0)	-0.06	-0.17	0.29	0.13	-0.1
Kurtosis Coefficients (N=3)	5.48	3.81	5.17	5.05	4.51
1% Left-tailed Critical Value (N= -2.33)	-3.05	-2.75	-2.83	-2.88	-2.90
2.5% Left-tailed Critical Value (N= -1.96)	-2.20	-2.38	-2.10	-2.25	-2.28
5% Left-tailed Critical Value (N= -1.65)	-1.50	-1.84	-1.51	-1.51	-1.68
1% Right-tailed Critical Value (N=2.33)	2.73	2.40	3.16	2.90	2.84
2.5% Right-tailed Critical Value (N=1.96)	2.12	2.02	2.44	2.25	2.15
5% Right-tailed Critical Value (N=1.65)	1.60	1.70	1.73	1.68	1.68
Number of Observations	2382	1236	2382	2382	2382

1. Denote R_t the t -th period stock index return. Also let μ and σ denote mean and standard deviation for index returns, respectively. Then return shock $e_t = (R_t - \mu) / \sigma$.

²¹ The TAIEX return distribution is an exception. For major industrial group index returns, the absolute values of the skewness coefficients are greater than or equal to 0.1.

²² For the left-tailed region, which is the focus of VaR estimation, both TAIEX-ELEC and TAIEX-EL&MACH return shocks have realized absolute 5% critical values slightly greater than those for the standard normal distributions. Moreover, for TAIEX, TAIEX-FIN and TAIEX-CONS, the corresponding absolute values are slightly less than that for the standard normal distributions.

2. N represents the measure for a standard normal distribution.
3. The return shocks for TAIEX slightly skew to the left. In contrast, TAIEX-ELEC (TAIEX-FIN and TAIEX-CONS) return shocks significantly skewed to the left (right).
4. The distribution for TAIEX-ELEC, which is slightly Lepto kurtic, is an exception. All the other industrial index shocks are significantly Lepto kurtic.
5. Surrounding the 1% and 2.5% regions, TAIEX and industrial return shocks all appear to be significantly fat-tailed on both sides as opposed to standard normal distributions. There exists less significant tail fatness, however, for the 5% regions on both tails.

On the ground that the distributions of unfiltered index return shocks are subject to non-normality, we use a two-regime setting for means and variances, with the Markov-chain process to capture regimes translations. The two-variance-regime setting serves to control Lepto kurtic and fat tail in return distribution.²³ Our modeling the two regimes with differential means, on the other hand, helps filtering skewness for our sample distributions.

Table 2 Skewness, Kurtosis, and 1%, 2.5%, 5% Critical Values for Taiwan's Index Return Shocks, with the Structural Change Components Being Filtered via MS Models

	TAIEX	TAIEX- ELEC	TAIEX- FIN	TAIEX- CONS	TAIEX- EL&MACH
Skewness Coefficients (N=0)	0.07	-0.05	0.03	-0.03	-0.01
Kurtosis Coefficients (N=3)	2.93	2.77	2.82	2.50	-2.92
1% Left-tailed Critical Value (N=-2.33)	-2.34	-2.24	-2.30	-2.12	-2.32
2.5% Left-tailed Critical Value (N=-1.96)	-1.97	-2.03	-1.95	-1.83	-2.00
5% Left-tailed Critical Value (N=-1.65)	-1.60	-1.70	-1.61	-1.57	-1.67
1% Right-tailed Critical Value (N=2.33)	2.32	2.15	2.26	2.09	2.29
2.5% Right-tailed Critical Value (N=1.96)	1.98	1.88	1.87	1.78	1.95
5% Right-tailed Critical Value (N=1.65)	1.66	1.62	1.62	1.58	1.60
Number of Observations	2382	1236	2382	2382	2382

N represents the measure corresponding to standard normal distribution. Also, let u_1 and u_2 (μ_1 and μ_2) denote regime 1 (regime 2) sample mean and standard error, respectively. Then the filtered stock index return shocks is:

$$e_t = P(s_t=1|I_T) \times (r_t - u_1) / \sigma_1 + P(s_t=2|I_T) \times (r_t - u_2) / \sigma_2,$$

²³ The notion is consistent with Duffie and Pan (1997) and Venkatarman (1997).

where $P(s_t / I_T)$, $s_t=1,2$, denotes the estimated smoothing probability for the full sample.

Table 2 presents the error term statistics for Taiwan's index return shocks, with the structural change components being filtered by MS models²⁴. Comparing Tables 1 and 2, we may conclude that, with structural changes being filtered via MS models, the error term non-normality drops significantly. For instance, with structural change being unfiltered, the 1% left-tailed critical value, skewness and kurtosis for TAIEX-FIN error term are -2.83 , 0.29 , and 5.17 , respectively. With the structural changes being filtered, in contrast, the corresponding measures are -2.30 , 0.03 , and 2.82 , respectively. Moreover, the filtering effect for tail-fatness for the 2.5% (1%) regions appears to be less (more) significant. The filtering effect for tail-fatness for the 5% region appears to be the least significant among the three alternatives. Note that we adopt a setting to which both of the two mean regimes are exogenous. Our setting, as compared with Venkatarman (1997), who set the first regime mean return to be zero, appears to be more generalized²⁵.

2. The Performance of the Competing Models in Mitigating Non-Normality Problems

There may be three alternatives to mitigate Lepto kurtic and fat tails in return distribution: the mixing normal, MS and GARCH models. In the mixture of normal distribution models, regimes switches are purely random, with no persistence in regimes cross time. In contrast, MS models adopt Markov chain process to incorporate the measure of regime persistence. The GARCH models differ from MS and mixing normal models in that the former setting parameters do not adjust for structural changes during the estimation periods and thus may over-estimate the volatility persistence.

To compare the performance of MS, the mixing normal and GARCH for Lepto kurtic and/or fat-tailed returns, we use the linear models as the benchmark settings and apply the three competing models to the probability density function

²⁴ We use OPTIMUM, a package program from GAUSS, and the built-in BFGS (Boyden, Fletcher, Goldfarb, and Shanno) algebra to derive the negative minimum likelihood (ML) function values of all the models. BFGS algebra is effective for deriving the maximum value of the non-linear likelihood functions. See Luenberger (1984). We randomly generate 100 sets of initial values. We then derive the ML function value for each of the 100 sets of initial values. The mapped converged measure with the greatest ML function value then serves as the estimate.

²⁵ The models we adopt, nevertheless, require an additional parameter as opposed to that in Venkatarman (1997).

of Taiwan Stock Exchange index returns.

Table 3 presents the relative strengths of the competing models in handling non-normality problems in the TAIEX return shocks²⁶. Our first horse race focuses on the Lepto kurtic problem. Note that the linear models result a large kurtosis coefficient of 5.48. In contrast, MS, Bayesian mixing normal and GARCH models all generate lower kurtosis coefficients (2.93, 2.38 and 4.66, respectively). Specifically, the kurtosis coefficient with MS model is closest to that of standard normal distribution. In contrast, the coefficient with respect to Bayesian mixing normal (GARCH) model appears to be lower (higher) as opposed to standard normal distributions. As to tail-fatness, the 1% left-tailed critical values for linear, MS, Bayesian mixing normal and GARCH models are -3.05, -2.34, -2.01 and -2.84, respectively. The results suggest that the latter three settings are all effective in filtering structural change components for fat-tailed returns. Nevertheless, as compared with standard normal distributions, for which the 1% left-tailed critical value is -2.33, MS models perform the best. Bayesian mixing normal (GARCH) models, relatively speaking, under-(over-) estimate the expected loss. As to the 2.5% and 5% left- and right-tailed regions, all the three remedial models are with decreasing magnitudes in reducing tail-fatness. These findings are consistent with our linear model results that tail-fatness is significant for the 1% region but less significant for the 2.5% and 5% regions on both sides. The relative performance of the competing models may be measured by the extent the statistics for filtered return shocks resemble those for standard Gaussian, not by the absolute measures of the statistics. In other words, the benchmark kurtosis coefficient, skewness coefficient, and left-tailed 1% critical value are 3, and 0, -2.33, respectively. With the corresponding estimates for kurtosis coefficient of 2.93, MS models appear to generate return shock statistics closer to the benchmark measure of 3.0 as opposed to Bayesian mixing-normal models (the corresponding estimates is 2.38.) Likewise, the left-tailed 1% critical value for MS is -2.34 and lies much closer to the benchmark of -2.33 as compared with Bayesian mixing-normal (the corresponding estimates is -2.01.)

Figure 1 presents the probability density function of TAIEX return shocks with the setting of no switches in mean or variance. As opposed to standard normal, there exist significant Lepto kurtic and tail fatness. Figures 2, 3 and 4 document the probability density functions of TAIEX return shocks for MS, Bayesian mixing normal and GARCH, respectively. Taking Figure 1 as our

²⁶ Table 2 shows that the TAIEX returns are not significantly skewed if we apply the linear model for the return shocks.

benchmark, Figures 2, 3 and 4 show that the three non-linear models help mitigate Lepto kurtic and tail-fatness problems.

Let us further examine Figures 2 to 4. Bayesian mixing normal models appear to *over-react* to Lepto kurtic and tail fatness. As shown in Figure 3, the tail areas differ significantly from that in the standard normal specification. GARCH models, on the other hand, *under-react* to Lepto kurtic and fat tails. In contrast, MS models, which generate less biased results, appear to be the most efficient in filtering structural changes. The error term with MS, as compared with those with the two competing models, dosely conforms to standard normal distributions. The finding is consistent with the notion that

Table 3 Effectiveness of the Competing Models in Mitigating Lepto kurtic, Tail-Fatness and Skewness Problems

	Linear	MS	Bayesian Mixing-Normal	GARCH (1,1)
Kurtosis Coefficients (N=3)	5.48	2.93*	2.38	4.66
Left-tailed 1% Critical Value (N=-2.33)	-3.05	-2.34*	-2.01	-2.84
Left-tailed 2.5% Critical Value (N=-1.96)	-2.20	-1.97*	-1.54	-2.17
Left-tailed 5% Critical Value (N=-1.65)	-1.50	-1.60	-1.36	-1.69*
Right-tailed 1% Critical Value (N=2.33)	2.73	2.32*	1.85	2.38
Right-tailed 2.5% Critical Value (N=1.96)	2.12	1.98	1.45	1.95*
Right-tailed 5% Critical Value (N=1.65)	1.60	1.66*	1.30	1.64*
Number of Observations	2,382	2,382	2,382	2,382

1. This table shows that with the linear models, TAIEX return shocks slightly skew to the left and are significantly Lepto kurtic with respect to critical probabilities of 1% and 2.5% for both tails. Also, there appears significant tail-fatness for both ends. However, tail fatness phenomena become less significant when the critical probability is set to 5% for both left- and right-tailed regions.
2. * denotes the measure closest to the statistics for standard normal distributions.

the general mixing normal models disregard the information for regimes persistence cross time, whereas GARCH models over-estimate the persistence in volatility. Namely, the prevalent normal specifications for the error term can benefit from incorporating MS models to control the structural change components in the estimation period.

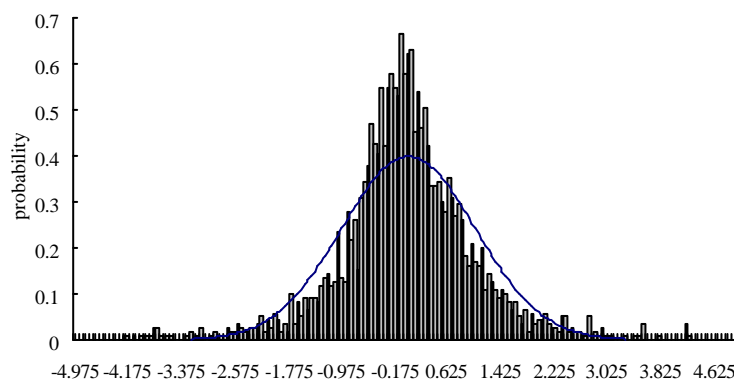


Figure 1. PDF for TAIEX Return Shocks with the Linear Models

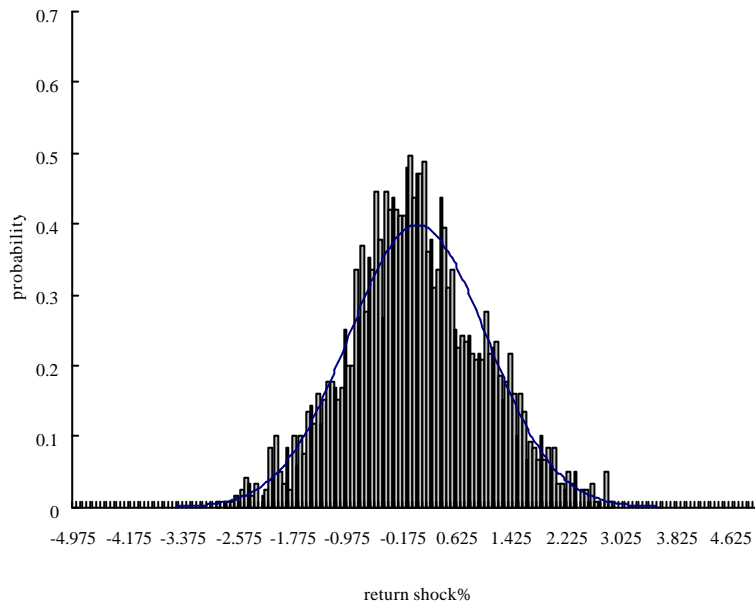


Figure 2. PDF for TAIEX Return Shocks with Markov-switching Models

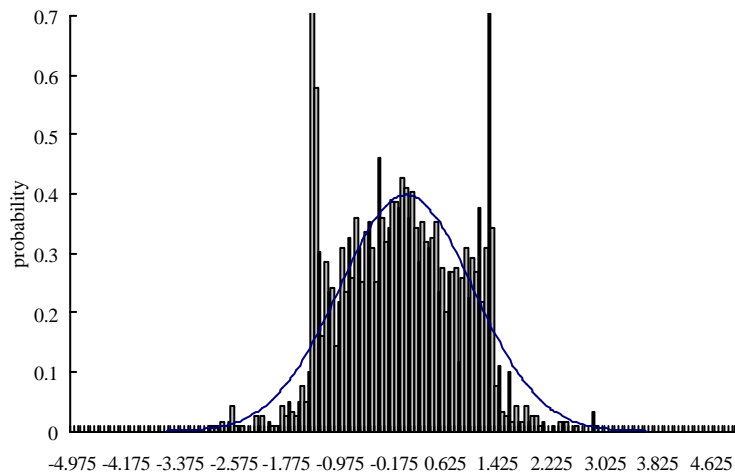


Figure 3. PDF for TAIEX Return Shocks with Mixture of Normal Models

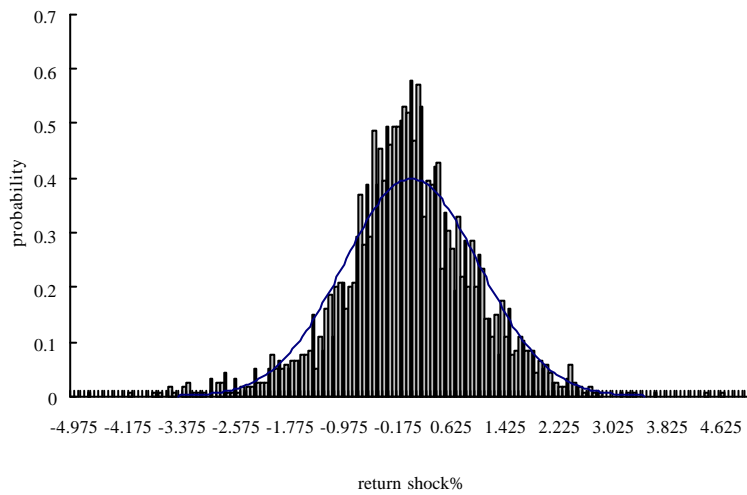


Figure 4. PDF for TAIEX Return Shocks with GARCH Models

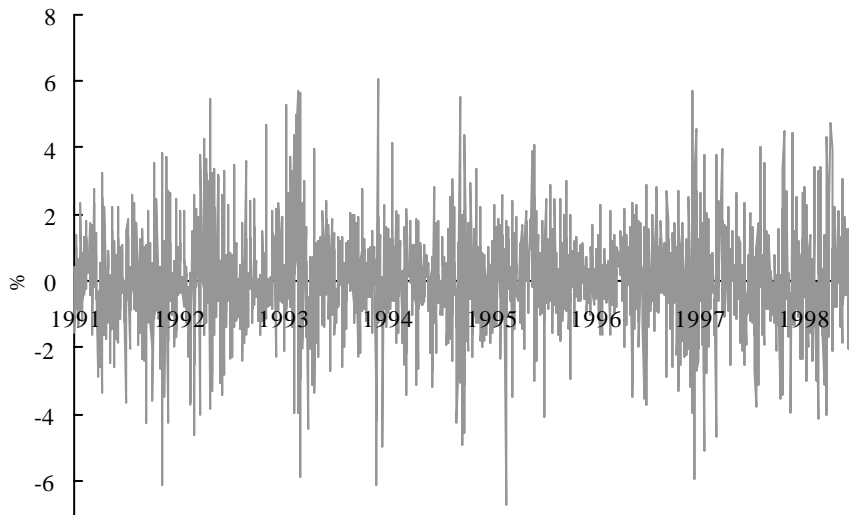


Figure 5. TAIEX Returns during the Sample Period

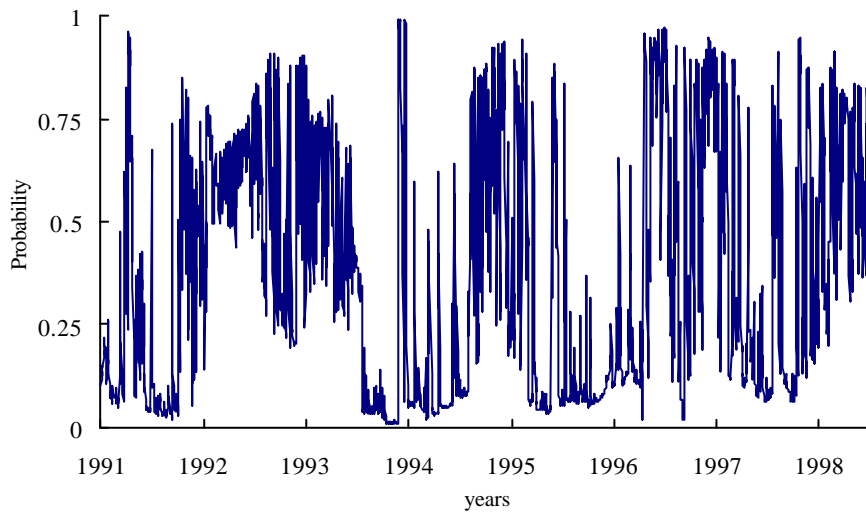


Figure 6. Regime 1 Predicting Probabilities for TAIEX Returns via MS Models

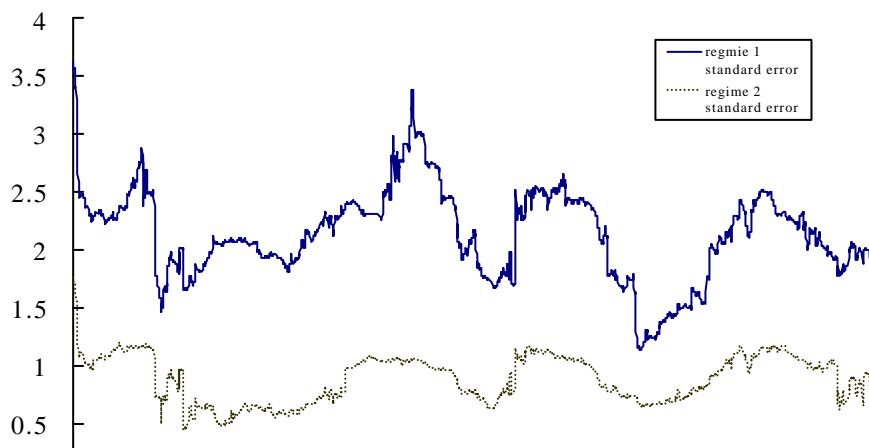


Figure 7. Estimated Regime 1 and Regime 2 Standard Errors for TAIEX returns via MS Models

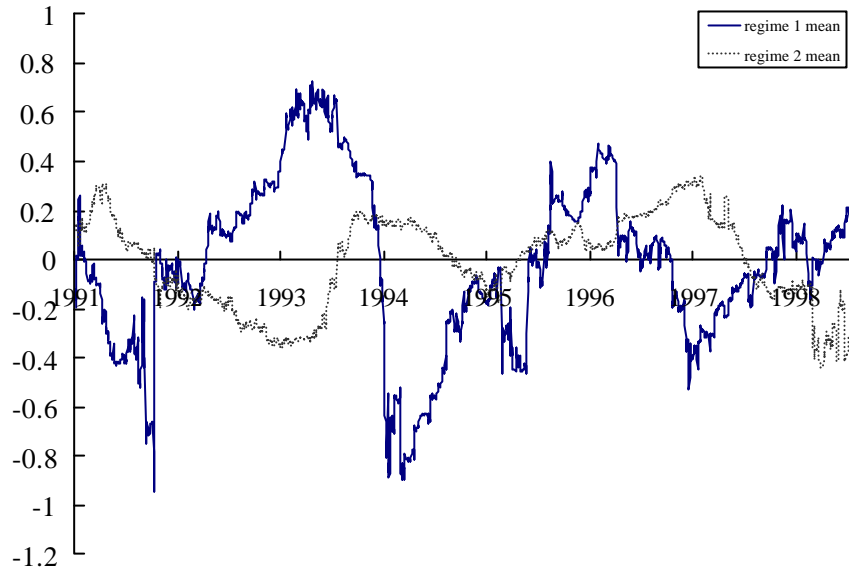
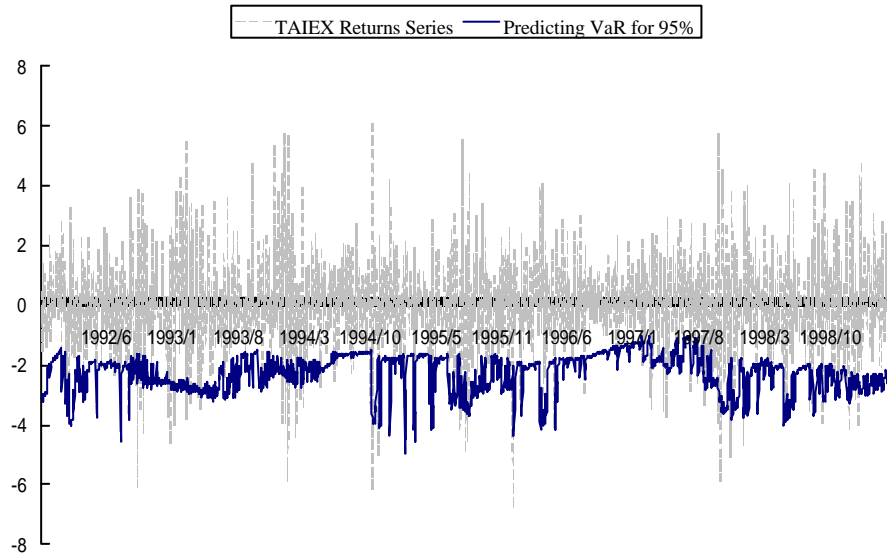
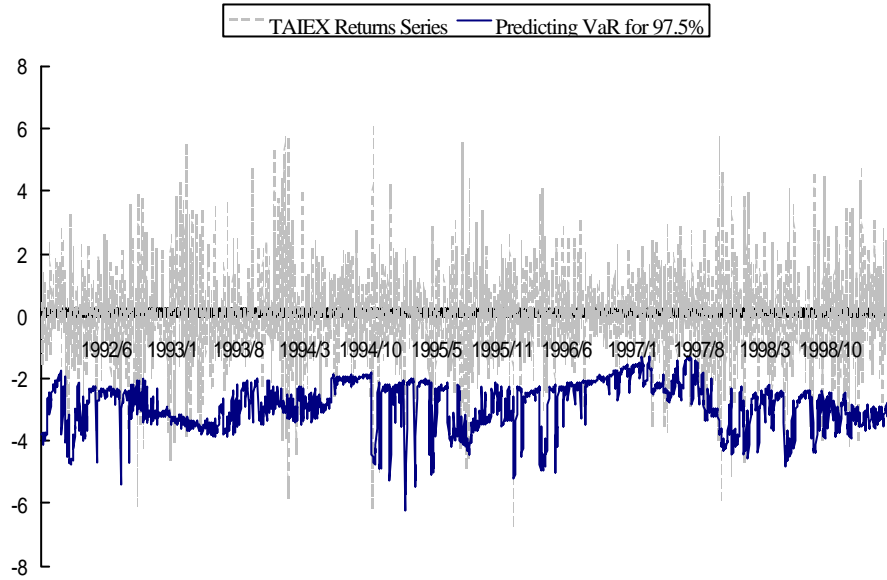


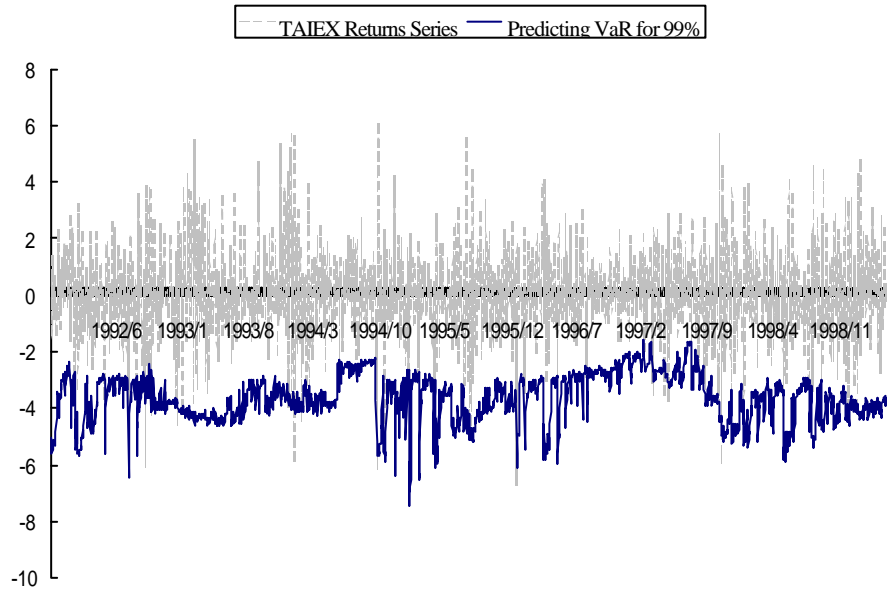
Figure 8. Estimated Regime 1 and Regime 2 Means for TAIEX Returns via MS Models



(a)The Predicting VaR for a Critical Interval of 95%



(b)The Predicting VaR for a Confidence Interval of 97.5%



(c)The Predicting VaR for a Confidence Interval of 99%

Figure 9. The Predicting VaR with 250-Prior-Trading-Day Windows via MS Model

3. Violation Rate Examinations

(1) Left-Tailed and Right-Tailed Violation Rates for Linear and MS Models with 250- and 500-Prior-Trading-Day Windows

We conduct horse races for both linear and non-linear (MS) models on TAIEX, TAIEX-ELEC, TAIEX-FIN, TAIEX-CONS and TAIEX-EL & MACH. We alternately use 250- and 500- prior-trading-day windows²⁷ in the rolling estimation process and measure the violation rates on both tails of the distributions.

(a) Estimating the Critical Values for Intervals with Negative Returns

Our research design may be illustrated via the tests with 250 pre-VaR daily observations. For each date t , we get the estimates from historical returns series $\{R_{i-t}\}_{i=1}^{250}$. Then our Monte Carlo algorithm generates 10,000 random numbers to simulate the distribution of index return R_t . For 1%, 2.5% and 5% left-tailed regions, we further estimate VAR_t and make comparisons between VAR_t and stock returns R_t . If $R_t < VAR_t$ ($R_t > VAR_t$), the number of violations would increase by one (would remain unchanged). The violation rate measures then serve as our indicator of forecast accuracy.²⁸ For estimation practices, we adopt the ex ante predicting probabilities $P(s_t|y_{t-1}, y_{t-2}, \dots)$ to weigh the two normal distributions with different means and variances.²⁹

To illustrate with our TAIEX returns series, the sample period is between January 3, 1991 and May 28, 1998, with 2,382 observations. For our tests with 250 prior trading days as the learning stage, we have 2,132 violation rate observations. Figures 5 and 6 present TAIEX returns and the predicting probabilities with respect to regime 1, respectively. Moreover, Figure 7 presents the estimates for daily standard deviation of returns corresponding to regimes 1 and 2. Furthermore, Figure 8 presents the estimated mean returns for regimes 1 and 2.

²⁷ In our rolling estimation process, we move one step further and introduce one more observation for each time point. To facilitate convergence in our estimation process, we use the estimates for each period as the initial values in our non-linear estimation for the immediately subsequent period.

²⁸ The closer the percentage violation to the critical probability, the higher is the forecast accuracy.

²⁹ For specific applications, conditional probability $p(s_t|y_{t+q}, y_{t+q-1}, \dots)$ may be tailored for different applications with simply substituting a different measure of q . Specifically, when $q=0$, it is a filtering probability. If $q<0$, it is an ex ante predicting probability. Whereas when $q>0$, it is an ex post smoothing probability.

Since the standard deviation for Regime 1 (Regime 2) returns is significantly greater (less)³⁰, Regime 1 (Regime 2) is labeled as the high (low) volatility regime³¹. Figure 9 presents the VaR estimates corresponding to MS models with 250-prior-trading-day windows for confidence intervals 95%, 97.5% and 99%.

The left-tailed region is the focus of concurrent VaR studies. Panel A of Table 4 presents the violation rates associated with TAIEX and Taiwan's industrial returns when the left-tailed critical probability is set to 1%. In terms of violation rate, the two MS models rank the first and the second in forecast accuracy. The findings lend support to the superiority of MS models. MS model with a 500 prior-trading-day learning window excels with respect to TAIEX, TAIEX-ELEC, TAIEX-CONS and TAIEX-FIN. It also ranks the second for TAIEX-EL&MACH. In contrast, MS with a 250 pre-VaR-day learning window performs the best for TAIEX-FIN as well as TAIEX-EL&MACH and ranks the second for TAIEX, TAIEX-ELEC and TAIEX-CONS.

Panel B of Table 4 presents the results for the setting with critical probability equal to 2.5%. The linear models with a 500-day learning window perform the best for TAIEX-CONS and rank the second for TAIEX-FIN. The linear models with a 250-trading-day learning period rank the second for TAIEX-ELEC. MS models with a 500-day learning period perform the best for TAIEX, TAIEX-FIN as well as TAIEX-EL&MACH and rank the second for TAIEX-CONS. The MS models with a 250-day window perform the best for TAIEX-ELEC and rank the second for TAIEX-EL&MACH and TAIEX-ELEC.

Panel C of Table 4 documents the horse race results regarding the 5% left-tailed critical probabilities. The linear models with a 500 pre-VaR observations perform the best for TAIEX. The linear models with a 250 trading-day windows perform the best for estimating TAIEX-ELEC volatility and ranks the second for TAIEX and TAIEX-EL&MACH. MS models with previous 500 daily return as the learning period perform the best for TAIEX-FIN, TAIEX-CONS, TAIEX-EL&MACH. MS models with a 250-day window rank the second for TAIEX-ELEC, TAIEX-FIN, and TAIEX-CONS.

³⁰ Namely, $\sigma_1 > \sigma_2$

³¹ In contrast, there exist mixed results as to the order of sample means with respect to regimes 1 and 2.

Table 4 Left-Tailed Violation Rates for the MS and the Linear Models with Differential Learning Windows for TAIEX Market and Industrial Returns

Panel A Critical Probability = 1%

Stock Index Returns	Linear Model with 250 Daily Return Obs.	Linear Model with 500 Daily Return Obs.	MS Model with 250 Daily Return Obs.	MS Model with 500 Daily Return Obs.
TAIEX	2.11%	2.07%	1.97% *	1.38% **
TAIEX-ELEC	2.64%	3.80%	1.93% *	1.77% **
TAIEX-FIN,	1.88%	1.59%	1.13% **	1.22% *
TAIEX-CONS	2.11%	2.13%	1.92% *	1.43% **
TAIEX- EL&MACH	2.53%	2.60%	1.36% **	1.43% *

Panel B Critical Probability = 2.5%

Stock Index Returns	Linear Model with 250 Daily Return Obs.	Linear Model with 500 Daily Return Obs.	MS Model with 250 Daily Return Obs.	MS Model with 500 Daily Return Obs.
TAIEX	3.28%	3.51%	3.19% *	2.60% **
TAIEX-ELEC	3.65% *	5.84%	3.25% **	4.89%
TAIEX-FIN	2.77%	2.60% *	2.95%	2.44% **
TAIEX-CONS	2.95%	2.82% **	3.24%	2.92% *
TAIEX-EL&MACH	3.71%	4.04%	3.42% *	2.82% **

Panel C Critical Probability = 5%

Stock Index Returns	Linear Model with 250 Daily Return Obs.	Linear Model with 500 Daily Return Obs.	MS Model with 250 Daily Return Obs.	MS Model with 500 Daily Return Obs.
TAIEX	4.74% *	4.89% **	5.77%	5.31%
TAIEX-ELEC	5.78% **	7.07%	6.09% *	7.07%
TAIEX-FIN	4.22%	3.83%	5.25% *	5.10% **
TAIEX-CONS	4.16%	4.41%	5.35% *	5.10% **
TAIEX-EL&MACH	5.39% *	5.90%	5.91%	5.37% **

1. We adopt both 250- and 500-day learning windows of prior daily returns. For the violation rate tests with a 250-day window, there are 986 (2,132) observations for TAIEX-ELEC (the other indices). In our tests with a 500-day window, the number of observations for TAIEX-ELEC (the other indices) is 736 (1,882).

2. ** (*) denotes the most accurate (the second most accurate) measure in terms of violation rate.

To sum up, for users with critical probability of 1%, MS with a 500-day learning window dominates the competing models in terms of violation rate. As to the 2.5% critical probability, MS with a 500 prior-trading-day window still beats the competing models but with smaller differences. As the critical probability is set to be 5%, there are even smaller differences. As to the choice of the learning window, generally speaking, the setting with 500 (250) pre-VaR observations generates more (less) accurate estimates.

(b) Estimating the Critical Value for the Interval with Positive Returns

To gain insights as to the strengths/weaknesses of the competing models as well as the statistical properties of TAIEX index returns, we also apply the VaR algorithm to study the right-tailed region. Specifically, we examine the violation rate, the measure of forecast accuracy, associated with each of our combinations of competing models and learning windows for TAIEX market and industrial returns. Panels A, B, and C in Table 5 present results corresponding to critical probabilities of 1%, 2.5% and 5%, respectively. In general, MS models beat the alternatives for both 1% and 2.5% right-tailed regions in terms of violation rate. When the critical probability is set to 5%, however, the MS models no longer outperform the linear models. Moreover, with MS models, we find the shorter 250-day learning window alternative outperforms the 500-day setting. A potential explanation is that incorporating more historical data helps gain the insights for the complete cycle but drives the model to be less sensitive to regime switches and thus over-estimate the persistence.

Let us summarize the findings in this section. For both tails, MS dominates the popular linear models in depicting the tailed regions in return distributions³². The difference, however tend to decrease as we move towards the center of the distribution. Moreover, our empirical results show that, to depict the left- as opposed to the right-tailed region, one needs to incorporate more historical data. For the busts, one would need a longer window of 500 prior daily returns (approximately two-year data) to capture the odd events with large negative losses. On the other hand, a learning window of 250 prior observations (nearly one-year data) would be sufficient to depict the booms.

³² When the right-tailed probability is set to 1% or 2.5%.

Table 5 Right-Tailed Violation Rates for the MS and the Linear Models with Differential Learning Windows for TAIEX Market and Industrial Returns

Panel A Critical Probability = 1%

Stock Index Returns	Linear Model with 250 Daily Return Obs.	Linear Model with 500 Daily Return Obs.	MS Model with 250 Daily Return Obs.	MS Model with 500 Daily Return Obs.
TAIEX	2.02%	1.86%	1.17% **	1.28% *
TAIEX-ELEC	1.83%	2.85%	0.91% **	1.49% *
TAIEX-FIN	3.10%	3.19%	1.69% **	1.86% *
TAIEX-CONS	2.39%	2.60%	1.41% *	1.28% **
TAIEX- EL&MACH	1.88%	2.07%	0.94% *	1.01% **

Panel B Critical Probability = 2.5%

Stock Index Returns	Linear Model with 250 Daily Return Obs.	Linear Model with 500 Daily Return Obs.	MS Model with 250 Daily Return Obs.	MS Model with 500 Daily Return Obs.
TAIEX	3.19%	3.19%	2.67% **	2.92% *
TAIEX-ELEC	3.55%	4.89%	2.64% **	2.72% *
TAIEX-FIN	4.41%	4.46%	3.61% **	4.04% *
TAIEX-CONS	3.94%	3.72%	2.91% **	3.45% *
TAIEX- EL&MACH	3.42%	3.72%	2.67% **	2.82% *

Panel C Critical Probability = 5%

Stock Index Returns	Linear Model with 250 Daily Return Obs.	Linear Model with 500 Daily Return Obs.	MS Model with 250 Daily Return Obs.	MS Model with 500 Daily Return Obs.
TAIEX	5.21% **	5.21% **	5.72%	5.42% *
TAIEX-ELEC	6.19%	7.74%	5.38% **	5.71% *
TAIEX-FIN	6.14%	5.74% **	6.14%	6.06% *
TAIEX-CONS	5.91%	5.47% **	5.82% *	6.11%
TAIEX-EL&MACH	5.96%	5.90%	4.83% **	5.37% *

When the critical probability is set to 1% or 2.5%, MS models is the most accurate alternative in terms of violation rate. In contrast, when the critical probability is set to be 5%, MS models are no longer superior to the competing settings.

(2) The Performance of the Competing Models in Terms of Left-Tailed Violation Rates

The previous section demonstrates that the MS models effectively help mitigate the excessive violation rate in traditional VaR estimation with linear models. With a focus on the left-tailed region, this section further explores the relative strengths and weaknesses of the competing non-linear models including MS, GARCH and Bayesian mixing normal³³. In order to incorporate in the information sets rare but damaging events, we adopt a learning window of 500 prior trading days.

Table 6 presents the left-tailed violation rates associated with the alternative models for major TAIEX index returns. Since there exists no strictly dominant model, we also present in this table the mean absolute errors for violation rates (hereafter MAE) associated with each of the models. MAE, the average squared difference between each critical probability level and the associated observed percentage of violations over the five indices, serves to proxy forecast accuracy.

Panel A of Table 6 shows that with 1% critical probability, the MS models stand out in terms of overall MAE. Slightly lower down the pecking order is Bayesian mixing normal. The less accurate specifications are GARCH and the linear models. Furthermore, the MS models appear to excel with respect to all the major indices except for the TAIEX Index³⁴. For the latter, Bayesian mixing normal models rank the first. One potential explanation is the ever-changing membership on TAIEX during the sample period. Namely, new companies, particularly the electronic firms with large market capitalization, kept getting listed on the Taiwan Stock Exchange in these years. Such changes in industry and company profiles not only violate the persistence assumption for the MS models but also make Taiwan's overall market return measure less relevant for investment decisions. Accordingly, analysts should not over-emphasize the victory of Bayesian mixing normal over the MS models for TAIEX series.

Panel B of Table 6 presents the violation rates with respect to the 2.5% tailed probability. GARCH (MS) comes out on top (comes a close second) on the

³³ Please refer to Hamilton (1991).

³⁴ Bayesian-mixing Normal slightly outshines MS models for TAIEX-FIN. The violation rates are 1.12% and 1.22%, respectively. Nevertheless, note that Monte Carlo simulation generates random numbers to resemble the return distributions with measurement errors despite that we have as many as 10,000 random draws in our setting. As compared with the potential noises, the 0.1% difference in violation rate appears to be statistically insignificant.

criterion of overall MAE. The bottom measures are Bayesian mixing normal and the linear models. Let us further investigate the performance of MS versus GARCH. In terms of MAE, GARCH and MS have a negligibly small difference of 0.04%. GARCH is the best model for TAIEX-ELEC and TAIEX-CONS, whereas MS is the most accurate (the second most accurate) measure for TAIEX-FIN and TAIEX-EL&MACH (for TAIEX, TAIEX-ELEC and TAIEX-CONS.) Namely, MS, which at worst ranks the second, serves as a more reliable specification. In contrast, the relative performance of GARCH varies significantly. The findings suggest that analysts adopt the MS models to estimate VaR for the 2.5% region.

Panel C of Table 6 presents the 5% region violation rates for the respective models. Again, there is no strictly winning alternative. Generally speaking, nevertheless, in terms of MAE, Bayesian mixing normal and MS perform the best.³⁵ The runners-up are GARCH and the linear models. Also interestingly, the traditional linear models are the losers for 1%, 2.5% and 5% regions but the MAE corresponding to the linear models decreases with the tailed probability.³⁶ This finding provides supports to the notion that the tail-fatness is significant only towards both extremes of the returns distributions. Namely, there exists a diminishing threat of high violation rates due to underestimating VaR as analysts' focus moves closer to the mean region.

To sum up, our findings back the superiority of MS. For both 2.5% and 5% regions, the MS models are not the strictly dominant but are generally the most robust measure. Moreover, to audience such as commercial bank investors or creditors, who most typically have a 1% or lower left-tailed critical probability, the MS models significantly outperform all the others. The results are consistent with significant persistence for rare but extremely damaging events. On the one hand, we propose that the major factor for the outperformance of MS over GARCH is the emergence of structural changes due to these extreme events. On the other hand, the significance of the persistence is likely to explain our findings that the MS models outperform Bayesian mixing normal.

4. Conclusion

This paper serves as one of the first systematic studies on applying

³⁵ The 0.04% difference in MAE between these two models, again, is trivial.

³⁶ With the linear models, the MAE's for the three cut-off criteria are 1.44%, 1.26% and 0.97%, respectively.

Table 6 Left-Tailed Violation Rates Associated with Linear, GARCH, Bayesian Mixing Normal and Markov-switching Models for TAIEX Market and Industrial Returns

Panel A Critical Probability = 1%

Stock Index Returns	Linear Model	GARCH(1,1)	Bayesian Mixing Normal	MS Model
TAIEX	2.07%	1.91%	1.06% ^{**}	1.38% [*]
TAIEX-ELEC	3.80%	2.17% [*]	3.26%	1.77% ^{**}
TAIEX-FIN,	1.59%	1.28%	1.12% ^{**}	1.22% [*]
TAIEX-CONS	2.13%	1.91%	1.70% [*]	1.43% ^{**}
TAIEX- EL&MACH	2.60%	1.86%	1.59% [*]	1.43% ^{**}
MAE	1.44%	0.83%	0.72% [*]	0.45% ^{**}

Panel B Critical Probability = 2.5%

Stock Index Returns	Linear Model	GARCH(1,1)	Bayesian Mixing Normal	MS Model
TAIEX	3.51%	2.92%	2.55% ^{**}	2.60% [*]
TAIEX-ELEC	5.84%	3.67% ^{**}	5.30%	4.89% [*]
TAIEX-FIN	2.60% [*]	2.23%	2.76%	2.44% ^{**}
TAIEX-CONS	2.82% ^{**}	2.82% ^{**}	2.82% ^{**}	2.92% [*]
TAIEX-EL&MACH	4.04%	3.29%	3.19% [*]	2.82% ^{**}
MAE	1.26%	0.59% ^{**}	0.80%	0.63% [*]

Panel C Critical Probability = 5%

Stock Index Returns	Linear Model	GARCH(1,1)	Bayesian Mixing Normal	MS Model
TAIEX	4.89% ^{**}	5.53%	5.21% [*]	5.31%
TAIEX-ELEC	7.07%	6.39% ^{**}	6.66% [*]	7.07%
TAIEX-FIN	3.83%	3.88%	4.94% ^{**}	5.10% [*]
TAIEX-CONS	4.41%	4.14%	5.21% [*]	5.10% ^{**}
TAIEX-EL&MACH	5.90%	5.37% ^{**}	5.63% [*]	5.37% ^{**}
MAE	0.97%	0.85%	0.55% ^{**}	0.59% [*]

** (*) denotes the most accurate (the second most accurate) measure in terms of violation rate.

Markov-switching (MS) models to VaR estimation. Prior VaR studies most typically encountered non-normality problems in return distributions. Our findings suggest that the emergence of structural changes during the estimation periods serves as one explanation to these results. The traditional linear models, unfortunately, do not adjust for the problems.

Among the alternative non-linear models, we propose that MS models outperform the competing specifications. First, contrast with general mixing normal, MS models incorporate the information as to persistence in regimes. Second, as compared with GARCH, MS models more effectively control the structural changes.

Our empirical results lend supports to the superiority of MS models. We show that, Lepto kurtic, fat tail and significant skewness exist in the return distributions of Taiwan's major indices. Our empirical results lend supports to the notion that TAIEX, TAIEX-Construction, TAIEX-Finance, TAIEX-Electronic, and TAIEX-Electric&Machinery returns are significantly skewed with fat tails. Furthermore, tail fatness becomes more significant as we move towards both extremes. Also consistently, we find that MS models outperform the other alternatives in mitigating the adverse impacts of non-normality. In contrast, the mixing normal and GARCH models tend to over- and under-react, respectively, to these problems. Moreover, despite that the superiority of MS is insignificant when the critical probability is greater than or equal to 5%, MS models appear to be the best alternative to financial institutions, for which regulators and the risk managers often focus on the 1% or more extreme left-tailed region. To reasonably conservative market participants, the more complex MS models are with greater accuracy in measuring market risks. Furthermore, as for the choice of learning windows, we find that for capturing the odd but influential events, one needs to move more (less) dated back for the historical data to depict the left-(right-) tailed returns.

References

1. Beder, T. S., 1995, "VAR: Seductive but dangerous," *Financial Analysts Journal*, 12-24.
2. Bollerslev, T., 1986, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 28, 307-327.

3. Cai, J., 1994, "A Markov model of unconditional variance in ARCH," *Journal of Business and Economic Statistics*, 12, 309-316.
4. Duffie D. and J. Pan, 1997, "An overview of value at risk," *Journal of Derivatives*, 4, 7-49.
5. Estrella, A., D. Hendricks, J. Kambhu, S. Shin and S. Walter, 1994, "The price risk of options positions: Measurement and capital requirements," *Federal Reserve Bank of New York Quarterly Review*, 27-43.
6. Engel, R. F., 1982, "Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation," *Econometrica*, 50, 987-1007.
7. Hamilton J. D., 1989, "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57, 357-384.
8. Hamilton J. D., 1991, "A quasi-Bayesian approach to estimating parameters for mixtures of normal distributions," *Journal of Business and Economic Statistics*, 9, 27-39.
9. Hamilton, J. D. and R. Susmel, 1994, "Autoregressive conditional heteroscedasticity and changes in regime," *Journal of Econometrics*, 64, 307-333.
10. Hamilton, J.D., 1994, *Time series analysis*, (Princeton University Press, Princeton, NJ).
11. Ho, C. Y. and H. W. Lin, 1999, "Explores the relative strength and weakness of Two-Limit Tobit model," *Working Paper*.
12. Li, M. Y. and H. W. Lin, 1999, "Examining TSE security return variability via a three-volatility-regime Markov-switching model," *The Seventh Conference on Pacific Basic Finance, Economics and Accounting*.
13. Liao, Y. C., 1998, "Managing the market risk: using the technique of σ -orthogonal GARCH of value at risk," *Working paper*.
14. Jackson, P., D. J. Maude and W. Peerraudin, 1997, "Bank capital and value at risk," *Journal of Derivatives*, 73-89.
15. Jorion, P., 1997, "Value at Risk: The new benchmark for controlling market risk, Irwin."
16. J. P. Morgan, RiskMetrics, 1995, Available on the web site: www.jpmorgan.com.

17. Kupiec, P. H., 1995, "Techniques for verifying the accuracy of risk measurement models," *Journal of Derivatives*, 73-84.
18. Simons, K., 1996, "Value at risk – New approach to risk management," *New England Economic Review*, 3-13.
19. Uong, T. Y. and W. H. Chen, 1998, "A quantitative approach "VAR" to measure the market risk of bank currency positions," *Working Paper*.
20. Venkataraman, S., 1997, "Value at risk for a mixture of normal distributions: the use of quasi-Bayesian estimation techniques," *Economic Perspectives*, 2-13.
21. Wang, J. J. and W. H. Chen, 1998, "Introduction and Applications on Value-at-Risk," *Working Paper*.